UNIVERSITÉ DE NEUCHÂTEL

The Multidimensional Nature of Poverty in Upper-Middle-Income Countries

PART II

Ph.D. Thesis/Dissertation

Author: A. Tugce BEYCAN

Supervisor: Christian Suter

Department of Sociology, University of Neuchâtel

Faculty of Letters and Human Sciences

December 18, 2020

UNIVERSITÉ DE
NEUCHÂTEL

Faculté des Lettres et Sciences Humaines
Espace Tilo-Frey 1
CH - 2000 Neuchâtel
Tél : +41 (0)32 718 17 00
Fax : +41 (0)32 718 17 71
secretariat.lettres@unine.ch

IMPRIMATUR

La Faculté des lettres et sciences humaines de l'Université de Neuchâtel, sur les rapports de M. Christian Suter, directeur de thèse, professeur, Université de Neuchâtel ; Mme Filomena Maggino, professore, Dipartimento di Scienze Statistiche, Sapienza Università di Roma ; Mme Valerie Møller, Professor Emeritus, Quality of Life Studies, Institute of Social and Economic Research, Rhodes University, Grahamstown, South Africa ; M. Serge Paugam, Directeur de recherche au CNRS, Directeur d'études à l'EHESS, Directeur du Centre Maurice Halbwachs (CNRS/EHESS/ENS), Paris autorise l'impression de la thèse présentée par Madame Aysegül Tugce Beycan en laissant à l'autrice la responsabilité des opinions énoncées.

Neuchâtel, le 18 décembre 2020

Le doyen
Pierre Alain Mariaux

P.O.

Table of Contents

6. Comparative analysis 13
6.1 Introduction 13
6.2 Comparison between South Africa, Turkey, Mexico, and Brazil 24
7. Conclusion 47
8. Bibliography 63
9. Appendix 117
9.1 Extensive literature review of poverty concept 117
Subsistence, assistance 117
Monetary poverty 122
Culture of poverty, cycle of deprivation, underclass 131
Working poor 151
Vulnerability, poverty dynamics, vulnerability to poverty 158
Subjective poverty 170
9.2 Comparative summary of poverty concepts 175
9.3 Statistical background: data mining, probability theory, classification trees 180
Data mining and knowledge discovery 180
Introduction to statistical terms: some basic principles of probability theory 187
Statistical method: classification trees 213
Mechanism of decision trees 213
Introduction to classification trees 216
Entropy function, Information gain 225
Gini index of diversity 240
Misclassification 243
Pruning trees 245
Predictive accuracy of classifier 247
Independent test set: training (or called learning) and test sets 247
k-fold cross-validation process 248
N-fold cross-validation process 249
Leave-one-out cross-validation 249
Stratified cross-validation 249
Evaluation measures and Confusion Matrix 249
Missing values 253
Advantages of classification trees 254
Statistical software 254
9.4 Case study of Turkey 255
Introduction 255

First step (step1 analysis): detecting multidimensional poverty patterns 267
Second step (step2 analysis): detecting dimensional poverty profiles of poverty groups 310
Third step (step3 analysis): detecting various profiles of poverty groups 320
Fourth step (step4 analysis): detecting multidimensional poverty profiles of specific groups ... 331
9.5 Case study of Mexico 340
Introduction 340
First step (step1 analysis): detecting multidimensional poverty patterns 353
Second step (step2 analysis): detecting dimensional poverty profiles of poverty groups 384
Third step (step3 analysis): detecting various profiles of poverty groups 394
Fourth step (step4 analysis): detecting multidimensional poverty profiles of specific groups ... 398
9.6 Case study of Brazil 400
Introduction 400
First step (step1 analysis): detecting multidimensional poverty patterns 411
Second step (step2 analysis): detecting dimensional poverty profiles of poverty groups 435
Third step (step3 analysis): detecting various profiles of poverty groups 440
Fourth step (step4 analysis): detecting multidimensional poverty profiles of specific groups ... 445
9.7 Additional comparative analysis 449
Comparison between South Africa, Turkey and Brazil 451
Comparison between South Africa, Turkey and Mexico 453
Comparison between South Africa, Mexico and Brazil 455
Comparison between Turkey, Mexico and Brazil 456
Comparison between South Africa and Turkey 458
Comparison between South Africa and Brazil 459
Comparison between South Africa and Mexico 461
Comparison between Turkey and Brazil 462
Comparison between Turkey and Mexico 464
Comparison between Mexico and Brazil 465
9.8 List of national statistical agencies of select upper-middle-income countries 467
9.9 Quality report of variables, South Africa GHS 2011 468

Table of Figures

Figure 1. Data mining and its relationship with other domains. Source: Han et al. 2010 [2012, 2000], 23. 181
Figure 2. Taxonomy of data mining (from Maimon and Rokach, 2010 [2005], 5-6) 184
Figure 3. The process of prediction in supervised machine learning techniques. 187
Figure 4. An example of the basic notions of probability 191
Figure 5. An example of basic notions of probability 192
Figure 6. Probability density function 199
Figure 7. The probability density function of a normal distribution based on randomly generated 100 observations 200
Figure 8. Union of two events in the case of no intersection between them by Venn diagram 203
Figure 9. Intersection of two events shown by Venn diagram 203
Figure 10. Example of Bayes theorem with the set illustration 210
Figure 11. An example of a decision tree and a classification (don't come vs. come) based on the sample in Table 27 218
Figure 12. Basic algorithm for inducing a decision tree from training tuples. Source: Han et al. (2012 [2000], 333) 223
Figure 13. Splitting partioning scenarios. Source: Han et al. (2012 [2000], 334) 227
Figure 14. Splitting partitioning scenarios Cont'd. Source: Han et al. (2012 [2000], 334) 228
Figure 15. Classification tree example with five terminal nodes and four splitting criteria 229
Figure 16. The partitioning of the sample into five groups, corresponding to the five terminal nodes 229
Figure 17. An example of visualization of the three after the split on descriptive variable (A) yielding the best split 242
Figure 18. Iris data classification tree 245
Figure 19. CAP112 pruned tree, data: Turkey SILC 2011 278
Figure 20. CAP113 pruned tree, data: Turkey SILC 2011 282
Figure 21. CAP114 tree, data: Turkey SILC 2011 284
Figure 22. CAP116 pruned tree, data: Turkey SILC 2011 288
Figure 23. CAP117 pruned tree, data: Turkey SILC 2011 292
Figure 24. CAP119 pruned tree, data: Turkey SILC 2011 296
Figure 25. Deprivation per centages represented by different colors in terms of variables emerging in the patterns by group within each model, data: Turkey SILC 2011 305
Figure 26. Frequency and composition of classification patterns per model, data: Turkey SILC 2011. 308
Figure 27. CAP112 pruned tree, data: Mexico MCS 2012 365
Figure 28. CAP113 pruned tree, data: Mexico MCS 2012 369
Figure 29. CAP114 tree, data: Mexico MCS 2012 372
Figure 30. CAP117 pruned tree, data: Mexico MCS 2012 375
Figure 31. Deprivation per centages represented by different colors in terms of variables emerging in the patterns by group within each model, MCS Mexico 2012 382
Figure 32. Frequency and composition of classification patterns per model, Mexico MCS 2012 383
Figure 33. CAP111 pruned tree, data: Brazil PNAD 2011 421
Figure 34. CAP112 tree, data: Brazil PNAD 2011 425
Figure 35. CAP115 tree, data: Brazil PNAD 2011 428
Figure 36. Deprivation per centages represented by different colors in terms of variables emerging in the patterns by group within each model, data: Brazil PNAD 2011 433
Figure 37. Frequency and composition of classification patterns per model, data: Brazil PNAD 2011 434
Figure 38. CAP112DEC tree model, data: Brazil PNAD 2011 444

Figure 39. Benefits received by employees and domestic workers by GR1, GR2, GR3. Note: F=food, T=transport, H=health, HS=housing. Column numbers are in per centage, data: Brazil PNAD 2011 448

Table of Illustrations

Table 1. Dimensions across South Africa GHS 2011, Turkey SILC 2011, Mexico MCS 2012, and Brazil PNAD 2011 used in overall deprivation index and decomposition analysis 15
Table 2. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011 25
Table 3. Illustration 1 of ESCAP_C1 profiles among South Africa, Turkey, Mexico and Brazil 26
Table 4. Illustration 2 of ESCAP_C1 profiles among South Africa and Turkey 26
Table 5. Illustration 2 of ESCAP_C1 profiles among Mexico and Brazil 27
Table 6. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for economically active 18+ population 30
Table 8. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for elderly population (65+) 33
Table 9. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for 18-24 population 34
Table 10. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for urban areas 36
Table 11. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for rural areas 36
Table 12. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for female 18+ population 37
Table 13. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for male 18+ population 38
Table 14. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for households with a head not economically active 39
Table 15. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for households with an educationally deprived head 40
Table 16. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for households with a female head 41
Table 17. Frequency (in per centage) of decomposition groups by indicators across South Africa, Turkey, Mexico, and Brazil, from Comparison 1 44
Table 18. Poverty gap index and Dalton's transfer principle. Source: Haughton and Khandker (2009, 71) 130
Table 19. Main advantage and disadvantage of each poverty approach described in the literature review 177
Table 20. Example of data for 10 persons 193
Table 21. A fictitious example of a sample collected by a researcher for a cancer analysis 194
Table 22. Example of throwing two dices 195
Table 23. An example of probability distribution of a discrete random variable 197
Table 24. An example of classification and probabilities 201
Table 25. The effective, joint and marginal distribution of a pair of discrete random variables X, Y (extracted from Travelling data- see Table 24) 205
Table 26. The effective, joint and marginal distribution of a pair of discrete random variables Price, Satisfaction 212
Table 27. An example of sample for showing a classification tree 217
Table 28. Training set for the computations of the information gain on split 231
Table 29. Second example of computation of entropy, information gain, impurity 234
Table 30. Evaluation measures of classification trees. Source: Han et al. (2012 [2000], 365) 250
Table 31. Confusion matrix of Japanese tourists' classification 251
Table 32. Confusion matrix of domestic workers classification. 251
Table 33. True and false positives and negatives 252

Table 34. Deprivation lines and composition of variables for model 1: overall deprivation index and multidimensional poverty patterns, data: Turkey SILC 2011 268
Table 35. Ratio of unique profiles to the number of persons in the deprivation group (based on the variable DEP_GRS and 28 variables), data: Turkey SILC 2011 274
Table 36. Per centage of groups in the sample of classification target groups, data: Turkey SILC 2011 275
Table 37. Training vs. test samples in CAP112 analysis, data: Turkey SILC 2011..... 276
Table 38. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP112 pruned tree, data: Turkey SILC 2011 277
Table 39. Coverage of GR3's patterns among the total population of GR3 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP112 pruned tree, data: Turkey SILC 2011 277
Table 40. Training vs. test samples in CAP113 analysis, data: Turkey SILC 2011..... 279
Table 41. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP113 pruned tree, data: Turkey SILC 2011 281
Table 42. Coverage of GR4's patterns among the total population of GR4 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP113 pruned tree, data: Turkey SILC 2011 281
Table 43. Training vs. test samples in CAP114 analysis, data: Turkey SILC 2011..... 283
Table 44. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP114 tree, data: Turkey SILC 2011 285
Table 45. Coverage of GR5's patterns among the total population of GR5 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP114 tree, data: Turkey SILC 2011 285
Table 46. Training vs. test samples in CAP116 analysis, data: Turkey SILC 2011..... 286
Table 47. Coverage of GR2's patterns among the total population of GR2 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP116 pruned tree, data: Turkey SILC 2011 289
Table 48. Coverage of GR4's patterns among the total population of GR4 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP116 pruned tree, data: Turkey SILC 2011 289
Table 49. Training vs. test samples in CAP117 analysis, data: Turkey SILC 2011..... 290
Table 50. Coverage of GR2's patterns among the total population of GR2 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP117 pruned tree, data: Turkey SILC 2011 293
Table 51. Coverage of GR5's patterns among the total population of GR5 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP117 pruned tree, data: Turkey SILC 2011 293
Table 52. Training vs. test samples in CAP119 analysis, data: Turkey SILC 2011..... 294
Table 53. Coverage of GR3's patterns among the total population of GR3 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP119 pruned tree, data: Turkey SILC 2011 297
Table 54. Coverage of GR5's patterns among the total population of GR5 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP117 pruned tree, data: Turkey SILC 2011 297
Table 55. Main confusion matrix statistics per model, data: Turkey SILC 2011..... 309
Table 56. Dimensional profiles of GR3 based on the CAP112 model, data: Turkey SILC 2011 314
Table 57. Dimensional profiles of GR4 based on the CAP113 model, data: Turkey SILC 2011 317

Table 58. Dimensional profiles of GR5 based on the CAP114 model, data: Turkey SILC 2011 319
Table 59. Deprivation lines and composition of variables for model 1: overall deprivation index and multidimensional poverty patterns, data: Mexico MCS 2012 353
Table 60. Ratio of unique profiles to the number of persons in the deprivation group (based on the variable DEP_GRS and 32 variables), data: Mexico MCS 2012 360
Table 61. Per centage of groups in the sample of classification target groups, data: Mexico MCS 2012 361
Table 62. Training vs. test samples in CAP112 analysis, data: Mexico MCS 2012 361
Table 63. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP112 pruned tree, data: Mexico MCS 2012 363
Table 64. Coverage of GR3's patterns among the total population of GR3 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP112 pruned tree, data: Mexico MCS 2012 363
Table 65. Training vs. test samples in CAP113 analysis, data: Mexico MCS 2012 365
Table 66. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP113 pruned tree, data: Mexico MCS 2012 367
Table 67. Coverage of GR4's patterns among the total population of GR4 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP113 pruned tree, data: Mexico MCS 2012 367
Table 68. Training vs. test samples in CAP114 analysis, data: Mexico MCS 2012 369
Table 69. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP114 tree, data: Mexico MCS 2012 371
Table 70. Coverage of GR5's patterns among the total population of GR5 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP114 tree, data: Mexico MCS 2012 371
Table 71. Training vs. test samples in CAP117 analysis, data: Mexico MCS 2012 373
Table 72. Coverage of GR2's patterns among the total population of GR2 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP117 pruned tree, data: Mexico MCS 2012 375
Table 73. Coverage of GR5's patterns among the total population of GR5 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP117 pruned tree, data: Mexico MCS 2012 376
Table 74. Main confusion matrix statistics per model, data: Mexico MCS 2012 381
Table 75. Dimensional profiles of GR3 based on the CAP112 model, data: Mexico MCS 2012 388
Table 76. Dimensional profiles of GR4 based on the CAP113 model, data: Mexico MCS 2012 391
Table 77. Dimensional profiles of GR5 based on the CAP114 model, data: Mexico MCS 2012 394
Table 78. Economic activity profiles of GR3 based on the CAP112 model, data: Mexico MCS 2012 395
Table 79. Economic activity profiles of GR4 based on the CAP113 model, data: Mexico MCS 2012 396
Table 80. Economic activity profiles of GR5 based on the CAP114 model, data: Mexico MCS 2012 397
Table 81. Deprivation lines and composition of variables for step 1: overall deprivation index and multidimensional poverty patterns, data: Brazil PNAD 2011 411
Table 82. Ratio of unique profiles to the number of persons in the deprivation group (based on the variable DEP_GRS and 17 variables), data: Brazil PNAD 2011 417
Table 83. Per centage of groups in the sample of classification target groups, data: Brazil PNAD 2011 417

Table 84. Training vs. test samples in CAP111 analysis, data: Brazil PNAD 2011 418
Table 85. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP111 pruned tree, data: Brazil PNAD 2011..... 419
Table 86. Coverage of GR2's patterns among the total population of GR2 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP111 pruned tree, data: Brazil PNAD 2011..... 420
Table 87. Training vs. test samples in CAP112 analysis, data: Brazil PNAD 2011 421
Table 88. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP112 pruned tree, data: Brazil PNAD 2011..... 423
Table 89. Coverage of GR3's patterns among the total population of GR3 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP112 pruned tree, data: Brazil PNAD 2011..... 423
Table 90. Training vs. test samples in CAP115 analysis, data: Brazil PNAD 2011 425
Table 91. Coverage of GR2's patterns among the total population of GR2 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP115 tree, data: Brazil PNAD 2011..... 427
Table 92. Coverage of GR3's patterns among the total population of GR3 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP115 tree, data: Brazil PNAD 2011..... 427
Table 93. Main confusion matrix statistics per model, data: Brazil PNAD 2011 434
Table 94. Dimensional profiles of GR2 based on the CAP115 model, data: Brazil PNAD 2011 438
Table 95. Dimensional profiles of GR3 based on the CAP112 model, data: Brazil PNAD 2011 440
Table 96. Indicators used in overall deprivation index and decomposition analysis (except the main variables used in Fourth Model, individual characteristics and household composition related variables) of 4 country cases: South Africa GHS 2011, Turkey SILC 2011, Mexico MCS 2012, and Brazil PNAD 2011. 450
Table 97. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, and BR PNAD 2011..... 452
Table 98. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, and MX MCS 2012 453
Table 99. Essential capabilities and material situation profiles across SA GHS 2011, MX MCS 2012, and BR PNAD 2011..... 455
Table 100. Essential capabilities and material situation profiles across TR SILC 2011, MX MCS 2012, and BR PNAD 2011..... 457
Table 101. Essential capabilities and material situation profiles across SA GHS 2011, and TR SILC 2011..... 458
Table 102. Essential capabilities and material situation profiles across SA GHS 2011, and BR PNAD 2011..... 459
Table 103. Essential capabilities and material situation profiles across SA GHS 2011, and MX MCS 2012..... 461
Table 104. Essential capabilities and material situation profiles across TR SILC 2011, and BR PNAD 2011..... 462
Table 105. Essential capabilities and material situation profiles across TR SILC 2011, and MX MCS 2012..... 464
Table 106. Essential capabilities and material situation profiles across MX MCS 2012, and BR PNAD 2011..... 465
Table 107. List of national statistical agencies of selected upper-middle income countries (accessed 8 January 2020) 467

6. Comparative analysis

In this chapter, we will start off by providing a review on technical issues should be taken in consideration while doing a comparative multidimensional poverty analysis. In light of this information, we will compare the results of our analysis across the 4 countries, Brazil, Mexico, South Africa, and Turkey.

6.1 Introduction

An accurate comparison between two or more countries in case of a multidimensional object requires fulfillment of many steps. First of all, comparison should be based on the same indicators. Dimensions are generally used for indicating the general theme of the indicator; for instance, if for the same indicator the two different methodologies employ different dimensions it won't create a problem for comparison regarding the indicator. However, in case of a multidimensional index or another measure including different indicators takes in consideration weighting, then the determination of the dimension for the indicator will be crucial. Suppose that an indicator is used in two country analysis but in one analysis the indicator is considered in a dimension X and in other analysis it is considered in a dimension Y. The dimension X has 4 indicators and the dimension Y has 2 indicators. Regarding nested-weighting application, both the dimensions X and Y will be associated with a weight of 0.5 (50%), and the indicator in question will be associated with a weight of 0.25 (25%) in one analysis based on the dimension Y vs. weight of 0.125 (12.5%) in other analysis based on the dimension X. This simple example shows that the same indicator has not the same importance in two analysis. In addition to weight issue, there are also a number of technical points should be considered in a comparison. In a broad sense, for being able to compare the two outputs coming from the two-country analysis, the same methodology in computing the output should be applied in both examinations. Depending on the subject of the analysis, methodological points may vary. Poverty from a multidimensional perspective is not always measured in a same way; for instance, there is a usage of composite index for which there is not only one method-different methodologies exist for building a composite index, or a dashboard can be employed as well. As we explained in multidimensional poverty section in the theoretical framework, a dashboard is characterized by containing several indicators and the comparison between two countries is realized per each indicator; in other words, if a dashboard contains 12 indicators then the comparison between countries will be done 12 times. Dashboard based comparison seems easier than comparison via composite indices. Even in such case, it is extremely important to look at carefully how indicators are created. An indicator is created from one or more questions existing in the survey questionnaire. Survey questionnaires may be different from one country to another. Even if the general structure of questionnaires remains the same, the questions may be asked in different manner. Each question contains a subject, answer options/categories and most importantly a target group. In general household or labour force surveys, economic activity questions are directly asked to individuals aged at least 18 years old. Or the material situation of a household may contain different items from one survey to another, or a dwelling roof material question may include different answer options/categories from one survey to another. It is extremely important to understand very well the questionnaire and the way the questions are asked; this is the basis for any multidimensional or social analysis. But also, the way that the indicator is created becomes essential. If more than one indicator is used for creating a composite index, then which indicators are exactly used and how they are combined should be understood in detail.

On the other hand, in case of composite index, the methodology becomes more complex compared to dashboard because the former does not only contain at least two indicators but most importantly a mathematical function is used for compounding the scores of the indicators to obtain a single (composite) score-for instance, in case of two metric indicators, the average function may be used as a way of compounding the two indicators (i.e., the average of two indicators as a composite result). More specifically, these issues should be considered when the results of two different composite indices are compared: (i) definition of the indicators, survey questions behind them and the target group they cover; (ii) in which dimensions, the indicators are considered and the weight associated with each indicator; (iii) the unit of analysis in the sense that whether the indicator addresses the same unit-for instance, if the indicator concerns a household or an individual; (iv) the mathematical function used for compounding together the scores of indicators; (v) if there are some thresholds used as a benchmark for evaluation, then when they are used in the methodology and their criteria; (vi) adjustment of the indicator with another indicator or a score (e.g., adjusted headcount ratio-adjustment of the incidence of poverty by its intensity); and (vi) whether the score of the index is decomposed by specific units such as setting (urban vs. rural, male vs. female, age groups). Even if we look at the statistical details, the survey sampling methodology used for building the sample of the survey emerges as one of the fundamental elements. When we use national surveys prepared and carried out by national statistical offices of countries, as a social analysist we don't generally examine the details of the sampling methods behind the surveys by just assuming that the national surveys represent adequately the country. Moreover, in analyzing the data, especially in case of multiple indicators, some records/observations may emerge as missing values. In other words, some records in the data may not have the answers for some questions (e.g., someone prefers not giving answer to a question or, he or she does not know the answer). In case of dropping some records from the data because of missing values, then it is important to check the per centage of the loss or the per centage of the initial data to be used for the analysis after removing missing values. For preserving the representativity of the sample, the per centage of missing values should be remained at an optimal level. As a solution to missing values, several imputation options are proposed in the literature; nevertheless, imputations cannot bring the exact answer and the imputed answers should be verified by the national statistical office for data dissemination issues or the already existing imputed answers by the national statistical office should be used for the analysis.

In our four case studies, we applied the same steps however as the availability of indicators vary from one survey to another, we cannot directly compare the results of Brazil, Mexico, South Africa, and Turkey. The two tables below show the dimensions and indicators used in 4 country case studies. In Table 1 below, first column represents dimensions with their short name in parenthesis. The code 'yes' is used to mean that the dimension is employed in the analysis, otherwise no. For instance, the dimension of social assistance exists in South Africa case study based on GHS 2011; but it does not exist in Turkey, Mexico, and Brazil case studies.

Table 1. Dimensions across South Africa GHS 2011, Turkey SILC 2011, Mexico MCS 2012, and Brazil PNAD 2011 used in overall deprivation index and decomposition analysis

Dimensions	**South Africa GHS 2011**	**Turkey SILC 2011**	**Mexico MCS 2012**	**Brazil PNAD 2011**
Dwelling situation (DWS)	Yes	Yes	Yes	Yes
Material situation (MS)	Yes	Yes	Yes	Yes
Educational attainment (EDUC)	Yes	Yes	Yes	Yes
Social assistance (SD)	Yes	*No*	*No*	*No*
Environment (ENV)	Yes	Yes	*No*	*No*
Health (HEALTH)	Yes	Yes	Yes	No
Economic activity (ECOA)	Yes	Yes	Yes	Yes
Financial situation (FS)	*No*	Yes	*No*	*No*
Nutrition (NUT)	Yes	*No*	Yes	*No*

Note: The dimensions of social assistance, and economic activity are used for decomposition analysis, and the others are part of the overall deprivation index and dimensional analysis[1].

As we can see from the table above, the dimensions of dwelling situation, material situation, educational attainment, and economic activity exist all of them in the four-country analysis. The dimension of social assistance exists only in South Africa, the dimension of environment exists only in South Africa and Turkey, the dimension of health does not exist only in Brazil, the dimension of financial situation exists only in Turkey, and the dimension of nutrition exists only in South Africa and Mexico. The first table in Appendix 9.7 shows the indicators per national survey. First column shows the indicators with their code names, second column shows the dimensions with their code names, third column indicates whether the indicator is used for the construction of the overall deprivation index (step1 and step2 analysis) or for decomposition analysis (step3 analysis). The last four columns concern the surveys of South Africa, Turkey, Mexico, and Brazil respectively. The code 'yes' is used to mention that the indicator is used for the case study; for instance, the indicator DWE_D is only used in South Africa case study based

[1] We note that not all dimensions mentioned in section 4.2 (see the indicators tables for countries) are used for constructing overall deprivation index. Some dimensions are used for building the overall deprivation index and therefore for the step1 analysis; otherwise, the other dimensions are used for decomposition analysis, which means for the step3 analysis (e.g., social assistance in case of South Africa). In addition, dimensions employed in overall deprivation index (step1 analysis) are also included in step2 analysis, which aims at focusing more specifically on the most relevant profiles within each dimension of the overall deprivation index.

on GHS 2011 but not in Turkey, Mexico, Brazil case studies. We also note that if the code name of the indicator includes '_D' this means that the indicator is used in overall deprivation index (D means deprivation).

The dimensions of dwelling situation, material situation, educational attainment, environment, health, financial situation, and nutrition are used in the construction of the overall deprivation index. However, the dimensions of social assistance and economic activity are used for decomposition analysis as the target group for these two dimensions is not the same compared to those of other dimensions. Not everyone among 18+ population receive a social grant and we cannot really consider deprived someone who received a grant because there might be other persons wishing receiving a grant due to need but not receiving it actually or someone may receive a grant even if he or she does not really need it. In addition, not everyone among 18+ population is actually employed and generally the rate of unemployment emerges quite low when we consider the questions of seeking a job and availability during the reference week. Underemployment or discouraged workers represent an important part of population in developing countries. Moreover, dimensions of household composition, geographic situation, and individual characteristics are used for decomposition analysis for bringing supplementary information. Concerning geographic situation, it does not only exist in Mexico MCS 2012.

We point out that it is not possible to do directly a comparison between 4 countries based on the overall deprivation index scores due to different indicators (see first table in Appendix 9.7). By taking in consideration the common variables, we can observe that the following comparisons between the countries are possible (country codes are mentioned in brackets, and the list of indicators can be used for the comparison is given with indicators' codes):

- {SA, TR, BR, MX}: SAN_D, WATD_D, TV_D, REF_D, TEL_D, MPHO_D, EDU_D, ES1

For instance, for a comparison between South Africa, Turkey, Brazil, and Mexico by using GHS 2011, SILC 2011, PNAD 2011, and MCS 2012 surveys, the indicators of SAN_D, WATD_D, TV_D, REF_D, TEL_D, MPHO_D, EDU_D, and ES1 can be used.

- {SA, TR}: ROFC_D, WALC_D, RO_D, KIT_D, SAN_D, WATD_D, ENG_H_D, TV_D, REF_D, TEL_D, MPHO_D, EDU_D, NOI_D, ES1

- {SA, BR}: ROF_D, WAL_D, RO_D, SAN_D, WATD_D, ENG_C_D, ENG_L_D, TV_D, RAD_D, REF_D, TEL_D, EDU_D, RUB_D, ES1

- {SA, MX}: ROF_D, WAL_D, KIT_D, SAN_D, WATD_D, ELC_D, ENG_C_D, TV_D, RAD_D, REF_D, TEL_D, INT_D, MPHO_D, EDU_D, RUB_D, MC_D, ES1

- {TR, BR}: RO_D, SAN_D, WATD_D, TV_D, REF_D, TEL_D, EDU, ES1, WM_D, CAR_D, COMPINT_D, MPHO_D, ES3, WH, SC

- {TR, MX}: KIT_D, SAN_D, WATD_D, TV_D, REF_D, TEL_D, EDU_D, ES1_D, HWS_D, HEATS_D, AIRC_D, WM_D, CAR_D, COMPINT_D, MPHO_D

- {BR, MX}: ROF_D, WAL_D, SAN_D, WATD_D, ENG_C_D, TV_D, RAD_D, REF_D, TEL_D, EDU_D, RUB_D, ES1, WM_D, CAR_D, COMPINT_D, MPHO_D, DR_D, JOB

- {SA, TR, BR}: RO_D, SAN_D, WATD_D, TV_D, REF_D, TEL_D, MPHO_D, EDU_D, ES1
- {SA, TR, MX}: KIT_D, SAN_D, WATD_D, TV_D, REF_D, TEL_D, MPHO_D, EDU_D, ES1
- {SA, BR, MX}: ROF_D, WAL_D, SAN_D, WATD_D, ENG_C_D, TV_D, RAD_D, REF_D, TEL_D, MPHO_D, EDU_D, RUB_D, ES1

- {TR, BR, MX}: SAN_D, WATD_D, TV_D, REF_D, TEL_D, EDU_D, ES1, WM_D, CAR_D, COMPINT_D, MPHO_D

After detected the common indicators, it is extremely important to check the content of indicators or in other words, what is the specific object of the indicator. We will see that the theme of an indicator may be the same for two countries, but the content of the indicator may vary. Now, we will focus on this issue to finalize which indicators are suitable for which comparison. For that, we will list the common indicators expressed in possible comparison (see above) and we will explain the situation of comparison in terms of their content.

- *We start off with the indicators of the dimension of dwelling situation (DWS).*

SAN: In SA GHS, sanitation facilities were asked only in terms of toilet. In TR SILC, sanitation facilities were asked both in terms of bathroom/shower (SAN1) and toilet (SAN2). In MX MCS, both SAN1 and SAN2 were asked. In TR SILC, it is asked whether *bathroom/shower* exist or not in the dwelling, or if it is shared with other dwellings whilst in MX MCS, it is asked whether *bathroom/shower* exist or not in the dwelling without taking in consideration the question of share between different dwellings. Under this condition, we can say that deprivation in terms of bathroom/shower risks to be underestimated in case of substantial number of shared bathroom/showers. On the other hand, concerning *toilet*, in MX MCS it is asked whether it exists or not in the dwelling without taking in consideration the question of share, in TR SILC the question of toilet is asked in a same way of bathroom/shower question, in BR PNAD sanitation facilities question is asked whether a toilet or a bathroom exist in the dwelling or if it is shared by other dwellings, and in SA GHS it is asked whether there is a toilet in the dwelling or if it is shared and whether toilet is chemical, pit latrine or bucket. The type of toilet is only asked in case of SA GHS. In these conditions, we can say that BR PNAD cannot be used for a comparison regarding both toilet and bathroom/shower because it is not possible to understand which one exists or not in the dwelling. Moreover, if we include SA GHS then it risks having a higher deprivation rate in terms of toilet compared to TR and MX case as the type of toilet is used for deprivation issue. Herewith, two comparisons are possible: **(i) whether a shower/bathroom exists or not in the dwelling without taking in consideration the question of sharing, in case of TR SILC and MX MCS; (ii) whether a toilet exists or not in the dwelling without taking in consideration the question of sharing and material, in case of TR SILC, MX MCS, and SA GHS.**

WATD: In SA GHS, 7 different topics related to drinking water are covered: whether the water comes from the piped (tap) water system, whether the source of water is inside the dwelling, whether water is safe to drink, whether water is clear, whether water is good in taste, whether water is free from bad smells, whether water s treated by the household members before using

it. In BR PNAD, whether the water comes from the piped (tap) water system, whether the source of water is inside the dwelling are asked. In TR SILC, whether a piped water system in the dwelling exists is asked. And in MX MCS, whether a piped water system in the dwelling exists is asked but also whether the water is available daily or not. In these conditions, we can observe that this comparison is possible: **whether there is a piped water system exists in the dwelling, in case of BR PNAD, TR SILC, MX MCS, and SA GHS.**

ROFC: In SA GHS, it is asked whether roof condition of the dwelling is very weak, weak, need minor repairs, good, or very good. In TR SILC, it is asked whether there are any problems regarding roof, wall and window. And there is no question on that in BR PNAD and MX MCS. As the item of window is considered in the question in TR SILC, there is a risk of overestimation of deprivation compared to SA GHS. **But also, answer options are not the same in two surveys for instance, for someone from SA GHS may find no problem if roof condition needs minor repairs or if it is weak. In these conditions, it is not accurate do a comparison for that indicator.**

WALC: **Same comment as ROFC.**

ROF: In SA GHS, the following answer options are used: bricks, cement block/concrete, corrugate iron/zinc, wood, plastic, cardboard, mixture of mud and dement, wattle and daub, tile, mud, thatching, asbestos. In BR PNAD, the answer options of tile, stone, slab, rigged wood, zinc, harnessed wood, straw are used. And in MX MCS, the answer options of waste material, cardboard sheet, sheet metal, asbestos, palm or straw, wood, rooftop with beams, floor framing, brick red, specific floor or joists with small vault are used. **We can see that the answer options are not exactly the same, which makes not adequate the comparison** (except the comparison may be done for common options).

WAL: In SA GHS, the following answer options are used: bricks, cement block/concrete, corrugate iron/zinc, wood, plastic, cardboard, mixture of mud and dement, wattle and daub, tile, mud, thatching, asbestos. In BR PNAD, the answer options of masonry, rigged wood, uncoated wall of mud, harnessed wood, straw are used. In MX MCS, the answer options of waster material, cardboard sheet, asbestos or metal foil, reed, bamboo or palm, wattle, daub, wood, adobe, partition wall, brick, block, stone, quarry, cement, concrete are used. **We can see that the answer options are not exactly the same, which makes not adequate the comparison** (except the comparison may be done for common options).

RO: In SA GHS, BR PNAD, and in TR SILC, the number of rooms per person in the dwelling includes bedroom, dining room, sitting room, TV room, open plan dining/sitting/TV room and exclude kitchen, bathroom, and toilet. **This variable can be used for a comparison between SA GHS, BR PNAD, and TR SILC** (given that the variable is not computed in MX MCS because it was not clear if it includes kitchen, bathroom, toilet).

KIT: In SA GHS, it is asked whether a kitchen exists in the dwelling, and in TR SILC and MX MCS it is asked whether an independent kitchen exists in the dwelling (and there is no question on kitchen in BR PNAD). The definition of an independent kitchen was not given in questionnaires, but we can suppose that an independent kitchen is different than a kitchenette or a kitchen corner in a room. We cannot really know whether people living in a dwelling having a kitchenette or a small kitchen corner inside a main room (e.g., studio) or a dining room or a TV room or a sitting room are answered no to the question of independent kitchen in TR SILC

and MX MCS. It is a very subtle point. Nevertheless, **we will employ this indicator in comparison of SA GHS, TR SILC, and MX MCS.**

ENG_H: In SA GHS, the following answer options are used: no energy, electricity from mains, electricity from generator, gas, paraffin, wood, coal, candles, animal dung, solar energy. Whilst, in TR SILC, the answer options of wood, coal, natural gas, fuel-oil, diesel oil (gasoil), electricity, dried cow dung are used (of course, we can also compute if there is no energy). **We can see that the answer options are not exactly the same, which makes not adequate the comparison (**except the per centage of people using wood, coal that exist in both surveys can be compared).

ENG_C: In SA GHS, the following answer options are used: no energy, electricity from mains, electricity from generator, gas, paraffin, wood, coal, candles, animal dung, solar energy. In MX MCS, the answer options of firewood, coal, tank gas, natural gas or pipe, electricity are used. In BR PNAD, the answer options of piped gas, electricity, gas cylinder, firewood, coal are used. **We can see that the answer options are not exactly the same, which makes not adequate the comparison** (except the per centage of people using wood, coal that exist in both surveys can be compared).

ENG_L: In SA GHS, the following answer options are used: no energy, electricity from mains, electricity from generator, gas, paraffin, wood, coal, candles, animal dung, solar energy Whilst, in BR PNAD, the answer options of electric, canister oil, kerosene, gas are used. **We can see that the answer options are not exactly the same, which makes not adequate the comparison.**

ELC: In SA GHS, it is asked if the household has no connection to the mains electricity supply. In MX MCS, it is asked how the electric in the dwelling is obtained via public service, or particular plant, or solar panel, or another source, or there is no electricity. The structure of these two questions is not the same. In SA GHS, the source of electricity is directly related to the mains electricity supply. In South Africa, state provides free basic electricity to poor households and generally electricity is supplied via general network called mains electricity supply. **By focusing on the fact whether a dwelling has an access or not to electricity no matter its source, the comparison is possible between SA GHS and MX MCS** (given that per centages of people using a particular plant and a solar panel are relatively low). **Or the per centage of people using public service for electricity in Mexico can be compared with those answered yes to the mains electricity supply question in South Africa case.**

HWS: In TR SILC, whether hot water system exists or not in the dwelling is asked without specifying the source. In MX MCS, whether a boiler/hot water system (calentador de agua) exist or not in the dwelling is asked. Calentador de agua is a specific tool for obtaining hot water in the dwelling. **As TR SILC does not specific a source for obtaining hot water, then we can say that a proper comparison cannot be done in this regard**.

HEATS: In TR SILC, what is the heating system in the dwelling is asked and the answer options are: stove (coal, gas, natural gas, electricity, etc.), radiator (joint or central heating), radiator (heating system for only a flat/combi boiler), air conditioner. In MX MCS, whether a heating system (caleffaccion) exists or not in the dwelling is asked. **The definition of caleffaccion is not very clear in the questionnaire-if it does mean radiator or air**

conditioner or include also stove (coal, gas.etc.). For that reason, it is not adequate the variable HEATS for a comparison between TR SILC and MX MCS.

RUB: In SA GHS, it is asked how is the refuse or rubbish of this household collected or removed and the answer options are: (1) removed by local authority/private company at least once a week; (2) removed by local authority/private company less often than once a week; (3) removed by community members, contracted by the municipality at least once a week; (4) removed by community members, contracted by the municipality less often than once a week; (5) removed by community members at least once a week; (6) removed by community members less often than once a week; (7) communal refuse dump/communal container; (8) own refuse dump; (9) dump or leave rubbish anywhere. In MX MCS, same question is asked with these answer options: (1) picked up by a truck or garbage cart; (2) thrown in the public dump; (3) thrown in a container or deposit; (4) burn; (5) buried; (6) thrown in a field wasteland or street; (7) thrown in the ravine or crevice thrown into the river, lake or sea. In BR PNAD, same question is asked with these answer options: (1) collected directly; (2) indirectly collected; (3) burnt or buried in property; (4) thrown on wasteland or patio; (5) thrown in river, lake, or sea. **We can see that the answer options are not exactly the same, which makes not adequate the comparison.**

DR: In MX MCS, the question of which of the following drainage system is used in the dwelling is asked with the following answer options: connection to public network, a sceptic tank, a pipe connected to a ravine or crevice, a pipe connected to a river, lake or sea, or no drainage. In BR PNAD, the following answer options are used: sewer or pluvial collection network, septic tank connected to the collecting network sewage or rain, septic tank not collected to the collecting network sewage or rain, rudimentary pit, ditch, directed to the river, lake or sea. The answer options are not the same however they are extremely similar. **But the main problem is in BR PNAD, drainage system is only asked for bathroom/toilet, and does not include kitchen. In these conditions, the comparison between BR PNAD and MX MCS cannot be done.**

- *The indicators of the dimension of material situation (MS)[2]:*

TV: It is asked whether the household owns a television. **The comparison between 4 countries can be done.**

REF: It is asked whether the household owns a refrigerator. **The comparison between 4 countries can be done.**

TEL: It is asked whether the household owns a landline telephone. **The comparison between 4 countries can be done.**

RAD: It is asked whether the household owns a radio. **The comparison between SA GHS, BR PNAD, and MX MCS can be done.**

[2] In TR SILC case, material situation related questions are asked in a way that whether the household cannot afford the material item due to financial reason/financial shortage, or the household does not own the item for another reason. In SA GHS, MX MCS, and BR PNAD, material situation related questions are asked not including financial aspect of the household but only asking whether the household own the item or not. So that, for making enable comparison between countries, we consider deprived the household no matter it is due to financial reason or not if she does not own the item in case of TR SILC.

INT: In MX MSC, it is asked whether the household owns an internet connection. In SA GHS, several answer options are used including internet connection in the household, internet in a library or at a school/university. However, **by just focusing on internet connection in the household, the comparison between MX MCS and SA GHS can be done.**

WM: It is asked whether the household owns a washing machine. T**he comparison between TR SILC, MX MCS, and BR PNAD can be done.**

CAR: It is asked whether the household owns a car. **The comparison between TR SILC, MX MCS, BR PNAD can be done.**

COMPINT: In SA GHS there is no question on computer, there is only a question for access to internet. In MX MCS, it is asked whether the household owns a computer. In TR SILC and BR PNAD, the question of computer is combined with access to internet. In TR SILC, it is asked whether the household having a computer have also an internet connection. The internet question then was not asked to people having not a computer in their household. For that reason, it is not convenient to compare TR SILC in terms of access to internet in the dwelling for the household, but it is possible concerning whether the household owns a computer. Similarly, in BR PNAD, it is asked whether the household own a computer and an internet for the computer in the dwelling. So that, BR PNAD is not also suitable for comparison in terms of internet as it is only asked to people having a computer in their dwelling. Thus, **the comparison in terms of computer will be done between MC MCS, BR PNAD, and TR SILC.** An internet available in the dwelling where the household lives may be used not only by a computer but also by other tools such as mobile phone, or tablet; for that reason, the way it is asked in BR PNAD and TR SILC is not very practical.

MPHO: It is asked whether the household owns a mobile/cellular telephone. **The comparison between 4 countries can be done.**

AIRC: It is asked whether the household owns an air conditioner. T**he comparison between TR SILC and MX MCS can be done.**

- *The indicators of the dimension of education (EDUC):*

EDU: The highest educational attainment is asked in 4 country cases. Even if the answer options are different, **by taking in consideration 12 grade (equivalent of a high school degree) as a threshold (i.e., individual not having at least 12 grades is considered deprived in terms of educational attainment), the comparison can be done.**

- *The indicators of the dimension of economic activity (ECOA):*

ES1: This indicator shows the current economic activity status of persons. In some data, it might be that the variable is already computer by the national statistical office or in other cases, it can be computed following the ILO criteria based on the available employment, seeking job, and availability for working during the reference week questions. In TR SILC, detailed employment and unemployment questions were available then we constructed ES1 based on them regarding the ILO criteria. It is the same for SA GHS given that availability for working question has different answer options. In BR PNAD, the question of availability for working and seeking a job are not asked, then we have only information for being active vs. not being active on the labour market but no information for unemployment status. In MX MCS, the ES1 variable already exist in the data including different answer options such as employed, employed but

temporarily absent, unemployed, retired etc. **Unemployment comparison can be done between TR SILC, SA GHS, and MX MCS. And, employment comparison can be done between the 4 country cases.**

ES3: In TR SILC and BR PNAD, the question of individual's employment status in the main job is asked. However, the answer options are not the same. For instance, the option of 'worker in production for own consumption' and 'construction worker for own use' exist in BR PNAD but not in TR SILC. **In these conditions, it is not adequate a comparison between BR PNAD and TR SILC regarding ES3.**

WH: Working hours per week indicator is available in TR SILC and BR PNAD. In TR SILC, we computed the indicator based on the metric question on working hours and ILO criterion. However, in BR PNAD, the indicator is already available in the data provided by the national statistical institute but composed with different working hours categories. **Thus, the comparison between TR SILC and BR PNAD is not possible regarding WH.**

SC: In BR PNAD, it is asked whether the person contributes to the social security due to her/his main job during the reference week. In TR SILC, it is asked whether the individual is registered or not to the national social security system due to her/his main job during the reference week. The social security system may be different from one country to another including different types of benefits. However, it is very important to know if an economically active person is registered or not to national security system. From this broad perspective, **TR SILC and BR PNAD can be compared in terms of SC.**

JOB: In MX MCS, it is asked how many jobs the person had last month (one job vs. more than one job), given that the reference week is last month. In BR PAND, it is asked whether the person has had more than one job during the reference week (one job, two job, more than two jobs). In these conditions, **this indicator can be used for a comparison between BR PNAD and MX MCS in terms of one job vs. more than one job for economically active population during the reference period of the survey.**

- *The indicators of the dimension of health (HEALTH):*

MC: In SA GHS, it is asked whether the person is covered by a medical aid or a medical benefit scheme or other private health insurance. In MX MCS, it is asked whether the person is affiliated or enrolled to receive medical care from any institution, including Popular Insurance. **Without focusing on the type of institution, the comparison can be done between SA GHS and MX MCS.**

- *The indicators of the dimension of environment (ENV):*

NOI: In SA GHS, it is asked whether the person experience excessive noise/noise pollution in her/his community or neighborhood. In TR SILC, it is asked whether there is noise in the dwelling due to neighbors or streets. **These two questions are very similar even if the way it is asked is not exactly the same, the comparison can be done between TR SILC and SA GHS.**

- *The indicators of the dimension of nutrition (NUT):*

In terms of nutrition/food situation of the household, in GHS SA the indicators of CHN, AND, NU1, NU2, NU3, and NU4 are created and in MX MCS the indicators of AF1, AF2, and AF3

are built. The variables AF1, AF2, AF3 focus on the fact that whether the household has experienced deprivation regarding access to food (AF1: run out of food, AF2: worry about the food running, AF3: run out of money or resources to get a healthy and varied diet in the household) *due to lack of money or resources*. The indicators ADN, and CHN ask whether any child and adult went hungry *because there was not enough food in the household*. Comparison by employing these indicators is not possible given that in MX MCS case accent is on financial aspect whilst in SA GHS there is not a specific precision on financial aspect. **The only possible comparison may be done by using AF1 from MX MCS and ADN, CHN from SA GHS supposing that run out food means go hungry and due to lack of money or resources a household runs out of food.**

Besides, the other indicators NU1, NU2, NU3, and NU4 from SA GHS, focus on whether the household did run out of money to buy food, did the household cut the size of meals because there was not enough food, did the household skip any meals because there was not enough food, and did someone in the household eat a smaller variety of food because there was not enough food in the house. **In this context, only option for a comparison between SA GHS and MX MCS is using the indicators NU1 from SA GHS and AF1 from MX MCS.**

Moreover, in TR SILC, the question of "at least 3 days a week, when required or needed, is the household able to afford a meal with meat, chicken, fish (equivalent food for vegetarians)?" is asked. However, this question does not focus on go hungry or run out of food but it is more related to a certain type of variety of food or a certain type of a healthy diet but still it cannot be used for a comparison with SA GHS in terms of NU4 neither with MX MCS in terms of AF3 because NU4 and AF3 does not precisely include the amount of meat, chicken, fish or vegetarian equivalent.

In light of the information mentioned previously, we can summarize that the following comparisons can be done- country names and deprivation items are indicated with their codes.

- {SA, TR, BR, MX}: WATD, TV, REF, TEL, MPHO, EDU, ES1
- {SA, TR}: RO, KIT, SAN, WATD, TV, REF, TEL, MPHO, EDU, NOI, ES1
- {SA, BR}: RO, SAN, WATD, TV, RAD, REF, TEL, EDU, ES1
- {SA, MX}: KIT, SAN, WATD, ELC, TV, RAD, REF, TEL, INT, MPHO, EDU, MC, ES1
- {TR, BR}: RO, WATD, TV, REF, TEL, EDU, ES1, WM, CAR, COMPINT, MPHO, SC
- {TR, MX}: KIT, SAN, WATD, TV, REF, TEL, EDU, ES1, AIRC, WM, CAR, COMPINT, MPHO
- {BR, MX}: WATD, TV, RAD, REF, TEL, EDU, ES1, WM, CAR, COMPINT, MPHO, DR, JOB
- {SA, TR, BR}: RO, WATD, TV, REF, TEL, MPHO, EDU, ES1
- {SA, TR, MX}: KIT, SAN, WATD, TV, REF, TEL, MPHO, EDU, ES1

- {SA, BR, MX}: WATD, TV, RAD, REF, TEL, MPHO, EDU, ES1

- {TR, BR, MX}: WATD, TV, REF, TEL, EDU, ES1, WM, CAR, COMPINT, MPHO

The analysis will be realized in 3 steps: (i) first, we will find out the essential capabilities profile within each comparison model based on the dimensions of dwelling situation, education, nutrition, health, and environment; (ii) second, we will find out the material situation profile within each comparison model; (iii) third, we will continue with some decomposition analysis including economic activity situation, elderly people, youth population, setting (urban vs. rural), gender (male vs. female), households headed by a not economically active person, households headed by educationally deprived person, and households headed by a female. We note that decomposition analysis will be only done for the first comparison model, which includes the 4 countries at the same time.

The essential capabilities profile is the combination of the available common variables from the dimensions of dwelling situation, education, nutrition, health, and environment respectively. Each indicator is binary, taking the value of 0 for not being deprived and the value of 1 for being deprived regarding the variable. Tables below show the distribution of the most relevant profiles (with at least 100 observations) among the countries in the comparison model. Frequency (in per centage) of each profile in the data is shown in parenthesis next to the profile.

We remark that all figures shown in tables in following pages are non-weighted.

And we note that due to the length of the thesis, the main comparison (comparison between the four countries at the same time) is explained here. The other comparisons are in Appendix 9.7.

6.2 Comparison between South Africa, Turkey, Mexico, and Brazil

First comparison model includes the four countries, South Africa, Turkey, Mexico, and Brazil. Taking in consideration the common variables (see above), the essential capabilities profile is the combination of the variables WATD and EDU, respectively (ESCAP_C1= {WATD, EDU}). On the other hand, the material situation profile is the combination of the variables TV, REF, TEL, and MPHO respectively (MSCAP_C1= {TV, REF, TEL, MPHO}). And, the complete profile is composed of WATD, EDU, TV, REF, TEL, MPHO, respectively (CAP_C1= {WATD, EDU, TV, REF, TEL, MPHO}).

Each variable represents a functioning such as WATD means access to drinking water. The outputs of functionings are *deprivation* or *non-deprivation*. The situation of deprivation is coded 1 and the situation of non-deprivation is coded 0. For instance, for ESCAP_C1, the profile {11} corresponds to deprivation regarding both WATD and EDU. Extracted profiles reflect (i) combination of deprivations regarding all functionings at the same time ('Total poverty') such as {11} in case of two functionings, or (ii) combination of non-deprivations regarding all functionings at the same time ('Total prosperity' or 'Zero poverty') such as {00}in case of two functionings, or (iii) combinations of deprivation and non-deprivation regarding the variables ('In-between poverty') such as {10}, {1,0} in case of two functionings. These combinations demonstrate different configurations of poverty. In other words, the discovered configurations of poverty inform about the capabilities patterns.

Table 2. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011

	South Africa (SA GHS 2011) 5.8% missing values	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C1)	11 (43.98), 01 (21.76), 00 (21.72), 10 (12.54) Total info=100%	01 (70), 00 (27.4), 11 (2.5) Total info=100%	01 (37.83), 00 (31.92), 11 (25.11), 10 (5.14) Total info=100%	01 (46.11), 00 (35.92), 11 (13.58), 10 (4.39) Total info=100%
Material situation profiles (MSCAP_C1)	0010 (58.24), 0000 (14.86), 1110 (10.56), 0110 (6.51), 1010 (3.22), 1111 (2.5), 0011 (2.18) Total info=98%	0000 (51.4), 0010 (42), 0001 (3.13) Total info=97%	0010 (33.9), 0000 (27.9), 0011 (10.85), 0001 (10.75), 0111 (4.77), 0110 (4.76), 1111 (2.9) Total info=96%	0010 (48), 0000 (38.14), 0011 (6.1), 0001 (2.26) Total info=95%
Complete profiles (CAP_C1)	110010 (23.99), 010010 (13.5), 000010 (11.64), 100010 (9.10), 000000 (9.04), 111110 (8.57), 010000 (4.86), 110110 (4.65) Total info=85.37%	010000 (33.33), 010010 (31.2), 000000 (17.85), 000010 (9.3), 010001 (3) Total info=94.5%	000000 (16.5), 010010 (14), 000010 (9.7), 010000 (9.6), 110010 (7.84), 010001 (5.66), 110011 (4.77), 010011 (4.36), 110111 (3.23), 000001 (3.07), 111111 (2.37), 100010 (2.32), 110110 (2.27) Total info=86%	010010 (22.5), 000000 (18.85), 010000 (16.89), 000010 (15.47), 110010 (7.4), 010011 (3), 100010 (2.7), 110011 (2.2) Total info=89%

For illustrative purpose, the figure below (called *Illustration 1 of ESCAP_C1 profiles*) can be used (e.g., for social reports to inform public and nonexperts) to display the frequency of the four ESCAP_C1 profiles among the four countries. This figure shows the *ranking of the countries per each relevant combination of two functionings* (WATD and EDU) in terms of deprivation and non-deprivation. The first column indicates deprivation or non-deprivation regarding the access to drinking water, the second column indicates deprivation or non-deprivation regarding educational attainment, and the last column shows the incidence of profiles in each country (in terms of per centage). The first row represents deprivation in both drinking water and educational attainment and the countries are ranked from the worst to the best in last column of this first row. For instance, 43.98 per cent of 18+ population in South Africa experience deprivation regarding both access to drinking water (WATD) and educational attainment (EDU). The second row represents non-deprivation in drinking water and deprivation in educational attainment and the countries are ranked from the worst to the best in last column of this second row. The third row represents deprivation in drinking water and non-deprivation in educational attainment and the countries are ranked from the worst to the best in last column of this third row. The fourth row represents non-deprivation in both drinking water and educational attainment and the countries are ranked from the best to the worst in last column of this fourth row. The light color means non-deprivation and the dark color means deprivation

regarding the indicator. And, SA is for South Africa, TR is for Turkey, MX is for Mexico, BR is for Brazil.

The other two figures below (called *Illustration 2 of ESCAP_C1 profiles*) display the *country profiles* in terms of frequency of the four ESCAP_C1 profiles in South Africa, Turkey, Mexico, Brazil as a function of their relevance (i.e., from the most occurred to the least occurred profiles in the country). For instance, in South Africa the profile of deprivation in both drinking water and educational attainment is the most occurred with 43.98% whilst in Turkey the profile of deprivation in educational attainment and non-deprivation in drinking water is the most occurred with 70%. Again, here also the light color means non-deprivation and the dark color means deprivation regarding the indicator.

Table 3. Illustration 1 of ESCAP_C1 profiles among South Africa, Turkey, Mexico and Brazil

WATD (Access to drinking water)	**EDU** (Educational attainment)	**Frequency of profiles (in per centage)**
DEPRIVATION	**DEPRIVATION**	1. SA: 43.98% 2. MX: 25% 3. BR: 13.58% 4. TR: 2.5%
NON-DEPRIVATION	**DEPRIVATION**	1. TR: 70% 2. BR: 46.11% 3. MX: 37.83% 4. SA: 21.76%
DEPRIVATION	**NON-DEPRIVATION**	1. SA: 12.54% 2. MX: 5.14% 3. BR: 4.39% 4. TR: 0%
NON-DEPRIVATION	**NON-DEPRIVATION**	1. BR: 35.92% 2. MX: 31.92% 3. TR: 27.4% 4. SA: 21.72%

Note: The light color means non-deprivation and the dark color means deprivation regarding the indicator. And WATD: Access to drinking water, EDU: Educational attainment.

Table 4. Illustration 2 of ESCAP_C1 profiles among South Africa and Turkey

South Africa (SA)

WATD	EDU	Frequency
DEPRIVATION	**DEPRIVATION**	43.98%
NON-DEPRIVATION	**DEPRIVATION**	21.76%
NON-DEPRIVATION	**NON-DEPRIVATION**	21.72%
DEPRIVATION	**NON-DEPRIVATION**	12.54%

Turkey (TR)

WATD	EDU	Frequency
DEPRIVATION	**DEPRIVATION**	2.5%
NON-DEPRIVATION	**DEPRIVATION**	70%
NON-DEPRIVATION	**NON-DEPRIVATION**	27.4%
DEPRIVATION	**NON-DEPRIVATION**	0%

Note: The light color means non-deprivation and the dark color means deprivation regarding the indicator. And WATD: Access to drinking water, EDU: Educational attainment.

Table 5. Illustration 2 of ESCAP_C1 profiles among Mexico and Brazil

Mexico (MX)			Brazil (BR)		
WATD	EDU	Frequency	WATD	EDU	Frequency
DEPRIVATION	DEPRIVATION	37.83%	DEPRIVATION	DEPRIVATION	46.11%
NON-DEPRIVATION	DEPRIVATION	31.92%	NON-DEPRIVATION	DEPRIVATION	35.92%
NON-DEPRIVATION	NON-DEPRIVATION	25.11%	NON-DEPRIVATION	NON-DEPRIVATION	13.58%
DEPRIVATION	NON-DEPRIVATION	5.14%	DEPRIVATION	NON-DEPRIVATION	4.39%

Note: The light color means non-deprivation and the dark color means deprivation regarding the indicator. And WATD: Access to drinking water, EDU: Educational attainment.

In terms of ESCAP profiles, we can say that: (i) South Africa is suffering the most regarding deprivation in both access to drinking water (WATD) and educational attainment (EDUC), concerning 44 per cent of 18+ population, which is 1.75 times higher than Mexico, 3.2 times higher than Brazil and 17.6 times higher than Turkey; (ii) however, regarding non-deprivation in both WATD and EDUC, Brazil has the best score covering 36 per cent of 18+ population, which is 1.12 times higher than Mexico, 1.65 times higher than South Africa, and 1.3 times higher than Turkey; (iii) regarding deprivation in educational attainment (and not in WATD), Turkey has the worst score covering 70 per cent of 18+ population, which is 1.5 times higher than Brazil, 1.85 times higher than Mexico, and 3.2 times higher than South Africa; (iv) and regarding deprivation in access to drinking water (and not in EDUC), South Africa has the worst score covering 12.54 per cent of 18+ population, which is 2.4 times higher than Mexico, 2.85 times higher than Brazil, given that this situation was not found in Turkey. As they are 2 indicators, deprivation regarding both 2 items (1,1) can be considered as the poorest, non-deprivation regarding both 2 items (0,0) as the most prosperous, and deprivation regarding one of 2 items (1,0), (0.1) as a combination of poverty and prosperity situation. In terms of total prosperity, Brazil has the highest score (35.92), then Mexico (31.92), Turkey (27.4), and South Africa (21.72) come. In terms of total poverty, South Africa (43.98) has the highest score, then Mexico (25.11), Brazil (13.58), and Turkey (2.5) come. In terms of intermediate poverty, Turkey has the highest score (70), then Brazil (50.5), Mexico (43), and South Africa (34.3) come.

In terms of MSCAP profiles, we can observe that: (i) Turkey has the best score regarding non-deprivation in all items, concerning 51.4 per cent of 18+ population, which is 1.35 times higher than Brazil, 1.84 times higher than Mexico, and 3.45 times higher than South Africa; (ii) the profile of deprivation only in TEL but not in other items appears as one of the most frequented profile concerning 58.24 per cent of 18+ population in South Africa, which is 1.2 times higher than Brazil, 1.36 times higher than Turkey, and 1.3 times higher than Mexico cases. There are also other profiles concerning not the 4 countries at the same time but concerning only 1 or 2 or 3 countries. Moreover, in terms of total prosperity, Turkey has the highest score (51.4), then Brazil (38.14), Mexico (27.9), and South Africa (14.86) come. In terms of total poverty, Mexico and South Africa has a score under 3 per cent, and in case of Turkey and Brazil it is zero or near zero. The ranking of intermediate poverty is the same of total prosperity. If we take into account intensity in terms of number of deprivations for intermediate poverty case, 3 deprivations should be considered the heaviest, 2 deprivations as the middle level, and 1

deprivation as the lowest one. So that, in terms of intermediate poverty with 3 deprivations, South Africa appears as the worst then Mexico arrives; in terms of 2 deprivations, South Africa, Mexico, Brazil come; in terms of 1 deprivation, South Africa, Brazil, Mexico, Turkey come respectively[3].

In terms of CAP profiles, it becomes more difficult to get information for a substantial share of the population. For instance, 15 per cent of information was not captured in South Africa case; in other words, a relevant heterogeneity of combination of ESCAP and MSCAP variables appears in South Africa case. However, we can conclude that (i) regarding deprivation only in EDUC, Turkey has the worst score concerning 33.3 per cent of 18+ population, which is 2 times higher than Brazil, 3.47 times higher than Mexico, and 6.85 times higher than South Africa; (ii) regarding deprivation only in TEL, Brazil has the worst score concerning 15.5 per cent of 18+ population, which is 1.33 times higher than South Africa, 1.6 times higher than Mexico, and 1.66 times higher than Turkey; (iii) regarding deprivation in EDUC and TEL and no deprivation in other items at the same time, Turkey has the worst score concerning 31.2 per cent of population, which is 1.4 times higher than Brazil, 2.2 times higher than Mexico, and 2.3 times higher than South Africa cases; (iv) and regarding no deprivation in all items at the same time, Brazil has the best score concerning 18.85 per cent of 18+ population, which is quite at similar level with Mexico and Turkey, and 2 times higher than South Africa cases. In terms of total prosperity, Brazil has the highest score (18.5), then Turkey (17.85), Mexico (16.5) and South Africa (9) come.

At an approximative level, we can say that in terms of material deprivation and access to drinking water, Turkey has the best performance but at the same time the worst performance regarding educational attainment. South Africa has the worst performance regarding access to drinking water and material deprivation. Mexico and Brazil have quite similar performance.

However, from the results above we can remark that comparison between several countries risks becoming complicated and even non meaningful in case there is a substantial heterogeneity of poverty profiles. Comparison between countries by an indicator from the basket of indicators/dashboard appears feasible; however, in case of many variables the comparison should be done in the number of variables appearing in the basket. Such process makes complicated to obtain synthetic comparison results; on the other hand, if comparison should be based on a particular item of poverty (e.g., educational attainment), then such way will be efficient. There is also a practice of ranking countries as a function of their aggregate/composite poverty index score; however, this method does not give a proper context to the content of the index (we cannot see the combination of variables and compensation effect reduces the reliability of the index). In this perspective, we can recommend that if the researcher wants a comparison between countries regarding one item of poverty for a specific reason, then comparison based on the headcount ratio of deprivation regarding that variable is an efficient solution. On the other part, if the aim is to compare countries in terms of their poverty profiles

[3] However, it is important to remark for such ranking a complete information (100%) is necessary, for other cases such as between 90 and 98 per cent of information, some robustness tests may be required in case of high heterogeneity (a lot of profiles with small per centages) or some programming can be done to obtain directly the headcount ratio of different levels of intermediate poverty statistics. For a country case study, detailed intermediate poverty statistics is not relevant; instead of that, the relevant profiles should be detected. By focusing on the relevant intermediate poverty profiles, a substantial effort on poverty reduction can be achieved.

from a multidimensional perspective based on many variables, then decomposition of the global poverty profile into different dimensional poverty profiles (e.g., educational attainment profile, material situation profile) or combination of certain dimensions to obtain a basic or another capabilities profile is a good way of reducing heterogeneity and then making enable comparison of countries based on the relevant emerging profiles. In singular country analysis, we used classification tree method to find out the most relevant poverty patterns however in case of more than one country the performance of classification algorithm risks to be fewer.

In traditional multidimensional poverty studies such as the Multidimensional Poverty Index (MPI), a threshold is used to determine who is poor. This threshold is called poverty line. For instance, in case of two functionings WATD and EDU (see ESCAP_C1 profiles above) if the poverty line is determined as deprivation regarding 2 variables at the same time, then in the poverty rate in South Africa will be 43.98%, in Turkey it will be 2.5%, in Mexico it will be 25% and in Brazil it will be 13.58%. However, our results show that the other in-between poverty configurations are important for reducing poverty, such as 70% in Turkey are deprived in one indicator (educational attainment). Or, in Brazil 50.5% are deprived regarding educational attainment or access to drinking water. And, as we can see the rankings of the countries change as well depending on the configurations. South Africa has the worst performance in terms of deprivation regarding both drinking water and educational attainment, and in terms of deprivation regarding drinking water and non-deprivation regarding educational attainment. But Turkey has the worst performance in terms of deprivation regarding educational attainment and non-deprivation regarding drinking water. And, even if the rankings are the same for instance regarding 2 configurations, the incidence of configuration varies, and it is important to consider it. In the framework of capabilities, Sen defines capabilities as the combinations of functionings. Our methodology, therefore, corresponds to this definition. Instead of focusing on only an overall poverty figure (e.g., headcount ratio such as 43.98% in South Africa based on the poverty threshold mentioned in the beginning of the paragraph), in our study we focus on the incidence and most importantly the nature of capabilities by respecting the criteria of diversity. In terms of ranking, for robust and transparent results it is important to take in consideration the multifaceted side of poverty and to rank countries regarding each poverty configuration. We argue that the number itself alone is not informative unless its content is not clearly revealed. Especially, we suggest that governments do not need to know how many poor people living in the country, but they need to know (i) how many configurations of poverty exist and (ii) which of them are the most relevant and (iii) what is their nature and (iii) they affect how many people in the country.

By distinguishing dimensions of poverty such as basic capabilities, material situation, etc. at country level the analysis can be done by respecting the following steps: (i) determine the *basic capabilities* items such as an adequate access to water, healthy food in terms of quantity and quality, electricity in the dwelling, proper toilet service in the dwelling (not shared with other households), proper bathroom in the dwelling (not shared with other households), proper kitchen with adequate energy source for cooking, proper heating source, proper lighting in the dwelling, availability of drainage system in the dwelling, and proper way of throwing out rubbish of the dwelling; (ii) obtain basic capabilities profiles and their headcount ratio in the population; (iii) regarding basic capabilities, compute head count ratio of total prosperity, headcount ratio of total poverty, and headcount ratio of intermediate poverty (between of total prosperity and total poverty) given that the latter category can be also divided into two sub-groups, upper vs. lower intermediate poverty in terms of quartiles; (iv) find out other

dimensional profiles and their frequency for total prosperity, total poverty, intermediate poverty groups of basic capabilities. Such way avoids poverty line, which is often set arbitrarily and takes into account the relative distribution of items in the population (i.e., a unique poverty line is not convenient for any comparison between countries in case distribution of deprivations vary significantly among countries); and (v) find out geographical situation information for each relevant basic capabilities profile, which may help to reduce poverty in a more efficient way by focusing on region, municipality, urban/rural areas.

Concerning *decomposition* analysis, we will find out ESCAP and MSCAP profiles of economically active vs, not economically active, elderly, youth (18-24) population, and in terms of setting (urban vs. rural), gender (male vs. female), households headed by a not economically active person, households headed by educationally deprived person, and households headed by a female.

- Economically active population (18+)-EAC[4]

The two tables below show ESCAP, MSCAP, and CAP profiles of economically active 18+ and economically not active 18+ population in 4 countries.

Table 6. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for economically active 18+ population

	South Africa (SA GHS 2011)[5]	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C1)	00 (34.17), 11 (30.76), 01 (19.97), 10 (15.11) Total info=100%	01 (62.5), 00 (34.58), 11 (2.78) Total info=100%	01 (34.83), 00 (34.17), 11 (25.64), 10 (5.36) Total info=100%	00 (42), 01 (39), 11 (14), 10 (5) Total info=100%
Material situation profiles (MSCAP_C1)	0010 (58.47), 0000 (20), 1110 (8.79), 0110 (5.47), 1010 (2.16) Total info=95%	0000 (50.63), 0010 (45.22) Total info=96%	0010 (36.16), 0000 (27.5), 0011 (10.2), 0001 (9.11), 0110 (5.29), 0111 (4.82), 1111 (2.8) Total info=96%	0010 (50.17), 0000 (38), 0011 (5.17), Total info=93%

[4] According to questionnaire and ILO definition, work without remuneration is also considered a sort of being economically active on the market. In such case, unpaid workers, volunteers can be examined as a separate case if the questionnaire allows it. Unemployed, seeking job, and other economically not active people are included here.

[5] As we have mentioned before SA data contains 5.8% missing values for this model given the available information for 18+ population regarding the variables. In addition, 3 per cent do not have information for economic activity.

Complete profiles (CAP_C1)	000010 (17.96), 110010 (16.5), 000000 (14.93), 010010 (12.92), 100010 (11.09), 111110 (6.5), 010000 (4.12), 110110 (3.35) Total info=87%	010010 (31), 010000 (28.4), 000000 (21.9), 000010 (12.4) Total info=94%	000000 (17.34), 010010 (14.33), 000010 (11.11), 010000 (8.36), 110010 (8.26), 110011 (4.65), 010001 (4.2), 010011 (3.82), 110111 (3.3), 000001 (2.9), 110110 (2.5), 100010 (2.46), 111111 (2.3) Total info=86%	000000 (22), 010010 (21), 000010 (18.36), 010000 (13.55), 110010 (7.7), 100010 (3), 110011 (2.2), 010011 (2) Total info=90%

- Economically not active population (18+)-NEAC

Table 7. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for economically not active 18+ population

	South Africa (SA GHS 2011)	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C1)	11 (52.24), 01 (22.69), 00 (13.69), 10 (11) Total info=100%	01 (76.8), 00 (20.8), 11 (2.31) Total info=100%	01 (42.66), 00 (28.3), 11 (24.25), 10 (4.8) Total info=100%	01 (59), 00 (24.7), 11 (13), 10 (3) Total info=100%
Material situation profiles (MSCAP_C1)	0010 (57.82), 1110 (11.8), 0000 (11.37), 0110 (7.28), 1010 (3.92), 1111 (2.89), 0011 (2.59) Total info=98%	0000 (52.11), 0010 (39.18), 0001 (4.6) Total info=96%	0010 (30.3), 0000 (28.6), 0001 (13.4), 0011 (11.88), 0111 (4.69), 0110 (3.9), 1111 (3) Total info=96%	0010 (44.2), 0000 (38.25), 0011 (7.8), 0001 (4.26) Total info=94.5%
Complete profiles (CAP_C1)	110010 (28.58), 010010 (13.7), 111110 (9.97), 100010 (7.96), 000010 (7.57), 110110 (5.57), 010000 (5.26), 000000 (5.22), 111010 (3), 111111 (2.52) Total info=89%	010000 (37.8), 010010 (31.36), 000000 (14), 010001 (4.4), 000010 (6.4) Total info=94%	000000 (15.18), 010010 (13.5), 010000 (11.7), 010001 (8), 000010 (7.5), 110010 (7.2), 010011 (5.2), 110011 (5), 000001 (3.3), 110111 (3.11), 111111 (2.5), 100010 (2) Total info=84%	010010 (25.2), 010000 (23), 000000 (13), 000010 (10.2), 110010 (6.86), 010011 (4.87), 010001 (3.42), 110011 (2.22) Total info=89%

ESCAP profiles

Concerning economically active 18+ population, Brazil (42) has the best performance in terms of non-deprivation in WATD and EDU at the same time, then Mexico (34.17) and South Africa (34.17) with same per centage and then Turkey (34.6) come. On the other hand, in terms of

deprivation in WATD and EDU at the same time, Turkey has the best performance concerning only 2.78 per centage of population, then Brazil (14), Mexico (25.64) and South Africa (30.76) come. In terms of deprivation only in WATD, South Africa has the worst score (15.11), then Brazil and Mexico with around 5 per cent of score come given that such case was not found in Turkey. And in terms of deprivation only in EDU, Turkey has the highest score (62.5), then Brazil and Mexico come (39, 34.8), and at the end South Africa emerges with around 20 per cent.

Concerning not economically active 18+ population, in terms of non-deprivation in WATD and EDU at the same time, Mexico (28) has the best score, then Brazil (24.7), Turkey (20.8) and South Africa (13.7) come respectively. Compared to economically active population, we can say that in4 countries, the proportion of total prosperity in terms of ESCAP is lower among economically not active population. For instance, for Brazil, the per centage is reduced from 42 to 24.7. Notably, discrepancy is higher for South Africa and less for Mexico. In terms of deprivation in WATD and EDU at the same time, Turkey has the best performance (2.3), South Africa has the worst performance indicating that half of the population are deprived in both 2 items-this proportion is much higher compared to economically active population case. Otherwise, for Brazil and Mexico, the proportion of people lacking in both 2 items is quite the same for both economically active and not economically active population. In terms of deprivation only in WATD, South Africa has the worst performance concerning 11 per cent of economically not active population; however, this per centage is slightly better compared to economically active population case. And in terms of deprivation only in EDU, Turkey has the worst score (76.8), which is 14.3 per centage points higher compared to economically active population situation; for Brazil, the per centage of economically not active population deprived regarding only EDU is 20 per centage points higher than economically active population case; for Mexico, the per centage of economically not active population deprived regarding only EDU is 7.83 per centage points higher than economically active population case; and for South Africa the per centage of deprivation only in education is slightly higher for economically not active population compared to economically active.

MSCAP profiles

In terms of non-deprivation in 5 items at the same time, the difference between economically active and economically not active population is very low or quite close in case of Turkey, Brazil, and Mexico however, for South Africa the difference is higher about 9 per centage points in detriment of economically not active population. In terms of deprivation in 5 items at the same time, there is not much difference between economically and economically not active populations. Concerning TV and refrigerator, the proportion of people not having at least one of them in their dwelling is higher for economically not active compared to economically active population in South Africa. Concerning mobile phone, the proportion of people not having it is also higher for economically not active compared to economically active population in 4 countries.

- Elderly (65+)-ELD

The table below shows ESCAP, MSCAP, and CAP profiles of elderly population (65+) in 4 countries.

Regarding ESCAP profiles, in terms of deprivation in both WATD and EDU, the per centages are notably higher in South Africa, Mexico, and Brazil for elderly population compared to total 18+ population. In terms of non-deprivation in both WATD and EDU, the situation of elderly decreases clearly in 4 countries compared to the situation of total 18+ population, for instance in South Africa such case was not found. In terms of deprivation only in WATD, the situation is quite the same for elderly as for the total 18+ population in South Africa. And, in terms of deprivation only in EDU, the situation becomes significantly worse such as in Turkey 91.5 per cent of elderly population was found deprived in that aspect.

Regarding MSCAP profiles, the share of elderly population not having experienced deprivation in 4 items at the same is quite similar to the per centages found for 18+ total population in Turkey, Mexico, and Brazil. In terms of deprivation in 4 items at the same time, Mexico has the highest score with 6.3 per cent and then South Africa comes with 4.37 per cent. Deprivation regarding TEL appears here also as the most important material deprivation; for instance, in terms of deprivation only in TEL, South Africa has the worst performance concerning 47.44 per cent of elderly population, then in Brazil it is 30.9, in Turkey it is 22.83 per cent in Mexico it is 18 per cent. However, deprivation in TV and/or REF is still an issue for elderly population in South Africa and Mexico. Regarding deprivation MPHO, the situation of elderly emerges worse compared to 18+ population in 4 countries. We can observe that deprivation regarding communication tools, TEL and MPHO, is a common problem among 4 countries.

Table 7. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for elderly population (65+)

	South Africa (SA GHS 2011)	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C1)	11 (56.95), 01 (30.45), 10 (11.56) Total info=99%	01 (91.5), 00 (5.56), 11 (2.97) Total info=100%	01 (57.5), 11 (31.6), 00 (10.4) Total info=100%	01 (73), 11 (17), 00 (9.5), Total info=100%
Material situation profiles (MSCAP_C1)	0010 (47.44), 0000 (21.89), 1110 (9.12), 0110 (4.42), 1111 (4.37), 1010 (4.24), 0011 (3.8), 0001 (2.98) Total info=98%	0000 (51.74), 0010 (22.83), 0001 (16.35), 0011 (4.37) Total info=95%	0000 (24.3), 0001 (21.8), 0010 (18), 0011 (15.4), 1111 (6.3), 0111 (5.7), 1011 (2.55), 0110 (2.34) Total info=96%	0000 (40.8), 0010 (30.9), 0011 (12.2), 0001 (9.6) Total info=93.5%
Complete profiles (CAP_C1)	110010 (31.58), 010010 (13.36), 010000 (12.17), 111110 (8.64), 000000 (8.46), 111111 (4.11), 110110 (4), 111010 (3.7), 110011 (2.39) Total info=88%	010000 (46.47), 010010 (21.3), 010001 (16), 010011 (4), 000000 (5) Total info=93%	010000 (16.37), 010001 (16.22), 010010 (11.2), 010011 (8), 110011 (7), 000000 (6.5), 110010 (5.8), 111111 (5.44), 110111 (4.18), 110001 (3.12), 000001 (2.48) Total info=86%	010000 (31), 010010 (22), 010001 (8.2), 010011 (8.2), 110010 (7.5), 000000 (7.4), 110011 (4), 110000 (2) Total info=90%

- Youth population (18-24 yrs.)-YOUTH

The table below shows ESCAP, MSCAP, and CAP profiles of youth population (18-24 yrs.) in 4 countries.

Regarding ESCAP profiles, in terms of deprivation in both WATD and EDU, South Africa has the highest score concerning 45.5 per cent of youth population, then Mexico arrives with 19 per cent and then Brazil with 12 per cent and then Turkey comes with 3.4 per cent. On the other hand, in terms of non-deprivation in both WATD and EDU, Mexico has the best score concerning 46 per cent of youth population, then Brazil comes with 39.5 per cent, then Turkey with 44.19, and South Africa with around 19.5 per cent. In terms of deprivation only in WATD, South Africa has the highest score (16.5), then Mexico (12.25) and Brazil (7) come respectively. In terms of deprivation only in EDU, Turkey has significantly highest score concerning 52.26 per cent of youth population, then Brazil comes with 42 per cent, then Mexico with 22.7 per cent and South Africa with 18.5 per cent.

Regarding MSCAP profiles, in terms of non-deprivation in 4 items, Turkey has the highest score concerning 46.18 per cent of youth population, then Brazil (34.15), Mexico (26.6) and South Africa (9.82) arrive. In terms of intermediate poverty, the most frequent profile is being deprived only in TEL, which is found 62.6 per cent in South Africa, 54 per cent in Brazil, 50 per cent in Mexico, and 37.8 per cent in Mexico. Besides, in South Africa, it is found that 11.66 per cent of youth population is deprived in TV, REF, and TEL. In terms of deprivation in MPHO, Mexico has the highest score concerning 24.4 per cent of youth population; for instance, 9 per cent of youth population in Mexico is deprived regarding both TEL and MPHO. Moreover, this last situation is around 4.8 per cent in Brazil.

Table 8. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for 18-24 population

	South Africa (SA GHS 2011)	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C1)	11 (45.52), 00 (19.47), 01 (18.5), 10 (16.5) Total info=100%	01 (52.26), 00 (44.19), 11 (3.4) Total info=100%	00 (45.74), 01 (22.7), 11 (19.3), 10 (12.25) Total info=100%	01 (41.7), 00 (39.5), 11 (11.88), 10 (7) Total info=100%
Material situation profiles (MSCAP_C1)	0010 (62.6), 1110 (11.66), 0000 (9.82), 0110 (7.69), 1010 (3.66) Total info=95%	0010 (50.2), 0000 (46.18) Total info=96%	0010 (37.8), 0000 (26.6), 0011 (8.99), 0001 (8), 0110 (6.64), 0111 (4.8), 1111 (2.6) Total info=95%	0010 (54), 0000 (34.15), 0011 (4.8), 0110 (2.2) Total info=95%
Complete profiles (CAP_C1)	110010 (25.8), 010010 (12.7), 000010 (12.24), 100010 (11.84), 111110 (9), 000000 (6), 110110 (5.2),	010010 (31.6), 010000 (18.6), 000000 (27.2), 000010 (16.3), 110010 (2) Total info=96%	000000 (20.4), 000010 (15.8), 010010 (10.6), 110010 (6.2), 100010 (5.3), 000001 (4.4), 010000 (4.2),	000010 (21.65), 010010 (21.4), 010000 (16.15), 000000 (15.6), 110010 (6.5), 100010 (4.5) Total info=86%

	010000 (3), 111010 (2.5) Total info=88%		110011 (3.12), 110111 (2.65), 110110 (2.48), 010011 (2.3), 000011 (2) Total info=79.5%	

- Setting (urban vs. rural)-URBAN vs. RURAL

Tables below show ESCAP, MSCAP, and CAP profiles of urban and rural population in 4 countries. We recall that in the data of Mexico, information related to setting was not found; so that, decomposition analysis at this level could not be realized for Mexico.

Regarding ESCAP profiles, in terms of non-deprivation in WATD and EDU, per centage is quite higher for urban settlers in South Africa, Turkey, and Brazil compared to rural settlers. In terms of deprivation in WATD and EDU, per centage is quite higher for rural settlers in South Africa and Brazil. In terms of deprivation only in WATD, per centage is higher for rural settlers in South Africa and Brazil. And, in terms of deprivation only in EDU, the per centage is lower for rural settlers in South Africa and Brazil. Rural settlers in South Africa and Brazil are mainly characterized by deprivation in WATD and EDU at the same whilst rural settlers in Turkey are mainly characterized by deprivation only in EDU. Urban settlers in Turkey do not suffer accessibility problems to drinking water whilst 6.7 per cent of rural settlers do suffer. In South Africa and Brazil, both rural and urban settlers suffer accessibility problems to drinking water, but the situation is much worse for rural settlers (34 per cent for urban settlers vs. 93 per cent for rural settlers in South Africa and 9.4 per cent for urban settlers vs. 73.36 per cent for rural settlers in Brazil), and it is the same situation for educational attainment as well.

Regarding MSCAP profiles, in terms of non-deprivation in all items, per centage is higher among rural settlers in Brazil, Turkey, and South Africa however the difference between urban and rural settlers is significantly lower in Turkey compared to South Africa and Brazil. The material situation of rural households is quite better in Turkey compared to South Africa and Brazil, given that 5 per cent are found deprived only in MPHO and 2 per cent are found deprived in TEL and REF. In terms of deprivation in all items, per centages are quite low in South Africa and Brazil. Rural settlers are mainly characterized by deprivation only in TEL in South Africa, Turkey, and Brazil. Among urban settlers also, deprivation only in TEL appears as one of the most frequent profiles. Otherwise, 18 per cent of rural settlers in South Africa are deprived regarding TV, REF, and TEL. In addition, in Brazil, 20 per cent of rural settlers are deprived in TEL and MPHO. In South Africa, for rural settlers, deprivation in TV and/or REF should be considered for reducing material poverty, and for urban settlers, even if the magnitude of such deprivation is considerably low these aspects of material deprivation should be taken into account. And in case of Brazil, we can see that deprivation in MPHO is around 30 per cent among rural settlers vs. 5 per cent among urban settlers. But also, deprivation in TV and/or REF should be considered for rural settlers in Brazil.

Table 9. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for urban areas

	South Africa (SA GHS 2011)	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C1)	00 (33.34), 01 (32.49), 11 (24.99), 10 (9.19) Total info=100%	01 (64.9), 00 (34.9), Total info=100%	-	01 (50.4), 00 (40.56), 11 (6.09), 10 (3.31) Total info=100%
Material situation profiles (MSCAP_C1)	0010 (58.67), 0000 (22.69), 1110 (5.94), 0110 (4.55), 0011 (2.39) Total info=94%	0000 (55.92), 0010 (40.32), 0001 (2.12) Total info=98%	-	0010 (47), 0000 (43.25), 0011 (3.92) 0001 (2.44) Total info=97%
Complete profiles (CAP_C1)	010010 (19.95), 000010 (17.74), 110010 (14.42), 000000 (14), 010000 (7.67), 100010 (6.56), 111110 (3.98), 110110 (2.47) Total info=87%	010000 (32.9), 010010 (28.5), 000000 (23), 000010 (11.6), 010001 (2) Total info=98%	-	010010 (24), 000000 (21.6), 010000 (19.3), 000010 (17.3), 110010 (3.6), 010011 (2.78) Total info=89%

Table 10. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for rural areas

	South Africa (SA GHS 2011)	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C1)	11 (75.2), 10 (18), 01 (4.11), 00 (2.63) Total info=100%	01 (79.4), 00 (13.6), 11 (6.7) Total info=100%	-	11 (61.76), 01 (20.8), 10 (11.33), 00 (6.1) Total info=100%
Material situation profiles (MSCAP_C1)	0010 (57.54), 1110 (18.16), 0110 (9.74), 1010 (5.75), 1111 (3.77), 0000 (2) Total info=97%	0010 (45.36), 0000 (42.99), 0001 (5.02) 0110 (2.15) Total info=95.5%	-	0010 (54.6), 0011 (20.12), 1111 (5.33), 0000 (5.23), 0111 (4.5), 0110 (3.97) Total info=94%

Complete profiles (CAP_C1)	110010 (39.74), 111110 (16.13), 100010 (13.26), 110110 (8.24), 111010 (4.83), 111111 (3.53), 010010 (2.92) Total info=89%	010010 (36.16), 010000 (33.9), 000000 (8.4), 010001 (4.8), 110010 (4), 000010 (5), Total info=92%	-	110010 (31.6), 110011 (13.18), 010010 (11.6), 100010 (7.5), 010011 (4.7), 111111 (4.7), 000010 (3.9), 110111 (3.5), 110110 (2.68) Total info=83.5%

- Gender (female vs. male)-FEM vs. MALE

Tables below show ESCAP, MSCAP, and CAP profiles of female and male population in 4 countries.

Regarding ESCAP profiles, in terms of deprivation in both WATD and EDU, there is not much difference between male and female in 4 countries. In terms of non-deprivation in WATD and EDU, there is not much difference between male and female in South Africa, Mexico and Brazil. In Turkey, the per centage of female not deprived in WATD and EDU is 21.5 per cent whilst it is 33.8 per cent among male. In terms of deprivation only in EDU, the per centage of female is around 12 points higher compared to male in Turkey whilst in other 3 countries the difference is not much significant. In terms of deprivation only in WATD, for Mexico and Brazil, there is not much difference between male and female whereas in South Africa, for male the per centage is 12.7 and for female it is 21.4.

Regarding MSCAP profiles, in terms of deprivation in all items, it emerges in South Africa and Mexico with similar per centages for both male and female population. In terms of non-deprivation in all items, there is not much difference between female and male in 4 countries. Deprivation only in TEL appears as one of the most frequent profiles in 4 countries given that there is not much difference between male and female. In South Africa, deprivation in TV, REF, and TEL affects 10 per cent of female and 11 per cent of male population. As we found before, deprivation regarding TV, REF, and TEL should be considered in South Africa, and their percentages are quite similar for both male and female populations. The situation of material capability is clearly better in Turkey compared to other 3 countries for both male and female populations. In Mexico, deprivation in MPHO appears as the most important one apart TEL among material items such that it concerns 29 per cent of male and female populations. Deprivation regarding REF, which is around 13 per cent for both male and female populations should be considered in Mexico for increasing capabilities. And in Brazil, apart TEL, deprivation in MPHO appears affecting 8 per cent of female vs. 6 per cent of male populations.

Table 11. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for female 18+ population

	South Africa (SA GHS 2011)	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C1)	11 (45.10), 01 (21.5), 00 (21), 10 (21.4) Total info=100%	01 (75.75), 00 (21.5), 11 (2.68) Total info=100%	01 (39), 00 (31.3), 11 (25), 10 (5) Total info=100%	01 (46), 00 (37.17), 11 (12.26), 10 (4.5) Total info=100%

Material situation profiles (MSCAP_C1)	0010 (59.7), 0000 (14.5), 1110 (10), 0110 (6.3), 1010 (3.3), 0011 (2.18), 1111 (2.05) Total info=98%	0000 (51.3), 0010 (41.46), 0001 (3.65) Total info=96.5%	0010 (33.5), 0000 (28.6), 0001 (11.3), 0011 (10.7), 0111 (4.6), 0110 (4.6), 1111 (2.8) Total info=96%	0010 (47.55), 0000 (39.3), 0011 (5.84), 0001 (2.56) Total info=95%
Complete profiles (CAP_C1)	110010 (25.68), 010010 (13.22), 000010 (11.46), 100010 (9.34), 000000 (8.6), 111110 (8.3), 010000 (5), 110110 (4.6), 111010 (2.4) Total info=89%	010000 (36.5), 010010 (33), 000000 (14.5), 000010 (7), 010001 (3.55) Total info=94%	000000 (16.4), 010010 (14.2), 010000 (10.3), 000010 (9.3), 110010 (7.84), 010001 (6), 110011 (4.7), 010011 (4.4), 000001 (3.15), 110111 (3.13), 111111 (2.3), 100010 (2.2), 110110 (2.2) Total info=86%	010010 (21.84), 000000 (19.4), 010000 (17.5), 000010 (16.17), 110010 (6.75), 010011 (3), 100010 (2.8)

Table 12. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for male 18+ population

	South Africa (SA GHS 2011)	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C1)	11 (42.6), 00 (22.6), 01 (22), 10 (12.7) Total info=100%	01 (63.66), 00 (33.8), 11 (2.4) Total info=100%	01 (36.55), 00 (32.63), 11 (25.5), 10 (5.4) Total info=100%	01 (46.16), 00 (34.54), 11 (15), 10 (4.25) Total info=100%
Material situation profiles (MSCAP_C1)	0010 (56.5), 0000 (15.3), 1110 (11.15), 0110 (6.8), 1010 (3.12), 1111 (3), 0011 (2.18) Total info=98%	0000 (51.5), 0010 (42.74), 0001 (2.57) Total info=97%	0010 (34.35), 0000 (27.2), 0011 (11), 0001 (10.15), 0111 (5), 0110 (5), 1111 (3) Total info=96%	0010 (48.6), 0000 (36.85), 0011 (6.4) Total info=92%
Complete profiles (CAP_C1)	110010 (21.94), 010010 (13.87), 000010 (11.87), 000000 (9.57), 111110 (8.88), 100010 (8.8), 110110 (4.7), 010000 (4.7), 111111 (2.6), 111010 (2.2) Total info=89%	010000 (29.75), 010010 (29.3), 000000 (21.45), 000010 (11.88), 010001 (2.3) Total info=95%	000000 (16.63), 010010 (13.4), 000010 (10.2), 010000 (8.84), 110010 (7.85), 010001 (5.2), 110011 (5), 010011 (4.33), 110111 (3.35), 000001 (3), 111111 (2.5), 100010 (2.4) Total info=83%	010010 (23.23), 000000 (18.24), 010000 (16.23), 000010 (14.7), 110010 (8.11), 010011 (3), 100010 (2.6), 110011 (2.5) Total info=88.6%

- Households with a head not economically active-H1

The table below shows ESCAP, MSCAP, and CAP profiles among 18+ population living in households headed by economically not active person in 4 countries. We recall that in the data of South Africa, the detection of household head was found problematic due to data internal inconsistency; so that, decomposition analysis at this level could not be realized for South Africa.

Regarding ESCAP profiles, in terms of non-deprivation in both WATD and EDU, Mexico and Brazil have the highest performances concerning around 30 per cent of 18+ population whereas it is about 23 per cent in Turkey. In terms of deprivation in both WATD and EDU, Turkey has the best performance, then it is 11 per cent in Brazil, and Mexico has the worst performance concerning 21 per cent of 18+ population. In terms of deprivation only in WATD, the per centages are 3.5 and 2.8 for Mexico and Brazil respectively. And, in terms of deprivation only in EDU, Turkey has the worst performance concerning around 74.34 per cent of 18+ population, then Brazil comes with 57 per cent, and Mexico emerges with 46 per cent. We can see that compared to WATD, deprivation in EDU appears significantly important for 3 countries.

Regarding MSCAP profiles, in terms of non-deprivation in all items, Turkey has the best performance concerning 55 per cent of 18+ population, then Brazil comes with 41.5 per cent and Mexico appears with 30.6 per cent. Deprivation in TEL appears as the most frequent profile in 3 countries. Concerning deprivation in MPHO, the per centage is 6.6 per cent in Turkey, 36 per cent in Mexico, and 12.74 per cent in Brazil. And in Mexico, 9.6 per cent of 18+ population emerges as being deprived regarding REF. We can say that in terms of material deprivation, Turkey appears as having the best performance followed very closely by Brazil, which suffers slightly more than Turkey in terms of MPHO. And, given deprivation regarding TV and/or REF, which was not found in Turkey neither in Brazil, Mexico has the worst performance.

Table 13. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for households with a head not economically active

	South Africa (SA GHS 2011)	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C1)	-	01 (74.34), 00 (23.10), 11 (2.5) Total info=100%	01 (46), 00 (30), 11 (21), 10 (3.5) Total info=100%	01 (57.13), 00 (29), 11 (11), 10 (2.8) Total info=100%
Material situation profiles (MSCAP_C1)	-	0000 (55.43), 0010 (32.69), 0001 (6.6) Total info=95%	0000 (30.6), 0010 (26.2), 0001 (17.2), 0011 (11.8), 0111 (3.8), 1111 (2.9), 0110 (2.9) Total info=95%	0000 (41.5), 0010 (40.8), 0011 (7.84), 0001 (4.9) Total info=95%

Complete profiles (CAP_C1)	-	010000 (38.34), 010010 (25.54), 000000 (16.9), 010001 (6.3), 000010 (5.6) Total info=93%	000000 (16.24), 010000 (12.65), 010010 (12.6), 010001 (10.45), 000010 (6.3), 010011 (6), 110010 (5.8), 000001 (4.66), 110011 (4.3), 110111 (2.5), 111111 (2.35) Total info=84%	010010 (23.44), 010000 (22.14), 000000 (17), 000010 (10.15), 110010 (5.6), 010011 (5.14), 010001 (3.85) Total info=87.5%

- Households with a head educationally deprived-H2

The table below shows ESCAP, MSCAP, and CAP profiles among 18+ individuals living in households headed by educationally deprived person in 4 countries. We recall that in the data of South Africa, the detection of household head was found problematic due to data internal inconsistency; so that, decomposition analysis at this level could not be realized for South Africa. Regarding ESCAP profiles, in terms of deprivation in WATD and EDU, Turkey has the best performance concerning 3 per cent of 18+ population, then Brazil comes with around 20 per cent and Mexico appears with 34 per cent. In terms of non-deprivation in WATD and EDU, Brazil has the best performance concerning 16.6 per cent of 18+ population followed by Turkey with 16.3 per cent, and Mexico with 13 per cent. In terms of deprivation only in WATD, performances of Brazil and Mexico are quite similar with 2.88 and 4.25 per centages. And, in terms of deprivation only in EDU, Turkey has the worst performance concerning 80.5 per cent of 18+ population, followed by Brazil with 61 per cent and then by Mexico by 48 per cent. Regarding MSCAP profiles, in terms of non-deprivation in all items, Turkey has the best score concerning 48.3 per cent of 18+ population followed by Brazil with 30.7 per cent and then by Mexico with 18.75 per cent. Deprivation only in TEL appears as one of the most frequent profiles given that it concerns 50.5 per cent of 18+ population in Brazil, 44 per cent in Turkey, and 34.6 per cent in Mexico. In addition, in Brazil, 11.5 per cent of 18+ population are found deprived in MPHO vs. 36 per cent in Mexico. Deprivation in TV and/or REF appears also in Mexico such that 16 per cent of 18+ population are deprived in REF.

Table 14. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for households with an educationally deprived head

	South Africa (SA GHS 2011)	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C1)	-	01 (80.5), 00 (16.3), 11 (3) Total info=100%	01 (48.2), 11 (34.2), 00 (13.2), 10 (4.25) Total info=100%	01 (60.88), 11 (19.64), 00 (16.6), 10 (2.88) Total info=100%
Material situation profiles (MSCAP_C1)	-	0000 (48.27), 0010 (43.95), 0001 (3.73) Total info=96%	0010 (34.6), 0000 (18.75), 0011 (13.9), 0001 (12), 0111 (6.4), 0110 (5.6), 1111 (3.9) Total info=95%	0010 (50.5), 0000 (30.7), 0011 (8.5), 0001 (2.96) Total info=93%

Complete profiles (CAP_C1)	-	010000 (37.5), 010010 (36.3), 000000 (10.35), 010001 (3.6), 000010 (5.7) Total info=93.5%	010010 (17.88), 010000 (11.4), 110010 (10.5), 010001 (7.5), 110011 (6.6), 010011 (6), 000000 (5.4), 110111 (4.5), 000010 (4.4), 111111 (3.3), 110110 (3.1), 110001 (2.2) Total info=83%	010010 (30.5), 010000 (20.4), 110010 (10.6), 000000 (8), 000010 (7.6), 010011 (4.6), 110011 (3.33), 010001 (2.4) Total info=87.5%

- Households with a female head-H3

The table below shows ESCAP, MSCAP, and CAP profiles among 18+ individuals living in households headed by a female in 4 countries. We recall that in the data of South Africa, the detection of household head was found problematic due to data internal inconsistency; so that, decomposition analysis at this level could not be realized for South Africa.

Regarding ESCAP profiles, in terms of deprivation in WATD and EDU, Mexico has the worst performance concerning 21.7 per cent of 18+ population, followed by Brazil with 9.8 per cent and by Turkey with 3 per cent. In terms of non-deprivation in WATD and EDU, Brazil has the best performance concerning 36.85 per cent of 18+ population, followed by Mexico with 33.7 per cent and by Turkey with 27.3 per cent. In terms of deprivation only in WATD, performances of Brazil and Mexico are quite similar concerning 3.66 and 4.8 per cent respectively. In terms of deprivation only in EDU, Turkey has the worst performance concerning 69.7 per cent of 18+ population, followed by Brazil with 50 per cent and by Mexico with 40 per cent. Regarding MSCAP profiles, in terms of non-deprivation in all items, Turkey has the best score concerning 51.2 per cent of 18+ population, followed by Brazil with 40 per cent and by Mexico with 29 per cent. Deprivation only in TEL appears as one of the most frequent profiles in 3 countries with 47 per cent in Brazil, 35 per cent in Turkey, and 32 per cent in Mexico. In terms of deprivation in MPHO, 7 per cent of 18+ population in Turkey in such case, in Brazil it is 8.7 per cent and in Mexico it is 30 per cent. Deprivation regarding TV and/or REF does not appear in Turkey and Mexico whereas in Mexico, 10.8 per cent were found deprived in REF.

Table 15. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, MX MCS 2012, and BR PNAD 2011, for households with a female head

	South Africa (GHS 2011)	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C1)	-	01 (69.7), 00 (27.3), 11 (2.9) Total info=100%	01 (39.8), 00 (33.7), 11 (21.7), 10 (4.8) Total info=100%	01 (49.7), 00 (36.85), 11 (9.8), 10 (3.66) Total info=100%

Material situation profiles (MSCAP_C1)	-	0000 (51.2), 0010 (35), 0001 (7.17) Total info=94%	0010 (32.12), 0000 (29), 0001 (12.9), 0011 (10.4), 0110 (4.2), 0111 (3.8), 1111 (2.8) Total info=95%	0010 (47), 0000 (39.8), 0011 (5.74), 0001 (2.96) Total info=95%
Complete profiles (CAP_C1)	-	010000 (33.4), 010010 (24.6), 000000 (17.8), 010001 (6.77), 000010 (9) Total info=92%	000000 (16.8), 010010 (13.7), 010000 (10.5), 000010 (9.8), 010001 (7), 110010 (6.6), 010011 (4.6), 000001 (4), 110011 (4), 110111 (2.4), 111111 (2.3), 100010 (2) Total info=83.5%	010010 (23.9), 000000 (19.3), 010000 (18.14), 000010 (15.8), 110010 (5.25), 010011 (3.4), 010001 (2.2), 100010 (2.14) Total info=90%

In next paragraphs, we will mention the main results of the comparison analysis between the four countries at the same time.

The table 17 below summarizes the scores of 4 countries in terms of only deprivation in WATD (WATD_D), only deprivation in EDU (EDU_D), deprivation in both WATD and EDU (WATD_D+EDU_D), no deprivation in both WATD and EDU (NO WATD_D+EDU_D), deprivation in both TV and REF (TV_D+REF_D), deprivation in TEL (TEL_D), deprivation in MPHO (MPHO_D), and deprivation in both TEL and MPHO (TEL_D+MPHO_D) according to ESCAP and MSCAP profiles for the 11 decomposition groups. The first row provides per centages for 18+ population in order to do comparison of decomposition groups profiles within each country. For instance, it is found that in South Africa 12.5 per cent of 18+ population are living deprived only in WATD according ESCAP profile whilst 15 per cent of economically active 18+ population (EAC) are in such case. The figures highlighted in bold in table reflect some interesting results: (1) the rate of youth population deprived only in WATD according to ESCAP profile is found slightly higher compared to those of 18+ population in South Africa, Mexico, and Brazil; (2) the rate of youth population deprived only in EDU according to ESCAP profile is found lower compared to those of 18+ population in South Africa, Turkey, and Brazil except in Mexico the difference is much higher; (3) the rate of youth population deprived in both EDU and WATD according to ESCAP profile is found lower compared to those of 18+ population in Mexico; (4) the rate of youth population not deprived in EDU and WATD according to ESCAP profile is found higher compared to those of 18+ population in Mexico; (5) the rate of youth population deprived only in TEL according to MSCAP profile is found higher compared to those of 18+ population in Turkey and Brazil- which may due to the use of mobile phone rather than landline telephone nowadays; (6) the rate of youth population deprived in MPHO according to MSCAP profile is found lower compared to those of 18+ population in 4 countries; (7) the rate of rural settlers deprived only in WATD according to ESCAP profile is found higher compared to those of 18+ population and those of urban settlers in South Africa, and Brazil; (8) the rate of rural settlers deprived only in EDU according to ESCAP profile is found lower compared to those of 18+ population and those of urban settlers in South Africa and Brazil because the rate of rural settlers deprived in both

WATD and EDU according to ESCAP profile is found much higher compared to those of 18+ population and of urban settlers in Turkey, but most importantly in South Africa and Brazil; (9) the rate of rural settlers not deprived in both WATD and EDU according to ESCAP profile is found much lower compared to those of 18+ population and those of urban settlers in South Africa, Turkey, and Brazil; (10) the rate of rural settlers deprived in both TV and REF according to MSCAP profile is found higher compared to those of 18+ population and those of urban settlers in Turkey, Mexico, but most importantly in South Africa; (11) the rate of rural settlers deprived in TEL according to MSCAP profile is found higher compared to those of 18+ population and those of urban settlers in South Africa, Turkey, and Brazil; (12) the rate of rural settlers deprived in MPHO according to MSCAP profile is found much higher compared to 18+ population and those of urban settlers in Brazil; (13) the rate of urban settlers deprived only in EDU according to ESCAP profile is found higher compared to those of 18+ population and those of rural settlers in South Africa because the rate of urban settlers deprived in both EDU and WATD according to ESCAP profile is lower compared to 18+ population and those of rural settlers; (14) the rate of urban settlers deprived only in EDU and the rate of urban settlers deprived in both EDU and WATD according to ESCAP profile are found lower compared to those of 18+ population and those of rural settlers in Turkey; (15) the rate of urban settlers deprived only in EDU according to ESCAP profile is found higher compared to those of 18+ population and those of rural settlers in Brazil because the rate of urban settlers deprived in both EDU and WATD according to ESCAP profile is lower compared to those of 18+ population and of those rural settlers; (16) the rate of urban settlers not deprived in both WATD and EDU according to ESCAP profile is found higher compared to those of 18+ population and those of rural settlers in South Africa, Turkey, and Brazil; (17) the rate of urban settlers deprived in both TV and REF, and the rate of urban settlers deprived in TEL according to MSCAP profile are found quite lower compared to those of 18+ population and those of rural settlers in South Africa; (18) the rate of female deprived only in WATD according to ESCAP profile is found much higher compared to those of 18+ population and those of male population in South Africa; (19) the rate of female not deprived in both WATD and EDU according to ESCAP profile is found lower compared to 18+ population and most importantly to male population in Turkey; (20) the rate of 18+ population deprived only in EDU according ESCAP profile and living in a household headed by a person not economically active is found higher compared to those 18+ population in Mexico and Brazil; (21) the rate of 18+ population not deprived in both WATD and EDU according to ESCAP profile and living in a household headed by a person not economically active is found lower compared to those of 18+ population in Brazil; (22) the rate of 18+ population deprived in MPHO according to MSCAP profile and living in a household headed by a person not economically active is found higher compared to those of 18+ population in Mexico; (23) the rate of 18+ population deprived only in EDU according to ESCAP profile and living in a household headed by a person educationally deprived is found higher compared to those of 18+ population in Turkey, Mexico, and Brazil; (24) the rate of 18+ population deprived in both WATD and EDU according to ESCAP profile and living in a household headed by a person educationally deprived is found higher compared to those of 18+ population in Mexico and Brazil; (25) the rate of 18+ population not deprived in both WATD and EDU according to ESCAP profile and living in a household headed by a person educationally deprived is found lower compared to 18+ population in Turkey, Mexico, and Brazil; (26) the rate of 18+ population deprived in MPHO and the rate of 18+ population deprived in TEL and MPHO according to MSCAP profile and living in a household headed by

a person educationally deprived are found higher compared to those of 18+ population in Mexico.

The results shown in Table 17 below and the other results mentioned in chapter 6 and in appendix 9.7 indicate that ranking of countries based on the overall poverty index (which is based on the poverty line) is not suitable as it hides information. For that reason, we suggest that ranking of countries should be done per relevant configuration of poverty.

Table 16. Frequency (in per centage) of decomposition groups by indicators across South Africa, Turkey, Mexico, and Brazil, from Comparison 1

Decomposition group	Country	WATD_D	EDU_D	WATD_D +EDU_D	NO WATD_D +EDU_D	TV_D+ REF_D	TEL_D	MPHO_D	TEL_D+ MPHO_D
18+ **(Reference group)**	SA	12.5	21.8	44	21.7	13	83	4.7	4.7
	TR	0	70	2.5	27.4	0	43	3	0
	MX	5	38	25	32	2.9	57	29.3	18.5
	BR	4.4	46	13.6	36	0	54	8.4	6.1
EAC	SA	15	20	30.8	34	11	75	0	0
	TR	0	62.5	2.78	34.58	0	45	0	0
	MX	5.4	35	25.6	34	2.8	59	27	18
	BR	5	39	14	42	0	55	5	5
NEAC	SA	11	22.7	52	13.7	14.7	28.5	5.5	5.5
	TR	0	76.8	2.3	20.8	0	39	5.6	0
	MX	5	42.6	24	28.3	3	53.5	33	19.6
	BR	3	59	13	24.7	0	52	12	7.8
ELD	SA	11.6	30.45	57	0	13.5	73.4	8.2	8.2
	TR	0	91.5	3	5.5	0	27	20.7	20.7
	BR	0	57.5	31.6	10.4	6	48	52	52
	MX	0	73	17	9.5	0	43	22	22
YOUTH	SA	**16.5**	**18.5**	45.5	19.5	11.7	85.6	**0**	0
	TR	**0**	**52.3**	3.4	44.2	0	**50.2**	**0**	0
	MX	**12.25**	**23**	**19.3**	**45.8**	2.6	60.8	**24.4**	17
	BR	**7**	**42**	12	39.5	0	**61**	**5**	5
RURAL	SA	**18**	**4.1**	**75.2**	**2.6**	**22**	**95**	3.8	3.8
	TR	**0**	**79.4**	**6.7**	**13.6**	**0**	**47.5**	5	0

	MX	-	-	-	-	.	-	-	-
	BR	**11.3**	**20.8**	**61.8**	**6.1**	**5.3**	**88.5**	**30**	30
URBAN	SA	9.2	**32.5**	**25**	**33.3**	**6**	**71.5**	2.4	2.4
	TR	0	**65**	**0**	**34.9**	0	40.3	2	0
	MX	-	-	-	-	-	-	-	-
	BR	3.3	**50.4**	**6**	**40.6**	0	51	6.4	4
FEM	SA	**21.4**	21.5	45	21	12	83.5	4.2	4.2
	TR	0	75.75	2.7	**21.5**	0	41.5	3.65	0
	MX	5	39	25	31	2.8	22.7	29	18
	BR	4.5	46	12.3	37	0	59	8.4	5.8
MALE	SA	12.7	22	42.6	22.6	14	83	5.2	5.2
	TR	0	63.66	2.4	**33.8**	0	42.7	2.6	0
	MX	5.4	36.5	25.5	32.7	3	58	29	19
	BR	4.2	46.2	15	34.5	0	55	6.4	6.4
H1	SA	-	-	-	-	-	-	-	-
	TR	0	74.34	2.5	23.10	0	32.7	6.6	0
	MX	3.5	**46**	21	30	3	47.6	**35.7**	18.5
	BR	2.8	**57**	11	**29**	0	48.7	12.7	7.8
H2	SA	-	-	-	-	-	-	-	-
	TR	0	**80.5**	3	**16.3**	0	44	3.7	0
	MX	4.2	**48.2**	**34.2**	**13.2**	4	**30**	**36.2**	**24.2**
	BR	3	**60.9**	**19.7**	**16.6**	0	59	11.5	8.5
H3	SA	-	-	-	-	-	-	-	-
	TR	0	65.7	3	27	0	**35**	7.2	0
	MX	4.8	40	22	34	3	**21.2**	30	17
	BR	3.7	50	10	37	0	**5.7**	8.7	5.7

Note: WATD_D: deprivation only in WATD (ESCAP model); EDU_D: deprivation only in EDU (ESCAP model); WATD_D+EDU_D: deprivation in both EDU and WATD (ESCAP model); NO WATD_D+EDU_D: no deprivation in both WATD and EDU (ESCAP model); TV_D+REF_D: deprivation in both TV and REF (MSCAP model); TEL_D: deprivation in TEL (MSCAP model); MPHO_D: deprivation in MPHO (MSCAP model); MPHO_D+TEL_D: deprivation in both MPHO and TEL (MSCAP model); EAC: Economically active 18+ population; NEAC: Economically not active 18+ population; ELD: Elderly 65+ population; YOUTH: 18-24 yrs. old young population; URBAN: 18+ population living in urban areas; RURAL: 18+ population living in rural areas; FEM: 18+ female

population; MALE: 18+ male population; H1: Households headed by a person not economically active; H2: Households headed by a person educationally deprived; H3: Households headed by a female.

Profiles based computations shown above are particularly suitable to capture combinations of functionings in order to understand in a comprehensive way the capability of a person. However, in such method, the delicate point is not to employ many variables at the same time. Plurality and diversity of different capabilities intrinsically leads to employing several indicators for reflecting as much as possible diverse living conditions of a person in her environment. Nevertheless, in case there are many variables it becomes very complex and even impossible to get a synthetic and interpretable picture of combinations of functionings. For that reason, we proposed to use classification trees analysis. But in case classification trees method cannot be applied (e.g., the method of classification trees is efficient when there are many variables. For the comparison of the 4 countries at the same time, we have only 6 variables, which is not much), we suggest extracting dimensional poverty profiles as a product of combinations of a relatively small number of functionings. We considered access to drinking water, healthy and enough food, having an adequate shelter including an adequate proper space for everyone living in the dwelling as a household, availability of electricity, good energy sources for heating, cooking, lighting, appropriate sanitation facilities (bathroom and toilet), kitchen, drainage system, rubbish removal system of the dwelling, educational attainment, living in security, having a medical coverage as the essential functionings, which should be taken into account to understand the level of essential capabilities the person has in her/his life. This conception corresponds to basic capabilities of Sen proposed in his framework to examine poverty in developing countries. The functionings of this sort of capabilities should be provided in an accurate way to every single person to have decent, basic living standards not only to survive but also to being able to have other varieties of capabilities such as being active on labour market, enriching her/his material situation, realizing different social activities. The content of essential capabilities, and of other types of capabilities might be long as well generating long combinations of functionings. Herewith the solution is to create as much as possible different dimensions of poverty, in other words focusing on different types of capabilities composed of relatively low number of items.

7. Conclusion

To conclude the thesis, we will mention the following points: why it is important to measure poverty, why capabilities approach is important for measuring poverty in upper-middle-income countries, social attachment profile and capabilities, relative vs. absolute characteristics of capabilities, comparative analysis, suggestions for policy makers in measuring and comparing poverty across countries, data limitations and subjective indicators of poverty, summary of our methodology including its limits, and at the end some improvements for further research on multidimensional poverty.

As we have mentioned in the literature review, there are different conceptual approaches to define the phenomenon of poverty. The most popular and common way over years has been the monetary approach using consumption, expenditure patterns and income thresholds to identify who is poor in society. This purely economic perspective is particularly known as an easy way for comparing different countries at international platform without being dependent on the structure of social surveys-as it is quite difficult to find out common social indicators among countries. However, this approach is also known as being an indirect way of measuring poverty, as it is not concretely able to take in consideration non-monetary items and the items that are not available or do not exist in the market such as dignity, trust in people, respect, affection, or access to various social services, and social relationships. At this point, multidimensional look at poverty becomes crucial, which is totally able to deal with non-monetary aspect of life. From this angle, life is recognized as an ensemble of different dynamics and of multiple living conditions. For instance, besides to financial capacity, the dimensions such as education, health, economic activity, leisure, social participation, environment, security, dwelling conditions are typically included in multidimensional poverty analysis. Within the multidimensional view, several conceptualizations of poverty are proposed by scholars in the literature such as capabilities, marginalidad, relative deprivation, social exclusion, social disqualification, working poor, subjective poverty.

A very general understanding of poverty is the hard-living conditions of very disadvantaged people living in extreme or severe poverty in their society. However, such regard at poverty do not much help to obtain a comprehensive picture on it given that societies are intrinsically complex composed of different social classes coupled with complex daily interactions among people. For that reason, it is tremendously important to detect how different social groups are distinguished in terms of life elements and what are the profiles of people from upper to bottom classes. Notably, associating poverty with a certain group of people makes these people feel more vulnerable, or fragile, or even excluded in their society. Poverty, as a social phenomenon, should be investigated in a broader aspect aiming at discovering profiles existing in society. Given the plurality and diversity of different life elements and conditions, there is not a particular form of poverty. With this reality, it becomes clear that to understand the magnitude of severe living conditions including food shortage, bad dwelling conditions, homeless, no access to drinking water, etc., first we have to examine deeply the society as a whole. Qualitative studies generally bring a lot of description on poverty circumstances providing detailed information on geographical situation (where the person lives and in which living standards/dwelling and neighborhood conditions), working conditions, social life of interviewed individuals on their daily routine. However, this type of studies generally covers a relatively small target population given that detailed interviews are not possible to be realized in large samples. At this stage, to reflect multidimensionality of poverty, quantitative approach

based on large representative national household socio-economic, or living conditions surveys is the best way to deal not only with large population but also it allows for the creation of various social indicators such as educational attainment, whether a person is covered or not by social security, whether a person has access or not to medical services when it is needed, whether a person suffers or not food shortage, in which sector the person is employed, the conditions of roof, wall, floor of the dwelling, possession of materials in the household, etc.

A definition of poverty is certainly necessary as a first step in order to have a first comprehension of it; however, it is not sufficient to have a concrete idea on it in terms of in which aspects poverty manifests itself in a given society. Thus, applied studies as quantitative or qualitative, both of them can demonstrate the existence of poverty in practice. As we have said qualitative method is more suitable for small scale observations. On the other hand, quantitative methods employing various social indicators show the extent of poverty in a given society by numbers. Notably, for policy makers and governments, poverty statistics are vital that help them to design and implement poverty eradication programs. The numbers are not only the proof of existence of poverty; but also, they are helpful for informing both public and governments about difficulties, needs, and prosperities of citizens regarding different elements of life. Social indicators are often at the core of contemporary relative deprivation, capabilities, social exclusion based multidimensional poverty studies. But of course, the indicators might be varying according to theoretical framework behind the approaches.

Amartya Sen's (1998 Nobel Prize Winner in Economics) capabilities approach is seen as the broadest theoretical framework amongst the other multidimensional approaches to poverty. Capabilities approach focuses on doings and beings called functionings that the person has at a given time and on the role of these functionings in accomplishing other functioning(s). In other words, combinations of functionings of a person at a given time correspond to his or her capability of realizing other functionings. In that vein, if the person achieves to obtain what he or she has valued and has right to value, then he or she is considered capable of accomplishing what he or she valued. In case of non-achievement regarding a functioning, the person can be seen as poor in that aspect of life. Here, the most important feature of the capability framework is the consideration of real liberty as the main indicator of poverty. If the person is capable of achieving what he or she values and has right to value, so that he or she got the real liberty in realizing his or her valued functioning. In opposition to classical monetary approaches, Sen in his framework wanted to emphasize the fundamental place of freedom/real liberty in increasing quality-of-life of people in any society (i.e., which is called 'development as freedom'). The main motivation behind this idea is that money and freedom do not necessarily live together in the sense that there is no rule that money brings intrinsically freedom. As the first principle of universal human rights and as part of natural rights, freedom of each single individual should be fulfilled by respecting human rights of all individuals in the society. Greater capability means greater real liberty. Greater an individual is capable of realizing various functionings, greater the person is prosperous and suffers less poverty. Even the capabilities approach seems fuzzy at first glance, it does in fact touch upon the most important thing that a person needs in his or her life from his or her arrival to the world. Capabilities approach has certainly broken the dominance of monetary view on poverty and has shed light on human development and freedom. Sen has also promoted the use of social indicators in the United Nations annual human development reports to measure poverty via capabilities, and most recently his approach has led to the emergence of internationally recognized Multidimensional Poverty Index (MPI).

Sen has also proposed the concept of basic capabilities to emphasize deprivation regarding the essential items of life including shelter, drinking water, food, clothes, dwelling conditions as well as some health and education issues such as child mortality, life expectancy, and educational attainment. This is especially very important in case of least developed countries such as in the region of Sub-Saharan Africa, which suffers significantly in terms of food shortage, hunger, starvation, child mortality, HIV, low life expectancy, no access to safe drinking water, no access to electricity, no appropriate energy sources, and vulnerability to natural disasters. The other more advanced developing countries in the Global South do less suffer regarding unmet basic needs, but they experience substantial deprivations regarding some other issues such as access to various public services, working conditions, informal employment, social security, safety, etc. In our study, we focus specifically on upper-middle-income countries that are supposed to be close to the situation of rich countries in some respects (e.g., material quality of life) but also at the same time they still witness under-development in some other aspects of life (e.g., safety issues, informal employment). This situation brings us immediately to accept that combinations of deprivations and non-deprivations are naturally part of life of people in these countries. So that, it is particularly interesting to observe both the progress and lags of such societies, which do comprehensively reflect complexity and diversity of poverty. The capabilities approach, therefore, flourishes as the most convenient framework to understand the nature and dynamics of poverty. Upper-middle-income countries are recognized as being good performers in terms of income per capita and industrial activities compared to other countries in the Global South; however, freedom of people is still limited in multiple aspects, and combinations of various deprivations and non-deprivations do clearly appear in the life of people such as not being deprived regarding many electronical materials but being deprived regarding educational attainment, social security at the same time.

Poverty is generally seen as a negative phenomenon making people feel poorer or disadvantaged in their society, and also it is seen as something symbolically reflecting the worst thing might happen to a person in his or her life. As many facts show, severe poverty and destitution are the living situations that no one deserves; but we have to also be realistic vis à vis the nature of livings. People generally spend their lives experiencing achievements, non-achievements, advantages, disadvantages. In the agenda of governments, people living in non-decent and hard conditions should be in the forefront; this is something, which cannot be ignored. But, on the other hand, society should also take her responsibility as a whole. In other words, not having capability of realizing something is not always of an individual failure. Related to that, in this study we aimed at highlighting the importance of social context and macro level functionings that correspond to the social facts concerning many people in society. As a human being is a social organism living in society with other people and the society is a dynamic process composed of individuals, hence each single individual encounters one way to another behaviors of others, different social acts, and therefore different social products. We assume that various social facts correspond to macro level functionings, which are part of individuals' capabilities set. In other words, social environment defines directly or indirectly the person's accessibility to different sources and means, and his or her way of processing in realizing functionings. If we don't follow same people over a certain period of time to observe their lives (their biography over time), we can't exactly know how their capabilities are configurated by macro level functionings. But macro-level functionings certainly reflect the nature of social environment such as household, family, community, country situation in which the person lives.

In the light of these explanations, we proposed categorizing functionings: micro-level functionings realized by the person herself for herself, macro-level functionings realized by society via perpetual complex interactions, and mixed macro-micro level functionings realized by a group of people for an individual or realized by an individual for a group of people. This typology puts at center the society, perceives individuals as social beings living in a certain dynamic in their society and comprehends poverty as a social phenomenon rather than an individual level non-achievement. Disadvantages resulting from macro, mixed macro-micro level functionings should be examined in a societal spirit rather than taking as an individual problem to be solved by the individual herself (i.e., the outputs of macro, mixed macro-micro level functinings cannot be seen as individual non-achievement or achievement). Societies can be efficiently developed in case every single individual takes her social responsibility, recognizes the rights, protection, and needs of others, and understands the situation of society in which they live. Investigating poverty for different social groups and from upper to bottom stratum will also help society to see clearly common poverty experiences as well as the different ways of experiencing poverty between distinct groups.

In the original framework of capabilities, Sen emphasizes much more economic factors in the determination of real liberty of people and he considers social interactions as an indirect factor having potential impacts on real liberty of people. However, in this thesis, we explained that social context is not an indirect driver but in fact it is a direct and the most fundamental element of real liberty. Our main argument is that the individual has bonds with the society and the society has bonds with the individual. For instance, in case a person is part of a manifestation for defending human rights, she connects with society where she lives (but also with the world in a broad sense via media- at global level). The output of this participation might be advantage or disadvantage for her but also for the rest of society in a broader sense. This activity corresponds to mixed macro-micro level functioning in the sense that she realizes a functioning for society in a context of a societal event (except the fact she is able to participate in such event is a personal achievement for her. Being able to be present physically at the manifestation place is a micro level functioning resulting in an achievement for her-achieving to be at the manifestation place. But, by realizing this activity she becomes part of a societal event destined for the whole society and not for her or another particular person. So, in this second perspective this functioning takes beyond personal micro aspect, and this is why we call it mixed macro-micro level functioning). Another example, we suppose a person finds herself unemployed for about 3 months due to pandemic. In this second case, the state in which the person finds herself at a given time results in a disadvantage for the person because she loses her job for 3 months and the event of pandemic is not a personal event or is not an event realized by a particular person or by a group of people only for her. Pandemic is a result of social interactions between individuals, and it is a societal event having impact on society as a whole. Under these characteristics, pandemic effect is a macro level functioning. If we look at profile of the person in terms of her capabilities at a given time (i.e., or it can be seen as a life situation/living conditions of the person at a given time), we can see that the person has different sorts of relationships with her entourage such as workplace, family, household, friends, public places such as streets, public transports, administration offices, etc. These relationships are the indicator of the nature of person's attachment with her society and therefore her social attachment profile shows us how she lives in terms of which activities and actions she is doing or did, in which places she visits or visited, the purposes of her actions and activities, people she encounters for job, administrative, family, friendship purposes. The way she is attached to

her society describes the way she lives, and as Sen defines capability as being able to live the lifestyle the person values; so that, we can deduce that person's social attachments reflect the capabilities of the person at a given time. Ignoring or taking as indirect tool the social bonds, therefore, will be leading to make a monumental error in well-being, quality-of-life, poverty studies. This is why we created a new typology of functionings to not only differentiate achievements, non-achievements, advantages, disadvantages, but most importantly to integrate social context via bonds into the capabilities and multidimensional poverty analysis.

As we have seen in the literature review that Paugam (1991; 2005; 2015 [2007], 37) shows different social attachment configurations as the indicators of how people are socially recognized and protected by society. These configurations include 4 types of social bonds that are: *organic participation bond* reflecting the relationship between the individual and her work, *parentage bond* focusing on family relationships, *elective participation bond* covering the relationship outside the family, with other people such as within institutions, religious community, diverse leisure groups, neighborhood, friendship, conjugal bond , and *citizenship bond* including the relationship between the person and the State functions such as accessibility to law, to rights, trust in people and trust in legal system. We argued that social context is the fundamental element of capabilities of people and to concretize this idea we proposed a new typology of functionings. At this point, we argue that Paugam's social bonds typology and social attachment approach support our intention to integrate social context into the capabilities analysis and also is the greatest way of showing the link between social context and capabilities. In the literature social exclusion is explained by some scholars as one of the results of lack of capabilities; however, as we have mentioned the link between human rights and socially perceived necessities should be revised carefully in developing countries and not everyone may feel socially excluded because she cannot afford socially perceived necessities. For these two reasons, Paugam's social attachment approach is more convenient because it brings a broader eye on social context including multiple types of social relationships the person encounters at a given time. In multidimensional poverty analysis: (i) organic participation bond is part of the dimensions of economic activity, labour market conditions; (ii) parentage bond is part of the dimensions of family relationship, household composition; (iii) elective participation bond is part of the dimensions of social life and social participation; (iv) citizenship bond is part of the dimensions of access to legal system, safety, human rights, social and economic rights, for instance. By combining indicators reflecting these 4 types of bonds, Paugam has found 3 main social attachment profiles that he described as integrated poverty, marginal poverty, and disqualified poverty. Inspired by Paugam's framework in terms of social attachment and poverty configurations, in this thesis we used a great number of indicators and we discovered the most relevant poverty configurations from a multidimensional perspective. Thus, in one way we applied an extended version of Paugam's social attachment framework by taking in consideration various additional dimensions such as dwelling conditions, educational attainment, material deprivation, health, financial situation to reflect how the person is in fact connected with society and what is the position of the person in society in terms of living conditions. And, in other way we showed that in fact social bonds are intrinsically part of capabilities. To improve capabilities of people, we have to understand how people are connected with society in terms of what society offers to them and what they have and what they do at a given time. Paugam's framework is the perfect way of operationalizing this idea that constitutes the foundation of our multidimensional poverty analysis. Current multidimensional poverty studies are number oriented focusing on headcount ratio and intensity

of poverty. However, poverty means nothing without any clear description of its roots as it is an abstract concept. This thesis, therefore, addresses this methodological gap in the poverty literature via the link between social bonds and capabilities, and by using patterns and profiles extraction statistical methods.

Currently, the most popular measure of capabilities is the Global Multidimensional Poverty Index used in annual UN human development reports. This index is based on Alkire and Foster (2007, rev. 2009b, 2011a) methodology (AF), which uses also counting process; however, the AF is focused on computing headcount ratio of poverty, adjusted head count ratio of poverty, intensity of poverty, and severity of poverty. These outputs are very number oriented, and they do not provide information on types of poverty. Contribution of each variable to the index is generally computed but it does not show the simultaneous combinations of variables. We argue that knowing the number of poor in society does not much help public, policy makers, governments because most importantly they have to know how the poor experience poverty (i.e., in terms of which elements of life and what is the social environment of the poor involving many items at the same time). Notably, for reducing inequalities between groups, governments need a comprehensive look at poverty configurations. For instance, deprivation regarding educational attainment might be a common problem for all groups; on the other hand, no access to drinking water, no access to electricity might be the concern of only the poorest group, and many other groups might experience non-deprivation regarding many material elements but might not be registered to social security at the same time.

As in the case of the Global Multidimensional Poverty Index, multidimensional poverty is often analyzed based on arbitrarily defined poverty lines that simply distinguish the society into two groups: poor vs. not poor. Differently from this binary approach, we proposed to focus on whole society through different poverty groups. Thanks to that, we were able to detect different types of poverty that each group experiences, and it helped also to see how the situation from upper to bottom classes gets worse. Regarding our methodology, we started off by building an additive composite index by counting the number of deprivations for each single individual in the data; then, as a function of the distribution of the index scores, different poverty groups were obtained. After that, we followed the following 4 steps: (i) step1 analysis: detecting multidimensional poverty patterns via Breiman et al.'s (1984) classification trees method- we found out the most relevant capabilities patterns (combinations of functionings) for each group. Based on the discovered combinations, the importance of relevant functionings was also computed; (ii) step2 analysis: detecting dimensional poverty profiles of poverty groups; (iii) step3 analysis: detecting various profiles of poverty groups; and (iv) step4 analysis: detecting multidimensional poverty profiles of specific groups. We showed all details about computations including R codes in the exemplary case study of South Africa (and the results of other three case studies are given in the appendix). In step1 analysis we obtained the multidimensional poverty configurations index (MPCI), which shows both the content and the frequency of each relevant configuration of poverty by 5 poverty groups. For that, we applied ten classification trees, namely *CAP111* for discovering the relevant configurations distinguishing GR1 and GR2, *CAP112* for GR1 vs. GR3, *CAP113* for GR1 vs. GR4, *CAP114* for GR1 vs. GR5, *CAP115* for GR2 vs. GR3, *CAP116* for GR2 vs. GR4, *CAP117* for GR2 vs. GR5, *CAP118* for GR3 vs. GR4, *CAP119* for GR3 vs. GR5, and *CAP1110* for GR4 vs. GR5. Given that the poorest group is GR5 and the least poor group is GR1, this analysis showed inequality between the groups in the sense that what are the main configurations distinguishing from each other. Thus, we also found out the common deprivations between the groups and how these deprivations differ in

terms of magnitude across the groups. In our empirical analysis, we used socio-economic surveys from the four upper-middle-income countries, namely South Africa, Turkey, Mexico, and Brazil to measure multidimensional poverty configurations. We applied the same methodology in four case studies. The results from South Africa, Turkey, Mexico, and Brazil analysis showed that in each group there is at least one form of poverty, and the majority cases are the combinations of deprivations and non-deprivations. But also, the poorest group in each country analysis is mainly distinguished from other groups by a combination of at least two deprivations.

Herewith, we showed that in case of many indicators reflecting multiple aspects of life, Breiman et al.'s (1984) classification trees method can be used to obtain the most relevant capabilities patterns. Moreover, we also showed that in case of less indicators such as dimensional analysis (step2 analysis), decomposition analysis (step3, step4 analysis), the profiles can be extracted as well without proceeding with classification trees. Profiles are simply the combinations of string indicators that have non-numerical characters, such as 0 is used for designating non-deprivation and 1 is used for deprivation. For instance, in case of two variables, namely educational attainment (EDU) and access to drinking water (WATD), we can obtain the following profiles: 00, 11, 10, 01 (e.g., the profile 00 means non-deprivation regarding both EDU and WATD). Given that step1 analysis is the essential of poverty examination, then the other steps involving extraction of the dimensional and decomposition profiles of the poverty groups enrich the analysis in terms of content. For instance, decomposition analysis in terms of geographic situation profiles shows where different poverty groups live as well as the distribution of discovered multidimensional poverty configurations by settings and regions. Or decomposition analysis in terms of social grants profiles shows types of grants the poverty groups receive (i.e., from the least to poorest groups).

In the literature, whether the capabilities approach is in an absolute or a relative term is much discussed. Sen clearly defines that poverty analysis in developing countries should include basic capabilities concept, which emphasizes the questions of hunger, starvation, access to food. However, Sen points out that he is not against to determine a list of functionings as a function of country/society situation and he is in favor of updating the list. Herewith, we can say that the notion of capabilities or in other words the real liberty is the absolute term in evaluation of poverty. According to capabilities approach, the criteria of poverty is the nature and extent of real liberty. In each society, human rights should be fulfilled; so, this is the absolute aspect of capabilities approach. In our analysis, we created many variables as much as possible as a function of the available questions in the questionnaires in order to reflect diversity of living conditions and living aspects. However, to define different groups of poverty from the least to the poorest, we took in consideration the relative distribution of total number of deprivations. To define indicators, we did not proceed with public opinions such as in the case of socially perceived necessities; but the dimensions defined in the basic needs approach (e.g., education, health, food, drinking water, electricity) are covered. The additional dimensions (e.g., material deprivation, household financial situation, various dwelling conditions, economic activity, social grants, social leisure) are determined as a function of questionnaire. Surveys we used contain mostly objective indicators and they do not include specific cultural matter questions such as habits, traditions, customs. To define deprivation threshold, we followed the universal human rights chart requirements. Thus, we can say that our analysis includes both absolute and relative aspects.

After showing different configurations of poverty by the four case studies, we also made comparison analysis. Although we created many indicators in each case study, the common variables to make a proper comparison between the four countries (South Africa, Turkey, Mexico, Brazil) are very few. For that reason, we did not apply classification trees method but instead we extracted profiles based on the indicators available in the surveys. First, we started by the comparison between the four countries at the same time, which is the main comparative analysis explained in chapter 6. Then, we continued with other comparisons such as the comparison between South Africa, Turkey and Brazil, the comparison between South Africa, Turkey and Mexico (which are shown in the appendix, and in total there are 11 comparisons). The main comparison model between the 4 countries called ESCAP_C1 (as the reflection of Sen's basic capabilities) contains only the variables of access to drinking water and educational attainment. The results showed that in terms of deprivation in both 2 indicators, South Africa has the worst performance followed by Mexico, Brazil, and Turkey. However, when it comes to non-deprivation in both 2 indicators, Brazil has the best performance followed by Mexico, Turkey, and South Africa. These results show that Turkey has the worst performance regarding 1 indicator compared to Brazil and Mexico. At this point we can say that the weak performance of Turkey regarding educational attainment plays a crucial role in displacing Turkey from 1st to 3rd position in the ranking of non-deprivation in 2 indicators at the same time. In terms of deprivation in only access to drinking water, Turkey has zero observation, Mexico and Brazil's performance are quite the same with around 5 per cent of the 18+ population concerned, and South Africa has the worst performance with around 12.5 per cent. On the other hand, in terms of deprivation in only educational attainment, South Africa has the best performance, Mexico and Brazil have close performances with 8 per centage points difference, and Turkey has the worst performance with 70 per cent of the 18+ population concerned. These results indicate us that: (i) the relationship between individuals of 18+ at the moment of the interview and the educational services in Turkey is quite weak (and it was the case when they were child); notably if the person could not go to school due to household reasons and/or no availability of schools in the neighborhood (in terms of both quality and quantity) should be considered by policy makers in the country; (ii) the relationship of 18+ population with access to public drinking water services is quite weak in South Africa, given that access to safe and clean drinking water is one of the services that the government should provide to their citizens. This result is probably due to environmental functionings and infrastructural macro level functionings. Access to education via public services (corresponding to macro level functionings) should also be provided by the government to their citizens as part of the universal human rights chart requirements. We cannot know what the exact causes behind these results as the causality analysis is out of scope of this thesis. We can say that for further research it is important for governments to explore the perceptions of citizens for access to drinking water in South Africa notably, and the public opinion on the importance of education in the 4 countries notably in Turkey. For instance, at present in Turkey it is obligatory the 12th level education but for next generations it will be very important to check the relationships/approaches of citizens to public educational services. South Africa's population is approximately 4 times less than Turkey, which may make easier the construction and implementation of public-school services in South Africa. In this perspective, implementation of water services could be also easier in South Africa compared to Turkey. But there are 2 main questions arise here in this discussion: on the one hand, ecological factors in South Africa making difficult to implement safe and clean public water system into dwellings (e.g., ecological factors may require specific installment systems with high costs) and/or habitation in bad dwellings in terms of infrastructure or massive

habitation in sowetos, townships, remote areas, suburbs far from cities and deprived of adequate infrastructure are the important factors for such situation in South Africa. On the other hand, Turkey has excellent performance in terms of access to drinking water as part of infrastructure system but at the same time implementation of public schools may have been remained weak in rural or suburbs areas of urban areas. In recent years, as our results show, there is a movement from rural to urban areas in Turkey. So that, it becomes important to provide necessary educational services in poor neighborhoods and it will be beneficial to know recent public opinion on educational attainment notably for girls. To increase the bond between public and government in terms of public educational services, the government needs information on why children cannot go to school. In Brazil and Mexico also, access to education is significantly more problematic compared to access to drinking water. In these 2 countries also given high population, infrastructural systems and trust in government for public education should be reviewed, notably among rural and suburbs of urban areas. In short term, no access to safe and clean drinking water will cause health problems (so that increase of costs of health services) and immediate diminishment of capability of leading a healthy and decent life. In long term, no access to education or no educational attainment will cause low capability of earning well and of obtaining high-paid jobs and diminishment of human capital in the country. As we can see, the results are not only at micro level but also at macro level having potential negative impact on society as a whole (i.e., macro level functionings generating disadvantages). Thus, these results support the importance of social context as one of the direct factors in capability analysis. We also note that as a function of the available common indicators, the more elaborated profile analysis could be done to demonstrate in more detail the differences between countries from social, economic, cultural perspectives but the common indicators are very few in this study. For instance, we don't have information on perception of parents on schooling of their children, how parents prefer to spend their money, whether instead of going to school the children getting married or are performing an economic activity, teaching quality. We, therefore, suggest designing household, living conditions survey on the basis of different social bonds the person might have with his or her entourage as we explained in our theoretical framework.

Our comparison analysis showed how it can be complicated to compare countries based on their national household surveys in case of the surveys are not harmonized in terms of questionnaire. Notably due to variations regarding the way the questions were asked, the target group of questions, the answer options of the questions in surveys, comparison becomes more delicate. We suggest that for accurate international comparisons, some essential questions such as access to drinking water, access to food, access to sanitation facilities (in terms of both toilet and shower), energy sources used for cooking, heating, access to electricity should be asked in each survey and in the same way. Additional questions specific to country context can be asked as well for obtaining a comprehensive multidimensional poverty examination at national level. Given that fulfilment of human rights is the absolute element of capabilities approach, we suggest designing poverty, living conditions, quality-of-life, general household, socio-economics type surveys according to articles defined in the universal human rights chart. This process requires operationalization of human rights (i.e., creating indicators of capabilities on the basis of universal human rights chart and other international recommendations and conventions to live a decent life). In line with Sen's ideas, we argue that basic capabilities including the most important elements of life (e.g., basic needs approach) should be included in poverty analysis especially concerning developing countries. Access to safe and clean drinking water in the dwelling, access to adequate daily food, living in adequate dwelling

conditions including roof, wall, floor, kitchen, bathroom (private shower and private toilet both), having adequate electricity, heating system, hot water system, energy source for cooking, energy source for lighting, rubbish removal system, tv, radio, telephone in the dwelling, living in safety (no conflict zone, no war, no physical, sexual, oral violence, no harassment), access to educational system (no child labour, no forced marriage for children), living in environment with access to health institutions/services, public transports, roads, legal services, public institutions are all of them basic functionings that every single human being needs and should have access to. From this perspective, surveys should include appropriate questions to detect statistics of these absolute functionings. Because of this, international collaboration with national statistical agencies and United Nations (as the most global institution on development, peace and human rights) is crucial to not only develop rich, adequate national surveys of capabilities and poverty but also to make fair multidimensional poverty comparisons across countries. Notably, response options concerning roof, wall, floor materials, type of dwelling, type of heating system, hot water system, type of energy sources for cooking, lighting vary between countries due to different practices or available sources (cultural, economic, geographical factors). Elaborated discussions at international level are needed to determine the same criteria regarding such response options for comparisons across countries.

In our analysis we created a substantial number of indicators, which is vital to reflect the multiple aspects of life. However, the data of 4 countries were quite limited in terms of subjective indicators. We employed general household, living conditions surveys that are generally more adequate to create objective indicators. As the commission of Stiglitz-Sen-Fitoussi has mentioned, for further research on quality-of-life it is important to employ both objective and subjective indicators. Subjective indicators allow to detect personal perception, personal evaluation, personal ideas regarding a functioning. For instance, life satisfaction surveys focus on different domains of life including general life satisfaction (overall life satisfaction), or satisfaction on a particular domain such as education, health, finance, democracy, safety situation of the country. Socially perceived necessities can be found also via subjective indicators in the sense that we can learn how importance citizens give to different items of life (e.g., whether having access to internet is important or not). More specifically, how people perceive poverty or whether the person feels being socially excluded from the mainstream of society or whether the person feels poor can be learnt via subjective indicators. In our Turkey analysis, for instance, we created two subjective indicators on health status and financial situation of the household. In the literature, poverty is generally measured based on objective indicators such as highest educational attainment, monthly household income level, being economically active or not, having an electricity at home or not, etc. As we have mentioned in the appendix 9.1, the approach of subjective poverty is quite recent; for instance, South Africa national statistical agency has published subjective poverty report in 2012 (which is not on regular basis). Measuring poverty based on only subjective indicators might be problematic in developing countries because of fear of answering (because of corruption, violence, be afraid of government or army), or due to social pressure via household relations (dominance of father figure or eldest male member of the family), customs, traditions, prevalent social beliefs, or people sometimes adapt themselves to situations and they perceive normal certain deprivations and violation of rights, or people do not have exact information on how important human rights, social and economic rights are. Thus, poverty risks to be under-estimated in such conditions. However, to capture expectations, observations, perceptions of people on different social phenomena and different aspects of life, subjective indicators are

important. We recommend that subjective poverty indicators can be used as complementary to objective indicators. For instance, having information on whether the person has access to legal system/services is important for ameliorating infrastructure system, reducing crime and violence, increasing safety, increasing democracy; however, it is also important to know whether people find necessary or not legal services. If many people do not find necessary legal services, government might follow the request of citizens and such situation will result in augmentation of violence in different forms (e.g., robbery, gender-based violence). Subjective indicators are quite used in election polls to identify requests, ideas, satisfactions, dissatisfactions of citizens. In a bad governance, policy makers and governments may use some results of subjective indicators to avoid investment in certain domains or to not spend money for reducing poverty if the majority of citizens are satisfied with their life. Moreover, subjective indicators can also be used for checking reliability or consistency of objective indicators in the data; for instance, the simple question on whether the person is covered by a medical aid, a medical help scheme or a private health insurance might be not sufficient in case of medical services are not of good quality. So that, indicators on the quality of medical/health services can be used for getting additional information on health coverage system. Being covered by a health insurance makes no sense if there are no hospitals, clinics, doctors, nurses, materials. Subjective approach to poverty is not in the scope of this thesis but we definitely suggest development of multidimensional poverty methodologies including both objective and subjective indicators in terms of selection, combination, use of indicators in the measurement process.

Through the extracted profiles, it is important to obtain a high level of information meaning that profile information of at least 90 per cent of the target population should be obtained. This in particular depends on the structure of data. Generally, with low number of items, we have less combinations that we can interpret almost 100 per cent of population, e.g., ESCAP profiles in our comparative analysis. But, despite a high number of items, if there are some dominant profiles meaning that there is a certain homogeneity in population, we will still be able to catch an important level of information. As we have shown in case studies, if there are a lot of items to be considered and the data structure is complex, then classification tree method can be used for discovering the most relevant patterns of capabilities. We found that there are not much available common variables between 4 countries; so that, dimensional poverty profile analysis has emerged as a suitable way of dealing with comparison. In some cases, we obtained a good level of information such as at least 93 per cent but in some other cases we got information for around 80 per cent of population. If we think that elimination of poverty is not an imminent act, we can say that even detecting 70 per cent of target population can be very helpful. At this point, however, there is one technical issue should be delicately considered in the analysis: if the researcher/analyst would like to obtain an overall importance of an item within the dimension for a comparison between countries/groups, he or she should take in consideration the amount of information obtained through all extracted profiles within the dimension. For instance, if the difference between 2 countries regarding an item is quite close, then per centage of unobtained information becomes crucial. Moreover, to deal with loss of information due to heterogeneity of profiles, groups of deprivation can be created according to quantiles, then dimensional poverty profiles can be proceeded within each group, and comparison of profiles between countries per poverty group can be done given that the indicators and quantiles criterion should be the same in each country.

To sum up, in this thesis we focused on multidimensional nature of poverty in upper-middle-income countries. We proposed a methodology composed of 4 steps to measure multidimensional configurations of poverty by using classification trees and profiles extraction methods. The idea of configurations of poverty is in fact already discussed in the literature by Piachaud (1981) and Paugam (2005). However, the authenticity of this thesis is about the methodology we proposed. In opposition to current poverty studies, we did not use the poverty line and we did not focus on the overall poverty rate (headcount ratio). We examined poverty in whole society, and we searched for the relevant configurations distinguishing the groups (from the least poor to the poorest in society). We demonstrated the nature of multidimensional poverty in terms of combinations of functionings and we provided the incidence of each relevant combination. Sen (1981, 1984, 1987a, 1992, 1999, 2009) has clearly mentioned that capabilities correspond to the combination of functionings. However, current capabilities-based multidimensional poverty studies including the most popular Alkire and Foster's multidimensional poverty index (2007, rev. 2009b, 2011a) focus on overall poverty rate and do not focus on the nature of these combinations. In this sense, the methodology we proposed in this thesis is a new way of measuring multidimensional poverty based on the capabilities approach. Recently, the profile analysis is already proposed by Fattore et al. (2011b, 2012), Fattore and Maggino (2014), Fattore (2016). The methodology proposed by the authors (POSET onwards) is based on the partial order theory and analyzes poverty (multidimensional ordinal deprivations) situation of individuals via the evaluation process, which takes the poverty line as the benchmark and produces a poverty degree to each single profile in the data. Beycan and Suter (2017) have applied this methodology (via the R package PARSEC- see Fattore and Arcagni 2014) to examine multidimensional poverty in Switzerland. Beycan and Suter (2017) have showed that POSET is very useful to better understand the complexity and extent of poverty, notably regarding deprivation situations of people above the poverty line, and combinations of deprivations and non-deprivations. Compared to the popular Alkire and Foster (called AF) (2007, rev. 2009b, 2011a) counting approach that is used to produce the global multidimensional poverty index (MPI), the POSET gives full information on the available profiles in the data. The AF methodology takes only into account people below the poverty line and censors all other information, which does not allow to see the other poverty situations in the data. But also, the AF methodology does not show the profiles of the poor. However, the POSET methodology is not very suitable in case of many variables because it becomes very difficult to interpret a large variety of profiles, and poverty degrees may change when the poverty line changes. In our study, we employ classification trees to find out the most relevant combinations of deprivations and non-deprivations as many profiles exist due to many indicators of poverty. We extract profiles in steps 2, 3, and 4 of the methodology- where the number of indicators is significantly less compared to step 1. But in the future as further research, by setting the poorest situation as the benchmark, POSET can be used in our step2, and step3 analysis.

According to our methodology, we recommend policy makers, governments applying the following steps to measure multidimensional poverty: (1) defining poverty as a lack of capabilities emphasizing the role of real liberty; (2) as a function of the structure of survey (questions in the survey questionnaire), designating multiple social indicators with the aim of reflecting as much as possible different aspects of life (but paying attention to not use similar indicators in terms of content, it is important that each indicator reflects a different aspect of life); (3) given that each indicator represents a particular functioning the person got at the time

of the survey interview or during the reference week of the survey, topologizing the selected functionings in terms of our proposed theoretical framework (i.e., macro level functionings, micro level functionings, and mixed macro-micro level functionings); (4) determining a deprivation line for each functioning as a function of the universal human rights chart. Deprivation lines designate in which conditions the person is considered deprived regarding the functioning; (5) on the proper data (eliminated from inconsistencies and missing values), creating/computing the defined social indicators and computing an additive index by counting the number of deprivations for each individual in the data. Then, each person will have a total/overall deprivation index score, and it is important to keep in mind that all indicators are weighted the same (of the same importance) as each of them reflects a peculiar aspect of life; (6) computing different poverty groups as a function of the distribution of the overall deprivation index scores in the data; (7) employing classification trees method, for instance the statistical package Rpart in R environment, for finding out the most relevant capabilities patterns that correspond to the most important combinations of functionings distinguishing the best the poverty groups (e.g., GR1 vs. GR2, GR2 vs. GR3, etc.); (8) as a function of classification performance via confusion matrix statistics, selecting the best comparison model for each poverty group; (9) computing also the overall importance of each functioning within each poverty group through the discovered patterns; (10) examining how deprivations groups differ between them in terms of experience of poverty (total poverty: combination of only non-achievements and/or disadvantages; intermediate poverty: combination of achievements, advantages, non-achievements, disadvantages; total prosperity: combination of only achievements and/or advantages), and how the capabilities situation gets worse from the least to the poorest group (i.e., poorest group is the group having relatively the highest overall deprivation index scores in the data). So that, we have information in terms of content and frequency of poverty; (11) carrying out dimensional poverty analysis focusing on each dimension separately: for each poverty group, finding out the most relevant profiles (combinations of functionings) within each dimension based on the discovered capabilities patterns; (12) carrying out decomposition analysis for obtaining a comprehensive picture of poverty, such as finding out the most relevant profiles within each poverty group regarding economic activity, geographical situation, social grants, household composition, youth population, elderly population, disabled people, rural areas based on the discovered capabilities patterns; (13) completing the analysis with the other macro level functionings computed by international organizations or national statistical offices. The aim is to take in consideration also the aspects that do not exist in the surveys employed in the empirical analysis. The results of this extern research cannot be directly combined with the empirical results, but they can be used as complementary information in evoking the country/region/community socio-economic situation; (14) preparing the poverty configurations report as a function of the obtained empirical results but also taking in consideration poverty alleviation and eradication programs implemented at national level to combat poverty in the country. Herewith, the aspects considered in the empirical analysis may be compared with the aspects included in the programs, but also the results may help to improve the content of programs. Moreover, in case of comparison between countries, regions, communities, the researcher needs to define first of all the common indicators by considering attentively the way the questions were asked in the survey questionnaires including target population, and available answer categories, before proceeding with any comparison. Then, as a function of available common functionings, repeating the steps mentioned above or just focusing on poverty profiles without carrying out classification trees in case there is not much common functionings.

We conclude by mentioning limits of this study and some suggestions for further research. From technical point of view, we faced data limitations in terms of subjective indicators (the 4 surveys we used for the analysis mostly contain questions to create objective indicators); for instance, satisfaction of citizens regarding some public services or legal services or safety, or personal evaluation of some public service qualities. Besides, in terms of objective indicators, statistics on some essential needs such as toilet, shower, food access could not be obtained due to lack of appropriate questions depending on the surveys. Especially, it is very important to ask separately toilet and shower and also ask whether sanitation facilities are shared or not in the dwelling where many households live together. In terms of educational attainment, the highest grade was the most reliable indicator; however, indicators related to teaching and material quality, and to distance between school and home, and to the transport used between school and home are crucial. The analysis covered people aged at least 18 yrs. because there were not enough questions to detect child poverty profiles. We emphasized the importance of social bonds and social attachment profile of the person to understand fully his or her capabilities. In this perspective, questionnaires were lacking questions related to the relationship with different institutions such as evaluation of contact with social assistance, or with other private charities, contact with employer (e.g., whether the person encounters harassment, violence, or dangerous activity at work), the behavior of household chief/head to other members, distribution of monetary and material resources between members of the household, trust in people, trust in community, trust in justice. For an accurate comparison between countries, the questions should be asked in the same way (including target population, reference group, answer options). Due to few available common indicators, our essential/basic capabilities profile comparisons remained limited.

From conceptual point of view, we conducted the analysis based on capabilities, which have a dynamic aspect. But, to detect the evolution of capabilities of a person over a certain period of time, we need to have detailed information on his or her goals in the life. The general household, living conditions surveys do not include such particular questions. For further research, we suggest combining qualitative and quantitative studies. Notably, ex-ante (before realization of the goal) and ex-post (after realization of the goal) capabilities analysis can be done through interviews. There are no specific national capabilities surveys conducted by national statistical agencies. For comprehensive multidimensional poverty analysis, tailored capabilities surveys are necessary. Capabilities especially put real liberty at the heart of the evaluation of well-being, quality-of-life of people. According to that, the fulfillment of human rights is extremely important. The questionnaires should be designed as a function of the universal human rights chart and at international platform these universal questions should exist in each national survey. In practice, the comparison of poverty across countries and the measurement of poverty in a country are two different purposes. On the one hand, national studies might include a larger variety of indicators to provide detailed sub-group analysis for addressing specific national policies. Particular country-context variables can be determined based on targets of specific national policies or the surveys on cultural activities, citizens' perception on different items of life are needed. In our study we did not have much country context indicators because the surveys we used are basically general household surveys aiming at showing some statistics on various living standards and income. On the other hand, international comparisons include in general less indicators as the particular country context variables remain out of scope. For this second case, we suggested employing the universal human rights chart to create indicators; however, an international dialogue is needed for solving the potential conflicts on the

determination of answer options such as regarding roof, wall, floor materials, dwelling type, energy source used for cooking, heating, lighting, toilet type, shower type. National statistical agencies might insist on different response categories depending on country's cultural practices and available sources. However, we did not elaborate in detail such technical issue as it does not enter into the core of this thesis and such topic is more suitable for a statistical dissertation.

Among upper-middle-income countries, very few of them apply actually multidimensional perspective to measure poverty. We showed in this thesis that poverty profiles are quite complex in upper-middle-income countries and given such diversity the multidimensional approach ensures a complete look on poverty in terms of both nature and extent. Actual multidimensional poverty studies employ poverty lines to separate poor vs. not poor and focus on the number of poor people in society. Nevertheless, we demonstrated that poverty is multi-faceted and different configurations of it emerge in each stratum. We suggest that further poverty research should fill this gap by aiming at discovering content of poverty rather than publishing headcount ratio of poor. We argue that instead of headcount ratio of poverty, the headcount ratio of poverty configurations should be published. For instance, detailed sub-group analysis revealing poverty configurations experienced by different social groups such as assisted people, youth population, unemployed, disabled, child, informal sector workers, etc., will be beneficial to understand how people are in relationship with their entourage. We have emphasized the substantial place of social context in the framework of capabilities, which considers social interactions as one of the indirect factors of capabilities. By typologizing in depth the functionings, we have explained that social bonds shape the capabilities of the person on daily basis. Capabilities surveys should be designed for covering not only deprivations vs. non-deprivations but also various features of social relationships/social attachment profile in extenso people have at a given time (i.e., friends, family members, colleagues, neighbors, institutions, public places). Specific qualitative studies focusing on daily social interactions of people and how they proceed to realize various functionings will be also contributive in the field of capabilities notably for designing specific capabilities surveys, creating a wide range of capabilities indicators, detecting the evolution of capabilities within the family from one generation to another as well as within the same community between different households or individuals.

8. Bibliography[6]

Abel-Smith, Brian, and Peter Townsend. 1965. *The Poor and the Poorest: A New Analysis of the Ministry of Labour's Family Expenditure Surveys of 1953–54 and 1960.* London, UK: Bell.

Abraham , Rosa A., and K. S. Kavi Kumar. 2008. "Multidimensional Poverty and Vulnerability." *Economic and Political Weekly* 43 (20): 79-87.

Abrahamson , Peter, and F. K. Hansen. 1996. *Poverty in the European Union, Report to the European Parliament* . Roskilde, DK: Centre for Alternative Social Analysis, Roskilde University .

Addison, Tony, David Hulme, and Ravi Kanbur, eds. 2009a. *Poverty Dynamics: Interdisciplinary Perspectives.* New York, NY: Oxford University Press.

Addison, Tony, David Hulme, and Ravi Kanbur. 2009b. "Poverty Dynamics: Measurement and Understanding from an Interdisciplinary Perspective." In *Poverty Dynamics: Interdisciplinary Perspectives*, edited by Tony Addison, David Hulme and Ravi Kanbur, 3-28. New York, NY: Oxford University Press.

Agarwal, Bina. 1994. *A Field of One's Own: Gender and Land Rights in South Asia.* Cambridge, UK: Cambridge University Press.

Alcock, Peter. (1993) 2006. *Understanding Poverty.* Basingstoke, UK: Palgrave Macmillan.

Alkire, Sabina. 2002. *Valuing Freedoms: Sen's Capability Approach and Poverty Reduction.* New York, NY: Oxford University Press.

—. 2005. "Why the Capability Approach?" *Journal of Human Development and Capabilities* 6 (1): 115-135.

—. 2007a. "Choosing Dimensions: The Capability Approach and Multidimensional Poverty." *Chronic Poverty Research Centre (CPRC) Working Paper* 88.

—. 2007b. "The Missing Dimensions of Poverty Data: Introduction to the Special Issue." *Oxford Development Studies* 35 (4): 347-359.

—. 2008a. "Choosing Dimensions: The Capability Approach and Multidimensional Poverty." In *The Many Dimensions of Poverty*, edited by Nanak Kakwani and Jacques Silber, 89-119. London, UK: Palgrave Macmillan.

—. 2008b. "Concepts and Measures of Agency." *OPHI University of Oxford Working Paper* 9.

—. 2008c. "Using the Capability Approach: Prospective and Evaluative Analyses." In *The Capability Approach: Concepts, Measures and Application*, edited by Flavio Comim, Mozaffar Qizilbash and Sabina Alkire, 26-49. Cambridge, UK: Cambridge University Press.

—. 2015. "Capability Approach and Well-being Measurement for Public Policy." *OPHI University of Oxford Working Paper* 94.

Alkire, Sabina, and Mauricio Apablaza. 2016. "Multidimensional Poverty in Europe 2006–2012: Illustrating a Methodology." *OPHI University of Oxford Working Paper* 74.

[6] Bibliography is organized according to the Chicago Manual of Style Sixteenth Edition (The University of Chicago Press 2010, 785-810). In case the full name of author is not available, then only the first letter of his or her name is mentioned in capital. And the bibliography includes the references mentioned in the Appendix as well.

Alkire, Sabina, Mauricio Apablaza, and Euijin Jung. 2014. "Multidimensional Poverty Measurement for EU-SILC." *OPHI University of Oxford Research in Progress Series* 36c.

Alkire, Sabina, and Rufus Black. 1997. "A Practical Reasoning Theory of Development Ethics: Furthering the Capabilities Approach." *Journal of International Development* 9 (2)*:* 263-279.

Alkire, Sabina, Adriana Conconi, Gisela Robles, José M. Roche, Maria Emma Santos, Suman Seth, and Ana Vaz . 2016. "The Global Multidimensional Poverty Index (MPI): 5-year methodological note." *OPHI University of Oxford Briefing* 38.

Alkire, Sabina, and Sévèrine Deneulin. 2009. "The Human Development and Capability Approach." In *An Introduction to Human Development and Capability Approach*, edited by Sévèrine Deneulin and Lila Shahani, Chapter 2. Abingdon, UK: Eartscan

Alkire, Sabina, and James E. Foster. 2007. "Counting and Multidimensional Poverty Measurement." *OPHI University of Oxford Working Paper* 7.

—. 2009a. "An Axiomatic Approach to Identification and Measurement of Multidimensional Poverty." *OPHI University of Oxford Research in Progress* 21.

—. 2009b. "Counting and Multidimensional Poverty Measurement (revised and updated)." *OPHI University of Oxford Working Paper* 32.

—. 2011a. "Counting and Multidimensional Poverty Measurement." *Journal of Public Economics* 95 (7-8)*:* 476-487.

—. 2011b. "Understandings and Misunderstandings of Multidimensional Poverty Measurement." *Journal of Economic Inequality* 9*:* 289-314.

—. 2016. "Dimensional and Distributional Contributions to Multidimensional Poverty." *OPHI University of Oxford Working Paper 100.*

Alkire, Sabina, James E. Foster, and Maria Emma Santos. 2011a. "Where Did Identification Go?" *Journal of Economic Inequality* 9: 501-505.

Alkire, Sabina, James E. Foster, Suman Seth, Maria Emma Santos, José Manuel Roche, and Paola Ballon. 2015. *Multidimensional Poverty Measurement and Analysis.* Oxford University Press.

Alkire, Sabina, and Gisela Robles. 2016. "Measuring Multidimensional Poverty: Dashboards, Union Identification, and the Multidimensional Poverty Index (MPI)." *OPHI University of Oxford Research in Progress* 46a.

Alkire, Sabina, José Manuel Roche, and Ana Vaz. 2017. "Changes over Time in Multidimensional Poverty: Methodology and Results for 34 Countries." *World Development* 94*:* 232-249.

Alkire, Sabina, José Manuel Roche, Maria Emma Santos, and Suman Seth. 2011b. "Multidimensional Poverty Index 2011: Brief Methodological Note." *OPHI University of Oxford Briefing* 07.

Alkire, Sabina, and Maria Emma Santos. 2009. "Poverty and Inequality Measurement." In *An Introduction to the Human Development and Capability Approach*, edited by Sévèrine Deneulin and Lila Shanai, Chapter 6. Abingdon, UK: Earthscan.

—. 2010. “Acute Multidimensional Poverty: A New Index for Developing Countries.” *OPHI University of Oxford Working Paper* 38.

—. 2013. "A Multidimensional Approach: Poverty Measurement & Beyond." *Social Indicators Research* 112 (2): 239-257.

—. 2014. "Measuring Acute Poverty in the Developing World: Robustness and Scope of the Multidimensional Poverty Index." *World Development* 59: 251-274.

Alkire, Sabina, Maria Emma Santos, Suman Seth, and Gaston Yalonetzky. 2010. "Is the Multidimensional Poverty Index Robust to Different Weights?" *OPHI University of Oxford Research in Progress* 22a.

Alkire, Sabina, and Suman Seth. 2015. "Multidimensional Poverty Reduction in India between 1999 and 2006: Where and How?" *World Development* 72*:* 93-108.

—. 2016. "Identifying Destitution through Linked Subsets of Multidimensionally Poor: An Ordinal Approach." *OPHI University of Oxford Working Paper 99.*

Alkire, Sabina, and Andy Sumner. 2013. "Multidimensional Poverty and the Post-2015 MDGs." *Development* 56 (1): 46-51.

Álvarez-Miranda, Berta. 2011. "In-Work Poverty among Immigrants." In *Working Poverty in Europe. Work and Welfare in Europe*, edited by Neil Fraser, Rodolfo Gutiérrez and Ramón Peña-Casas, 250-277. London, UK: Palgrave Macmillan.

Alwang, Jeffrey, Paul B. Siegel, and Steen L. Jorgensen. 2001. "Vulnerability: A View from Different Disciplines." *Social Protection Discussion Paper* 0115, Washington, D.C.: World Bank.

Anand, Paul, and Martin van Hees. 2006. "Capabilities and Achievements: An Empirical Study." *Journal of Socio-Economics* 35 (2): 268-284.

Anand, Paul, Graham Hunter, and Ron Smith. 2005. "Capabilities and Well-Being: Evidence based on the Sen–Nussbaum Approach to Welfare." *Social Indicators Research* 74: 9-55.

Anand, Paul, Graham Hunter, Ian Carter, Keith Dowding, Francesco Guala, and Martin van Hees. 2009. "The Development of Capability Indicators." *Journal of Human Development and Capabilities* 10 (1)*:* 125-152.

Anand, Paul, Cristina Santos, and Ron Smith. 2008. "The Measurement of Capabilities." In *Poverty, Capabilities and Measurement*, edited by Paul Anand, Cristina Santos and Ron Smith, Chapter 16. Oxford, UK: Oxford University Press.

Anand, Sudhir, and Christopher J. Harris. 1994. "Choosing a Welfare Indicator." *American Economic Review* 84 (2)*:* 226-231.

Anand, Sudhir, and Amartya K. Sen. 1993. "Human Development Index: Methodology and Measurement." *UNDP Human Development Report Office, Occasional Paper* 12.

—. 1994. "Sustainable Human Development: Concepts and Priorities." *UNDP Human Development Report Office 1994 Occasional Papers.*

—. 1997. "Concepts of Human Development and Poverty: A Multidimensional Perspective." In *Poverty and Human Development: Human Development Papers 1997*, edited by UNDP, 1-20. New York, NY: UNDP.

Anderson, Elijah. 1991. "Neighborhood Effects on Teenage Pregnancy." In *The Urban Underclass*, edited by Christopher Jencks and Paul E. Peterson , 375-398. Washington, D.C.: The Brookings Institution.

Anderson, Elizabeth S. 1999. "What is the Point of Equality?" *Ethics* 109 (2): 287-337.

Andreß , Hans-Jürgen, and Henning Lohmann. 2008. *The Working Poor in Europe : Employment, Poverty and Globalization*. Cheltenham, UK : E. Elgar.

Angulo, Roberto, Yadira Diaz, and Renata Pardo. 2016. "The Colombian Multidimensional Poverty Index: Measuring Poverty in a Public Policy Context." *Social Indicators Research* 127*:* 1-38.

Annoni, Paola, Rainer Bruggemann, and Lars Carlsen. 2015. " Multidimensional View on Poverty in the European Union by Partial Order Theory." *Journal of Applied Statistics* 42 (3): 535–554.

—. 2017. "Pecularities in Multidimensional Districtal Poverty." In *Partial Order Concepts in Applied Sciences*, edited by Marco Fattore and Rainer Bruggemann, 121–133. Berlin, DE: Springer.

Annoni, Paola, Marco Fattore, and Rainer Bruggemann. 2011. "A Multi-Criteria Fuzzy Approach for Analyzing Poverty Structure." *Statistica & Applicazioni Special Issue:* 7-30.

Antonopoulos, R., T. Masterson, and A. Zacharias. 2012. *The Interlocking of Time and Income Deficits: Revisiting Poverty Measurement, Informing Policy Response.* Undoing Knots, Innovating for Change Series. New York, NY: Regional Centre for Latin America and the Caribbean, United Nations Development Program.

Arndt, Christian, and Jurgen Volker. 2011. "The Capability Approach: A Framework for Official German Poverty and Wealth Reports." *Journal of Human Development and Capabilities* 12 (3): 311-337.

Asselin, Louis-Marie, and Vu Tuan Anh. 2008. "Multidimensional Poverty and Multiple Correspondance Anaylsis." In *Quantitative Approaches to Multidimensional Poverty Measurement*, edited by Nanak Kakwani and Jacques Silber, 80-103. London, UK: Palgrave Macmillan.

Atkinson, Anthony B. 1970. "On the Measurement of Inequality." *Journal of Economic Theory* 2 (3:) 244-263.

— . 1987. "On the Measurement of Poverty." *Econometrica* 55 (4)*:* 749-764.

—. 1992. "Measuring Poverty and Differences in Family Composition." *Economica*, 59: 1-16.

—. 1998. "Social Exclusion, Poverty and Unemployment." In *Exclusion, Employment and Opportunity*, edited by Anthony B. Atkinson and John Hills, 1-20. London, UK: Center for Analysis of Social Exclusion, London School of Academics.

—. 2003. "Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches." *Journal of Economic Inequality* 1 (1)*:* 51-65.

Atkinson, Anthony B., Bea Cantillon, Eric Marlier, and Brian Nolan. 2002. *Social Indicators. The EU and Social Inclusion.* Oxford, UK: Oxford University Press.

Atkinson, Anthony B. , and Francois Bourguignon. 2001. "Poverty and Inclusion from a World Perspective." In *Governance, Equity and Global Markets*, edited by Joseph Stiglitz and Pierre-Alain Muet. New York, NY: Oxford University Press.

Atkinson, Anthony B., and Eric Marlier. 2010a. *Analysing and Measuring Social Inclusion in a Global Context.* New York: Economic and Social Affairs. United Nations.

—, eds. 2010b. *Income and Living Conditions in Europe.* European Union.

Atkinson, Anthony B., Alan K. Maynard, and Chris G. Trinder . 1983. *Parents and Children: Incomes in Two Generations* . London, UK: Heinemann Educational Books.

Atkinson, Anthony B. , L. Rainwater, and T. M. Smeeding. 1995. "Income Distribution in OECD Countries." *OECD Social Policy Studies* 18.

Atkinson, Elizabeth J., and Terry M. Therneau. 2000. *An Introduction to Recursive Partitioning Using the RPART Routines*. Mayo Foundation. Available at (accessed 25 May 2020): https://www.mayo.edu/research/documents/rpartminipdf/doc-10027257

Auletta, Ken. (1982) 1983. *The Underclass*. New York, NY: Vintage Books.

Bader, Christoph, Sabin Bieri, Urs Wiesmann, and Andreas Heinimann. 2016. "A Different Perspective on Poverty in Lao PDR: Multidimensional Poverty in Lao PDR for the Years 2002/2003 and 2007/2008." *Social Indicators Research* 126: 483-502.

Baffoe, Gideon, and Hirotaka Matsuda. 2017. "Why Do Rural Communities Do What They Do in the Context of Livelihood Activities? Exploring the Livelihood Priority and Viability Nexus." *Community Development* 48 (5): 715-734.

Baker, Fred A. 1993. "Classification and Regression Tree Analysis for Assessing Hazard of Pine Mortality Caused by Heterobasidion Annosum." *Plant Disease* 77 (2): 136-139.

Bane, Mary Jo, and David T. Ellwood. 1986. "Slipping into and out of Poverty: The Dynamics of Spells." *Journal of Human Resources* 21 (1): 1-23.

Banfield, Edward C. 1968. *The Unheavenly City: The Nature and Future of our Urban Crisis*. Boston, MA: Little, Brown and Company.

Barnett, Henrietta A. 1886. "National Enemies and National Defences." *National Review* vii (July): 652-667.

Barrientos, Armando. 2013. "Does Vulnerability Create Poverty Traps?" In *Chronic Poverty: Concepts, Causes and Policy*, edited by Andrew Shepherd and Julie Brunt, 85-111. London, UK: Palgrave Macmillan.

Basu, K. 1987. "Achievements, Capabilities, and the Concept of Well-being." *Social Choice and Welfare* 4: 69-76.

Batchelder, Alan B. 1971. *Economics of Poverty*. John Wiley & Sons.

Batson, Edward. 1941. *The Poverty Datum Line*. Cape Town, ZA: University of Cape Town.

—. 1945. *The Poverty Line in Salisbury*. Cape Town, ZA: University of Cape Town.

—. 1950. *A Further Recalculation of the Poverty Datum Line* . Cape Town, ZA: University of Cape Town.

Battiston, Diego, Guillermo Cruces, Luis Felipe Lopez Calva, Maria Ana Lugo, and Maria Emma Santos. 2009. "Income and Beyond: Multidimensional Poverty in Six Latin American Countries." *OPHI University of Oxford Working Paper* 17.

Baulch, Bob, and John Hoddinott. 2000. "Economic Mobility and Poverty Dynamics in Developing Countries." *Journal of Development Studies* 36 (6): 1-24.

Baulch, Bob, and Edoardo Masset. 2003. "Do Monetary and Nonmonetary Indicators Tell the Same Story About Chronic Poverty? A Study of Vietnam in the 1990s." *World Development* 31 (3): 441-453.

Bayes, Thomas. 1763. "An Essay towards Solving a Problem in the Doctrine of Chances." *Philosophical Transactions of the Royal Society of London* 53: 370-418.

Becker, Howard S. 1963. *Outsiders: Studies in the Sociology of Deviance*. Glencoe, ILO: The Free Press.

Beegle, Kathleen, Joachim de Weerdt, and Stefan Dercon. 2006. "Adult Mortality and Consumption Growth in the Age of HIV/AIDS." *World Bank Policy Research Working Paper* 4082.

Bentham, Jeremy. (1789) 1948. *An Introduction to the Principles of Morals and Legislation*. New York, NY: Hafner Publishing Co.

Berger, Peter L., and Thomas Luckmann. 1966. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York, NY: Doubleday.

Bergman, Jos. 1995. "Social Exclusion in Europe: Policy Context and Analytical Framework." In *Beyond the Threshold*, edited by Graham Room, 10-28. Bristol, UK: The Policy Press.

Bettio, Francesca, and Alina Verashchagina. 2009. *Gender Segregation in the Labour Market: Root Causes, Implications and Policy Responses in the EU*. Luxembourg, LU : European Commission.

Beveridge, Sir William. 1942. *Social Insurance and Allied Services*. London: His Majesty's Stationery Office.

Beycan, Tugce, and Christian Suter. 2017. "Application of Partial Order Theory to Multidimensional Poverty Analysis in Switzerland." In *Partial Order Concepts in Applied Sciences*, edited by Marco Fattore and Rainer Bruggemann, 135-150. Springer.

Beycan, Tugce, B.P. Vani, Rainer Bruggemann, and Christian Suter. 2019. "Ranking Karnataka Districts by the Multidimensional Poverty Index (MPI) and by Applying Simple Elements of Partial Order Theory." *Social Indicators Research* 143: 173–200.

Bird, Kate. 2013. "The Intergenerational Transmission of Poverty: An Overview." In *Chromic poverty: Concepts, Causes and Policy*, edited by Andrew Shepherd and Julie Brunt, 60-84. London, UK: Palgrave Macmillan.

Bishop, Christopher M. (1995) 2006. *Pattern Recognition and Machine Learning*. Springer.

Blacker, Charles P. 1952. *Problem Families: Five Inquiries*. London, UK: Blackwell Publising Ltd.

Blaylock, James R., and David M. Smallwood. 1986. "An Alternative Approach to Defining and Assessing Poverty Thresholds." *Western Journal of Agricultural Economics*, Western Agricultural Economics Association 11 (1): 1-6.

Blum, R.L. 1982. "Discovery, Confirmation, and Incorporation of Causal Relationships from a Large Time-Oriented Clinical Database: the RX Project." *Comput Biomed Res.* 15 (2): 164-187.

Blumer, Herbert. (1969) 1986. *Symbolic Interactionism : Perspective and Method*. Berkeley, CA: University of California Press.

Boarini, Romina, and Marco Mira d'Ercole. 2006. "Measures of Material Deprivation in OECD Countries." *OECD Social, Employment and Migration Working Papers* 37.

Boggess, Scott, Mary Corcoran, and Stephen P. Jenkins. 1999. *Cycles of Disadvantage?* Wellington, N.Z.: Institute of Policy Studies.

Boltvinik, Julio. 1992. "El Método de Medición Integrada de la Pobreza. Una Propuesta para su Desarrollo." *Comercio Exterior* 42 (4): 354-365.

Bonoli, Giuliano. 1997. "Classifying Welfare States: A Two-Dimension Approach." *Journal of Social Policy* 26 (3): 351-372.

Booth, Charles. 1902-3. *Life and Labour of the People in London*, 17 vols. . London, UK: Macmillan.

Bosanquet, Helen. 1902. "Poverty: A Study of Town Life." *Charity Organisation Review* 11 (65): 260-266.

Bourdieu, Pierre. 1979. *La Distinction: Critique Sociale du Jugement* . Les Editions de Minuit.

—. 1993. *La Misère du Monde*. Editions du Seuil.

Bourguignon, Francois, and Satya. R. Chakravarty. 2003. "The Measurement of Multidimensional Poverty." *Journal of Economic Inequality* 1 (1): 25-49.

Bowley, Arthur L., and Alexander R. Burnett-Hurst. 1915. *Livelihood and Poverty: A Study in the Economic Conditions of Working Class Households in Northampton, Warrington, Stanley and Reading*. London, UK: G. Bell and Sons, Ltd.

Bowpitt, Graham. (2000) 2016. "Poverty and its Early Critics: the Search for a Value-Free Definition of the Problem." In *Getting the Measure of Poverty: The Early Legacy of Seebohm Rowntree*, edited by Jonathan Bradshaw and Roy Sainsbury. Routledge.

Boyden, Jo, and Elizabeth Cooper. 2009. "Questioning the Power of Resilience: Are Children up to the Task of Disrupting the Transmission of Poverty?" In *Poverty Dynamics: Interdisciplinary Perspectives*, edited by Tony Addison, David Hulme and Ravi Kanbur. Oxford, UK: Oxford University Press.

Bradshaw, Jonathan R. 1993. *Budget Standards for the United Kingdom, Studies in Cash & Care*. Avebury, UK: Aldershot.

—. 1996. "Family Policy and Family Poverty." *Policy Studies* 2 (17): 93-106.

Bradshaw, Jonathan, David Gordon, Ruth Levitas, Sue Middleton, Christina Pantazis, Sarah Payne, and Peter Townsend. 1998. *Perceptions of Poverty and Social Exclusion*. Bristol, UK: Townsend Center for International Poverty Research, University of Bristol.

Bradshaw, Jonathan R., Deborah Mitchell, and Jane Morgan. 1987. "Evaluating Adequacy: The Potential of Budget Standards." *Journal of Social Policy* 16 (2): 165-181.

Bradshaw, Jonathan R., and Carol Stimson. 1997. *Using Child Benefit in the Family Budget.* London, UK: The Stationery Office.

Brady, David. 2003. "Rethinking the Sociological Measurement of Poverty." *Social Forces* 81 (3): 715-751.

Brady, David. 2009. *Rich Democracies, Poor People: How Politics Explain Poverty*. New York, NY: Oxford University Press.

Brady, David, and Denise Kall. 2007. "Nearly Universal, but Somewhat Distinct: The Feminization of Poverty in Affluent Western Democracies, 1969–2000." *Luxembourg Income Studies Working Paper Series, Working Paper* 462.

Bramer, Max. 2007. *Principles of Data Mining*. Springer.

Brauer, Carl M. 1982. "Kennedy, Johnson, and the War on Poverty." *Journal of American History* 69 (1): 98-119.

Breiman, Leo. 1996. "Bagging Predictors." *Machine Learning* 24: 123-140.

—. 2001. "Random Forests." *Machine Learning* 45: 5–32.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. 1984. *Classification and Regression Trees*. Chapman & Hall/CRC.

Briggs, Asa. 1961. "The Welfare State in Historical Perspective." *European Journal of Sociology* 2: 221-258.

Briguglio, Lino. 1995. "Small Island States and their Economic Vulnerabilities." *World Development*, 23 (9): 1615-1632.

—. 1997. *Alternative Economic Vulnerability Indices for Developing Countries*. Report prepared for the Expert Group on Vulnerability Index, United Nations.

Briguglio, Lino, Gordon Cordina, Nadia Farrugia, and Stephanie Vella. 2009. "Economic Vulnerability and Resilience: Concepts and Measurements." *Oxford Development Studies* 37 (3): 229-247.

Briguglio, Lino, and Waldemar Galea. 2003. "Updating the Economic Vulnerability Index." *Malta Islands and Small States Occasional Papers on Islands and Small States* No. 2003-4.

Brown, Muriel, and Nicola Madge. 1982. *Despite the Welfare State: A Report on the SSRC/DHSS* Programme of Research into Transmitted Deprivation. London.

Buchanan, Bruce G., and Tom M. Mitchell. 1978. "Model-directed Learning of Production Rules." In *Pattern-directed Inference Systems*, edited by D.A. Waterman and Frederick Hayes-Roth, 297-312. Academic Press.

Buckner, John C. 1988. "The Development of an Instrument to Measure Neighborhood Cohesion." *American Journal of Community Psychology* 16: 771–791.

Budowski, Monica, and Christian Suter. 2002. " Living Arrangement and the Dynamics of Poverty: Lone Parenting and the Life Courses of Mothers". *Swiss Journal of Sociology* 28 (2): 319-338.

Budowski, Monica, Robin Tillmann, Wiebke Keim, and Michèle Amacker. 2010. "Conceptualizing "Precarious Prosperity" for Comparative Research." In *Inequality Beyond Globalization: Economic Changes, Social Transformations, and the Dynamics of Inequality*, edited by Christian Suter, 201-222. Zurich, CH: Lit Verlag GmbH & Co. KG Wien.

Bulmer, Martin, Kevin Bales, and Kathryn K. Sklar. 1991. *The Social Survey in Historical Perspective, 1880–1940*. Cambridge, UK: Cambridge University Press.

Burchardt, Tania. 2000. "Social Exclusion: Concepts and Evidence." In *Breadline Europe: The Measurement of Poverty*, edited by David Gordon and Peter Townsend. Bristol, UK: The Policy Press.

—. 2004. "Capabilities and Disability: The Capabilities Framework and the Social Model of Disability." *Disability & Society* 19 (7): 735-751.

Burchardt, Tania, Julian Le Grand, and David Piachaud. 1999. "Social Exclusion in Britain 1991-1995." *Social Policy Administration* 33 (3): 227-244.

Burchardt, Tania, Julian Le Grand, and David Piachaud. 2002. "Degrees of Exclusion: Developing a Dynamic, Multidimensional Measure." In *Understanding Social Exclusion*, edited by John Hills, Julian Le Grand and David Piachaud, 30-43. Oxford University Press.

Burchardt, Tania, and Polly Vizard. 2011. "Operationalizing' the Capability Approach as a Basis for Equality and Human Rights Monitoring in Twenty-first-century Britain." *Journal of Human Development and Capabilities* 12 (1): 91-119.

Burchi, Francesco, Pasquale De Muro, and Eszter Kollar. 2014. "Which Dimensions Should Matter for Capabilities? A Constitutional Approach." *Ethics and Social Welfare* 8 (3): 233-247.

Burgess, John, and Ian Campbell. 1998. "The Nature and Dimensions of Precarious Employment in Australia." *Labour & Industry: A Journal of the Social and Economic Relations of Work* 8 (3): 5-21.

Callan, Tim, Brian Nolan, B. J. Whelan, D.F. Hannan, and S. Creighton. 1989. "Poverty, Income and Welfare in Ireland." *ESRI Research Paper* 146.

Callan, Tim, and Brian Nolan. 1991. "Concepts of Poverty and The Poverty Line: A Critical Survey of Approaches to Measuring Poverty." *Journal of Economic Surveys* 5 (3): 243-262.

Calvo, Cesar. 2008. "Vulnerability to Multidimensional Poverty: Peru, 1998–2002." *World Development* 36 (6): 1011–1020.

Calvo, Cesar, and Stefan Dalco. 2005. "Measuring Individual Vulnerability." *University of Oxford Department of Economics Working Paper Series* 229.

Campbell, Angus. 1976. "Subjective Measures of Well-being." *American Psychologist* 31(2): 117–124.

—. 1981. *The Sense of Well-Being in America: Recent Patterns and Trends*. New York, NY: McGraw-Hill.

Campbell, Angus, Philip E. Converse, and Willard L. Rodgers. 1976. *Quality of American Life, The: Perceptions, Evaluations, and Satisfactions*. Russell Sage Foundation.

Canberra Group . 2011. *Handbook on Household Income Statistics*. Geneva: United Nations.

Cantril, Hadley. 1965. *The Pattern of Human Concern*. New Brunswick, NJ: Rutgers University Press.

Carletto, Gero, and Alberto Zezza. 2006. "Being Poor, Feeling Poorer: Combining Objective and Subjective Measures of Welfare in Albania." *Journal of Development Studies 42 (5): 739-760.*

Carré, Jacques. 2000. *Les Visiteurs du Pauvre: Anthologie d'Enquêtes Britanniques sur la Pauvreté Urbaine (XIX-XX siècles)*. Paris: Karthala.

Carter, Michael R., and Christopher B. Barrett. 2006. "The Economics of Poverty Traps and Persistent Poverty: An Asset-based Approach." *Journal of Development Studies* 46 (2): 178-199.

Carter, Michael R., and Julian May. 1999. "Poverty, Livelihood and Class in Rural South Africa." *World Development* 27 (1): 1-20.

Carter, Michael R., and Munenobu Ikegami. 2009. "Looking Forward: Theory Based Meausres of Chronic Poverty and Vulnerability." In *Poverty Dynamics: Interdisciplinary Perspectives*, edited by Tony Addison, David Hulme and Ravi Kanbur, 128-153. Oxford, UK: Oxford University Press.

Castel, Robert. 1978. "La "Guerre à la Pauvreté" aux États-Unis : Le Statut de l'Indigence dans une Société d'Abondance ." *Actes de la Recherche en Sciences Sociales* 19: 47-60.

—. 1990. "Le Roman de la Désaffiliation. A' propos de Tristan et Iseult." *Le Débat* 61: 152-164.

—. 1991. "De l'Indigence à l'Exclusion, la Désaffilitaion." In *Face à l'Exclusion, le Modèle Francais*, edited by Jacques Donzelot, 137-168. Paris: Editions Esprit.

—. 1994. "La Dynamique des Processus de Marginalisation: De la Vulnérabilité à la Désaffiliation." *Cahiers de Recherche Sociologique* 22: 11-27.

—. 1995a. *Les Métamorphoses de la Question Sociale. Une Chronique du Salariat*. Librairie Arthème Fayard.

—. 1995b. "Les Pièges de l'Exclusion." *Lien social et Politiques* 34: 13-21.

—. 1996. "Les Marginaux dans l'Histoire." In *L'Exclusion: l'Etat des Savoirs*, edited by Serge Paugam. Paris: Editions La Découverte.

—. 2003. *L'Insécurité Sociale. Qu'est-ce qu'Etre Protégé?* Paris: Seuil-La Républiques des idées.

—. 2009. *La Montée des Incertitudes : Travail, Protections, Statut de l'Individu*. Paris: Editions du Seuil.

Castilla, Carolina. 2012. "Subjective Well-being and Reference-Dependence: Insights from Mexico." *Journal of Economic Inequality* 10: 219–238.

Cellini, Stephanie Riegg, Signe-Mary McKernan, and Caroline Ratcliffe. 2008. "The dynamics of Poverty in the United States: A Review of Data, Methods, and Findings." *Journal of Policy Analysis and Management* 27 (3): 577-605.

Cerioli, Andrea, and Sergio Zani. 1990. "A Fuzzy Approach to the Measurement of Poverty." In *Income and Wealth Distribution, Inequality and Poverty. Studies in Contemporary Economics*, edited by Camilo Dagum and Michele Zenga, 272-284. Heidelberg, DE: Springer.

Chakravarty, Satya R. 1983. "A New Index of Poverty." *Mathematical Social Sciences* 6 (3): 307-313.

Chakravarty, Satya R., Diganta Mukherjee, and Ravindra R. Ranade. 1998. "On the Family of Subgroup and Factor Decomposable Measures of Multidimensional Poverty." *Research on Economic Inequality* 8: 175-194.

Chakravarty , Satya R., and Jacques Silber. 2008. "Measuring Multidimensional Poverty: The Axiomatic Approach." In *Quantitative Approaches to Multidimensional Poverty Measurement*, edited by Nanak Kakwani and Jacques Silber, 192-209. London, UK: Palgrave Macmillan.

Chambers, Robert. 1989. "Vulnerability, Coping and Policy." *IDS Bulletin* 20 (2): 1-7.

—. 2006. "What is Poverty? Who Asks? Who Answers?" In *Poverty in Focus, What is Poverty? Concepts and Measures*, edited by Dag Ehrenpreis, 3-4. International Poverty Centre, United Nations Development Programme.

Chambers, Robert, and Gordon R. Conway. 1991. "Sustainable Rural Livelihoods: Practical Concepts for The 21st Century." *IDS Discussion Paper* 296.

Chaudhuri, Shubham, Jyotsna Jalan, and Asep Suryahadi. 2002. "Assessing Household Vulnerability to Poverty from Cross-sectional Data: A Methodology and Estimates from Indonesia." *Department of Economics Discussion Papers, Columbia University* 0102-52.

Cheli, Betti, and Achille Lemmi. 1995. "A Totally Fuzzy and Relative Approach to the Multidimensional Analysis of Poverty." *Economic Notes* 24: 115-133.

Chen , Shaohua, and Martin Ravallion. 2008. "The Developing World is Poorer than We Thought, but No Less Successful in the Fight Against Poverty." *The World Bank Policy Research Working Paper* 4703.

Chen, Martha A. 1983. *A Quiet Revolution: Women in Transition in Rural Bangladesh*. Cambridge, MA: Schenkman.

Chen, Xiang, Minghui Wang, and Heping Zhang. 2011. "The Use of Classification Trees for Bioinformatics." *Wiley Interdiscip Rev Data Min Knowl Discov.* 1 (1): 55-63.

Chiappero-Martinetti , Enrica. 1996. "Standard of Living Evaluation Based on Sen's approach: Some Methodological Questions." *Politeia* 12: 43-44.

—. 2000. "A Multidimensional Assessment of Wellbeing Based on Sen's Functioning Approach." *Rivista Internazionale di Scienze Sociali* 108 (2): 207-239.

Chiappero-Martinetti, Enrica, and José Manuel Roche. 2009. "Operationalization of the Capability Approach, from Theory to Practice: A Review of Techniques and Empirical Applications." In *Debating Global Society: Reach and Limits of the Capability Approach*, edited by Enrica Chiappero-Martinetti, 157-204. Milan: Fondazione Giangiacomo Feltrinelli.

Chicago Commission on Race Relations. 1922. *The Negro in Chicago: A Study of Race Relations and A Race Riot.* Chicago, IL: The University of Chicago Press.

Chicago Manual of Style Sixteenth Edition. 2010. Chicago, IL and London, UK: The University of Chicago Press.

Christiaensen, Luc J., and Kalanidhi Subbarao. 2005. "Towards an Understanding of Household Vulnerability in Rural Kenya." *Journal of African Economies* 14 (4): 520-558.

Chronic Poverty Research Centre. 2005. *The Chronic Poverty Report 2004-2005*. Manchester, UK: The Chronic Poverty Research Centre.

Chronic Poverty Research Centre. 2009. *The Chronic Poverty Report 2008-09: Escaping Poverty Traps*. The Chronic Poverty Research Centre.

Cicourel, Aaron V. 1974. *Cognitive Sociology: Language and Meaning in Social Interaction*. New York, NY: Free Press.

Cimadamore, Alberto D., Gabriele Koehler, and Thomas Pogge. 2016. "Poverty and the Millennium Development Goals: A Critical Look Forward." In *Poverty and the Millennium Development Goals: A Critical Look Forward*, edited by Alberto D. Cimadamore, Gabriele Koehler and Thomas Pogge, 3-25. London: Zed Books Ltd.

Cingolani, Patrick. 2005. *La Précarité.* Paris, FR: PUF.

Citro, Constance F., and Robert T. Michael. 1995. *Measuring Poverty: A New Approach.* Washington, D.C.: National Academy Press.

Clark, David A. 2005. "Sen's Capability Approach and the Many Spaces of Human Well-being." *Journal of Development Studies* 41 (8): 1339-1368.

—. 2008. "The Capability Approach: Its Development, Critiques and Recent Advances." In *Development Studies, Volume II*, edited by R. Ghosh, K. R. Gupta and P. Maiti. New Delhi, IN: Atlantic Books and Distributors.

—. 2012. "Adaptation and Development — Issues, Evidence and Policy Relevance." In *Adaptation, Poverty and Development: The Dynamics of Subjective Well-being*, edited by David A. Clark, 1-31. Palgrave Macmillan.

Clark, Kenneth B. 1965. *Dark Ghetto: Dilemmas of Social Power*. New York, NY: Harper.

Clark, Kenneth B., and Richard P. Nathan. 1982. "The Urban Underclass." In *Critical Issues for National Policy: A Reconnaissance and Agenda for further Study*, edited by National Research Council, 33-53. Washington, D.C.: Committee on National Urban Policy, National Research Council.

Clark, Linda A., and Daryl Pregibon. 1992. "Tree-based Models." In *Statistical Models in S*, edited by John M. Chambers and Trevor .J. Hastie. Chapman & Hall/CRC

Clark, Stephen, Richard Hemming, and David Ulph. 1981. "On Indices for the Measurement of Poverty." *Economic Journal* 91: 515-526.

Coates, Ken, and Richard Silburn. 1970. *Poverty: The Forgotten Englishmen*. Penguin Books Ltd.

Cobb, Clifford W. 2000. *Measurement Tools and the Quality of Life*. San Francisco, CA: Redefining Progress.

Colasanto, Diane, Arie Kapteyn, and Jacques ven der Gaag. 1984. "Two Subjective Definitions of Poverty: Results from the Wisconsin Basic Needs Study." *Journal of Human Resources* 19 (1:) 127-137.

Comim, Flavio, Mozaffer Qizilbash, and Alkire Sabina. 2008. *The Capabilities Approach: Concepts, Measures and Applications*. Cambridge, UK: Cambridge University Press.

Comim, Flavio, and Miriam Teschl. 2005. "Adaptive Preferences and Capabilities: Some Preliminary Conceptual Explorations." *Review of Social Economy* 63 (2): 229-247.

Commission of the European Communities. 2003. Joint Report on Social Inclusion Summarising the Results of the Examination of the National Action Plans for Social Inclusion (2003-2005). Available at (accessed 25 May 2020): https://ec.europa.eu/employment_social/soc-prot/soc-incl/final_joint_inclusion_report_2003_en.pdf

CONAPO. 1990. *Desigualdad Regional y Marginacion Municipal en México, 1990*. México, D.F., MX: Consejo Nacional de Poblacion, Comision Nacional del Agua.

CONEVAL. 2010. *Methodology for Multidimensional Poverty Measurement in Mexico*. Mexico City: Consejo Nacional de Evaluación de la Política de Desarrollo Social.

CONPES. 2012. *Documento Conpes Social No. 150: Metodologias Oficiales y Arreglos Institucionales para la Medicion de la Probreza en Colombia.* Bogota: Consejo Nacional de Política Económica y Social. República de Colombia. Departamento Nacional de Planeación .

Cooley, Charles H. 1902. *Human Nature and the Social Order*. New York, NY: Charles Scribner's sons.

Corazzini, Luca, Lucio Esposito, and Francesca Majorano. 2010. "Exploring the Absolutist vs. Relativist Perception of Poverty Using Cross-country Questionnaire Survey." *Journal of Economic Psychology* 32(2): 273-283.

Corder, Clive K. 1997. "The Reconstruction and Development Programme: Success Or Failure?" *Social Indicators Research* 41: 183–203.

Cornelius, Wayne A. 1975. *Politics and the Migrant Poor in Mexico City*. Stanford: Stanford University Press.

Coromaldi, Manuela, and Carlo Drago. 2017. "An Analysis of Multidimensional Poverty: Evidence from Italy." In *Measuring Multidimensional Poverty and Deprivation. Global Perspectives on Wealth and Distribution*, edited by Roger White, 69-86. Cham: Palgrave Macmillan.

Coromaldi, Manuela, and Mariangela Zoli. 2012. "Deriving Multidimensional Poverty Indicators: Methodological Issues and an Empirical Analysis for Italy." *Social Indicators Research* 107 (1): 37-54.

Cottingham, Clement. 1982a. "Conclusion: The Political Economy of Urban Poverty." In *Race, Poverty and The Urban Underclass*, edited by Clement Cottingham. Aldershot, UK: Lexington Books.

—. 1982b. "Introduction." In *Race, Poverty, and the Urban Underclass*, edited by Clement Cottingham. Aldershot, UK: Lexington Books.

—, ed. 1982c. *Race, Poverty, and the Urban Underclass*. Aldershot, UK: Lexington Books.

Crais, Clifton. 2011. *Poverty, War, and Violence in South Africa*. New York, NY: Cambridge University Press.

Crettaz, Eric. 2011. *Fighting Working Poverty in Post-industrial Economies. Causes, Trade-offs and Policy Solutions*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Crettaz, Eric, and Christian Suter. 2013. "The Impact of Adaptive Preferences on Subjective Indicators: An Analysis of Poverty Indicators." *Social Indicators Research* 114: 139-152.

Crettaz, Eric, and Giuliano Bonoli. 2011. "Words of Working Poverty: National Variations in Mechanisms." In *Working Poverty in Europe: A Comparative Approach*, edited by Neil Fraser, Rodolfo Gutiérrez and Ramón Peña-Casas, 46-69. Palgrave Macmillan.

Crilly, Tony. 2008. *50 Maths Ideas You Really Need to Know*. Quercus.

Crompton, Rosemary. 1999. *Restructuring Gender Relations and Employment: The Decline of the Male Breadwinner*. Oxford, UK: Oxford University Press.

Crush, Jonathan, Alice Hovorka, and Daniel Tevera. 2010. *Urban Food Production and Household Food Security in Southern African Cities*. Queen's University and AFSUN, Urban Food Security Series 4.

Cummins, Robert A. 1996. "The Domains of Life Satisfaction: An attempt to Order Chaos." *Social Indicators Research* 38: 303–328.

Cutter, Susan L., Bryan J. Boruff, and W. Lynn Shirley. 2003. "Social Vulnerability to Environmental Hazards." *Social Science Quarterly* 84 (2): 242-261.

DAFF. n.d. *Fetsa Thala Integrated Food Production Initiative 2013/2014: An Overview*. Republic of South Africa.

Dahrendorf, Ralf. 1987. "The Erosion of Citizenship and Its Consequences for Us All." *New Statesman* 113: 12-15.

Dalton, Hugh. 1920. "The Measurement of the Inequality of Incomes." *The Economic Journal* 30 (119): 348-361.

DANE. 2012. *Metodologias Oficiales y Arreglos Institucionales para la Medicion de la Pobreza en Colombia*. Bogota, D.C., CO: Consejo Nacional de Política Económica y Social República de Colombia, Departamento Nacional de Planeación.

Danziger, Sheldon, Jacques van der Gaag, Michael K. Taussig, and Eugene Smolensky. 1984. "The Direct Measurement of Welfare Levels: How Much Does It Cost to Make Ends Meet?" *Review of Economics and Statistics* 66 (3): 500-505.

Dasgupta, Partha. 1990. "Well-being in Poor Countries." *Economic and Political Weekly* 25 (31): 1713-1715 and 1717-1720.

David, Gordon. 2006. "The Concept and Measurement of Poverty." In *Poverty and Social Exclusion in Britain*, edited by Christina Pantazis, David Gordon and Ruth Levitas, 29-69. Bristol, UK: The Policy Press.

Davie, Grace. 2015. *Poverty Knowledge in South Africa : A Social History of Human Science, 1855-2005*. New York, NY: Cambridge University Press.

Dávila Lárraga , Laura G. 2016. "How does Prospera Work?: Best Practices in the Implementation of Conditional Cash Transfer Programs in Latin America and the Caribbean." *Inter-American Development Bank Technical Not*e IDB-TN-971.

de Muro, Pasquale, Matteo Mazziotta , and Adriano Pareto. 2011. "Composite Indices of Development and Poverty: An Application to MDGs." *Social Indicators Research* 104: 1-18.

de Vos, Klaas, and Thesia I. Garner. 1991. "An Evaluation of Subjective Poverty Definitions: Comparing Results from the U.S. and the Netherlands." *Review of Income and Wealth* 37 (3): 267-285.

de Tocqueville, Alexis. 1835. "Mémoire sur le Paupérisme." In *Mémoires de la Société Royale Académique de Cherbourg, vol. 2*. Cherbourg, FR: Société Royale Académique de Cherbourg.

—. 1951. *Œuvres Complètes, t. XVI.* Paris, FR: Gallimard.

De'ath, Glenn. 2002. "Multivariate Regression Trees: A New Technique for Modeling Species Environment Relationships." *Ecology* 83 (4), 1105-1117.

De'ath, G., and K.E. Fabricius. 2000. "Classification and Regression Trees: A Powerful yet Simple Technique for Ecological Data Analysis." *Ecology* 81: 3178-3192.

Deaton, A., and S. Zaidi. 1998. "Guidelines for Constructing Consumption Aggregates for Welfare Analysis." *The World Bank LSMS Working Paper* 135.

Dehejia, Rajeev, and Roberta Gatti. 2002. "Child Labor: The Role of Income Variability and Access to Credit Across Countries." *NBER Working Paper* 9018.

Deleeck, Herman, Karel Van den Bosch, and L. De Lathouwer. 1991. *Indicators of Poverty and Adequacy of Social Security: Methodological Considerations and Comparative Results for Seven Countries*. Second Specific Community Action to Combat Poverty, Centre for Social Policy, University of Antwerp.

Deneulin, Sévèrine. 2008. "Beyond Individual Freedom and Agency: Structures of Living Together in Sen's Capability Approach to Development." In *The Capability Approach: Concepts, Measures and Application*, edited by Sabina Alkire, Flavio Comim and Muzaffer Qizilbash, 105-124. Cambridge, UK: Cambridge University Press.

Deneulin, Sévèrine. 2014. *Wellbeing, Justice and Development Ethics*. London, UK: Routledge.

Deneulin, Sévèrine, and Frances Stewart. 2002. "Amartya Sen's Contribution to Development Thinking." *Studies in Comparative International Development* 37 (2): 61-70.

Dercon, Stefan. 2001. *Assessing Vulnerability* . Oxford, UK: CSAE, Department of Economics, University of Oxford.

—. 2005. "Risk, Poverty and Vulnerability in Africa." *Journal of African Economies* 14 (4): 483-488.

Dercon, Stefan, and Pramila Krishnan. 2000. "Vulnerability, Seasonality and Poverty in Ethiopia." *Journal of Development Studies* 36(6): 25-53.

Desai, Meghnad. 1986. "Drawing the line: On Defining the Poverty Threshold." In *Excluding the Poor*, edited by P. Golding. London: Child Poverty Action Group.

—. 1991. "Human Development: Concepts and measurement." *European Economic Review* 35 (2-3): 350-357.

—. 1995. "Poverty and Capability: Towards an Empirically Implementable." In *Poverty, Famine and Economic Development: The Selected Essays of Meghnad Desai, Volume II*, 185-204. Aldershot: Edward Elgar Publishing Company.

Dezalay, Yves, and Bryant Garth. 1998. "Le "Washington Consensus". Contribution à une Sociologie de l'Hégémonie du Néolibéralisme." *Actes de la Recherche en Sciences Sociales* 121-122 (1): 3-22.

Diaz-Cayeros, Alberto, Federico Estevez, and Beatriz Magaloni. 2016. *The Logic of Poverty Relief : Electoral Strategies and Social Policy in Mexico*. Cambridge, UK: Cambridge University Press.

Dickens, Charles. 1837-1839. *Oliver Twist*. Richard Bentley.

Diener, Ed. 1984. "Subjective Well-being." *Psychological Bulletin* 95 (3): 542-575.

Diener, Ed, and Eunkook M. Suh. 2000. *Culture and Subjective Well-being*. Cambridge, MA: MIT Press.

Diener, Ed, Robert A. Emmons, Randy J. Larson , and Sharon Griffin. 1985. "The Satisfaction With Life Scale." *Journal of Personality Assessment* 49 (1): 71-75.

Diener, Ed, Weiting Ng, James Harter, and Raksha Arora. 2010. "Wealth and Happiness Across the World: Material Prosperity Predicts Life Evaluation, whereas Psychosocial Prosperity Predicts Positive Feeling." *Journal of Personality and Social Psychology* 99 (1): 52-61.

Diener, Ed, Shigehiro Oishi, and Richard E. Lucas. 2002. "Subjective Well-being: The Science of Happiness and Life Satisfaction." In The Handbook of Positive Psychology, edited by Shane J. Lopez and C.R. Snyder, 63-73. New York, NY: Oxford University Press.

Dodge, Rachel, Annette P. Daly, Jan Huyton, and Lalage D. Sanders. 2012. "The Challenge of Defining Wellbeing." *International Journal of Wellbeing* 2(3): 222-235.

Dotter, Caroline, and Stephan Klasen. 2014. "The Multidimensional Poverty Index: Achievements, Conceptual and Empirical Issues." *UNDP Human Development Report Office Occasional Paper*.

Drewnowski, Jan, and Wolf Scott. 1966. *The Level of Living Index*. UNRISD.

Drèze, Jean, and Amartya K. Sen. 1989. *Hunger and Public Action*. New York, NY: Oxford University Press.

DSD, SASSA and UNICEF. 2012. *The South African Child Support Grant Impact Assessment: Evidence from a Survey of Children, Adolescents and their Households*. Pretoria, ZA: UNICEF South Africa.

Du Bois, W.E.B. (William Edward Burghardt). 1899. *The Philadelphia Negro.* University of Pennsylvania Press.

Dubet, Francois. 1987. *La Galère: Jeunes en Survie*. Paris: Fayard.

Duclos, Jean-Yves, and Luca Tiberti. 2016. "Multidimensional Poverty Indices: A Critical Assessment." In *The Oxford Handbook of Well-Being and Public Policy*, edited by Matthew D. Adler and Marc Fleurbaey. Oxford University Press.

Duda, Richard O., Peter E. Hart, and David G. Stork. (1973) 2001. *Pattern Classification*. New York, NY: John Wiley & Sons.

Duffy, Katherine. 1995. *Social Exclusion and Human Dignity in Europe : Background Report for The Proposed Inititative by the Council of Europe*. Strasbourg: Council of Europe.

Dugard, Jackie. 2016. "Testing the Government's Emergency Relief Mechanism: What Happens when Poor Households Attempt to Access the Social Relief of Distress Grant?" In *Food Security in South Africa: Human Rights and Entitlement Perspectives*, edited by Sakiko Fukuda-Parr and Viviene Taylor, 190-209. UCT Press.

Duggan, Marie Christine. *Reserve Army of Labor and Migration*. Available at (accessed 25 May 2020): https://www.academia.edu/6222930/Reserve_Army_of_Labor_and_Immigration

Duncan, Greg, and S. D. Hoffman. 1991. "Teenage Underclass Behaviour and Subsequent Poverty: Have the Rules Changed?" In *The Urban Underclass*, edited by Christopher Jencks and Paul E. Peterson, 155-174. Washington, D.C.: Brookings Institution.

Duncan, Greg, Richard Coe, Mary Corcoran, Martha S. Hill, Saul Hoffman, and James N. Morgan. 1984. *Years of Poverty, Years of Plenty: The Changing Economic Fortunes of American Workers and Families*. The University of Michigan Press .

Durand, Jean-Pierre, and Robert Weil. 1997. *Sociologie Contemporaine*. Paris, FR: Editions Vigot.

Durkheim, Emile. (1912) 2003. *Les Formes Elémentaires de la Vie Réligieuse*. Paris, FR: PUF.

—. (1893) 2007. *De la Division du Travail Social*. Paris, FR: PUF.

—. (1897) 2007b. *Le Suicide*. Paris, FR: PUF.

—. (1894) 2009. *Les Règles de la Méthode Sociologique*. Paris, FR: Payot.

Eckstein, Susan Eva. 1977. *The Poverty of Revolution: Social Controls of Mexican Urban Poor*. Princeton, NJ: Princeton University Press.

ECLAC. 2009. *Income Poverty and Unsatisfied Basic Needs*. Mexico, MX: ECLAC, United Nations.

—. 2014. *Social Panorama of Latin America 2014*. Santiago, CL: ECLAC.

Eden, Frederick Morton. (1797) 1966, 3 vols. *The State of the Poor; or, An History of the Labouring Classes in England*. London, UK: Cass.

Edgar, Bill, Joe Doherty, and Henk Meert. 2002. *Access to Housing: Homelessness and Vulnerability in Europe*. Bristol, UK: The Policy Press.

EEC. 1985. "*On Specific Community Action to Combat Poverty*." Official Journal of the EEC 2/24.

Elias, Norbert. 1970. *Was ist Soziologie?* München: Juventa-Verlag.

—. 1987. *Die Gesellschaft der Individuen*. Suhrkamp Verlag.

Ellis, Frank. 2000. "The Determinants of Rural Livelihood Diversification in Developing Countries." *Journal of Agricultural Economics* 51 (2): 289-302.

Ellis, G.F.R. 1984. "The Dimensions of Poverty." *Social Indicators Research* 15: 229-253.

Engels, Frederick. (1845) 1887. *The Condition of the Working Class in England*. New York, NY: John W. Lovell Company.

Esping-Andersen, Gøsta. 1990. *The Three Worlds of Welfare Capitalism*. Cambridge, UK: Polity Press.

—. 1999. *Social Foundations of Postindustrial Economies*. New York, NY: Oxford University Press.

—. 2002. *Why We Need a New Welfare State*. New York, NY: Oxford University Press.

Etienne, Gilbert. 1977. "Foodgrain Production and Population in Asia: China, India and Bagladesh." *World Development* 5 (5-7): 425-440.

European Communities. 2008. N*ACE Rev. 2: Statistical Classification of Economic Activities in the European Community.* Luxembourg, LU: Office for Official Publications of the European Communities

European Union Commission. 1981 (15 December). *Final Report from the Commission to the Council on the First Programme of Pilot Schemes and Studies to Combat Poverty*. European Union Commission COM, (81) 769 final.

Eurostat. 2012. *Measuring Material Deprivation in the EU: Indicators for the Whole Population and Child-specific Indicators*. European Union.

Evans, Martin. 1998. "Behind the Rhetoric: The Institutional Basis of Social Exclusion and Poverty." *IDS Bulletin* 29 (1): 42-49.

Evans, Peter. 2002. "Collective Capabilities, Culture, and Amartya Sen's Development as Freedom." *Studies in Comparative International Development* 37: 54-60.

Fafchamps, Marcel, Christopher Udry, and Katherine Czukas. 1998. "Drought and Saving in West Africa: Are Livestock a Buffer Stock?" *Journal of Development Economics* 55 (2): 273-305.

Fagan, Colette, P. Urwin, and K. Melling. 2006. *Gender Inequalites in the Risks of Poverty and Social Exclusion for Disadvantaged Groups in Thirty European Countries*. Luxembourg: European Commission, Office for Official Publications of the European Communities.

Famine Inquiry Commission India. 1945. *The Famine Inquiry Commission. Final report*. Delhi, IN: The Manager of Publications Delhi.

FAO. Statistics Division. 1977. *FAO Production Yearbook 1976*. Rome, IT: FAO.

Fassin, Didier. 1996. "Exclusion, Underclass, Marginalidad: Figures Contemporaines de la Pauvreté Urbaine en France, aux Etats-Unis et en Amérique Latine." *Revue Francaise de Sociologie* 37 (1): 37-75.

Fattore, Marco. 2016. "Partially Ordered Sets and the Measurement of Multidimensional Ordinal Deprivation." *Social Indicators Research* 128: 835-858.

—. 2017. "Synthesis of Indicators: The Non-Aggregative Approach." In *Complexity in Society: From Indicators Construction to their Synthesis*, edited by Filomena Maggino, 193-212. Springer.

Fattore, Marco, and Alberto Arcagni. 2014. "PARSEC: An R Package for Poset-Based Evaluation of Multidimensional Poverty." In *Multi-indicator Systems and Modelling in Partial Order*, edited by Rainer Bruggemann, Lars Carlsen and Jochen Wittmann, 317-330. Berlin, DE: Springer.

Fattore, Marco, Rainer Bruggemann, and Jan Owsinski. 2011a. "Using Poset Theory to Compare Fuzzy Multidimensional Material Deprivation Across Regions." In *New Perspectives in Statistical Modeling and Data Analysis*, edited by Salvatore Ingrassia, Roberto Rocci and Maurizio Vichi. Heidelberg, DE: Springer.

Fattore, Marco, Filomena Maggino, and Emilio Colombo. 2012. "From Composite Indicators to Partial Orders: Evaluating Socio-economic Phenomena through Ordinal Data." In *Quality of Life in*

Italy: Researches and Reflections. Social Indicators Research Series, edited by Filomena Maggino and Giampaolo Nuvolati, 41-68. Amsterdam: Springer.

Fattore, Marco, Filomena Maggino, and F. Greselin. 2011b. "Socio-economic Evaluation with Ordinal Variables: Integrating Counting and Poset Approaches." *Statistica & Applicazioni*, Partial Orders in Applied Sciences (special issue) 31–42.

Fattore, Marco, and Filomena Maggino. 2014. "Partial Orders in Socio-economics: A Practical Challenge for Poset Theorists or a Cultural Challenge for Social Scientists?" In *Multi-indicator Systems and Modelling in Partial Order*, edited by Rainer Bruggemann, Lars Carlsen and Jochen Wittmann, 197-214. New York, NY: Springer.s

Feres, Juan Carlos, and Xavier Mancero. 2001. *El Método de las Necesidades Básicas Insatisfechas (NBI) y sus Aplicaciones en América Latina*. Santiago de Chile, CL: CEPAL.

Fergusson, David, Bev Hong, John Horwood, John Jensen, and Peter Travers. 2011. *Living Standards of Older New Zealanders*. Wellington: Ministry of Social Development.

Fernández-Aguirre, Karmele, María I. Landaluce, Ana Martin, and Juan I. Modroño. 2008. "Data Mining of an On-line Survey. A Market Research Application." In *Data Analysis, Machine Learning and Applications. Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation e.V., Albert-Ludwigs-Universität Freiburg, March 7–9, 2007* ,edited by Christine Preisach, Hans Burkhardt, Lars Schmidt-Thieme and Reinhold Decker, 183-191. New York, NY: Springer.

Ferrer-i-Carbonell, Ada. 2002. "Subjective Questions to Measure Welfare and Well-being: A Survey." *Tinbergen Institute Discussion Papers.*

Ferrer-i-Carbonell, Ada, and Bernard van Praag . 2001. "Poverty in Russia." *Journal of Happiness Studies* 2 (2): 147-172.

Flach, Peter. 2012. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data.* Cambridge, UK: Cambridge University Press.

Fleurbaey, Marc. 2006. "Capabilities, Functionings and Refined Functionings." *Journal of Human Development and Capabilities* 7 (3): 299-310.

—. 2002. "Development, Capabilities and Freedom." *Studies in Comparative International Development* 37: 71-77.

Food Security Working Group. 1997, November. Food Security Policy for South Africa. *A Discussion Document. Food Security Working Group for the Department of Agriculture and Land Affairs.*

Foster, James E. 1983. *On Economic Poverty: A Survey of Aggregate Measures*. Institute for Research in the Behavioral, Economic, and Management Sciences, Krannert Graduate School of Management, Purdue University.

—. 1998. "Absolute Versus Relative Poverty." *American Economic Review* 8 (2): 335-341.

—. 2009. "A Class of Chronic Poverty Measures." In P*overty Dynamics: Interdisciplinary Perspectives*, edited by Tony Addison, David Hulme and Ravi Kanbur, 59-76. Oxford, UK: Oxford University Press.

—. 2010. "A Report on Mexican Multidimensional Poverty Measurement." *OPHI University of Oxford Working Papers* 40.

Foster, James E. , Joel Greer, and Erik Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica* 52: 761-776.

Foster, James E. , and Anthony F. Shorrocks. 1988. "Poverty Orderings." *Econometrica* 56: 173-177.

—. 1991. "Subgroup Consistent Poverty Indices." *Econometrica* 59: 687–709.

Frawley, William J., Grgeory Piatetsky-Shapiro, and Christopher J. Matheus. 1991. "Knowledge Discovery in Databases: An Overview." In *Knowledge Discovery in Database*s, edited by Gregory Piatetsky-Shapiro and William J. Frawley. Menlo Park.

Frederick, Shane, and George Loewenstein. 1999. "Hedonic Adaptation." In *Well-being: The Foundations of Hedonic Psychology*, edited by Daniel Kahneman, Ed Diener and Norbert Schwarz, 302–329. Russell Sage Foundation.

Fried, Albert, and Richard M. Elman. 1969. *Charles Booth's London: A Portrait of The Poor at The Turn of The Century Drawn from His Life and Labour of The People in London*. Penguin Books.

Fudge, Judy, and Rosemary Owens. 2006. *Precarious Work, Women and The New Economy: The Challenge to Legal Norms*. Oxford, UK: Hart Publishing.

Fujii, Tomoki. 2016. "Concepts and Measurement of Vulnerability to Poverty and Other Issues: A Review of Literature." *ADBI Working Paper* 611.

Fukuda-Parr, Sakiko. 2003. "The Human Development Paradigm: Operationalizing Sen's Ideas on Capabilities." *Feminist Economics* 9(2-3): 301-317.

Fukuda-Parr, Sakiko, and A. K. Shiva Kumar. 2009. *Handbook of Human Development: Concepts, Measures and Policies*. Delhi, IN: Oxford University Press.

Fukuda-Parr, Sakiko, Terra Lawson-Remer, and Susan Randolph. 2015. *Fulfilling Social and Economic Rights*. New York, NY: Oxford University Press.

Fukuda-Parr, Sakiko, and Alicia Ely Yamin. 2015. *Millennium Development Goals, Capabilities and Human Rights: The Power of Numbers to Shape Agendas.* London, UK: Routledge.

Gaiha, Raghav. 1988. "On Measuring The Risk of Poverty in Rural India." In *Rural Poverty in South Asia*, edited by T.N. Srinivasan and Pranab K. Bardhan. New York, NY: Columbia University Press.

Gaiha, Raghav. 1989. "Are The Chronically Poor Also The Poorest in Rural India?" *Development and Change* 20 (2): 295-322.

Gaiha, Raghav, and Anil Deolalikar. 1993. "Persistent, Expected and Innate Poverty: Estimates for Semi-arid Rural South India, 1975-1984." *Cambridge Journal of Economics* 17 (4): 409-421.

Gaiha, Raghav, and Katsushi Imai. 2004. "Vulnerability, Shocks and Persistence of Poverty: Estimates for Semi-arid Rural South India." *Oxford Development Studies* 32 (2): 261-281.

Gammage, Sarah. 2010. "Time Pressed and Time Poor: Unpaid Household Work for Guatemala." *Feminist Economics*, 16 (3): 79-112.

—. 2011. "Conditional Cash Transfers and Time Poverty: An Example from Guatemala." International Seminar “Policies on Time, Time for Policies”, *Cuadragésima sexta reunión de la Mesa Directiva de la Conferencia Regional sobre la Mujer de América Latina y el Caribe.*

Gangas, Spiros. 2015. "From Agency to Capabilities: Sen and Sociological Theory." *Current Sociology* 64 (1): 22-40.

Gans, Herbert J. 1968. "Culture and Class in the Study of Poverty: An Approach to Anti-poverty Research." In *On Understanding Poverty: Perspectives from the Social Sciences*, edited by Daniel P. Moynihan. New York: Basic Books.

—. 1970. "Poverty and Culture: Some Basic Questions about Methods of Studying Life-styles of the Poor." In *The Concept of Poverty*, edited by Peter Townsend, 146-164. London, UK: Heinemann Educational Books Ltd.

—. 1990. "Deconstructing the Underclass: The Term's Dangers as a Planning Concept." *Journal of the American Planning Association* 56 (3): 271-277.

—. 1995. *The War against the Poor: The Underclass and Antipoverty Policy*. Basic Books.

Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, NJ : Prentice-Hall.

Garland, David. 2014. "The Welfare State: A Fundamental Dimension of Modern Government." *European Journal of Sociology* 55 (3): 327-364.

Gasparini, Leonardo, Walter Sosa-Escudero, Mariana Marchionni, and Sergio Olivieri. 2013. "Multidimensional Poverty in Latin America and the Caribbean: New Evidence from The Gallup World Poll." *Journal of Economic Inequality* 11: 195–214.

Gasper, Des. 1997. "Sen's Capability Approach and Nussbaum's Capabilities Ethic." *Journal of International Development* 9 (2): 281-302.

—. 2002. "Is Sen's Capability Approach an Adequate Basis for Considering Human Development?" *Review of Political Economy* 14 (4): 435-461.

Gauvin, Francois-Pierre. 2009. "What is a Deliberative Process?". *Canada National Collaborating Centre for Healthy Public Policy Fact Sheet* 1193.

Geremek, Bronislaw. 1976. *Les Marginaux Parisiens aux XIV et XV Siècles*. Paris, FR: Flammarion.

Germani, Gino. (1973) 1980. Marginality. New Brunswick, NJ: Transaction books.

Gertler, Paul, and Jonathan Gruber. 2002. "Insuring Consumption Against Illness." *American Economic Review* 92 (1): 51-70.

Ghai, Dharam P. , Azizur Rahman Khan, E.L.H. Lee, and Torkel Alfthan. 1977. *The Basic Needs Approach to Development: Some Issues Regarding Concepts and Methodology*. Geneva, CH: International Labour Office.

Ghailani, Dalila. 2009. "Gender Wage Gap as a Tension between Quality and Quantity of Jobs in the European Labour Market." In *Quality of Work in the European Union: Concept, Data and Debates from a Transnational Perspective*, edited by Ana M. Guillen and Svenn-Age Dahl, 117-139. Brussels, BE: P.I.E. PETER LANG S.A.

Gibert, Karina, J. Spate, Miquel Sanchez-Marre, Ioannis Athanasiadis, And Joaquim Comas. 2008. "Chapter Twelve Data Mining for Environmental Systems." *Developments In Integrated Environmental Assesment* 3: 205-228.

Giddens, Anthony. 1973. *The Class Structure of the Advanced Societies*. London, UK: Hutchinson University Library.

Giddens, Anthony, and Philip W. Sutton. 2014. *Essential Concepts in Sociology*. Polity Press .

Gilbert, Geoffrey. 1997. "Adam Smith on the Nature and Causes of Poverty." *Review Of Social Economy* 55 (3): 273-291.

Gillie, Alan. 1996. "The Origin of the Poverty Line." *Economic Histoty Review* XLIX: 715-730.

—. 2000. "Rowntree, Poverty Lines and School Boards." In *Getting the Measure of Poverty: The Early Legacy of Seebohm Rowntree* , edited by Jonathan Bradshaw and Roy Sainsbury, 85-108. Aldershot, UK: Ashgate.

Gladwin, Thomas. 1967. Poverty, U.S.A. Little Brown & Company.

Glasgow, Douglas G. 1968. *The Sons of Watts Improvement Association. "The Sons of Watts" : Analysis of Mobility Aspirations and Life-Styles in the Aftermath of the Watts Riot, 1965*. D.S.W. University of Southern California.

—. 1980. *The Black Underclass: Poverty, Unemployment, and Entrapment of Ghetto Youth*. San Francisco, CA: Jossey-Bass.

Glewwe, Paul, and Kwaku A. Twum-Baah. 1991. "The Distribution of Welfare in Ghana, 1987-88." *The World Bank Living Standards Measurement Study Working Paper* 75.

Goedhardt, Theo, Victor Halberstadt, Arie Kapteyn, and Bernard van Praag. 1977. "The Poverty Line: Concept and Measurement." *Journal of Human Resources* 12 (4): 503-520.

Goerne, Alexander. 2011. "A Comparative Analysis of In-Work Poverty in the European Union." In *Working Poverty in Europe. Work and Welfare in Europe*, by Neil Fraser, Rodolfo Gutiérrez and Ramón Peña-Casas, 15-45. London, UK: Palgrave Macmillan.

Goffman, Erving. 1961a. *Asylums*. New York, NY: Anchor Books.

—. 1961b. *Encounters*. The Bobbs-Merrill Company, Inc.

—. 1963a. *Behavior in Public Places: Notes on the Social Organization of Gatherings*. The Free Press.

—. 1963b. *Stigma, Notes on the Management of Spoiled Identity*. Englewood Cliffs, NJ: Prentice-Hall.

—. 1967. *Interaction Ritual. Essays on Face-to-Face Behavior*. New York, NY: Pantheon Books.

—. 1969. *Strategic Interaction*. Philadelphia, PA: University of Pennsylvania Press.

—. 1974. *Frame Analysis: An Essay on the Organization of Experience*. Cambridge, MA: Harvard University Press.

—. 1981. *Forms of Talk*. Philadelphia, PA: University of Pennsylvania Press.

—. (1956) 1990. *The Presentation of Self in Everyday Life*. London, UK: Penguin Books.

Goglin, Jean-Louis. 1976. *Les Misérables dans l'Occident Médiéval*. Paris, FR: Seuil.

Gordon, David. 1995. *Key Findings from the Welsh Omnibus Survey*. Bristol, UK: Unpublished research paper for Week In Week Out (BBC TV Wales, documentary on poor Wales), first broadcast on 5 March 1996.

—. 2000. "Measuring Absolute and Overall Poverty." In *Breadline Europe: The Measurement of Poverty*, edited by David Gordon and Peter Townsend, 49-78. Bristol, UK: The Policy Press.

—. 2010. "Theoretical Disputes about Poverty." In *The Peter Townsend Reader*, edited by Alan Walker, David Gordon, Nicola Yeates, Chris Philipson, Ruth Levitas, Peter Philimore and Margot E. Salomon, 227-251. Bristol, UK: Policy Press.

Gordon, David, Ruth Levitas, Christina Pantazis, Demi Patsios, Sarah Payne, Peter Townsend, Laura Adelman, Karl Ashworth, Sue Middleton, Jonathan Bradshaw, and Julie Williams. 2000a. *Poverty and Social Exclusion in Britain*. Joseph Rowntree Foundation.

Gordon, David, and Shailen Nandy. 2016. "Policy-relevant Measurement of Poverty in Low, Middle and High Income Countries." In *Poverty and Inequality in Middle Income Countries: Policy Achievements, Political Obstacles*, edited by Einar Braathen, Julian May, Marianne S. Ulriksen and Gemma Wright, 16-43. London, UK: Zed Books.

Gordon, David, and Christina Pantazis, eds. 1997. *Breadline Britain in the 1990s*. Bristol, UK: Summerleaze House Books.

Gordon, David, Christina Pantazis, and Peter Townsend. 2000b. "Absolute and Overall Poverty: A European History and Proposal for Measurement." In *Breadline Europe: The Measurement of Poverty*, edited by David Gordon and Peter Townsend, 79-106. Bristol, UK: The Policy Press.

Gordon, David, and Paul Spicker, eds. 1999. *The International Glossary on Poverty*. London, UK: Zed Books Ltd.

Grammenos, Stefanos. 2003. *Illness, Disability and Social Inclusion*. Luxembourg, LU: European Foundation for the Improvement of Living and Working Conditions. Office for Official Publications of the European Communities.

Griffin, James. 1986. *Well-being: Its Meaning, Measurement and Moral Importance*. Oxford, UK: Clarendon Press.

Grimm, Michael. 2010. "Mortality Shocks and Survivors' Consumption." *Oxford Bulletin of Economics and Statistics* 72 (2): 146-171.

Guarcello, Lorenzo, Fabrizio Mealli, and Furio Camillo Rosati. 2010. "Household Vulnerability and Child Labor: The Effect of Shocks, Credit Rationing, and Insurance." *Journal of Population Economics* 23 (1): 169-198.

Guillaumont, Patrick. 2009. "An Economic Vulnerability Index: Its Design and Use for International Development Policy." *Oxford Development Studies* 37 (3): 193-228.

Guio, Anne-Catherine, Alessio Fusco, and Eric Marlier. 2009. *A European Union Approach to Material Deprivation using EU-SILC and Euro-barometer Data*. IRISS WP 2009-19, IRISS.

Günther, Isabel, and Kenneth Harttgen. 2009. "Estimating Households Vulnerability to Idiosyncratic and Covariate Shocks: A Novel Method Applied in Madagascar." *World Development* 37 (7): 1222-1234.

Günther, Isabel, and Stephan Klasen. 2009. "Measuring Chronic Non-income Poverty." In *Poverty Dynamics: Interdisciplinary Perspectives*, edited by Tony Addison, David Hulme and Ravi Kanbur, 77-101. Oxford, UK: Oxford University Press.

Gutiérrez, Rodolfo, Marta Ibáñez, and Aroa Tejero. 2011. "Mobility and Persistence of In-Work Poverty." In *Working Poverty in Europe. Work and Welfare in Europe*, edited by Neil Fraser, Rodolfo Gutiérrez and Ramón Peña-Casas, 175-201. 2011: Palgrave Macmillan.

Hagenaars, Aldi J. M. 1986. *The Perception of Poverty*. North Holland.

Hagenaars, Aldi J. M., and Klaas de Vos. 1988. "The Definition and Measurement of Poverty." *Journal of Human Resources* 23 (2): 211-221.

Hagenaars, Aldi J. M., Klaas de Vos, and M. Asghar Zaidi. 1994. *Poverty Statistics in the Late 1980s: Research based on Micro-data*. Luxembourg, LU: Office for Official Publications of the European Communities.

Hagenaars, Aldi J. M., and Bernard van Praag. 1985. "A Synthesis of Poverty Line Definitions." *The Review of Income and Wealth* 31 (2): 139-154.

Hagerty, Michael R., Robert A. Cummins, Abbott L. Ferriss, Kenneth Land, Alex C. Michalos, Mark Peterson, Andrew Sharpe, Joseph Sirgy, and Joachim Vogel. 2001. "Quality of Life Indexes for National Policy: Review and Agenda for Research." *Social Indicators Research* 55: 1-96.

Hagerty, Michael, Joachim Vogel, and Valerie Møller. 2002. "Introduction: Assessing National Quality of Life and Living Conditions: The State of the Art". *Social Indicators Research* 58 (1/3): 1-11.

Haig, R. M. 1921. "The Concept of Income-economic and Legal Aspects." In *The Federal Income Tax*, edited by R. M. Haig, 1-28. New York, NY: Columbia University Press.

Halleröd, Björn. 1994. *Poverty in Sweden: A New Approach to Direct Measurement of Consensual Poverty*. Umeå, SE: Dept. of Sociology, Umeå University.

—. 1995a. "Making Ends Meet: Perception of Poverty in Sweden." *Scandinavian Journal of Social Welfare* 4: 174-189.

—. 1995b. "The Truly Poor: Indirect and Direct Measurement of Consensual Poverty in Sweden." *Journal of European Social Policy* 5 (2): 111-129.

—. 1998. "Poor Swedes, Poor Britons: A Comparative Analysis of Relative Deprivation." In *Empirical Poverty Research in a Comparative Perspective*, 283-311. Aldershot, UK: Ashgate.

—. 2000. "Poverty, Inequality and Health." In *Breadline Europe: The Measurement of Poverty*, edited by David Gordon and Peter Townsend, 165-190. Bristol, UK: The Policy Press.

—. 2006. "Sour Grapes: Relative Deprivation, Adaptive Preferences and the Measurement of Poverty." *Journal of Social Policy* 35: 371-390.

Halleröd, Björn, Daniel Larsson, David Gordon, and Veli- Matti Ritakallio. 2006. "Relative Deprivation: A Comparative Analysis of Britain, Finland and Sweden." *Journal of European Social Policy* 16 (4): 328-345.

Han, Jiawei, Micheline Kamber, and Jian Pei. (2000) 2012. *Data Mining: Concepts and Techniques.* San Francisco, CA: Morgan Kaufmann.

Hanandita, Wulung, and Gindo Tampubolon. 2016. "Multidimensional Poverty in Indonesia: Trend Over the Last Decade (2003–2013)." *Social Indicators Research* 128: 559-587.

Hannerz, Ulf. 1969. *Inquiries into Ghetto Culture and Community*. The University of Chicago Press.

Hanoman, Jacqueline. 2018. *Hunger and poverty in South Africa: The Hidden Faces of Food Insecurity*. Abingdon, UK: Routledge.

Haq, Mahbub ul. 1995. *Reflections on Human Development*. New York, NY: Oxford University Press.

Harrington, Michael. 1981. *The Next America: The Decline and Rise of the United States*. New York, NY: Touchstone.

—. 1984. *The New American Poverty*. New York, NY: Henry Holt.

—. (1962) 2012. *The Other America: Poverty in the United States*. New York, NY: Scribner.

Harrison, Lawrence E. 1985. *Underdevelopment is a State of Mind: The Latin American Case.* Cambridge: University Press of America or the Center for International Affairs, Harvard University.

Harrison, Lawrence E., and Samuel P. Huntington. 2001. C*ulture Matters: How Values Shape Human Progress*. Basic Books.

Hastie, Trevor, Robert Tibshirani, and Jerome H. Friedman. (2001) 2009. T*he Elements of Statistical Learning*. New York, NY: Springer-Verlag.

Haughton, Jonathan, and Shahidur Khandker. 2009. *Handbook on Poverty and Inequality.* Washington, D.C.: The World Bank.

Hennock, Ernest Peter. 1991. "Concepts of Poverty in the British Social Surveys from Charles Booth to Arthur Bowley." In *The Social Survey in Historical Perspective 1880-1940*, edited by Martin Bulmer, Kevin Bales and Kathryn Kish Sklar, 189-216. Cambridge, UK: Cambridge University Press.

Hess, K.R., M.C. Abbruzzese, R. Lenzi, M.N. Raber, and J.L. Abbruzzese. 1999. "Classification and Regression Tree Analysis of 1000 Consecutive Patients with Unknown Primary Carcinoma." *Clinical Cancer Research* 5 (11): 3403–3410.

Hick, Rod. 2012. "The Capability Approach: Insights for a New Poverty Focus." *Journal of Social Policy*, 41 (2): 291-308.

—. 2014. "Poverty as Capability Deprivation: Conceptualising and Measuring Poverty in Contemporary Europe." *European Journal of Sociology* 55 (3): 295-323.

Hicks, Norman, and Paul Streeten. 1979. "Indicators of Development: The Search for a Basic Needs Yardstick." *World Development* 7 (6): 567-580.

Highes, Everett C. (1971) 1984. *The sociological Eye: Selected Papers*. New Brunswick, CA: Transaction Books.

Hirai, Tadashi. 2017. *The Creation of the Human Development Approach*. Palgrave Macmillan.

Hoddinott, John, and Agnes Quisumbing. 2003. "Methods for Microeconometric Risk and Vulnerability Assessments." *The World Bank Social Protection Discussion Paper Series* 0324.

Holman, Robert. 1978. *Poverty: Explanations of Social Deprivation.* New York, NY: St. Martin's Press.

Holzmann, Robert, and Steen Jorgensen. 1999. "Social Protection as Social Risk Management." *The World Bank SP Discussion Paper* 9904.

Howard, Joanna, and Jennifer Leavy. 2013. *What Matters Most? Evidence from 84 Participatory Studies with those Living with Extreme Poverty and Marginalisation.* Brighton, UK: IDS.

Hübinger, Werner. 1996. *Prekärer Wohlstand: Neue Befunde zu Armut und sozialer Ungerechtigkeit.* Lambertus.

Hulme, David. 2015. *Global Poverty: How Global Governance Is Failing the Poor*. New York, NY: Routledge.

Hulme, David, and Andrew McKay. 2005. "Identifying and Measuring Chronic Poverty: Beyond Monetary Measures." *The Chronic Poverty Research Centre Working Paper* 30.

—. 2008. "Identifying and Measuring Chronic Poverty: Beyond Monetary Measures?" In *The Many Dimensions of Poverty*, edited by Nanak Kakwani and Jacques Silber, 187-214. New York, NY: Palgrave Macmillan.

Hulme, David, and Andrew Shepherd. 2003. "Conceptualizing Chronic Poverty." *World Development* 31(3): 403-423.

Hulme, David, Karen Moore, and Andrew Shepherd. 2001. "Chronic Poverty: Meanings and Analytical Frameworks." *Chronic Poverty Research Centre Working Paper* 2.

Ibrahim, Solava. 2006. "From Individual to Collective Capabilities: the Capability Approach as a Conceptual Framework for Self-help." *Journal of Human Development* 7(3): 397-416.

ILO. 1976. *Employment, Growth and Basic Needs: A One-world Problem*. New York: Praegers Publishers.

—. 2016. *Non-standard Employment around the World: Understanding Challenges, Shaping Prospects*. Geneva, CH: ILO.

ILO/IMF/OECD/UNECE/Eurostat/The World Bank. 2004. *Consumer Price Index Manual: Theory and Practice*. Geneva: International Labour Office.

INEC. 2015. *Indice de Pobreza Multidimensional (IPM) Metodologia*. San José, CR: Costa Rica. Instituto Nacional de Estadística y Censos.

Izenman, Alan J. 2008. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. Springer.

Jacobs, Jerry A., and Kathleen Gerson. 2004. *The Family and Public Policy. The Time Divide: Work, Family, and Gender Inequality*. Harvard University Press.

Jahan, Selim. 2003. "Evolution of the Human Development Index." In *Readings in Human Development: Concepts, Measures and Policies for a Development Paradigm*, edited by Sakiko Fukuda-Parr and A.K. Shiva Kumar. New Delhi, IN: Oxford University Press.

Jalan, Jyotsna, and Martin Ravallion. 1998. "Determinants of Transient and Chronic Poverty : Evidence from Rural China." *The World Bank Policy Research Working Paper* 1936.

—. 2001. "Household Income Dynamics in Rural China." *The World Bank Policy Research Working Paper* 2706.

—. 2002. "Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China." *Journal of Applied Econometrics* 17 (4): 329-346.

James, Gareth, Trevor Hastie, Daniela Witten, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning: With Applications in R*. Springer.

Jargowsky, Paul A., and Mary Jo Bane. 1991. "Ghetto Poverty in the United States, 1970-1980." In *The Urban Underclass*, edited by Christopher Jencks and Paul E. Peterson, 235-273. Washington, D.C.: The Brookings Institution.

Jencks, Christopher. 1989. "What is the Underclass-and Is It Growing?" *Focus* 12 (1): 14-26.

—. 1992. *Rethinking Social Policy: Race, Poverty, and the Underclass*. Harvard University Press.

Jencks, Christopher, and Paul E. Peterson. 1991. *The Urban Underclass*. Washington, D.C.: The Brookings Institution.

Jenkins, Stephen P. 1990. "The Distribution of Wealth: Measurement and Models." *Journal of Economic Surveys* 4 (4): 329-360.

Jensen, John, Matt Spittal, Sarah Crichton, Sathi Sathiyandra, and Vasantha Krishnan. 2002. *Direct Measurement of Living Standards: The New Zealand Elsi Scale*. Wellington: The Ministry of Social Development .

Jordan, Bill. 1974. *Poor Parents: Social Policy and the 'Cycle of Deprivation'*. London, UK: Routledge and Kegan Paul.

Joseph, Keith. 1972. "The Cycle of Family Deprivation." In *Caring for People*, edited by Keith Joseph, 29-46. London, UK: Conservative Political Centre.

Juppenlatz, Morris. 1970. *Cities in Transformation; The Urban Squatter Problem in the Developing World*. St. Lucia, Queensland: Univ. Queensland Press.

Kabacoff, Robert I. (2011) 2015. *R in Action: Data analysis and graphics with R*. Manning Publications.

Kahneman, Daniel, Norbert Schwartz, and Ed Diener. 1999. *Well-being: The Foundations of Hedonic Psychology*. New York, NY: Russell Sage.

Kakwani, Nanak. 1980a. *Income Inequality and Poverty: Methods of Estimation and Policy Applications*. Oxford University Press (published for the World Bank).

—. 1980b. "On a Class of Poverty Measures." *Econometrica* 48: 437-446.

—. 1984. "Welfare Ranking of Income Distributions." *Adv Econ.* 3: 191-213.

—. 2006. "Poverty and Wellbeing." In *Poverty in Focus: What is Poverty? Concepts and Measures*, edited by Dag Ehrenpreis, 20-21. International Poverty Centre, United Nations Development Programme.

Kakwani, Nanak, and Jacques Silber. 2008. *Quantitative Approaches to Multidimensional Poverty Measurement*. Palgrave Macmillan.

Kamanou, Gisele, and Jonathan Morduch. 2002. "Measuring Vulnerability to Poverty." *WIDER Discussion Paper* 2002/58.

Kapteyn, Arie, Peter Kooreman, and Rob Willemse. 1988. "Some Methodological Issues in the Implementation of Subjective Poverty Definitions." *Journal of Human Resources* 23: 222-242.

Kasarda, John D. 1985. "Urban Change and Minority Opportunities." In *The New Urban Reality*, edited by Paul E. Peterson, 33-68. Washington, D.C.: Brookings Institution.

—. 1989. "Urban Industrial Transition and the Underclass." *Annals of the American Academy of Political and Social Science* 501: 26-47.

Kass, G.V. 1980. "An Exploratory Technique for Investigating Large Quantities of Categorical Data." *Applied Statistics* 29: 119–127.

Katz, Michael B. 1986. *In the Shadow of the Poorhouse: A Social History of Welfare in America.* New York, NY: Basic Books.

—. 1989. *The Undeserving Poor: From the War on Poverty to the War on Welfare*. New York, NY: Pantheon Books.

—. 1993. *The "Underclass" Debate: Views from History*. Princeton University Press.

—. 1995. *Improving Poor People: The Welfare State, the 'Underclass', and Urban Schools as History.* Princeton University Press.

—. 2013. *The Undeserving Poor: America's Enduring Confrontation with Poverty*. Oxford University Press.

Kazianga, Harounan, and Christopher Udry. 2006. "Consumption Smoothing? Livestock, Insurance and Drought in Rural Burkina Faso." *Journal of Development Economics* 79 (2): 413-446.

Kelleher, John D., Brian Mac Namee, and Aoife D'Arcy. 2015. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. Cambridge, MA: The MIT Press.

Kelley, Allen C. 1991. "The Human Development Index: "Handle with Care"". *Population and Development Review* 17 (2): 315-324.

Kent, Raymond. 1981. *The History of British Empirical Sociology*. Aldershot, UK: Gower.

Kerbo, Harold R. 2006. *World Poverty: Global Inequality and the Modern World System*. New York, NY: McGraw-Hill.

Khan, Atta Ullah, Abdul Saboor, Abid Hussain, Shumalia Sadiq, and Abdul Qayyum Mohsin. 2014. "Investigating Multidimensional Poverty across the Regions in the Sindh Province of Pakistan." *Social Indicators Research* 119 (2): 515-532.

Khan, Atta Ullah, Abduk Saboor, Ikram Ali, Wasim Shahid Malik, and Khalid Mahmood. 2016. "Urbanization of Multidimensional Poverty: Empirical Evidences from Pakistan." *Quality & Quantity* 50 (1): 439-4.

Kingdon, Geeta Gandhi, and John Knight. 1199-1224. "Subjective Well-being Poverty vs. Income Poverty and Capabilities Poverty?" *Journal of Development Studies* 42 (7): 2006.

Klanfer, Jules. 1965. *L'Exclusion Sociale: Etude de la Marginalité dans les Sociétés Occidentales.* Paris, FR: Bureau de recherces sociales, ATD.

Klasen, Stephen. 1997. "Poverty, Inequality and Deprivation in South Africa: An Analysis of the 1993 SALDRU Survey." *Social Indicators Research* 41: 51–94.

—. 2000. "Measuring Poverty and Deprivation in South Africa." *Review of Income and Wealth* 46 (1): 33-58.

Klasen, Stephan, Tobias Lechtenfeld, and Felix Povel. 2013. "Using Household Surveys to Capture Vulnerability: Issues and Challenges." In *Vulnerability to Poverty: Theory, Measurement and Determinants, with Case Studies from Thailand and Vietnam*, edited by M. Grimm, M. Waibel and S. Klasen, 80-114. London, UK: Palgrave Macmillan.

Klasen, Stephan, and Felix Povel. 2013. "Defining and Measuring Vulnerability: State of the Art and New Proposals." In *Vulnerability to Poverty: Theory, Measurement and Determinants, with Case Studies from Thailand and Vietnam*, edited by Stephan Klasen and Hermann Waibel, 17-49. London, UK: Palgrave Macmillan.

Kolmogorov, Andrey N. (1933) 2018. *Foundations of the Theory of Probability*. Dover Publications.

Kornblum, W. 1991. "Who is the Underclass? Contrasting Approaches. A Grave Problem." *Dissent*, Spring: 202-211.

Kovalerchuk, Boris, and Evgenii Vityaev. 2005. "Data Mining for Financial Applications." In *Data Mining and Knowledge Discovery Handbook*, edited by Oded Maimon and Lior Rokach, 1203-1224. Boston, MA: Springer.

Kuhn, Max. 2008. Building Predictive Models in R Using the caret Package. J*ournal of Statistical Software* 28 (5): 1-26.

Kuhn, Lisa, Karen Page, John Ward, and Linda Worrall-Carter. 2014. "The Process and Utility of Classification and Regression Tree Methodology in Nursing Research." *Journal of Advanced Nursing* 70 (6): 1276–1286.

Kuhnert, Petra, and Bill Venables. 2005. *An Introduction to R: Software for Statistical Modelling & Computing*. Cleveland, Australia: CSIRO Mathematical and Information Sciences .

Kuklys, Wiebke. 2005. *Amartya Sen's Capability Approach: Theoretical Insights and Empirical Applications*. Berlin, DE: Springer.

Kulig, Judith C., Dana S. Edge, Ivan Townshend, Nancy Lightfoot, and William Reimer. 2013. "Community Resiliency: Emerging Theoretical Insights." *Journal of Community Psychology* 41 (6): 758-775.

Kumar, T. Krishna, Anil P. Gore, and V. Sitaramam. 1996. "Some Conceptual and Statistical Issues on Meausurement of Poverty." *Journal of Statistical Planning and Inference* 49 (1): 53-71.

Langley, Pat. 1996. *Elements of Machine Learning*. San Francisco, CA: Morgab Kaufmann.

Lansley, Stewart, and Joanna Mack. 2015. *The Rise of Mass Poverty*. Oneworld Publications.

Lawson, David, Andrew McKay, and Karen Moore. 2003. *Panel Datasets in Developing and Transitional Countries*. Manchester, UK: Chronic Poverty Research Centre.

Lawson, David, David Hulme, and James Muwonge. 2008. "Combining Quantitative and Qualitative Research to further our Understanding of Poverty Dynamics: Some Methodological Considerations." *International Journal of Multiple Research Approaches* 2(2): 191-204.

Leacock, Eleanor B. 1971a. "Introduction." In *The Culture of Poverty: A Critique* , edited by Eleanor B. Leacock, 9-40. New York, NY: Simon and Schuster.

—, ed. 1971b. *The Culture of Poverty: A Critique.* New York, NY: Simon and Schuster.

Leeds, Anthony. 1971. "The Concept of the Culture of Poverty: Conceptual, Logical, and Empirical Problems, with Perspectives from Brazil and Peru." In *The Culture of Poverty: A Critique,* edited by Eleanor Leacock, 226-284. New York, NY: Simon and Schuster.

Lemann, Nicholas. 1986. "The Origins of the Underclass." *The Atlantic Monthly* 258 (1), June: 54-68.

Lemert, Edwin M. 1951. *Social Pathology: A Systematic Approach to the Theory or Sociopathic Behavior*. New York, NY: McGraw-Hill Book Company.

Lemmi, Achille, and Gianni Betti. 2006. *Fuzzy Set Approach to Multidimensional Poverty Measurement*. Springer.

Lenoir, René. 1974. Les Exclus : Un Français sur *Dix*. Paris, FR: Editions du Seuil.

Levitan, Sar A. 1969. *The Great Society's Poor Law; A New Approach to Poverty*. Baltimore, MD: Johns Hopkins University Press.

Levitas, Ruth. 1996. "The Concept of Social Exclusion and the new 'Durkheimian' Hegemony." *Critical Social Policy* 16 (1): 5-20.

—. (1998) 2000. The Inclusive Society? Social Exclusion and New Labour. Palgrave Macmillan.

Levy, Santiago. 2006. *Progress Against Poverty : Sustaining Mexico's Progresa-oportunidades Program*. Washington, D.C.: Brookings Institution Press.

Lewis, Jane, Mary Campbell, and Carmen Huerta. 2008. "Patterns of Paid and Unpaid Work in Western Europe: Gender, Commodification, Preferences and the Implications for Policy." *Journal of European Social Policy* 18 (1): 21-37.

Lewis, Jane. 2009. *Work–Family Balance, Gender and Policy*. Edward Elgar Publishing.

Lewis, Oscar. 1944 . "Social and Economic Changes in a Mexican Village: Tepoztlan 1926-44." *América Indigena* 4: 281-314.

—. 1950. "An Anthropological Approach to Family Studies." *American Journal of Sociology* 55 (5): 468-475.

—. 1958. *Village Life in Northern India*. Vintage Books.

—. 1959. *Five Families: Mexican Case Studies in the Culture of Poverty*. New York, NY: Basic Books.

—. 1961. *The Children of Sanchez: Autobiography of a Mexican Family*. New York, NY: Vintage Books.

—. 1963 (originally published in a clothbound edition in 1951). *Life in a Mexican Village: Tepotzlan restudied*. Urbana, IL: University of Illinois Press.

—. 1964. *Pedro Martinez: A Mexican Peasant and his Family*. London, UK: Secker & Warburg.

—. 1966. "The Culture of Poverty." *Scientific American* 215 (4): 19-25.

—. (1965, 1966) 1967. *La Vida: a Puerto Rican Family in the Culture of Poverty-San Juan & New York*. London, UK: Secker & Warburg.

—. 1968. *A Study of Slum Culture: Backgrounds for La Vida*. New York, NY: College Department, Random House.

Lewis, Roger J. 2000. "An Introduction to Classification and Regression Tree (CART) Analysis." *Paper Presented at the 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Francisco*, CA.

Liaw, Andy, and M. Wiener. 2012. "Random Forest: Breiman and Cutler's Random Forests for Classification and Regression." *R package version 4.6-3*.

Liebow, Elliot. 1967. *A Study of Negro Streetcorner Men*. Boston, MA: Little, Brown & Co.

Ligon, Ethan, and Laura Schechter. 2004. "Evaluating Different Approaches to Estimating Vulnerability." *The World Bank Social Protection Discussion Paper Series* 0410.

Likert, Rensis. 1932. "A Technique for the Measurement of Attitudes." *Archives of Psychology* 22 (140): 5-55.

Linton, Ralph. 1936. *The Study of Man: An Introduction*. New York, NY: Appleton-Century-Crofts, Inc.

Lipton, Michael. 1983. "Poverty, Undernutrition, and Hunger." *The World Bank Staff Working Papers* 597.

—. 1988. "The Poor and the Poorest: Some Interim Findings." *The World Bank Discussion Papers.*

Lister, Ruth. 2004. *Poverty*. Cambridge: Blackwell/Polity Press.

—. 1990. *The Exclusive Society: Citizenship and the Poor*. London, UK: Child Poverty Action Group.

Loh, Wei-Yin. 2014. "Fifty Years of Classification and Regression Trees." *International Statistical Review* 82(3): 329-348.

Lohmann, Henning, and Ive Marx. 2008. "The Different Faces of In-work Poverty across Welfare State Regimes." In *The Working Poor in Europe,* edited by Hans-Jürgen Andreß and Henning Lohmann, Chapter 1. Edward Elgar Publishing.

Lomnitz, L. 1975. Como Sobreviven los Marginados. Mexico, MX: Siglo XXI.

Lustig, Nora. 2011. "Multidimensional Indices of Achievements and Poverty: What Do We Gain and What Do We Lose? An introduction to JOEI Forum on Multidimensional Poverty." *Journal of Economic Inequality* 9: 227-234.

Luthar, Suniya S. 2003. *Resilience and Vulnerability: Adaptation in the Context of Childhood Adversities*. New York, NY: Cambridge University Press.

Luzzi, Giovanni Ferro, Yves Flückiger, and Sylvain Weber. 2008. "A Cluster Analysis of Multidimensional Poverty in Switzerland." In *Quantitative Approches to Multidimensional Poverty Measurement*, edited by Nanak Kakwani and Jacques Silber, 63-79. Palgrave Macmillan.

Lynn, Richard, Caner Sakar, and Helen Cheng. 2015. "Regional Differences in Intelligence, Income and Other Socio-economic Variables in Turkey." *Intelligence* 50: 144–149.

Mack, Joanna, and Stewart Lansley. 1985. *Poor Britain*. London, UK: George Allen & Unwin (Publishers) Ltd.

Maggino, Filomena. 2016. "Challenges, Needs and Risks in Defining Wellbeing Indicators." In *A Life Devoted to Quality of Life: Festschrift in Honor of Alex C. Michalos*, edited by Filomena Maggino, 209-233. Springer.

—. 2017a. "Dealing with Syntheses in a System of Indicators." In C*omplexity in Society: From Indicators Construction to their Synthesis,* edited by Filomena Maggino, 115-137. Springer.

—. 2017b. "Developing Indicators and Managing the Complexity." In *Complexity in Society: From Indicators Construction to their Synthesis,* edited by Filomena Maggino, 87-114. Social Indicators Research Series, vol 70. Springer, Cham.

Maimon, Oded Z., and Lior Rokach. (2005) 2010. *Data Mining and Knowledge Discovery Handbook.* Springer.

Maindonald, John, and W. John Braun. 2010. *Data Analysis and Graphics Using R: An Example-Based Approach*. Cambridge University Press.

Maître, Eric, Brian Nolan, and Christopher T. Whelan. 2014. "L'Indicateur EU2020 de suivi de la Pauvreté et d'Exclusion." *Economie et Statistique* 469-470: 147-167.

Malthus, Thomas R. 1798. *An Essay on the Principle of Population, as It Affects the Future Improvement of Society. With Remarks on the Speculations of Mr. Godwin, M. Condorcet and Other Writers*. London, UK: J. Johnson.

Mangin, William 1967. "Latin American Squatter Settlements: A Problem and a Solution." *Latin American Research Review* 2 (3): 65-98.

Marlier, Eric, and Anthony B. Atkinson. 2010. "Indicators of Poverty and Social Exclusion in a Global Context." *Journal of Policy Analysis and Management* 29: 285-304.

Marlier, Eric, Anthony B. Atkinson, Bea Cantillon, and Brian Nolan. 2007. *The EU and Social Inclusion: Facing the Challenges.* Bristol, UK: The Policy Press.

Marshall, Roger John. 2001. "The Use of Classification and Regression Trees in Clinical Epidemiology." *Journal of Clinical Epidemiology* 54 (6): 603-609.

Marx, Karl. (1847) 1947. *Wage, Labour and Capital*. Moscow.

Marx, Karl. (1867) 1990. *Capital: A Critique of Political Economy (Volume I)*. New York, NY: Penguin Books.

Marx, Karl, and Frederick Engels. 1848. *Manifest der Kommunistischen Partei* . J E Burghard.

Maslow, Abraham H. 1962. *Toward a Psychology of Being*. Princeton, NJ: D. Van Nostrand Company.

Massey, Douglas S., and Nacy A. Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass.* Cambridge, MA: Harvard University Press.

May, Julian. 2000. *Poverty and Inequality in South Africa: Meeting the Challenge.* Cape Town, ZA and London, UK: David Phillip and Zed Press.

Mayer, Susan. E. 1995. "A Comparison of Poverty and Living Conditions in the United States, Canada, Sweden, and Germany." In Poverty, Inequality and the Future of Social Policy: Western States in the New World Order, edited by K. McFate, R. Lawson and W. J. Wilson, 109-151. New York, NY: Sage.

Mayhew, Henry. (1852) 2008. *London Labour and the London Poor*. Wordsworth Editions Limited.

Mazziotta, Matteo, and Adriano Pareto. 2016. "On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena." *Social Indicators Research* 127(3): 983-1003.

—. 2017. "Synthesis of Indicators: The Composite Indicators Approach." In *Complexity in Society: From Indicators Construction to their Synthesis*, edited by Filomena Maggino, 115-138. Cham, Switzerland: Springer International Publishing.

McDougall, Alexander. 1885. *Drink and Poverty*. Manchester.

Mead, George Herbert. 1934. *Mind, Self, and Society*. Chicago, IL: The University of Chicago Press.

Méndez, Floribel, and Juan Diego Trejos. 2004. "Costa Rica: Un Mapa de Carencias Criticas para el Ano 2000." In *Costa Rica a la Luz del Censo 2000*, edited by Instituto Nacional de Estadistica y Censos (INEC), 205–233. INEC.

Messenger, Robert, and Lewis Mandell. 1972. "A Modal Search Technique for Predictive Nominal Scale Multivariate Analysis." *Journal of the American Statistical Association* 67(340): 768-772.

Mestrum, Francine. 2002. *Mondialisation et Pauvreté: De l'Utilité de la Pauvreté dans le Nouvel Ordre Mondia*l. L'Harmattan.

Meulders, Danièle, and Sile Padraigin O'Dorchai. 2010. "Revisiting Poverty Measures towards Individualisation." *DULBEA Universite Libre de Bruxelles Working Papers* 10-03.

Meulders, Danièle, Sile Padraigin O'Dorchai, Janneke Plantenga, and Francesca Bettio. 2009. "Gender Segregation in the Labour Market: Roots, Implications and Policy Responses: A Comparative Review of Thirty European countries." *ULB Institutional Repository* 2013/92386.

Mfusi, Zanele E., and Krishna K. Govender. 2015. "Alleviating Poverty in South Africa – A Theoretical Overview of the Expanded Public Works Program." *Journal of Economics* 6 (2): 118-127.

Miceli, David. 1998. "Measuring Poverty using Fuzzy Sets." *University of Canberra National Centre for Social and Economic Modelling Discussion Paper* 38.

Michalski, R.S., and R.L. Chilausky. 1980. "Learning by Being Told and Learning from Examples: An Experimental Comparison of the Two methods of Knowledge Acquisition in the Context of Developing an Expert System for Soybean Disease Diagnosis." *Policy Analysis and Information Systems* 4: 125–160.

Michie, Donald, David Spiegelhalter, and Charles C. Taylor. 1994. *Machine Learning, Neural and Statistical Classification*. Ellis Horwood.

Miller, Walter B. 1958. "Lower Class Culture as a Generating Milieu of Gang Delinquency." *Journal of Social Issues* 14 (3): 5-19.

—. 1959. "Implications of Urban Lower-class Culture for Social Work." *Social Service Review* 33: 219-236.

Mitchell, Deborah. 1991. *Income Transfers in Ten Welfare States*. Aldershot, UK: Avebury.

Mitchell, Paul Mark, Tracy E. Roberts, Pelham M. Barton, and Joanna Coast. 2017. "Applications of the Capability Approach in the Health Field: A Literature Review." *Social Indicators Research* 133: 345-371.

Mitchell, Tom M. 1997. *Machine Learning*. New York, NY: McGraw Hill.

Mitra, Sophie, and Debra L. Brucker. 2014. "Poverty and Disadvantage through a Multidimensional Lens in the United States." *Fordham University Discussion Paper* 5.

Modigliani, Franco. 1986. "Life Cycle, Individual Thrift, and the Wealth of Nations." *American Economic Review* 76 (3): 297-313.

Moisen, G.G. 2008. "Classification and Regression Trees." *Encyclopedia of Ecology* 1: 582-588.

Mollat, Michel. 1978. *Les Pauvres au Moyen Age*. Hachette.

Møller, Valerie. 1998. "Quality of life in South Africa: Post-apartheid Trends." *Social Indicators Research* 43 (1-2): 27-68.

—. 2007a. "Quality of life in South Africa: the First Ten Years of Democracy." *Social Indicators Research* 81 (2): 181, 183-201.

—. 2007b. "Satisfied and Dissatisfied South Africans: Results from the General Household Survey in International Comparison." *Social Indicators Research* 81 (2): 389-415.

—. 2013. "South African Quality of Life Trends Over Three Decades, 1980–2010." *Social Indicators Research* 113 (3): 915-940.

Møller, Valerie, and Lawrence Schlemmer. 1983. " Quality of life in South Africa: Towards an instrument for the assessment of quality of life and basic needs". *Social Indicators Research* 12 (3): 225-289.

Moon, Mikyung, and Soo-Kyoung Lee. 2017. "Applying of Decision Tree Analysis to Risk Factors Associated with Pressure Ulcers in Long-Term Care Facilities." *Healthc Inform Research* 23 (1): 43-52.

Moore, Barrington Jr. 1966. *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*. Boston, MA: Beacon Press.

Morgan, James N., and John A. Sonquist. 1963 . "Problems in the Analysis of Survey Data and a Proposal." *Journal of the American Statistical Association* 58 (302): 415–434.

Morris, David. 1979. *Measuring the Condition of the World's Poor: The Physical Quality of Life Index.* New York, NY: Pergamon.

Morris, Lydia. 1994. *Dangerous Class: The Underclass and Social Citizenship*. Routledge.

Morse, Richard M. 1965. "Recent Research on Latin American Urbanization: A Selective Survey with Commentary." *Latin American Research Review* 1 (1): 35-74.

Morse, Stephen 2006. "Making Development Simple. The Genetic Deterministic Hypothesis for Economic Development." *Ecological Economics* 56: 79-88.

Mosteller, Frederick, Robert E.K. Rourke, and George B. Thomas, Jr. 1961. *Probability with Statistical Applications*. Addison-Wesley Publishing Company.

Moynihan, Daniel P. 1965. *The Negro Family: The Case for National Action*. Washington, D.C.: Office of Policy Planning and Research, U.S. Department of Labor.

Muffels, R., A. Kapteyn, A. de Vries, et al. 1990. *Poverty in the Netherlands: Report on The Dutch Contribution to an International Comparative Study on Poverty and the Financial Efficacy of the Social Security System.* The Hague : VUGA.

Murmis, Miguel. 1969. "Tipos de Marginalidad y Posición en el Proceso Productivo." *Revista Latinoamericana de Sociología* 2: 413-421.

Murray, Charles. 1984. *Losing Ground: American Social Policy, 1950-1980*. Basic Books.

—. 1987. "In Search of the Working Poor." *Public Interest* 89: 3-19.

—. 1990. *The Emerging British Underclass*. London, UK: Institute of Economic Affairs, Health and Welfare Unit.

—. 1994. "The Underclass: The Crisis Deepens." *IEA Choice in Welfare Series* 20.

Mushongera, Darlington, Precious Zikhali, and Phindile Ngwenya. 2017. "A Multidimensional Poverty Index for Gauteng Province, South Africa: Evidence from Quality of Life Survey Data." *Social Indicators Research* 130: 277–303 .

Myrdal, Gunnar. 1944. *An American Dilemma: The Negro Problem and Modern Democracy*. New York, NY: Harper & Bros.

—. 1963. *Challenge to Affluence*. London, UK: Victor Gollancz Ltd.

—. 1970. *The Challenge of World Poverty: A World Anti-poverty Program in Outline*. New York, NY: Pantheon Books.

Nandy, Shailen, and Marco Pomati. 2015. "Applying the Consensual Method of Estimating Poverty in a Low Income African Setting." *Social Indicators Research* 124: 693–726.

Narayan, Deepa, Robert Chambers, Meera K. Shah, and Petesch Patti. 2000. Voices of the Poor: Crying out for Change. New York, NY: Oxford University Press.

Narayan, Deepa, and Petesch Patti. 2007. *Moving Out of Poverty : Volume 1. Cross-Disciplinary Perspectives on Mobility*. Washington, DC: World Bank and Palgrave Macmillan.

Narayan, Deepa, Lant Pritchett, and Kapoor Soumya. 2009. *Moving Out of Poverty : Volume 2. Success from the Bottom Up*. Washington, DC: World Bank and Palgrave Macmillan.

Nardo, Michela, Michaela Saisana, Andrea Saltelli, Stefano Tarantola, Anders Hoffman, and Enrico Giovannini. 2005. "Handbook on Constructing Composite Indicators." *OECD Statistics Working Paper*.

Naude, Wim, Mark McGillivray, and Stephanie Rossouw. 2008. "Measuring the Vulnerability of Subnational Regions." *UNU-WIDER Working Paper* 54/2008.

Neff, Daniel. 2012. "Adaptation, Poverty and Subjective Well-Being: Evidence from South India." In *Adaptation, Poverty and Development: The Dynamics of Subjective Well-Being*, edited by David A. Clark, 137-157. Palgrave macmillan.

Newbold, Paul, William L. Carlson, and Betty M. Thorne. 2013. *Statistics for Business and Economics*, 8th Edition. Pearson Education.

Niemietz, Kristian. 2010. "Measuring Poverty: Context-Specific but not Relative." *Journal of Public Policy* 30 (3): 241-262.

Nisbet, Robert, John Elder, and Gary Miner. 2009. *Handbook of Statistical Analysis and Data Mining Applications*. New York, NY: Academic Press.

Noble, Michael, Miriam Babita, Helen Barnes, Chris Dibben, Wiseman Magasela, Stefan Noble, Phakama Ntshongwana, et al. 2006. *The Provincial Indices of Multiple Deprivation for South Africa 2001 Technical Report. Centre for the Analysis of South African Social Policy*. Department of Social Policy and Social Work. University of Oxford, UK.

Nolan, Brian, and Christopher T. Whelan. 1996. "Measuring Poverty Using Income and Deprivation Indicators: Alternative Approaches." *Journal of European Social Policy* 6 (3): 225-240.

—. 2007. "On the Multidimensionality of Poverty and Social Exclusion." In *Poverty and Inequality: New Directions*, edited by Stephen R. Jenkins and John Micklewright, 146-164. Oxford: Oxford University Press.

—. 2010. "Using Non-Monetary Deprivation Indicators to Analyze Poverty and Social Exclusion: Lessons from Europe?" *Journal of Policy Analysis and Management* 29 (2): 305-325.

—. 2011. *Poverty and Deprivation in Europe*. Oxford: Oxford University Press.

Nolan, Brian, and Ive Marx. 2000. "Low Pay and Household Poverty." In *Labour market Inequalities: Problems and Policies of Low-wage Employment in International Perspective*, edited by Mary Gregory, Wiemer Salverda and Stephen Bazen, 100-119. Oxford: Oxford University Press.

Noll, Heinz-Nerbert. 2004. "Social Indicators and Quality of Life Research: Background, Achievements and Current Trends." In *Advances in Sociological Knowledge: Over half a Century*, edited by Nikolai Genov, 151-181. Wiesbaden, DE: VS Verlag für Sozialwissenschaften.

Norris, Fran H., Susan P. Stevens, Betty Pfefferbaum, Karen F. Wyche, and Rose L. Pfefferbaum . 2008. "Community Resilience as a Metaphor, Theory, Set of Capacities, and Strategy for Disaster Readiness." *American Journal of Community Psychology* 41: 127–150.

Notenbaert, An, Stanley Nganga Karanja, Mario Herrero, Maute Felisberto, and Siboniso Moyo. 2013. "Derivation of a Household-level Vulnerability Index for Empirically Testing Measures of Adaptive Capacity and Vulnerability." *Regional Environmental Change* 13: 459–470.

Nowak, Daniel, and Christoph Scheicher. 2017. "Considering the Extremely Poor: Multidimensional Poverty Measurement for Germany." *Social Indicators Research* 133: 139-162.

Ntsalaze, Lungile, and Sylvanus Ikhide. 2018. "Rethinking Dimensions: The South African Multidimensional Poverty Index." *Social Indicators Research* 135: 195-213.

Nun, José. 1969. "Superpoblacion Relativa, Ejercito Industrial de Reserva y Masa Marginal." *Revista Latinoamericana de Sociologia* 5 (2): 178-235.

—. 1999. "El Futuro del Empleo y la Tesis de la Masa Marginal." *Desarrollo Economico* 38 (152): 985-1004.

Nussbaum, Martha C. 1988. "Nature, Functioning and Capability: Aristotle on Political Distribution." *Oxford Studies in Ancient Philosophy* 6: 145-184.

—. 1992. "Human Functioning and Social Justice. In Defense of Aristotelian Essentialism." *Political Theory* 20 (2): 202-246.

—. 1993. "Non-Relative Virtues: An Aristotelian Approach." In *The Quality of Life*, edited by Martha Nussbaum and Amartya K. Sen. Oxford: Clarendon Press.

—. 1995. "Human Capabilities, Female Human Beings." In *Women, Culture, and Development: A Study of Human Capabilities*, edited by Martha C. Nussbaum and Jonathan Glover, 61-104. Oxford University Press.

—. 1997. "Capabilities and Human Rights." *Fordham Law Review* 66 (2): 273-300.

—. 1999. *Sex and Social Justice*. Oxford University Press.

—. 2000. *Women and Human Development: The Capabilities Approach*. Cambridge University Press.

—. 2002a. "Capabilities and Social Justice." *International Studies Review* 4 (2): 123-135.

—. 2002b. "Women's Capabilities and Social Justice." In *Gender Justice, Development, and Rights*, edited by Maxine Molyneux and Shahra Razavi. Oxford University Press.

—. 2003. "Capabilities as Fundamental Entitlements: Sen and Social Justice." *Feminist Economics* 9 (2-3): 33-59.

—. 2004. "Beyond the Social Contract: Capabilities and Global Justice." *Oxford Development Studies* 32: 3-18.

—. 2006. "Poverty and Human Functioning: Capabilities as Fundamental Entitlements." In *Poverty and Inequality*, edited by David B. Grusky and Ravi Kanbur, 47-75. Stanford, CA: Stanford University Press.

—. 2008. "The Capabilities Approach and Social Justice: Human Rights and Political Participation in the 21st century." In *Democracy in a Global World*, edited by Deen K. Chatterjee, 111-146. Rowman and Littlefield Publishers, Inc. .

—. 2011. *Creating Capabilities*. Cambridge, MA: Harvard University Press.

—. 2013. *Creating Capabilities: The Human Development Approach*. Harvard University Press.

Nussbaum, Martha C., and Amartya K. Sen, eds. 1993. *The Quality of Life*. Clarendon Press.

Ó Gráda , Cormac. 2010. *Famine : A Short History*. Princeton, NJ: Princeton University Press.

O'Connor, Alice. 2001. *Poverty Knowledge: Social Science, Social Policy, and the Poor in Twentieth-Century U.S. History*. Princeton, NJ: Princeton University Press.

ODI. 1978. "Basic Needs." *ODI Briefing Paper* 5.

—. 1980. "The Brandt Commission." *ODI Briefing Paper* 2.

OECD. 1976. *Public Expenditure on Income Maintenance Programmes*. Paris, FR: OECD.

—. 1982. *The OECD List of Social Indicators*. Paris, FR: OECD.

—. 2008a. *Growing Unequal? Income Distribution and Poverty in OECD Countries*. Paris, FR: OECD.

—. 2008b. *Handbook on Constructing Composite Indicators*. Paris, FR: OECD.

—. 2012. *Perspectives on Global Development 2012: Social Cohesion in a Shifting World*. Paris, FR: OECD.

—. 2013. *OECD Guidelines on Measuring Subjective Well-being*. Paris, FR: OECD.

—. 2017. *How's Life? 2017. Measuring Well-being*. Paris, FR: OECD.

Ogien, Ruwen. 1983. *Théories Ordinaires de la Pauvreté*. Paris, FR: Presses Universitaires France.

Oldfield, Nina, and Autumn C. S. Yu. 1993. *The Cost of a Child. Living Standards for the 1990s*. London, UK: Child Poverty Action Group.

Ore, Oystein. 1960. "Pascal and the Invention of Probability Theory." *American Mathematical Monthly* 67 (5): 409-419.

Orshansky, Mollie. 1963. "Children of the poor." *Social Security Bulletin* 26 (7): 3-13.

—. 1965. "Counting the Poor: Another Look at the Poverty Profile." *Social Security Bulletin* 28 (1): 3-29.

—. 1969. "How Poverty is Measured." *Monthly Labor Review* 92 (2): 37-41.

Osmani, Siddiqur. 2009. "When Capabilities and Opportunities Don't Match: Understanding Chronic Poverty." In *Poverty Dynamics: Interdisciplinary Perspectives*, edited by Tony Addison, David Hulme and Ravi Kanbur. New York, NY: Oxford University Press.

Paes-Sousa, R. P., and Paulo de Martino Jannuzzi. 2016. "Going Beyond the Eradication of Extreme Poverty in Brazil: the Case for Sustainable Development Goals." In *Poverty and the Millennium Development Goals*, edited by Alberto Cimadamore, Gabriele Koehler and Thomas Pogge, 112-125. London, UK: Zed Books.

Pantazis, Christina, David Gordon, and Ruth Levitas, eds. 2006. *Poverty and Social Exclusion in Britain: The Millenium Survey*. Bristol, UK: The Policy Press.

Park, Robert E. 1921. *Introduction to the Science of Sociology, "Isolation"/ "Social Contacts"*. Chicago, IL: The University of Chicago Press.

—. 1928. "Human Migration and the Marginal Man." *American Journal of Sociology* 33 (6): 881-893.

Parker, Hermione. 1998. *Low Cost but Acceptable: A Minimum Income Standard for the UK*. Family Budget Unit: London.

Parsons, Talcott. 1967. *Sociological Theory and Modern Society.* New York, NY : The Free Press.

Pasqua, Silvia. 2008. "Wives' Work and Income Distribution in European countries." *European Journal of Comparative Economics* 5 (2): 197-226.

Paugam, Serge. 1991a. *La Disqualification Sociale. Essai sur la Nouvelle Pauvreté*. Presses Universitaires de France.

—. 1991b. "Les Statuts de la Pauvreté Assistée." *Revue Francaise de Sociologie* 32 (1): 75-101.

—. 1993. *La Société Francaise et des Pauvres. L'expérience du Revenu Minimum d'Insertion.* Presses Universitaires de France.

—, ed. 1996. *L'Exclusion, l'Etat des savoirs*. La Découverte.

—. 1997. "Poverty and Social Disqualification: A Comparative Analysis of Cumulative Social Disadvantage in Europe." *Journal of European Social Policy* 6 (4): 287-303.

—. 2000. *Le Salarié de la Précarité. Les Nouvelles Formes de l'Intégration Professionnelle*. Presses Universitaires de France.

—. 2005. *Les Formes Elémentaires de la Pauvreté.* Paris, FR: Presses Universitaires de France.

—, ed. (2007) 2015. *Repenser la Solidarité. L'Apport des Sciences Sociales*. Presses Universitaires de France.

—. 2016. "La Perception de la Pauvreté sous l'Angle de la Théorie de l'Attachement: Naturalisation, Culpabilisation et Victimisation." *Communications* 98: 125-146.

—. (2008) 2018. *Le lien Social*. Presses Universitaires de France.

Paugam, Serge, Tugce Beycan, and Christian Suter.2020. "What Attaches Individuals to Groups and Society: A European Comparison." *Swiss Journal of Sociology* 46 (1): 7-35.

Paugam, Serge, and Nicolas Duvoux. 2008. La Régulation des Pauvres. Presses Universitares de France.

Paugam, Serge, and Marion Selz. 2005. "La Perception de la Pauvreté en Europe depuis le Milieu des Années 1970. Analyse des Variations Structurelles et Conjoncturelles." *Economie et Statistique* 383-385: 283-305.

Peña-Casas, Ramón, and Dalila Ghailani. 2011. "Towards Individualizing Gender In-Work Poverty Risks." In *Working Poverty in Europe. Work and Welfare in Europe*, edited by Neil Fraser , Rodolfo Gutiérrez and Ramón Peña-Casas, 202-231. London, UK: Palgrave Macmillan.

Pena-Casas, Ramón, and Mia Latta. 2004. *Working Poor in the European Union*. Office for Official Publications of the European Communities.

Perlman, Janice E. 1976. *The Myth of Marginality: Urban Poverty and Politics in Rio de Janeiro.* Berkeley, CA: Univ. Calif. Press.

Phillips, David. 2006. *Quality of Life: Concept, Policy and Practice*. New York, NY: Routledge.

Philp, A. F., and Noel Timms. 1957. *The Problem of the 'Problem Family'* . London, UK: Family Service Units.

Piachaud, David. 1981. "Peter Townsend and the Holy Grail." *New Society Sept.* 10th: 419-421.

Piachaud, David. 1987. "Problems in the Definition and Measurement of Poverty." *Journal of Social Policy* 16 (2): 147-164.

Piatetsky-Shapiro, Gregory, and William J. Frawley. 1991. *Knowledge Discovery in Databases.* Menlo Park, CA: AAAI Press/MIT Press.

Pino-Mejias, Rafael, Mercedes Carrasco-Mairena, Antonio Pascual-Acosta, Maria-Dolores Cubiles-de-la-Vega, and J. Munoz-Garcia. 2008. "A Comparison of Classification Models to Identify the Fragile X Syndrome." *Journal of Applied Statistics* 35 (3): 233-244.

Plantenga, Janneka, and Chantal Remery. 2005. *Reconciliation of Work and Private Life: A Comparative Review of Thirty European countries*. European Commission.

Pogge, Thomas. 2002a. "Can the Capability Approach be Justified?" *Philosophical Topics* 30 (2): 167–228.

—. 2010. "A Critique of the Capability Approach." In *Measuring Justice: Primary Goods and Capabilities*, edited by Harry Brighouse and Ingrid Robeyns, 17-60. Cambridge, UK: Cambridge University Press.

—. 2016. "Are We Violating the Human Rights of the World's Poor?" In Ethical Issues in *Poverty Alleviation*, edited by Helmut P. Gaisbauer, Gottfried Schweiger and Clemens Sedmak, 17-42. Springer, Cham.

Pogge, Thomas, and Sanjay G. Reddy. 2003. "Unknown: The Extent, Distribution and Trend of Global Income Poverty." *Economic and Political Weekly* 41 (22): 2241-2247.

—. 2010. "How Not to Count the Poor." In *Debates on the Measurement of Global Poverty*, edited by Sudhir Anand, Paul Segal and Joseph E. Stiglitz, 42-85. Oxford: Oxford University Press.

Portes, Alejandro. 1985. "Latin American Class Structures: Their Composition and Change During the Last Decades." *Latin American Research Review* 20 (3): 7-39.

Posel, Dorrit, and Daniela Casale. 2011. "Relative Standing and Subjective Well-Being in South Africa: The Role of Perceptions, Expectations and Income Mobility." *Social Indicators Research* 104 (2): 195-223.

Poynter, John R. 1960. *Society and Pauperism*. London, UK: Routledge & Kegan Paul.

Pradhan , Menno, and Martin Ravallion. 2000. "Measuring Poverty using Qualitative Perceptions of Consumption Adequacy." *Review of Economics and Statistics* 82(3): 462-471.

Pritchett, Lant, Asep Suryahadi, and Sudarno Sumarto. 2000. "Quantifying Vulnerability to Poverty: A Proposed Measure, Applied to Indonesia." *The World Bank Policy Research Working Paper* 2437.

Qizilbash, Mozaffar. 1996. "Capabilities, Well-being and Human Development: A Durvey." *Journal of Development Studies* 33(2): 143-162.

—. 2002. "A Note on the Measurement of Poverty and Vulnerability in the South African Context." *Journal of International Development* 14(6): 757-772.

—. 2006. "Well-being, Adaptation and Human Limitations." *Royal Institute of Philosophy Supplements* 81: 83–110.

Qizilbash, Mozaffer, and David A. Clark. 2005. "The Capability Approach and Fuzzy Poverty Measures: An Application to the South African Context." *Social Indicators Research* 74: 103-139.

Quijano, Anibal. 1969. *Notas sobre la Marginalidad y Politica en America Latina*. CEPAL, United Nations.

—. 1974. "Marginal Pole and Marginal Labor Force in Latin America." *Economy and Society* 3(4): 393-428.

Quijano, Anibal, and Peggy Westwell. 1983. "Imperialism and Marginality in Latin America." *Latin American Perspectives* 10 (2-3): 76-85.

Quinlan, John Ross. 1986. "Induction of Decision Trees." *Machine Learning* 1 (1): 81-106.

—. 1987. "Simplifying Decision Trees." *International Journal of Man-Machine Studies* 27: 221-234.

—. 1988. "Decision Trees and Multi-valued Attributes." In *Machine Intelligence 11*, edited by J.E. Hayes, D. Michie and J. Richards. Oxford: Oxford University Press.

—. 1992. *C4.5. Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.

Quisumbing, Agnes R. 2009. "Investments, Bequests, and Public Policy: Intergenerational Asset Transfers and the Escape from Poverty." In *Poverty Dynamics: Interdisciplinary Perspectives*, edited by Tony Addison, David Hulme and Ravi Kanbur. Oxford, UK: Oxford University Press.

Rainwater, Lee. 1966. "Crucible of Identity: The Negro Lower-class Family." *Daedalus* 95: 176-216.

Rainwater, Lee, and William L. Yancey. 1967. *The Moynihan Report and the Politics of Controversy*. Cambridge, MA: MIT Press.

Rajadel, Tania. 2002. "Vulnerability and Participation to the Non-Agricultural Sector in Rural Pakistan." *Université Panthéon-Sorbonne, TEAM (Théories et Applications en Microéconomie et Macroéconomie) Working Paper.*

Ralston, Margaret, Enid Schatz, Jane Menken, Francesc Xavier Gómez-Olivé, and Stephen Tollman. 2016. "Who Benefits-or Does not-from South Africa's Old Age Pension? Evidence from Characteristics of Rural Pensioners and Non-pensioners." *International Journal of Environmental Research and Public Health* 13 (1): 85.

Ramphele, Mamphela, and Francis Wilson. 1989. *Uprooting Poverty: The South African Challenge*. New York, NY: Norton.

Ramprakash, Deo. 1994. "Poverty in the Countries of the European Union: a Synthesis of Eurostat's Statistical Research On Poverty." *Journal of European Social Policy* 4 (2): 117-128.

Ranci, Costanzo. 2010. "Social Vulnerability in Europe." In *Social Vulnerability in Europe : The New Configuration of Social Risks*, edited by Costanzo Ranci, 3-24. London UK: Palgrave Macmillan.

Ranci, Costanzo, and Mauro Migliavacca. 2010. "Social Vulnerability: A Multidimensional Analysis." In *Social Vulnerability in Europe: The New Configuration of Social Risks*, edited by Costanzo Ranci, 219-249. London, UK: Palgrave Macmillan.

Ranis, Gustav, and Frances Stewart. 2000. "Success and Failure in Human Development, 1970–2007." *Journal of Human Development and Capabilities* 13 (2): 167-195.

Ranis, Gustav, Frances Stewart, and Emma Samman. 2006. "Human Development: Beyond the Human Development Index." *Journal of Human Development and Capabilities* 7 (3): 323-358.

Ravallion, Martin. 1988. "Expected Poverty under Risk-Induced Welfare Variability." *The Economic Journal* 98: 1171-1182.

—. 1992. *Poverty Comparisons: A Guide to Concepts and Methods*. The World Bank.

—. 1996. "Issues in Measuring and Modleing Poverty." *The World Bank Policy Research Working Paper* 1615.

—. 1998. *Poverty Lines in Theory and Practice*. The World Bank.

—. 2003. "The Debate on Globalization, Poverty and Inequality: Why Measurement Matters." *International Affairs* 79(4): 739-754.

—. 2011. "On Multidimensional Indices of Poverty." *Journal of Economic Inequality* 9: 235-248.

—. 2012a. "Poor or Just Feeling Poor? On Using Subjective Data in Measuring Poverty." *The World Bank Policy Research Working Paper* 5968.

—. 2012b. "Why Don't We See Poverty Convergence?" *American Economic Review* 102 (1): 504-523.

Ravallion, Martin, and Shaohua Chen. 2009. "Weakly Relative Poverty." T*he World Bank Policy Research Working Paper* 4844.

Ravallion, Martin, Gaurav Datt, and Dominique Van de Walle. 1991. "Quantifying Absolute Poverty in the Developing World." Review of Income and Wealth 37 (4): 345-361.

Ravallion, Martin, and Michael Lokshin. 2002. "Self-rated economic welfare in Russia." *European Economic Review* 46: 1453-1473.

Rawls, John. 1958. "Justice as Fairness." *Philosophical Review* 67 (2): 164-194.

—. 1971. *A Theory of Justice*. The Belknap Press of Harvard University Press.

—. 1982. "Social Unity and Primary Goods." In *Utilitarianism and Beyond*, edited by Amartya K. Sen and B.A.O. Williams. Cambridge, New York: Cambridge University Press.

—. 1993. *Political Liberalism*. New York, NY: Columbia University Press.

—. 2001. *Justice as Fairness: A Restatement.* Harvard University Press.

Raworth, Kate, and David Stewart. 2003. "Academic Critiques of the Human Development Index: A Review." In *Readings in Human Development: Concepts, Measures and Policies for a Development Paradigm*, edited by Sakiko Fukuda-Parr and A.K. Shiva Kumar, 164-167. New Delhi, IN: Oxford University Press.

Reddy, Sanjay G., and Camelia Minoiu. 2007. "Has World Poverty Really Fallen?" *Review of Income and Wealth* 53: 484-502.

Redfield, Robert. 1955. *The Little Community: Viewpoints for the Study of a Human Whole.* Chicago, IL: The University of Chicago Press.

Reid, John. 1982. "Black America in the 1980s." *Population Bulletin* 37 (4).

Rein, Margaret 1970. "Problems in the Definition and Measurement of Poverty." In *The Concept of Poverty. Working Papers on Methods of Investigation and Life-styles of the Poor in Different Countries*, edited by Peter Townsend, 46-63. London, UK: Heinemann Educational Books.

Republic of South Africa. 1996. *The Constitution of the Republic of South Africa, 1996*. ISBN 978-0-621-39063-6.

—. 2002. *The Integrated Food Security Strategy for South Africa*. Pretoria, ZA: Department of Agriculture, Republic of South Africa.

—. 2007a. *ASGISA Annual Report 2007.* Pretoria, ZA: Republic of South Africa.

—. 2007b. *Expanding Economic Opportunity in the Presidential Poverty Nodes*. Department of Provincial and Local Government and Business Trust.

Ricardo, David. (1817) 2004. *The Principles of Political Economy and Taxation*. New York, NY: Dover Publications Inc. .

Ricketts, Erol R., and Isabel V. Sawhill. 1988. "Defining and Measuring the Underclass." *Journal of Policy Analysis and Management* 7 (2): 316-325.

Riffault, Hélène. 1991. "How Poverty is Perceived." In *Eurobarometer*, edited by K. Reif and R. Inglehart, 349-354. London, UK: Palgrava Macmillan.

Ringen, Stein. 1988. "Direct and Indirect Measures of Poverty." *Journal of Social Policy* 17(3): 351-365.

Rio Group. 2006. *Compendium of Best Practices in Poverty Measurement*. Rio de Janeiro, BR: Rio Group (Expert Group on Poverty Statistics).

Ripley, Brian D. 1996. *Pattern Recognition and Neural Networks*. Cambridge, UK: Cambridge University Press.

Rippin, Nicole. 2016. "Multidimensional Poverty in Germany: A Capability Approach." *Forum for Social Economics* 45 (2-3): 230-255.

Robbins, D., et al. 1994. *Observatory on National Policies to Combat Social Exclusion. Third annual report.* Brussels: Commission of the European Communities, DG V.

Roberts , Bryan R. 1978. *Cities of Peasants: The Political Economy of Urbanization in the Third World*. London, UK Sage.

Robeyns, Ingrid. 2003. "Sen's Capability Approach and Gender Inequality: Selecting Relevant Capabilities." *Feminist Economics* 9 (2/3): 61-92.

—. 2005a. "Assessing Global Poverty and Inequality: Income, Resources, and Capabilities." *Metaphilosophy* 36 (1-2): 30-49.

—. 2005b. "Selecting Capabilities for Quality of Life Measurement." *Social Indicators Research* 74 (1): 191-215.

—. 2005c. "The Capability Approach: A Theoretical Survey." *Journal of Human Development* 6 (1): 93-117.

—. 2006a. "Three Models of Education Rights, Capabilities and Human Capital." *Theory and Research in Education* 4 (1): 69-84.

—. 2006b. "The Capability Approach in Practice." *Journal of Political Philosophy* 14 (3): 351–76.

—. 2008. "Sen's Capability Approach and Feminist Concerns." In *The Capability Approach: Concepts, Measures, Applications*, edited by Sabina Alkire, Flavio Comin and Mozaffer Qizilbash, 82-104. Cambridge, UK: Cambridge University Press.

—. 2017. *Wellbeing, Freedom and Social Justice: The Capability Approach Re-examined*. Cambridge, UK: Open Book Publishers.

Robeyns, Ingrid, and Harry Brighouse. 2010. "Introduction: Social Primary Goods and Capabilities as Metrics of Justice." In *Measuring Justice: Primary Goods and Capabilities*, edited by Harry Brighouse and Ingrid Robeyns, 1-14. Cambridge, UK: Cambridge University Press.

Robinson, Peter, and Carey Oppenheim. 1998. *Social Exclusion Indicators: A Submission to the Social Exclusion Unit*. London, UK: IPPR.

Rodgers, Gerry, and Janine Rodgers. 1989. *Precarious Jobs in Labour Market Regulation: The Growth of Atypical Employment in Western Europe.* Geneva, CH: International Institute for Labour Studies.

Rogan, Michael. 2016. "Gender and Multidimensional Poverty in South Africa: Applying the Global Multidimensional Poverty Index (MPI)." *Social Indicators Research* 126 (3): 987-1006.

Rojas, Mariano. 2006. "Well-being and the Complexity of Poverty: A Subjective Well-being Approach." In *Understanding Human Well-being*, edited by M. McGillivray and M. Clarke, 182-206. United Nations University Press.

—. 2007. "A Subjective Well-being Equivalence Scale for Mexico: Estimation and Poverty and Income-distribution Implications." *Oxford Development Studies* 35(3): 273-293.

—. 2008. "Experienced Poverty and Income Poverty in Mexico: A Subjective Well Being Approach." *World Development* 36(6): 1078-1093.

—. 2009. "Enhancing Poverty-Abatement Programs: A Subjective Well-Being Contribution." *Applied Research in Quality of Life* 4(2): 179-199.

—. 2015. "Poverty and People's Wellbeing." In *Global Handbook of Quality of Life. International Handbooks of Quality-of-Life*, edited by Wolfgang Glatzer, Laura Camfield, Valerie Møller and Mariano Rojas, 317-350. Dordrecht: Springer.

Rojas, Mariano, and E. Jiménez. 2008. "Pobreza Subjetiva en México: El Papel de las Normas de Evaluación de Ingreso." *Revista Perfiles Latinoamericanos* 16 (32): 11-33.

Rokach, Lior, and Oded Maimon. 2010. "Classification Trees." In *Data Mining and Knowledge Discovery Handbook*, edited by Oded Maimon and Lior Rokach, 149-174. Springer.

—. 2008. *Data Mining with Decision Trees: Theroy and Applications*. River Edge, NJ: World Scientific Publishing Co., Inc.

Room, Graham. 1995. *Beyond the Threshold: The Measurement and Analysis of Social Exclusion.* Bristol University Press.

Rowntree, B. Seebohm. 1901. *Poverty: A Study of Town Life*. London, UK: Macmillan and Co.

—. 1941. *Poverty and Progress: A Second Social Survey of York. Poverty and Progress : A Second Social Survey of York*. London, UK: Longmans, Green.

—. No publisher or date printed; Briggs gives 1903 but British Library of Political and Economic Science catalogue gives 1901. *The 'Poverty Line': A Reply* .

Rowntree, B. Seebohm, and G. R. Lavers. 1951. *Poverty and the Welfare State*. London, UK: Longmans, Green and Co.

Ruggeri-Laderchi, Caterina, Ruhi Saith, and Frances Stewart. 2003. "Does it Matter that we do not Agree on the Definition of Poverty? A Comparison of Four Approaches." *Oxford Development Studies* 31 (3): 243-274.

—. 2006. "Does the Definition of Poverty Matter? Comparing four approaches." In *Poverty in Focus: What is Poverty? Concepts and Measures*, edited by Dag Ehrenpreis, 10-11. International Poverty Centre, United Nations Development Programme.

Runciman, Walter Garrison. 1972 . *Relative Deprivation and Social Justice: A Study of Attitudes to Social Inequality in Twentieth-century England*. Penguin Books (first published in 1966 by Routledge & Kegan Paul).

Russell, Stuart, and Peter Norvig. (1994) 2009. *Artifical Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice Hall.

Rutter, Michael, and Nicola Madge. 1976. *Cycles of Disadvanatge: A Review of Research*. London, UK: Heinemann.

Sacks, Harvey. (1964-1968, 1968-1972) 1992. *Lectures on Conversation, Vol. I & Vol. II*. Oxford, UK: Basil Blackwell.

Sagar, Ambuj D., and Adil Najam. 1998. "The Human Development Index: A Critical Review." *Ecological Economics* 25: 249–264.

Saith, Ruhi. 2001. "Capabilities: The Concept and its Pperationalisation." *QEH University of Oxford Working Paper Series* QEHWPS66.

Santos, Maria Emma, Maria Ana Lugo, Luis Felipe Lopez-Calva, Guillermo Cruces, and Diego Battiston. 2010. "Refining the Basic Needs Approach: A Multidimensional Analysis of Poverty in Latin America." In *Studies in Applied Welfare Analysis: Papers from the Third ECINEQ Meeting (Research on Economic Inequality, Vol. 18)*, edited by John A. Bishop, 1-29. Bingley: Emerald Group Publishing Limited.

Santos, Maria Emma, Pablo Villatoro, Xavier Mancero, and Pascual Gerstenfeld. 2015. "A Multidimensional Poverty Index for Latin America." *OPHI University of Oxford Working Paper* 79.

SAPRIN. 2002. *The Policy Roots of Economic Crisis and Poverty: A Multi-country Participatory Assessment of Structural Adjustmen*t. Washington, D.C.: SAPRIN.

Saunders, Peter, Yuvisthi Naidoo, and Megan Griffiths. 2008. "Towards New Indicators of Disadvantage: Deprivation and Social Exclusion in Australia." *Australian Journal of Social Issues* 43 (2): 175-194.

Saunders, Peter, and George Matheson. 1992. "Perceptions of Poverty, Income, Adequacy and Living Standards in Australia." *The University of New South Wales, Social Policy Research Center Reports and Proceedings* 99.

Schapire, Robert E. 1999. "A Brief Introduction to Boosting." *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*.

Schatz, Enid, Sangeetha Madhavan, and Jill Williams. 2011. "Female-headed Households Contending with AIDS-related Hardship in Rural South Africa." *Health Place* 17(2): 598–605.

Schlemmer, Lawrence, and Valerie Møller. 1997. "The Shape of South African Society and its Challenges." *Social Indicators Research* 41 (1-3): 15-50.

Schlesinger, Banjamin. 1963. *The Multi-problem Family*. University of Toronto Press.

Schuschny, Andres Ricardo, y Gilberto Gallopin. 2004. *La Distribución Espacial de la Pobreza en Relación a los Sistemas Ambientales en América Latina.* Santiago de Chile, CL: CEPAL, Naciones Unidas.

Schütz, Alfred. 1970. *On Phenomenology and Social Relations: Selected Writings*. Chicago, IL: The University of Chicago Press .

Scoones, Ian. 1998. "Sustainable Rural Livelihoods: A Framework for Analysis." *IDS Working Paper* 72.

Scott, Federick. 1889. "The Condition and Occupations of the People of Manchester and Salford." *Transactions of the Manchester Statistical Society* 93-116.

Seekings, Jeremy, and Nicole Nattrass. 2015. *Policy, Politics, and Poverty in South Africa*. London, UK: Routledge.

Seidl, Christian. 1988. "Poverty Measurement: A Survey." In *Welfare and Efficiency in Public Economics*, edited by Dieter Bös, Manfred Rose and Christian Seidl, 71-147. Springer.

Seligman, Ben. 1968. *Permanent Poverty: An American Syndrome*. Chicago, IL: Quadrangle Books.

Sen, Amartya K. 1976. "Poverty: An Ordinal Approach to Measurement." *Econometrica* 44 (2): 219-231.

—. 1977. "Starvation and Exchange Entitlements: A General Approach and Its Application to the Great Bengal Famine." *Cambridge Journal of Economics* 1(1): 33-59.

—. 1978. *Three Notes on the Concept of Poverty*. Geneva, CH: International Labour Organization.

—. 1979. "Issues in the Measurement of Poverty." *Scandinavian Journal of Economics* 81 (2): 285-307.

—. 1980. "Equality of What?" In *McMurrin S Tanner Lectures on Human Values, Volume 1*. Cambridge: Cambridge University Press.

—. 1981. *Poverty and Famines: An Essay on Entitlement and Deprivation*. Oxford: Clarendon Press.

—. 1982. *Choice, Welfare and Measurement*. Oxford: Basil Blackwell.

—. 1983. "Poor, Relatively Speaking." *Oxford Economic Papers* 35: 153-169.

—. 1984. *Resources, Values and Development*. Oxford: Basil Blackwell.

—. 1985. "Well-being, Agency and Freedom: the Dewey lectures 1984." *Journal of Philosophy* 82 (4): 169-221.

—. 1987a. *Commodities and Capabilities*. New Delhi, IN: Oxford University Press.

—. 1987b. "The Standard of Living: Lecture I, Concepts and Critiques." In *The Standard of Living*, edited by Geoffrey Hawthorn, 1-19. Cambridge University Press.

—. 1987c. "The Standard of Living: Lecture II, Lives and Capabilities." In *The Standard of Living*, edited by Geoffrey Hawthorn, 20-38. Cambridge University Press.

—. 1990. "Justice: Means versus Freedoms." *Philosophy and Public Affairs* 19 (2): 111-121.

—. 1992. *Inequality Reexamined.* Oxford: Clarendon Press.

—. 1993. "Capability and Well-being." In *The Quality of Life*, edited by Martha Nussbaum and Amartya K. Sen. Oxford: Clarendon Press.

—. 1995a. "Gender Inequality and Theories of Justice." In *Women, Culture and Development*, edited by Martha Nussbaum and Jonathan Glover. Oxford: Clarendon Press.

—. 1995b. "Rationality and social choice." *American Economic Review* 85 (1): 1-24.

—. 1996. "On the Foundations of Welfare Economics: Utility, Capability and Practical Reason." In *Ethics, Rationality and Economic Behavior*, edited by Francesco Farina, Frank Hahn and Stefano Vannucci. Oxford: Clarendon Press.

—. 1999. *Development as Freedom* . New York: Alfred Knopf.

—. 2003. "Development as Capability Expansion." In *Readings in Human Development: Concepts, Measures and Policies for a Development Paradigm*, edited by Sakiko Fukuda-Parr and A. K. Shiva Kumar. New Delhi and New York: Oxford University Press.

—. 2004. "Capabilities, Lists, and Public Reason: Continuing the Conversation." *Feminist Economics* 10 (3): 77-80.

—. 2005. "Capabilities, Lists, and Public Reason: Continuing the Conversation." In *Amartya Sen's Work and Ideas: A Gender Perspective*, edited by Bina Agarwal, Jane Humphries and Ingrid Robeyns, 335-338. Routledge.

—. 2006. "Conceptualizing and Measuring Poverty." In *Poverty and Inequality*, edited by David B. Grusky and Ravi Kanbur, 30-46. Stanford, CA: Stanford University Press.

—. 2009. *The Idea of Justice*. Harvard University Press.

Sessa, Margherita. 2016. "Measuring and Monitoring the Achievements of the Millennium Development Goals Through Dynamic Composite Indices." *Social Indicators Research* 127: 469-503.

Shannon, Claude E. July, October 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27: 379-423, 623-656.

Shannon, Claude E., and Warren Weaver. 1949. *The Mathematical Theory of Communication.* Urbana, IL: The University of Illinois Press.

Shepherd, Andrew. 2013. "An Evolving Framework for Understanding and Explaining Chronic Poverty." In *Chronic Poverty: Concepts, Causes and Policy*, edited by Andrew Shepherd and Julia Brunt, 7-37. London, UK: Palgrave Macmillan.

Sherrieb, Kathleen, Fran H. Norris, and Sandro Galea. 2010. "Measuring Capacities for Community Resilience." *Social Indicators Research* 99 (2): 227-247.

Shildrick, Tracy, and Jessica Rucell. 2015. *Sociological Perspectives on Poverty*. Joseph Rowntree Foundation.

Shorrocks, Anthony F. 1995. "Revisiting the Sen Poverty Index." *Econometrica* 63: 1225-1230.

Silberschatz, Avi, Michael Stonebraker, and Jeff Ullman, eds. October 1991. "Database Systems: Achievements and Opportunities." *Communications of the ACM* 34 (10): 110-120.

Silver, Hilary. 1994. "Social Exclusion and Social Solidarity: Three Paradigms." *International Labour Review* 133 (5-6): 531-578.

Silver, Hilary, and Frank Wilkinson . 1995. "Policies to Combat Exclusion: A French-british Comparison." In *Social Exclusion: Rhetoric, Reality and Responses*, edited by Gerry Rodgers, Gore Charles and Jose B. Figueiredo. International Institute for Labour Studies, UNDP.

Simmel, Georg. (1907) 1998. *Les Pauvres*. Paris, FR: Presses Universitaires France.

Simon, Jonathan. 1993. *Poor Discipline: Parole and the Social Control of the Underclass 1890-1990*. University of Chicago Press.

Simons, Henry. 1938. *Personal Income Taxation: The Definition of Income as a Problem of Fiscal Policy*. Chicago, IL: University of Chicago Press.

Singh, Ajit. 1979. "The 'Basic Needs' Approach to Development vs the New International Economic Order: The Significance of Third World Industrialization." *World Development* 7 (6) : 585-606.

Sirven, Nicolas. 2007. "De la Pauvreté à la Vulnérabilité : Évolutions Conceptuelles et Méthodologiques." *Mondes en Développement* 140: 9-24.

Skoufias, Emmanuel, and Agnes R. Quisumbing. 2005. "Consumption Insurance and Vulnerability to Poverty: A Synthesis of the Evidence from Bangladesh, Ethiopia, Mali, Mexico and Russia." *European Journal of Development Research* 17 (1): 24-58.

Smith, Adam. 1776, Vol. 2. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London, UK: Printed for W. Strahan & T. Cadell.

—. (1759) 1812. *The Theory of Moral Sentiments; or, an Essay towards an Analysis of the Principles by which Men Naturally Judge concerning the Conduct and Character, First of their Neighbours, and then of Themselves.* London & Oxford: Cadell & Davies.

Smith, David J. 1992. *Understanding the Underclass*. Policy Studies Institute .

Smith, Tom, Michael Noble, Stefan Noble, Gemma Wright, David McLennan, and Emma Plunkett. 2015. *The English Indices of Deprivation 2015*. Department for Communities and Local Government, Crown.

Soares, Fabio Veras, Rafael Perez Ribas, and Rafael Guerreiro Osorio. 2007. *Evaluating the Impact of Brazil's Bolsa Família: Cash Transfer Programmes in Comparative Perspective.* Brasilia DF, BR: International Poverty Centre, United Nations Development Programme.

Social Exclusion Unit. 1998a. *Bringing Britain Together: a National Strategy for Neighbourhood Renewal*. London, UK: The Stationery Office.

—. 1998b. *Rough Sleeping*. London, UK: The Stationery Office.

—. 1999. *Teenage Pregnancy*. London, UK: The Stationery Office.

Spencer, John C. 1970. "The Multi-problem Family." In *The Multi-problem Family: A Review and Annotated Bibliography*, edited by Benjamin Schlesinger, 3-54. Toronto, CA: University of Toronto Press.

Spicker, Paul. 1990. "Charles Booth: The Examination of Poverty." *Social Policy Administration* 24 (1): 21-38.

Spicker, Paul, Sonia Alvarez Leguizamon, and David Gordon, eds. 2007. *Poverty: An International Glossary*. Zed Books.

Srinivasan, T. N. 2010. "Irrelevance of the $1 a Day Poverty Line." In *Debates on the Measurement of Global Poverty*, edited by Sudhir Anand, Paul Segal and Joseph E. Stiglitz, 143-151. New York: Oxford University Press.

Statistics South Africa. 2000. *Measuring Poverty in South Africa*. Pretoria, ZA: Statistics South Africa.

—. 2011. *Statistical Release (revised version) General Household Survey 2010*. Pretoria, ZA: Statistics South Africa.

—. 2012a. *Statistical Release (revised) Census 2011*. Pretoria, ZA: Statistics South Africa.

—. 2012b. *Subjective Poverty in South Africa: Findings of the Living Conditions Survey 2008/2009*. Pretoria, ZA: Statistics South Africa.

—. 2012c. *Statistical Release General Household Survey 2011*. Pretoria, ZA: Statistics South Africa.

—. 2014. *The South Africa MPI: Creating a Multidimensional Poverty Index using Census Data.* Pretoria, ZA: Statistics South Africa.

—. 2016a. *Quest for Nodal Development. Evidence from Census 2001 and Census 2011*. Pretoria, ZA: Statistics South Africa.

—. 2016b. *Monitoring Economic Development in Urban Nodal and Urban Non-Nodal Areas using Data from Censuses 2001 and 2011*. Pretoria, ZA: Statistics South Africa.

—. 2018. *Men, Women and Children: Findings of the Living Conditions Survey 2014/15*. Pretoria, ZA: Statistics South Africa.

—. 2019. *National Poverty Lines*. Pretoria, ZA: Statistics South Africa.

Steinberg, Dan, and Phillip Colla. 1995. *CART: Tree-structured Non-parametric Data Analysis.* San Diego, CA: Salford Systems.

Stevenson, Betsey, and Justin Wolfers. 2013. "Subjective Well-being and Income: Is there any Evidence of Satiation?" *NBER Working Papers* 18992.

Stewart, Fiona, and Juan Yermo. 2009. "Pensions in Africa." *OECD Working Papers on Insurance and Private Pensions* 30.

Stewart, Frances. 1977. *Technology and Underdevelopment*. Boulder, CO: Westview Press.

—. 1980. "The Brandt Report." *ODI Review* 1.

—. 1985. *Basic Needs in Developing Countries* . Baltimore, MD: John Hopkins University Press.

—. 1996. "Basic Needs, Capabilities, and Human Development." In *In Pursuit of the Quality of Life*, edited by Avner Offer. Oxford: Clarendon Press.

Stiglitz, Joseph E., Amartya K. Sen, and Jean-Paul Fitoussi. 2009. *Report by the Commission on the Measurement of Economic Performance and Social Progress.* Commission on the Measurement of Economic Performance and Social Progress.

Stock, James H., and Mark W. Watson. 2007. *Introduction to Econometrics*. Pearson Education.

Stonequist, Everett V. 1961. *The Marginal Man: A Study in Personality and Culture Conflict*. Russell & Russell Pub .

Stouffer, S. A., E. A. Suchman, L. C. Devinney, S. A. Star, et R. M., Jr. Williams. 1949. T*he American Soldier: Adjustment during Army Life.* (Studies in Social Psychology in World War II). Princeton Univ. Press.

Strauss, Anselm L. (1959) 1992. *Miroirs et Masques: Une Introduction à l'Interactionnisme.* Paris, FR: Ed. Métailié.

Streeten, Paul. 1979. "Basic Needs : Premises and Promises." *Journal of Policy Modeling* 1 (1): 136-146.

—. 1981. "The Distinctive Feature of a Basic Needs Approach to Development." In *Development Perspectives*. London, UK: Palgrave Macmillan.

—. 1984. "Basic Needs: Some Unsettled Questions." *World Development* 12(9): 973-978.

—. 1994. "Human Development: Means and Ends." *American Economic Review* 84(2): 232-237.

—. 1995. "Foreword." In *Reflections on Human Development*, edited by Mahbub ul Haq. Oxford: Oxford University Press.

—. 2000. "Looking Ahead: Areas of Future Research in Human Development." *Journal of Human Development* 1 (1): 25-48.

—. 2003. "Shifting Fashions in Development Dialogue." In *Readings in Human Development*, edited by Sakiko Fukuda-Parr and A. K. Shiva Kumar, 72-75. Delhi, IN: Oxford University Press.

Streeten, Paul, and Shahid J. Burki. 1978. "Basic needs: some issues." *World Development* 6 (3): 411-421.

Streeten, Paul, Shahid J. Burki, Mahbub ul Haq, Norman Hicks, and Frances Stewart. 1981. *First Things First: Meeting Basic Needs in Developing Countries.* Oxford University Press.

Strengmann-Kuhn, Wolfgang. 2002. *Working Poor in Europe: A Partial Basis Income for Workers?* Available at SSRN: https://ssrn.com/abstract=386540 or http://dx.doi.org/10.2139/ssrn.386540.

Strobl, Carolin. 2014. "Discussions." *International Statistical Review* 82 (3): 349-352.

Sumner, Andy. 2016. *Global Poverty: Deprivation, Distribution, and Development since the Cold War*. Oxford University Press.

Sundaram, Jomo K. . 2016. "The Geography of Poverty." *Inter Press Service.* June 30. Accessed April 13, 2020. http://www.ipsnews.net/2016/06/the-geography-of-poverty/.

Suppa, Nicolai. 2015. "Towards a Multidimensional Poverty Index for Germany." *OPHI University of Oxford Working Paper* 98.

Suter, Christian. 1992. *Debt Cycles in the World-Economy: Foreign Loans, Financial Crises, and Debt Settlements, 1820-1990.* Boulder Colo.: Westview Press.

—. 1995. *Demokratisierung, Menschenrechte und Strukturanpassung in Lateinamerika: 1970-1993.* Einsiedeln: Nationales Forschungsprogramm 28.

—. 2012. "Debt Crises in the Modern World-System." In *Routledge International Handbook of World Systems Analysis*, edited by Salvatore Babones and Christopher Chase-Dunn. Oxford: Routledge.

Suter, Christian, Tugce Beycan, and Laura Ravazzini. 2017. "Sociological Perspectives on Poverty." In *The Cambridge Handbook of Sociology: Volume 1: Core Areas in Sociology and the Development of the Discipline*, edited by Kathleen Odell Korgen, 388-396. Cambridge University Press.

Suter, Christian, and Katia Iglesias. 2015. "Relative Deprivation and Well-being: Switzerland in a Comparative Perspective." In *Contemporary Switzerland: Revisiting the Special Case*, edited by Hanspeter Kriesi, Peter Farago, Martin Kohli and Milad Zarin-Nejadan. Houndmills: Palgrave Macmillan.

Suter, Christian, and Denise Paris. 2002. "Inequality and Deprivation: Switzerland in a three Country Comparison." *Swiss Journal of Sociology* 28(2): 217-240.

Taylor, Nikki M. 2005. *Frontiers of Freedom: Cincinnati's Black Community, 1802-1868*. Athens, Ga.: Ohio University Press.

Taylor, Viviene. 2014. "The Institutional and Development Impact of Cash Grants: Is South Africa Transforming Social Policy?" *Financial and Fiscal Commission*. Johannesburg, ZA.

—. 2016. "Achieving Food Security through Social Policies: Comprehensive Social Protection for Development." In *Food Security in South Africa: Human Rights and Entitlement Perspectives*, edited by Fukuda-Parr Sakiko and Viviene Taylor, 145-166. Claremont, ZA: UCP Press.

Therneau, Terry M., and Elizabeth J. Atkinson. 1997. *An Introduction to Recursive Partitioning using the RPART Routines*. Mayo Foundation. Available at (accessed 25 May 2020): https://www.mayo.edu/research/documents/biostat-61pdf/doc-10026699

Tienda, M., and H. Stier. 1991. "Joblessness and Shiftlessness: Labour Force Activity in Chicago's Inner City." In *The Urban Underclass*, edited by Christopher Jencks and Paul E. Peterson, 135-154. Washington, D.C.: Brookings Institution.

Tilak, J.B.G. 1991. "From Economic Growth to Human Development: A Commentary on Recent Indexes of Development." *International Journal of Social Economics* 19 (2): 31-42.

Time Magazine. 1977. "Minority Within a Minority: The Underclass." *The Time Magazine* August 29: 14-17.

Titmuss, Richard M. 1965. "Goals of Today's Welfare State." In *Towards Socialism*, edited by P. Anderson and R. Blackburn. London, UK: Fontana.

—. 1968. *Commitment to Welfare*. London, UK: Allen & Unwin Ltd.

—. (1958). 1969. *Essays on 'The Welfare State'*. George Allen & Unwin Ltd.

—. 1974. *Social Policy: An Introduction.* George Allen & Unwin Ltd.

Torgo, Luis. (2010) 2016. *Data Mining with R: Learning with Case Studies*. Chapman and Hall/CRC.

Townsend, Peter. 1954. "Measuring Poverty." *British Journal of Sociology* 5 (2): 130-137.

—. 1962. "The Meaning of Poverty." *British Journal of Sociology* 13 (3): 210-227.

—. 1970. "Measures and Explanations of Poverty in High Income and Low Income Countries: The Problems of Operationalizing the Concepts of Development, Class and Poverty." In *The Concept of Poverty*, edited by Peter Townsend, 1-45. Heinemann Educational Books.

—. 1979. *Poverty in the United Kingdom: A Survey of Household Resources and Standards of Living.* London, UK: Penguin Books and Allen Lane.

—. 1981. *An Alternative Concept of Poverty: How it Might be Applied in National Case Studies in Developing Countries*. Paris, FR: UNESCO.

—. 1983. "A Theory of Poverty and the Role of Social Policy." In *Social Policy and Social Welfare*, edited by M. Loney, D. Boswell and J. Clarke, 58-82. Milton Keynes, Open University Press.

—, ed. 1993. *The International Analysis of Poverty* . Routledge.

Townsend, Peter, and David Gordon. 1991. "What is Enough? New Evidence on Poverty Allowing the Definition of a Minimum Benefit." In *The Sociology of Social Security*, edited by M. Adler, C. Bell, J. Clasen and A. Sinfield, 35-69. Edinburgh, UK: Edinburgh University Press.

Townshend, Ivan, Olu Awosoga, Judith Kulig, and Hai Yan Fan. 2015. "Social Cohesion and Resilience across Communities that Have Experienced a Disaster." *Natural Hazards* 76 (2): 913-938.

Trejo, Guillermo, and Claudio Jones. 1998. "Political Dilemmas of Welfare Reform: Poverty and Inequality in Mexico." In *Mexico under Zedillo,* edited by Susan Kaufman Purcell and Luis Rubio, 67-99. Boulder, CO: Lynne Rienner.

Tsui, Kai-yuen. 2002. "Multidimensional Poverty Indices." *Social Choice and Welfare* 19 (1): 69-93.

Tucker, Deborah. 2002. "'Precarious' Non-standard Employment- A Review of the Literature." *New Zealand Department of Labour, Labour Market Policy Group Occasional Paper Series* 2002/2.

Turkey State Planning Organization. 2016. *Ninth Development Plan 2007-2013.* Turkish Grand National Assembly, Republic of Turkey.

Turner, J.F.C. 1966. *Uncontrolled Urban Settlements: Problems and Policies. Trabajo preparado para el Seminario Inter-Regional sobre Políticas de Desarrollo y Planeamiento en Relación a la Urbanización*. Pittsburgh: Universidad de Pittsburgh and Naciones Unidas.

Twine, Wayne, and Lori M. Hunter. 2011. "Adult Mortality and Household Food Security in South Africa: Does AIDS represent a Unique Mortality Shock?" *Development Southern Africa* 28(4): 431-444.

Tyrovolas, Stefanos, Yannis Tountas, Evangelos Polychronopoulos, and Demosthenes Panagiotakos. 2012. "A Classification and Regression Tree Analysis for the Evaluation of the Role of Nutritional Services on Cardiovascular Disease Risk Status of Older People Living in Greek Islands and Cyprus." *International Journal of Food Sciences and Nutrition* 63(5): 522-529.

UNDP. 1990. *Human Development Report 1990*. New York, NY: Oxford University Press.

—. 1997. *Human Development Report 1997*. New York, NY: Oxford University Press.

—. 2010. *Human Development Report 2010: The Real Wealth of Nations - Pathways to Human Development*. New York, NY: Palgrave Macmillan.

—. 2011. *Human Development Report 2011. Sustainability and Equity: A Better Future for All.* New York, NY: Palgrave Macmillan.

—. 2014. *Human Development Report 2014. Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience.* New York, NY: UNDP.

—. 2015. *Human Development Report 2015: Work for Human Development.* New York, NY: UNDP.

—. 2016. *Human Development Report 2016: Human Development for Everyone.* New York, NY: United Nations Development Programme.

—. 2018. *Human Development Indices and Indicators. 2018 Statistical Update.* New York, NY: United Nations Development Programme.

UNECE. 2017. *Guide on Poverty Measurement*. New York, NY: United Nations.

UNESCO. 1978. *Study in Depth on the Concept of Basic Human Needs in Relation to Various Ways of Life and its Possible Implications for the Action of the Organization.* UNESCO.

United Nations. 1995. *The Copenhagen Declaration and Programme of Action: World Summit for Social Development 6-12 March 1995.* New York, NY: United Nations Department of Publications.

—. 1996. *Report of the World Summit for Social Development*. New York, NY: United Nations.

—. 2000. *Resolution Adopted by the General Assembly. United Nations Millennium Declaration.* A/RES/55/2, United Nations.

—. 2015. *The Millennium Development Goals Report 2015*. New York, NY: United Nations.

—. 2016. *Report on the World Social Situation 2016*. New York, NY: United Nations.

UNRISD. 1994. *Structural Adjustment in a Changing World*. Geneva, CH: UNRISD.

Valentine, Charles A. 1968. *Culture and Poverty: Critique and Counter-Proposals*. Chicago, IL: The University of Chicago Press.

—. 1969. "Culture and Poverty: Critique and Counter-Proposals." *Current Anthropology* 10 (2/3): 181–201.

—. 1971. "The "Culture of poverty": Its Scientific Significance and its Implications for Action." In *The Culture of Poverty: A Critique*, edited by Eleanor B. Leacock, 193-225. New York, NY: Simon and Schuster.

van den Bosch, Karel. 1998. "Poverty and Assets in Belgium." *Review of Income and Wealth* 44 (2): 215-228.

—. 2001. *Identifying the Poor. Using Subjective and Consensual Methods.* Aldershot, UK: Ashgate.

—. 2002 (July). "Convergen in Poverty Outcomes and Social Income Transfers in Member States of the EU." *Paper for the XV Congress of Sociology in Brisbane, Australia.*

van den Bosch, Karel, Tim Callan, Jordi Estivill, Pierre Hausman, Bruno Jeandidier, Ruud Muffels, and John Yfantopoulos. 1993. "A Comparison of Poverty in seven European countries and Regions Using Subjective and Relative Measures." *Journal of Population Economics* 6: 235-259.

van der Berg, Servaas. 1997. "South African Social Security under Apartheid and Beyond." *Development Southern Africa* 14 (4): 481-503.

van Oorschot, Wim, and Loek Halman. 2000. "Blame or Fate, Individual or Social?" *European Societies* 2 (1): 1-28.

van Praag, Bernard. 1968. *Individual Welfare Functions and Consumer Behavior*. Amsterdam, NE: North-Holland Publishing Co.

—. 2003. "The Anatomy of Subjective Well-being." *Journal of Economic Behavior & Organization* 51 (1): 29-49.

van Praag, Bernard, Theo Goedhart, and Arie Kapteyn. 1980. "The Poverty Line: A Pilot Survey in Europe." *Review of Economics and Statistics* 62 (3): 461-465.

van Praag, Bernard, Aldi J. M. Hagenaars, and Hans Van Weern. 1982. "Poverty in Europe." *Review of Income and Wealth* 28 (3): 345-359.

Vayssieres, Marc, R.E. Plant, and Barbara H. Allen-Diaz. 2000. "Classification Trees: An Alternative Non-parametric Approach for Predicting Species Distributions." *Journal of Vegetation Science* 11 (5): 679-694.

Veit-Wilson, John. 1986. "Paradigms of Poverty: A Rehabilitation of B.S. Rowntree." *Journal of Social Policy* 15 (1): 69-99.

—. 1987. "Consensual Approaches to Poverty Lines and Social Security." *Journal of Social Policy* 16 (2): 183-211.

Venables, William N., and Brian D. Ripley. (1994) 2002. *Modern Applied Statistics with S*. Springer.

Vickery, Clair. 1977. "The Time-poor: A New Look at Poverty." *Journal of Human Resources* 12 (1): 27-48.

Vijay, Verma, Gianni Betti, and Francesca Gagliardi. 2017. "Fuzzy Measures of Longitudinal Poverty in a Comparative Perspective." *Social Indicators Research* 130: 435–454.

Villeneuve, André. 1984. "Construire un Indicateur de Précarité : Les Etapes d'une Démarche Empirique." *Economie et Statistique* 168: 93-105.

Villermé, Louis-René. 1840. *Tableau de l'Etat Physique et Moral des Ouvriers Employés dans les Manufactures de Coton, de Laine et de Soie*. Paris, FR: Jules Renouard et Cié, libraires.

Vizard, Polly. 2006. *Poverty and Human Rights: Sen's Capability Perspective*. Oxford, UK: Oxford University Press.

Vizard, Polly, Sakiko Fukuda-Parr, and Diane Elson. 2011. "Introduction: The Capability Approach and Human Rights." *Journal of Human Development and Capabilities* 12 (1:) 1-22.

Vizard, Polly, and Liz Speed. 2015. "Examining Multidimensional Inequality and Deprivation in Britain Using the Capability Approach." *Forum for Social Economics* 45: 1-31.

Vogel, Joachim. 1997. *Living Conditions and Inequality in the European Union*. Luxembourg: Eurostat Population and Social Conditions.

Vosko, Leah F. 2006. *Precarious Employment: Understanding Labour Market Insecurity in Canada.* Montreal and Kingston, CA: McGill-Queen's University Press.

Vosko, Leah F., Nancy Zukewich, and Cynthia Cranford. 2003. "Precarious Jobs: A New Typology of Employment." *Perspectives* October: 16-26.

Wacquant, Loic J.D. 1996. "The Rise of Advanced Marginality: Notes on its Nature and Implications." *Acta Sociológica* 39: 121-139.

Wagle, Udaya. 2005. "Multidimensional Poverty Measurement with Economic Well-being, Capability, and Social Inclusion: A Case from Kathmandu, Nepal." *Journal of Human Development* 6 (3):301-328.

—. 2008. "Multidimensional Poverty: An Alternative Measurement Approach for the United States?" *Social Science Research* 37 (2): 559-580.

—. 2009. "Capability Deprivation and Income Poverty in the United States, 1994 and 2004: Measurement Outcomes and Demographic Profiles." *Social Indicators Research* 94 (3): 509-533.

—. 2014. "The Counting-Based Measurement of Multidimensional Poverty: The Focus on Economic Resources, Inner Capabilities, and Relational Resources in the United States." *Social Indicators Research* 115 (1): 223-240.

Walker , Ed Alan, and Carol Walker. 1997. *Britain Divided: The Growth of Social Exclusion in the 1980s and 1990s*. London, UK: CPAG.

Wallerstein, Immanuel. 1974. *The Modern World-system I*. New York, NY: Academic Press.

Warner, William L. 1952. *Structure of American Life*. University Press.

—. (1949) 1960. *Social Class in America: A Manual of Procedure for the Measurement of Social Status*. Harper Collins.

Watts, Harold W. 1968. "An Economic Definition of Poverty." In *On Understanding Poverty*, edited by Daniel P. Moynihan. New York: Basic Books.

—. 1980. *New American Budget Standards: Report of the Expert Committee on Family Budget Revisions.* Madison, WI: Institute for Research on Poverty, University of Wisconsin.

Watts, H.L. 1967. "The Poverty Datum Line in three Cities and four Towns in the Republic of South Africa." *Institute for Social Research, University of Natal Fact Paper* 1.

Weber, Max. *1904-1905. Die Protestantische Ethik und der "Geist" des Kapitalismus*. Archiv für Sozialwissenschaft und Sozialpolitik.

Weisbrod, Burton A. 1965. *The Economics of Poverty: an American Paradox*. Prentice-Hall, Inc.

Welshman, John . 2006. *Underclass: A History of the Excluded, 1880-2000.* Hambledon Continuum.

Wessels, Daan. 1999. "South Africa's Reconstruction and Development Programme: "A Better Life for All"." *Journal of Social Sciences* 3(4): 235-243.

Westergaard, John, and Henrietta Resler. 1975. *Class in a Capitalist Society: A Study of Contemporary Britain.* London: Heinemann Books.

Weziak-Bialowolska, Dorota Maria, and Lewis Dijkstra. 2014. "Monitoring Multidimensional Poverty in the Regions of the European Union. Analysis of Situation in 2012 and 2007." *European Union Joint Research Centre* JRC91000.

Whelan, Christopher T., and Bertrand Maître. 2012. "Understanding Material Deprivation in Europe: A Multilevel Analysis." *Research in Social Stratification and Mobility* 30: 489–503.

White, Roger, and Stacy Yamasaki. 2017. "Multidimensional Poverty Among the Native- and Foreign-Born in the United States: Evidence from the 2010–2014 American Community Surveys." In *Measuring Multidimensional Poverty and Deprivation: Incidence and Determinants in Developed Countries*, edited by Roger White, 183-222. Cham, CH: Palgrave Macmillan.

Whiteford, A, D. Posel, and T. Kelatwang. 1995. *A Profile of Poverty, Inequality and Human Development*. Pretoria, ZA: Human Sciences Research Council.

Whiteford, Peter, and Willem Adema. 2007. "What Works Best in Reducing Child Poverty : A Benefit or Work Strategy?" *Documents de Travail de l'OCDE sur les Questions Sociales, l'Emploi et les Migrations* 51.

Wilson, Francis. 1996. "Challenges for the Post-Apartheid Economy." T*he American Economic Review* 86 (2): 322-325, Papers and Proceedings of the Hundredth and Eighth Annual Meeting of the American Economic Association San Francisco, CA.

Wilson, William Julius. 1982. "Race Oriented Programs and the Black Underclass." In *Race, Poverty and the Urban Underclass*, edited by Clement Cottingham, 113-132. Lexington Books.

—. 1985. "Cycles of Deprivation and the Underclass Debate." *Social Service Review* 59: 541-559.

—. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago, IL: The University of Chicago Press.

—. 1989. "The Underclass: Issues, Perspectives, and Public Policy." *Annals of the American Academy of Political and Social Science* 501: 182-192.

—. 1991. "Studying Inner-city Social Dislocations ." *American Sociological Review* 56 (1): 1-14.

—. 1993. *The Ghetto Underclass: Social Science Perspectives.* SAGE Publications.

—. 2006. "Social Theory and the Concept "Underclass"." In *Poverty and Inequality*, edited by David B. Grusky and Ravi Kanbur, 103-116. Stanford, CA: Stanford University Press.

Witten, Ian, Eibe Frank, Mark Hall, and Christopher Pal. (2000) 2017. *Data Mining: Practical Machine Learning Tools and Techniques*. Cambridge, MA: Morgan Kaufmann.

Wolff, Jonathan, and Avner De-Shalit. 2007. *Disadvantage*. Oxford University Press.

Wolfson, M., and J. Evans. 1989. "Statistics Canada's Low Income Cut-Offs: Methodological Concerns." *Statistics Canada Analytical Studies Branch Research Paper.*

World Bank. 1990. *World Development Report 1990*. Oxford University Press.

—. 2001. *World Development Report 2000/2001 : Attacking Poverty*. New York: Oxford University Press.

—. 2009. *World Development Report 2009 : Reshaping Economic Geography*. Washington, D.C.: World Bank.

—. 2017. *World Development Indicators 2017*. Washington, DC.: World Bank.

Wresinski, Joseph. 1987. "Grande Pauvreté et Précarité Economique et Sociale." *Avis et Rapports du Conseil Economique et Social, Journal Officiel de la République Francaise* 6.

Wright, Gemma. 2011. "Socially Perceived Necessities in South Africa: Comparing the Views of Sub-Groups of the Population". *CASASP Unversity of Oxford Working Paper* 9.

Xiaogang, Su, Andres Azuero, June Cho, Elizabeth Kvale, Karen M. Meneses, and M. Patrick McNees. 2011. "An Introduction to Tree-Structured Modeling with Application to Quality of Life (QOL) Data." *Nurs Research* 60 (4): 247-255.

Yannakoulia, Mary, Anastasia Lykou, Christina Maria Kastorini, and Eirini Saranti Papasaranti. 2016. "Socio-economic and Lifestyle Parameters Associated with Diet Quality of Children and Adolescents using Classification and Regression Tree Analysis: The DIATROFI study." *Public Health Nutrition* 19 (2): 339-347.

Yaqub, Shahin. 2002. "Poor Children Grow into Poor Adults: Harmful Mechanisms or Over-deterministic Theory." *Journal of International Development* 14: 1081-1093.

Yohannes, Yisehac, and John Hoddinott. 1999. *Classification and Regression Trees: An Introduction.* Washington, D.C.: International Food Policy Research Institute.

Yohannes, Yisehac, and Patrick Webb. 1999. *Classification and regression trees, CART: A User Manual for Identifying Indicators of Vulnerability to Famine and Chronic Food Insecurity.* Washington, D.C.: International Food Policy Research Institute .

Zheng, Buhong. 1997. "Aggregate Poverty Measures." *Journal of Economic Surveys* 11 (2): 123-162.

Zhang, Heping, and Burton Singer. 1999. *Recursive Partitioning in the Health Sciences.* New York, NY: Springer.

—. 2010. *Recursive Partitioning and Applications*. New York, NY: Springer.

Zhang, Heping, and Yuanging Ye. 2008. "A Tree-based Method for Modeling a Multivariate Ordinal Response." *Statistics and Its Interface* 1: 169-178.

9. Appendix

In Appendix, there are nine sections. The first section, 9.1, provides an extensive literature review of poverty concept for showing the evolution of conceptualization and measurement of poverty. The review includes the approaches of 'subsistence, assistance', 'monetary poverty', , 'culture of poverty', 'cycle of deprivation', 'underclass', 'working poor', 'vulnerability', 'poverty dynamics', 'vulnerability to poverty', and 'subjective poverty', respectively. The second section, 9.2, yields the comparison of poverty approaches in terms of advantages and disadvantages. The third section, 9.3, presents introductory information on data mining and knowledge discovery and explains some basic principles of probability theory. The fourth section, 9.4, includes the case study of Turkey starting with the quality report of variables as introduction and continues with the four steps of our methodology. The fifth section, 9.5, focuses on the case study of Mexico and the sixth section, 9.6, comprises the case study of Brazil. Case studies of Mexico and Brazil are also explained in the same way as in South Africa and Turkey analysis, starting by the quality report of variables and then continuing with the four steps of our methodology. The seventh section, 9.7, gives the complementary comparative analysis of South Africa, Turkey, Mexico and Brazil. In chapter 6 of the thesis ('Comparative analysis'), we focus on the main comparison that includes the four countries at the same time. However, there are ten other comparisons including two or three countries at the same. The eight section, 9.8, lists the websites of the national statistical agencies of select upper-middle-income countries. The websites mentioned in the list are the sources of information given in chapter 3 ('poverty in upper-middle-income countries'). And the final section, 9.9, contains some descriptive statistics.

9.1 Extensive literature review of poverty concept

In theoretical framework (chapter 2) we have already seen the capabilities and multidimensional poverty approaches, which we use in this thesis to conceptualize and measure poverty. In addition to these two approaches, we have also explained on the one hand the approaches of absolute poverty, basic needs, relative deprivation to clarify whether the capabilities framework is an absolute or a relative approach to poverty, on the other hand the approaches of social exclusion, social disqualification, and marginalidad to show the link between capabilities and social bonds. In the following sections, we will focus on other poverty concepts from the literature: subsistence, assistance, monetary poverty, culture of poverty, cycle of deprivation, underclass, working poor, vulnerability, poverty dynamics, vulnerability to poverty, and subjective poverty, respectively.

Subsistence, assistance

In recent years, the phenomenon of poverty has particularly become one of the most popular topics in development agenda of international organizations such as the United Nations, Word Bank and of non-governmental organizations. On the one hand, the phenomenon of poverty has been existing for centuries in the world -even probably it has emerged since human beings exist in the world, and on the other hand, it has conceptualized and measured by researchers from different perspectives based on various qualitative and quantitative tools. Following pages in this sub-section, we will put in evidence different concepts and measurement process developed by researchers to define and measure poverty through years, more specifically from the eighteenth century. However, we note that not necessarily all theoretical concepts of poverty have operationalised. First, we start by focusing on poverty studies in Victorian era realized by

British researchers during the eighteenth century. As far as we know, the very first well documented poverty investigations are due to this period, and particularly in London area.

In the United Kingdom, for a long time since the Victorian era[7], there has been the tradition of measuring poverty among the urban working poor, particularly in the urban areas of London at the beginning of the Industrial Revolution[8]. However, even before the Victorian era, some other poverty studies have been carried out as well such as those of Eden (1797, 1966, 3 vols.) and the first social surveys have been conducted in the early nineteenth century (see Kent 1981). Eden's major work (1797) on poverty is titled '*The State of the Poor: Or, an History of the Labouring Classes in England, from the Conquest to the Present Period'*. Eden, in particular, has focused on the situation of workhouses known as the obligatory work, which the indigents of Oxford were carrying out in 1795 (see chapter 1 in Carré 2000 (ed.)). At the time when Eden focused on poverty investigation, the United Kingdom was in financial crisis regarding trade and production chain because of the war with France. In addition, the nourishment crisis between 1794 and 1795 had an impact on the living conditions of the poorest people (e.g., indigents, paupers). We do also know, from various economic and demographic works such as of Malthus (1798), that there was a baby boom at that time in the United Kingdom. On the one side, financial crisis, on the other side the rapid growth rate amongst the poorest class have negatively affected the functioning of public assistance system-which was supposed to help indigents and paupers[9]. Eden, as he was working in the insurance sector, has been very brilliant in family budget analysis. His talent in statistics and risk analysis has had a positive impact on his works on poverty and assistance. The works of Eden have influenced a new movement called 'modern scientific based approach' to solve socio economic problems in Britain. Eden's analysis has included more than one hundred and ten parishes. The measured outputs (i.e., poverty items) have been expressed in terms of money. Nevertheless, it is worth to mention that his list of items has included different aspects of life such as heating, nutrition, clothes, furniture, debt[10].

Meanwhile, keeping in mind that assistant programs were mainly inefficient, the British parliament has realized a survey on the management of workhouses around London. One of the volumes of this survey's report has concentrated *on the state of mendicity in the Metropolis*[11]. Regarding the results of this detailed report, the British government has decided to assess who really needs charities' helps. It is very interesting to note that despite the report has been about the situation of indigents, the administrative staff of charity programs, police, people from the religious area, managers in the workhouses, or other different people from high position in institutions have been interviewed in this investigation. The main reason behind these

[7] In the history of United Kingdom, the Victorian era is the period between 1837 and 1901 under the reign of Queen Victoria. But the Victorian era is also known as a period started with the entry into force of the Reform Act 1832.

[8] In the history of United Kingdom, this period is considered from the end of the eighteenth century.

[9] "A pauper was recipient of assistance under the laws that existed for the poor prior to modern welfare systems. 'Pauperism' came, in political debates, to stand for poverty" (Poynter 1960 cited in Gordon and Spicker 1999, 99).

[10] The information given in this paragraph on Eden and his works is cited in chapter 1of Carré 2000 (ed.).

[11] Written by House of Commons Paper No 473, Session 1814-1815. Please refer to the link below for further information: http://www.cse.psu.edu/~deh25/post/Timeline_files/1810-1819-Parliamentary-Papers.html#1814-15 (accessed 14 April 2020).

interviews is that the poor had been seen as an '*image pollution'* in the urban areas and also as '*problem people'* whose caring costs had become very expensive to British government and society. This kind of reports have stimulated the New Poor Law (original name is *Poor Law Amendment Act*), which has realized in 1834. The New Poor Law (1834) has replaced the old law on assistance to the poor (*Act for the Relief of the Poor* known as *Poor Law* of 1601) from the Elizabethan era. Regarding the Poor Law, non-disabled poor men were sent to workhouses, while the elderly people, orphans, mothers of young children, and other vulnerable people were sent to special institutions (e.g., asylum, old age homes, orphanages)[12].

The New Poor Law has aimed at changing the poor relief system in the United Kingdom. At the time, the British parliament agreed that assistant services provided to the poor were costly and not efficient. The New Law was based on the two principles: first, *less eligibility*, second, *workhouse test*. Regarding first criterion, conditions in workhouses should be worse than the worst living conditions at the outside in order to being able to benefit assistance. Regarding second criterion, assistance should be provided only inside the workhouse. From many perspectives this New Law has ignored some population among the paupers; for instance, mothers of illegitimate children have received less support. The New Law was influenced mainly by three approaches: (i) first, *Malthusianism* (see Malthus 1798), the growing population was a huge concern for British government at that time. It was believed that the provided assistance would not be sufficient for growing population but also the poor more and more were relying on the assistant programs. To Malthus, poverty was clearly related to the surplus of labour class population, and the British society was producing the poor that they would like to help; (ii) second, *Ricardo's wage theory* (see Ricardo 2004 [1817]), which was supposed to balance the level of salaries in the labour market; (iii) third, *Bentham's theory of utilitarianism* (see Bentham 1948 [1789]), it was seen that the poor were stimulated to not work sufficiently when they received assistance. Besides, the British parliament was widely criticized because of the earlier reforms, which were thought having diminished the well-being of upper-class but were benefited the poor at the same time.

Especially, during the early years of the Industrial Revolution, poverty has seen as a result of personal fault, inappropriate attitudes of the poor (e.g., vagabondism, laziness, insobriety, improvidence). In spite of some social surveys developed in the nineteenth century focusing on 'deserving poor' (e.g., unemployment, low wages, irregular, informal jobs, illness, single mother, children, old age), the individual characteristics-based poverty perception has widely dominated social policies in Victorian era in Britain. Meanwhile, famous novel writers of the time such as Charles Dickens (1812-1870) has criticized the New Law. Orphan children were the most important population among indigents during the Queen Victoria's reign. In his book titled *Oliver Twist* (Dickens 1837-1839), we can witness the pauperism of orphan children and how the Poor Law had affected the life of vulnerable people. Other books such as those of Jack London (1876-1916) have also told the stories of indigenous people and their difficult living conditions and social inequalities from different perspectives in British society at the time. On the one hand, through assistance programs and charities, poverty was perceived as the miserable situation of indigenous people threating the British society. On the other hand, the detailed reports including extended interviews such as the parliament's reports on the workhouse conditions correspond to the pioneer works regarding the link between social assistance and poverty. But such investigations have stigmatized the society. Especially, the

[12] The information given in this paragraph on mendicity is cited in chapter 2 of Carré 2000 (ed.).

binary classification of the poor as *good poor vs. bad poor* (*deserved vs. undeserved poor*) and the promotion of forced labour in exchange of social assistance have had a negative bearing on the British society.

During the discussions on assistance, charities, and poverty in the United Kingdom in the early 1880s; in France, Tocqueville (1835, 1951) has come up with his famous work on pauperism. He was particularly interested in Britain's New Poor Law (1834), and in the function of charities towards poverty. He has argued that legal charity (i.e., general help system provided by government, organizations to paupers) was less effective and more problematic compared to private charity to solve pauperism. In a comparative perspective, based on his observations across some European countries, he has notably pointed out that there was a clear distinction between poverty in rich countries (i.e., industrial countries) and poverty in relatively poorer countries (i.e., agriculture dominated countries). In an opulent country such as Britain, paupers were appearing as a lowest class (i.e., lowest social status) people with their own socio-economic characteristics such as being severely deprived in terms of material well-being, and with a strong dependence to charities, help programs, and community. Especially, the permanent assistant status of the paupers become their signature, which deprived them from the full participation in society. Whilst, in poorer countries like Spain and Portugal, life was miserable for the majority population. Tocqueville has emphasized that paupers in Britain were also suffering morally at high level because of permanent humiliation, and sardonic look of others towards them. As a result of his comparative analysis, Tocqueville has linked pauperism to civilization process, and has questioned democracy in rich industrialized countries such as Britain where pauperism has appeared manifesting in worst conditions[13] (cited in chapter 1 in Paugam 2005).

We can also say that Tocqueville has pioneered relative perspective to poverty in terms of location and time (cited in Paugam 2005, 28). He has showed that needs of people, social conditions, living standards change from one society to another and these issues depend on conjuncture. Regarding him, poverty studies need to be done regarding the situation of whole society (cited in Paugam 2005, 28). His observations and remarks on pauperism and assistance are remarkable; however, his studies do not contain economic explanations such as the relationship between labourers and upper-class people[14], inequalities generated by capitalist system (e.g., property rights, exploitation of labour class) (cited in Paugam 2005, chapter 1, 39). In contrast, Marx has explained poverty from an economic perspective, as a function of the capitalist system. To Marx ((1847) 1947), poverty is a necessary condition for the continuum of the capitalist system. He has defined three groups (see Duggan 2013, 3; Marx [1867] 1990 cited in Duggan 2013,3): (i) 'the reserve army of labour' or called 'industrial reserve army' or 'relative surplus population'; (ii) 'the army of labour'; and (iii) 'lumpenproletariat (see Engels 1887 (originaly published in 1845), Marx and Engels 1848) referring to unemployed and underemployed workers, to working class people (i.e., by four categories: the stagnant pool, the floating reserves, the latent reserve and pauperdom), to the rest of society that are nor the industrial reserve neither the reserve army of labour. The 'lumpenproletariat' that includes criminals, delinquents, beggars, vagabonds, prostitutes, corresponds to the dangerous lowest strata of the society. Marx has argued that the 'lumpenproletariat' serves any function in the

[13] The information given in this paragraph about Tocqueville and his works is cited in chapter 1 in Paugam (2005).

[14] Considered that upper-class people dominate the production system.

capitalist system. Herewith, in his framework, the poor in the era of capitalism (for instance, industrial revolution in England) is 'the reserve army of labour'[15].

As we have seen previously, Tocqueville has pioneered the topic of assistance for explaining pauperism in the nineteenth century. After him, in the early twentieth century, the German sociologist Georg Simmel (2009, originally published in 1908) has constructed a comprehensive sociological theory on poverty on the basis of assistance and social bonds, and he has elaborated in detail the link between poverty and assistance. Simmel has taken the status of the poor as a specific social category. To him, the poor are who do receive assistance (e.g., help, social allowance, relief) or who should receive assistance regarding social norms in the society. This identification process of the poor has led to look at poverty as a social problem generated by the society. Compared to Marx and Tocqueville, Simmel has shed light on that the poor are part of society and has argued that the way that social system regulates (e.g., social institutions) makes the poor assisted to specific welfare programs and charities. As other members of society, the poor also do fulfill a function in society. But their function is characterized mainly by being assisted to their community. In fact, this function is given to them by the mechanisms of society. However, the poor have only one function that is being poor and assisted. Society does not allow them to have another status and being assisted is the only trajectory they can ever have in their life (see Paugam 2005, chapter 2, 40-53).

As the poor form the lowest strata in any society, they are generally seen as excess, useless, or dangerous. This social identity given to them by society leads them to be stigmatized in different ways ('status degradation')[16] (e.g., see Paugam 1991b, 87). In this perspective, they can be considered as socially excluded. In fact, their exclusion is part of the social system, which integrates the poor into society as exclusive assisted poor people. Simmel considers the poor as actors in social interactions. Herewith, the assisted poor are not the outsiders; in opposite they are included in society through assistance (i.e., inclusion corresponds to being dependent on social assistance). At first glance, it seems that the poor have right to assistance. But Simmel has emphasized that in fact the state has an obligation of assisting the poor and that the poor do not really have right of demanding assistance (because the poor do not have any right to appeal if their demand of assistance is refused). Simmel's explanations have showed us that assistance has a specific social *function*, which determines the trajectory and social relationships of the poor. As we can see Simmel's *functionalist* approach is beyond yielding a description of living conditions of the poor. He does not see poverty as an individual fact, and he does not attempt to conceptualize poverty on the basis of a particular list of necessities. His approach has focalized on how the society determines the identity, role and position of the poor as assisted and the poor's relationship with the society (see Paugam 1991b, 2005; Ogien 1983 for detailed explanations on Simmel's framework). For instance, through the sociology of gift, Simmel has highlighted that the poor's gifts between each other are significantly different than the types of gift offered amongst the upper strata.

[15] Debates on poverty in the era of capitalism, from economic perspective, have drawn attention to the 'reserve army of labour' class as the *real* poor living in miserable conditions, in opposition to the 'lumpenproletariat' (or known as the *paupers*).

[16] A heterogeneous group of isolated people (i.e., different life trajectories of isolation and rupture of social bonds).

Monetary poverty[17]

According to (Sen 1976), for measuring poverty, first we have to identify who is poor and then we have to determine the aggregation process, which will help us to find out the overall poverty rate in society. The identification process basically divides the population into two categories: poor vs. not poor. For distinguishing the poor from the non-poor, generally a *poverty line* (or called as a *poverty threshold* or *cutoff*) is constructed in identification process. The principle is that the individuals or households (or other type of units such as family) staying under the poverty line are considered poor, and vice versa (see Hagenaars and van Praag 1985, Hagenaars 1986, Callan and Nolan 1991, Groedhart et al. 1977 for a review of poverty line definitions, and Gillie 1996, Ravallion 1998, 2012b for the discussion on the origin of the poverty line). In monetary perspective, there are two main approaches for determining the poverty line: absolute vs. relative. The absolute monetary poverty line (e.g., World Bank, see below) reflects the minimum necessary income (see Haig 1921, Simons 1938 for the computation of income), or the minimum necessary consumption expenditure level (including purchased goods and services and one's own production for own consumption), without taking into account the distribution within population, for pursuing a non-poor life in society[18]. But, in opposite, the relative approach (i.e., the relative monetary poverty line) includes the distribution of income and/or the relative importance of consumption items, and of expenditure patterns for determining what should be the minimum income, or consumption expenditure level given the living standards, necessities, customs, and habits within the society. Ravallion (2003) qualifies relative monetary poverty indicators as kind of inequality indicators as he argues that there is a strong correlation between inequality indicators such as Gini coefficient and relative monetary poverty indicators. For instance, in terms of relative income poverty line, people having an income level below the defined threshold are considered relatively (monetary) poor given the income distribution in society.

Absolute monetary poverty lines are made upon the basket of (essential) goods and services, such as budget standards (see Bradshaw 1993, Rein 1970, Rowntree and Lavers 1951) or more specifically upon a certain number of subsistence items[19] (i.e., compute the minimum absolute poverty line in monetary terms for affording food[20], shelter, clothing) such as Ravallion's cost-

[17] In some studies, it is called as *financial poverty* (e.g., see Atkinson et al. 2002).

[18] Ringen (1988) qualifies it as 'direct measure' of poverty (i.e., low level of consumption), the consumption patterns approach; whilst as 'indirect measure' the income-based approach (i.e., low income).

[19] For instance, Rowntree (1901, 1941) has defined the absolute poverty line based on some subsistence items.

[20] In the literature, related to food, there are also two other approaches to measure poverty: calories consumed per person per day, and food consumption as a fraction of total expenditure. The first one identifies the poor as the people who are not able to consume the necessary minimum quantity calories per day and per person (see also Food Energy Intake). Haughton and Khandker (2009, 35) define calorie as "the energy required to heat 1 gram of water by 1°C. A kilocalorie, also referred to as a calorie, represents 1,000 calories", which is generally set at 2,100 calories per person, but it varies as a function of various parameters such as gender, age, working activities, illness. In the second one, for identifying the food expenditure share, it is very important to know household composition (child, elderly member of the household) and relative price of food. See also Engel for his findings on the relationship between food expenditure and household income per capita among Germany household. Engel curve shows that even household income per capita increases, the food expenditure increases less quickly. Moreover, Hagenaars (1986) mention that it is difficult to agree on levels of calories needed for different subgroups such as a function of age, sex, occupation, living standards.

of-basic needs approach (1998), which estimates and sum the cost of adequate nutrition[21] and of other basic needs such as food and shelter, or upon the food ratio poverty lines such as Canada's Low-Income Cut-Offs (i.e., focuses on the share of income spent to certain basic needs such as food, clothing, heating; see Wolfson and Evans 1989), or upon calorie income function (i.e., computing the minimum necessary income for being able to afford the minimum amount of calories per day and per person), or, it is just simply set by governments as the minimum income level given social, and fiscal legislation (e.g., subsistence income in Belgium, which is an income level guaranteed to each citizen in Belgium). The latter approach is referred as *statutory method* (see Atkinson et al. 2002); however, it is not necessarily an ideal situation because there is no guarantee that the minimum income or the social allowances will keep people out of poverty. Because national statutory lines do not necessarily cover a comprehensive amount of social necessities and norms to provide an adequate social inclusion to every citizen; they are rather influenced by policy decision making process and budget constraints. Among all these methods, the oldest is the budget standards (i.e., determining a weekly basket of goods leading to a weekly diet based on nutritionist's suggestions; for instance, since 1964 the U.S. has been applying it to produce official poverty rates). However, the main shortcoming of this method is that generally the selected items in the basket are determined based on value judgments of experts. Furthermore, it is important to note that in general the content of absolute monetary poverty lines is remained the same over time; they are just adjusted for inflation (e.g., U.S official poverty line).

Concerning technical aspect of constructing a poverty line, the process is that once the items necessary for leading a pleasant life are defined by governments, experts, researchers (the items are often qualified as arbitrarily defined based on value judgments), then by taking into account the relative prices of these items, the poverty line in terms of income or consumption expenditure is drawn up. Generally, developed countries have preference to use income whilst developing countries are more likely to use expenditure patterns of consumption for setting monetary poverty line. In developing countries, particularly in less-developed countries, it can be quite difficult to compute income given that informal employment, self-employment, home agricultural production is quite prevalent (in seasonal works, the income estimation becomes quite problematic as well). Veit-Wilson (1986) points out that generally the basket of goods is not an efficient view on measuring poverty as the baskets are determined by experts who are in many cases stranger to the poor's living conditions. To Veit-Wilson (1987), the consumption expenditure patterns of people are more realistic practices to measure poverty. He proposed a sociological approach to setting budget standards. He argued that non-necessary expenditures are also part of the life, and that determining income level sufficient for necessities (based on expenditure patterns) will be the most efficient solution such as how Rowntree included some non-necessary items in his computation of 'secondary poverty'. Concerning the computation of actual expenditure patterns of people, the Watts Committee in the U.S. has proposed some concepts based on statistical tools such as 'prevailing family standard'-median level, 'social minimum standard'- fifty per cent below the median, 'social abundance standard'- fifty per cent above the median (see Watts 1980). Bradshaw et al. (1987) applied this methodology in British context. In the mid-1990s, Bradshaw (see Bradshaw 1996, Bradshaw and Stimson 1997) has

[21] For instance, 2,100 calories per person per day. If price information is not available, the food energy intake method is used which "plots expenditure (or income) per capita against food consumption (in calories per person per day) to determine the expenditure (or income) level at which a household acquires enough food" (Haughton and Khandker 2009, 39).

pioneered to establish the Family Budget Unit (FBU) that has aimed at computing budget standards for different types of families based on expenditure patterns and consumption needs.

As mentioned by Haughton and Khandker (2009), both two approaches have some shortcomings. For instance, on the one hand, with income approach it is not always possible to capture directly welfare and the reporting period might not reflect the average income of the household. On the other hand, in expenditure approach, some consumptions are not on regular basis, it is difficult to estimate expenditure of durable goods and household consumption preferences may change such as a rich household may prefer to live a more modest way. In addition to absolute and relative approaches, there are also hybrid poverty lines combining both absolute and relative approaches (for instance, see Foster 1998, Atkinson and Bourguignon 2001, Ravallion and Chen 2009), and consensual poverty line (for instance, see van Praag et al. 1982, Van den Bosch 2001, Lansley et Mack 2015) that is decided upon the subjective opinions of the citizens itself (e.g., or of the target population) on the minimum necessary income, or consumption expenditure levels. Notably, the works of van Praag and his colleagues at the University of Leyden in the Netherlands (1982) have played an important role in development of consensual approach. They conducted several surveys in a number of European countries collecting information on: (i) what level of income people need to meet their needs and avoiding poverty; (ii) what level of cash income people would assess to various living standards, for later constructing poverty lines based on statistical analysis (see Leyden poverty line). As mentioned by Haughton and Khandker (2009, 60), a typical question might be "What income level do you personally consider to be absolutely minimal? That is to say, with less you could not make ends meet", or as mentioned by Atkinson et al. (2002, 86) "What is the lowest net monthly income, all included, that your household needs in order to make ends meet?" However, the structure of the question may have a significant effect on income thresholds expressed by the population (see Van den Bosch 2001). But also, in some survey cases there is no information available neither on the budget items passed through one's mind nor on living conditions under which the population answered the question.

Notably, from the end of twentieth century, the World Banks's absolute (monetary) poverty line approach (see World Bank 1990) has been one of the most used poverty measurement techniques. Regarding this approach, people having less than $1 per day are considered poor. It has been proposed by the Word Bank as a universal poverty line notably for facilitating cross-country comparisons (see Ravallion et al. 1991) for the origin of the $1-a-day poverty line). Initially, the World Bank (1990, 26) has defined poverty based on consumption as "the inability to attain a minimal standard of living" by taking in consideration nutrition, warmth, and shelter as the basic necessities of life, which was very similar to the concepts of absolute poverty, or basic needs. Herewith, $1 per day was considered sufficient for covering the daily necessary expenditures to "buy a minimum standard of nutrition and other basic necessities and a further amount that varies from country to country, reflecting the cost of participating in the everyday life of society" (1990, 26). Nevertheless, later, the World Bank used only the nutrition criterion for measuring poverty because it was estimated that it would be not possible to reach an objective argument regarding the selection of other items. For instance, "in some countries indoor plumbing is a luxury, but in others it is a necessity." (1990, 27). Following this decision, the poverty line was set to $370 per capita. People, therefore, having an income level per year less than $370 are considered *poor*, and those having an income level per year less than $275[22]

[22] $370 and $275 are based on purchasing power parity at 1985 prices.

are defined as *extremely poor*. In this light, currently, the World Bank uses a poverty line of US$1.25 a day as a lower level, or US$2 a day as a higher level.

The World Bank (see Haughton and Khandker 2009) measures world poverty by applying six steps that are respectively: setting a monetary poverty line in US. Dollars such as $1.25 per person per day in 2005 dollars, convert it to local currencies based on purchasing power parity exchange rates, adjust the local-currency poverty line for inflation using local consumer price indices for any given year, draw Lorenz curves from household survey data, and in the end, estimate an overall local poverty rate in the country. The World Bank's international comparison of monetary poverty rates has become particularly important with the United Nation's Millennium Development Goals (MDGs onwards) initiative[23] where the main goal of the MDGs was to halve, between 1990 and 2015, the share of people in the southern counties living on less than $1 per day (in terms of purchasing-power-parity). Within this purpose, two indicators were used to measure the progress in countries: (1) the proportion of the population below living $1 (1993 PPP) per day; (2) the poverty gap ratio (see below for the formula). Regarding Chen and Ravallion (2008), the best monetary poverty line is US$1.25 per person per day, in 2005 prices, for world poverty comparisons (see Haughton and Khandker 2009) for methodological process). However, Alkire and Santos (2009, 133) mention that "absolute poverty line shaves the advantage that they are comparable over time (with the appropriate price adjustments) and across countries (even when the food items differ, the set is supposed to provide a certain amount of calories)". These poverty lines, in fact, do not cover multiple deprivations other than food, such as the issue of real liberty (e.g., democracy, justice, human rights), and any social aspects of life including social relationship, social participation in society. As Gordon et al. (2000b, 84) have mentioned, the World Bank's poverty lines are first, cannot be considered as an international poverty lines because they are basically suitable for the least developed countries, second, they do not include regional, or country differentiations, the minimum value of poverty is underestimated, third, as a result, the number of the poor is underestimated. Pogge and Reddy (2006) also argued that the original $1 per day per person poverty line (in 1993 prices) is very low, and it does not include nor capabilities neither basic needs. They mentioned that with such a small amount per day per person, it is not possible to live decently in the U.S. or even in other local currencies as well. With an example of the period of War on Poverty in the U.S., Castel (1978) expressed that in 1965, 90 dollars on average was redistributed to the poor. Maybe, this amount can be considered as an amount of survival but certainly this is not covering dignity[24].

On the other side, in relative perspective, the monetary poverty line is generally set as per capita national income, or median or mean of disposable household income, or of equivalized disposable household income (see the OECD 1976, 2008a; Jenkins 1990; Mitchell 1991; Atkinson et al. 2002 for comparative aspect). Often, household income information comes from national household surveys (e.g., general household survey, household budget survey, socio-economic survey, living conditions survey) where the question of 'what is the household income or expenses' is asked to the head of household. Then, the household income (adjusted to economies of scale (i.e., household size) and/or household composition) is compared with

[23] Please see the link below for further information on MDGs (accessed 15 April 2020)
https://www.un.org/millenniumgoals/

[24] In his original work written in French (1978, 58), he mentions that "En 1965, grande année de la "guerre" (à la pauvreté), 90 dollars en moyenne ont été redistribués à chaque « pauvre » de la nation américaine. C'est peut-être le prix de la survie, ce n'est pas celui de dignité ».

the defined (monetary) poverty line. For instance, the most frequent relative monetary poverty lines are 40, 50 or 60 per cent of median of the disposable equivalized household income (these thresholds are often considered arbitrary). Herewith, households having a disposable equivalized income below the relative thresholds are considered poor (thereby, the individual living in poor household are considered poor). Among scholars, and international institutions, it is generally perceived that the relative monetary approach facilitates cross national comparison, but there is no international consensus neither on the choice of percentage level nor the choice of mean or median to be applied[25]. Of course, the number of poor will vary as a function of the percentage employed, for instance, if the percentage is low the poverty rate will tend to be high. At that point, it is worth to note that Atkinson et al. (2002) suggest particularly to distinguish the notions of poverty and income inequality. They mention that "it would, for example, make no sense to use the term 'poor' to apply to those below the bottom quintile, since there would by definition always by 20% of the population in poverty" (2002, 79).

The relative monetary approach has been employed by most of European Union countries, in collaboration with Eurostat. For instance, at the Stockholm Summit in March 2001, European Commission proposed, in the context of social inclusion, using the proportion of population below the poverty line (defined as 60% of national median equivalized income) before and after social transfers, and the proportion of population below the poverty line for three consecutive years (Atkinson et al. 2002). General view is relative approach will be more convenient for developed countries compared to developing countries. However, as mentioned by Alkire and Santos (2009, 134) "if all incomes in a society suddenly dropped sharply by exactly the same extent, leaving most of the society in a destitute situation from an absolute point of view, relative poverty would not change." In same line, Whiteford et Adema (2007) have showed that among the OECD countries, the absolute monetary poverty rates are mainly declined from the mid-1980s to 2000, however, relative monetary poverty rates are increased. Crettaz (2011, 17) also argues that "as a rule of thumb it can be said that the use of relative poverty lines disadvantages countries with high median incomes and large income inequalities, such as the U.S, whereas the use of absolute poverty lines adjusted with purchasing power parities tends to disadvantage countries in which many services are provided for free or at a very low cost by the state, such as Sweden." As we can see, monetary poverty lines are also not a pure solution to international comparison of poverty.

If, in contemporary poverty studies, more and more organizations and researchers have been recognising the multidimensional nature of poverty (i.e., not just focus on money, but takes into account social aspect of life, see capabilities and/or multidimensional poverty), still many countries' national official poverty numbers are based on monetary poverty approach. Atkinson et al. (2002) emphasize that measuring poverty in terms of income only do not capture multiple aspects of poverty. They argue that some households identified as poor may not be deprived in some indicators and not socially excluded from society. Multiple deprivation regarding living standards should be observed over a long period, rather than the week, month or year of reference of income (in household surveys). This is why, they point out the importance of combining both low income and (non-monetary) multiple deprivation indicators to find out socially excluded poor people at a given time. However, for example, in the United States, since August 1969, the official (absolute) poverty line (i.e., a minimum necessary income level

[25] Atkinson et al. (2002, 84) suggest that due to the limits of available good quality data, the threshold could be 40, 50, 60, or 70 per cent of mean or median equivalized income.

for not being considered poor), has been $3, 165 per month. This monetary poverty line was calculated by Mollie Orshansky[26] (from the U.S Social Security Administration) in 1963 (published in the July 1963 Social Security Bulletin, namely 'Children of the Poor') based on hybrid budget method during the *war on poverty* in U.S (see also Orshansky 1965) where she discusses methodological issues in how to compute monetary poverty based on income-food expenditure relationship, in how to find poverty profiles for families of different composition, children. And she provides detailed tabulations of poverty estimations in the U.S. in 1993). The hybrid budget method was developed by the US Department of Agriculture (USDA onwards) aiming at determining basic food expenditures given different types of households at different income levels. In total, four different budgets were computed. Each of budgets was corresponding to a nutritionally adequate diet with different details mainly because of different types of households. In terms of formula, this poverty line is obtained by multiplying three items together, namely, vector of prices ($\vec{P}$), array of foods ($\vec{X}$) and a multiplier (H). The multiplier is set 3 because regarding the USDA's consumption surveys' results in the 1960s, the proportion of income used for meeting the alimentary needs for low-income households was one-third. Orshansky (1969) argued that poor people correspond to who are living in households spending more than 30 per cent of their budget to food. Thus, based on real consumption expenditure patterns of various types of families, Orshansky determined 'income proxy' for measuring poverty. Meanwhile, despite an important number of works on working poor, underclass and actual living standards in the U.S. (see Citro and Michael 1995), notably since the mid-1990s in the U.S., neither Orshansky's method has been re-employed to compute a new poverty line nor a multidimensional poverty measurement has been officially adopted. Instead, the 1963 poverty line of $3, 165 has been adjusted to each year's inflation based on the Consumer Price Index[27].

In the literature, there are also various discussions on the purpose and technical details of monetary poverty measures. First of all, as many modern poverty studies show the money is not sufficient itself to capture all aspects of life. Especially, there are some non-monetary and non-material items such as social relationship, social participation, justice playing substantial role in the life of people, which cannot be solely reached by money. Mainly for this reason, income is seen as an indirect measure of poverty. Srinivasan (2010, 143) summarized the

[26] To Orshansky (1969, 37) "poverty, like beauty, lies in the eye of the beholder. Poverty is a value judgement; it is not something one can verify or demonstrate, except by inference or suggestion, even with a measure of error. To say who is poor is to use all sorts of value judgements." She clearly recognized that poverty is a multi-faceted social phenomenon. However, she mentioned that it is not possible to define a list of obligatory necessary goods and services arguing that "yet as yesterday's luxuries become tomorrow's necessities, who can define for today how much is enough?" and measured poverty in a very restricted way from a typical economic perspective, with a limited value judgment (1965, 5). She also mentions that (1969, 37) "we need benchmarks to distinguish the population group that we want to worry about. A benchmark should neither select a group so small, in relation to all the population, that it hardly seems to deserve a general program, nor so large that a solution to the problem appears impossible."

[27] Please see the link below for the Consumer Price Index computation in U.S. (accessed 15 April 2020):
https://www.bls.gov/cpi/
or International Labour Organization's handbook (ILO 2004) is available in the link below:
https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/presentation/wcms_331153.pdf

importance of multidimensional perspective, for instance, "whether or not an individual is deemed to be poor depends not merely on that individual's economic and non-economic endowments but also, equally, if not more importantly, the place of that individual in the socio-economic-political processes of the country of which he or she is a citizen." Besides, several studies show that income and other living conditions are weakly linked each other (Halleröd et al. 2006, Mayer 1995).

There are also some debates on a number of technical issues such as household equivalence scale[28] (i.e., which are used for obtaining an equivalized income poverty line after taking in consideration different household size and composition for adjusting household income level) reflecting the consumption per adult equivalent instead of consumption per capita , intra-distribution of household income and resources (i.e., generally, it is assumed that within household consumption of resources is shared equally amongst its members ('orthodox model' of egalitarian decision-making in household); but in fact, this assumption is quite far from the reality because not everyone has the same needs, and not in each household the decision is made upon democracy i.e., distribution of power may be not equal among household members). Alcock (2006 [1993], 106) highlights that some indirect costs of intra-household distribution such as "emotional dependency, loss of control and time poverty" are rarely used in monetary poverty studies. For instance, Lister (1990, 2004) has argued that time poverty (see Vickery 1977; Gammage 2011, 2010; Antonopoulos et al. 2012) does significantly appear in the lives of many women, which is hidden in income or expenditure based monetary poverty computations. Moreover, computation of own consumption (see Haughton and Khandker 2009) measuring incomes of self-employed, the use of mean or median for determining poverty line (see Townsend and Gordon 1991, Canberra Group 2011) are other debate topics in poverty measurement, and income measures are generally lack of wealth and credit information (Crettaz 2011).

Now, we will focus on some aggregate measures of monetary poverty[29]. As Sen has mentioned (1976, 219) for measuring poverty, there are two steps: "(i) identifying the poor among the total population; and (ii) constructing an index of poverty using the available information on the poor". Thus, the first step involves the identification process of the poor that we saw in previous paragraphs. Thanks to income or consumption expenditure-based poverty lines, poor people can be identified (i.e., the determination of a 'poverty line' (or criterion of poverty) such as in terms of real income per head). In this regard, those who fall below the 'poverty line' are considered poor, and vice versa. Now, we will focus on the second step that consists of

[28] OECD (1982) suggests using the weights of 1.0 for the first adult, 0.7 for other adults and 0.5 for children, to obtain the scale that is equal $1 + 0.7 * (N_A\text{-}1) + 0.5 * N_C$) in calculation of consumption per adult equivalent. Hagenaars, De Vos and Zaidi (1994) proposed the 'modified version of OECD scale' as 0.5 for other adults and 0.3 for children. Moreover, Atkinson et al. (1995) recommends the 'square root' scale, where the equivalence factor corresponds to the square root of the number of household members. The World Bank (see Haughton and Khandker 2009, 29) has used the technique of Glewwe and Twum-Baah (1991) in the context of LSMS survey. The employed formula is $(N_{adults} + \alpha\, N_{children})^{\theta}$ where α corresponds the cost of child, and θ (which is supposed to ≤ 1) reflects the effects of economies of scale. See also Deaton and Zaidi (1998) for more details.

[29] We will shortly look at the different aggregate measures mentioned in chapter 4 of the World Bank's Handbook on Poverty and inequality (Haughton and Khandker 2009). As already clarified in the handbook, we do also emphasize that when it comes to integrate weights (inference from data sample to population level), the formulas should include weight command. We don't enter in detail about weighting here.

aggregating all poor individuals together by employing mathematical formulas (e.g., index) in order to obtain an overall rate of poverty (e.g., the overall poverty indicator in the country). For instance, the traditional *headcount ratio* (or it is also known as *headcount index,* see (Rowntree 1901; Atkinson 1970; Townsend 1954, 1979; Weisbrod 1965 for examples) is the most popular one among others. The headcount ratio does simply correspond to the number of poor people divided by the total number of population (i.e., proportion of the poor in the population). In other words, the headcount ratio measures the incidence (i.e., frequency) of poverty as follows:

$$Headcount\ ratio\ of\ poverty\ (P_0\ or\ H) = \frac{N_p}{N}.$$

where N_p is the number of poor people (i.e., the number of people having an income/consumption expenditure level less than the income/consumption expenditure poverty line), and N is the total population (or sample). For example, if there are 80 poor people among 250 people sampled in a survey, then $P_0 = \frac{80}{250} = 0.32 = 32$ percent (i.e. 32 per cent of sample is poor).

Sen (1976, 219) considers headcount ratio as "a very crude index" because it violates monotonicity, and transfer axioms. Sen defines these two axioms (1976, 219): "Monotonicity axiom: Given other things, a reduction in income of a person below the poverty line must increase the poverty measure; Transfer Axiom: Given other things, a pure transfer of income from a person below the poverty line to anyone who is richer must increase the poverty measure". The transfer axiom is also known as Dalton's 'principle of transfers' in measuring inequality (Dalton 1920, and see also Atkinson 1970). Herewith, headcount poverty ratio does not include how poor the poor are in relation to poverty line (i.e., the headcount figure remains unchanged even if poor people become poorer). But another measure called *poverty gap* satisfies the monotonicity axiom, including the intensity (i.e., depth) of poverty (see Batchelder 1971). But still, the poverty gap violates the transfer axiom. There are two formulas of poverty gap, denoted as *poverty gap index 1* and *poverty gap index 2* respectively below. On the one hand, the first formula is seen rather as income gap among the poor, in the literature (see Haughton and Khandker 2009):

$$Poverty\ gap\ index\ 1 = \frac{1}{N_p}\sum_{i=1}^{N_p}\frac{(z-x_i)}{z}.$$

where the target population is the poor people, thereby, N_p is the number of poor people, i is for indicating each individual in the poor population, z is the poverty line, and x_i is the disposable income/consumption expenditure level of the ith individual.

On the other hand, the second formula of poverty gap is computed regarding the total population, as follows:

$$Poverty\ gap\ index\ 2\ (P_1) = \frac{1}{N}\sum_{i=1}^{N_p}\frac{(z-x_i)}{z}.$$

where N represents the total population, N_p is the number of poor people, z is poverty line, x_i is the disposable income of the ith individual, and the non-poor have zero income gap. We can also write that:

$$Poverty\ gap\ index\ 2\ (P_1) = poverty\ gap\ index\ 1\ \times headcount\ ratio\ of\ poverty.$$

meaning that the poverty gap is the product of income gap and headcount ratio of poverty. In the formula above, the term $(z - x_i)/z$ shows the proportionate distance of an individual's income/consumption expenditure level to the poverty line. If we sum all poverty gaps and divide it by the total number of population (or sample), then we will obtain the mean proportionate poverty gap (given that the non-poor have zero poverty gap). For example, we suppose that there are three persons A, B, C with an income level respectively 20, 40, 18. If the poverty line corresponds to 60 per cent of the median (that equals to 24), the persons A and C are below the poverty line. Their distance to poverty line, therefore, are respectively: 4 and 6. If we divide these numbers by poverty line ($(z - x_i)/z$), then we obtain respectively: 0.16 and 0.25. When we multiply by 100 these figures, we can do this interpretation: for the person A, her/his distance to poverty line is 16 per cent of the poverty line, and for the person C, her/his distance to poverty line is 25 per cent of the poverty line. We can see that the intensity of the distance is higher for the person C. By summing 0.16 and 0.25, then dividing by 3 (=N; the total number of population in the example), we obtain 0.136. This last figure is the poverty gap index (i.e., mean proportionate poverty gap) meaning that about 13.6 per cent of the poverty line the poor are below the poverty line. In monetary terms, in average, we need 3.264 unit for compensating the gap of the poor regarding the poverty line. This measure can also be seen as the income transfer indicator from non-poor to poor in a society. In average, the amount necessary for transferring all poor people up to the poverty line is equal 3.264 monetary unit in our example. Or in other words, the poverty gap index corresponds the minimum cost of eliminating poverty in a society. However, it is important to note that this measure is not perfect as well because it still violates Dalton's transfer principle. We will show it through the example given by Haughton and Khandker (2009, 71).

Table 17. Poverty gap index and Dalton's transfer principle. Source: Haughton and Khandker (2009, 71)

Poverty Gap Poverty Rates in A and B, Assuming Poverty Line of 125						
	Expenditure for each individual in country				Poverty gap rate	Headcount index
Expenditure in country A	99	101	150	150	0.10	50%
Expenditure in country B	79	121	150	150	0.10	50%

From the table above, we can see that the two countries A and B have the same poverty gap rate, but country B faces more deprivation as one of its members has a severe poverty.

Moreover, by taking square of the poverty gap index 2, the severity of poverty can be obtained. This measure includes inequality among the poor that is as follows:

$$Squared\ poverty\ gap\ index\ (P_2) = \frac{1}{N}\sum_{i=1}^{N_p}\left[\frac{(z-x_i)}{z}\right]^2 .$$

As mentioned by Haughton and Khandker (2009), it is not easy to interpret the squared poverty gap index results, this is why it is not very often used in the literature.

There are also other types of poverty indices such as Foster-Greer-Thoerbecke (FGT onwards) class of poverty measures developed by Foster et al. in 1984. The FGT poverty index is as follows:

$$FGT = \frac{1}{N_p}\sum_{i=1}^{N_p}\left[\frac{(z - x_i)}{z}\right]^{\alpha}$$

where, (i) if the parameter $\alpha = 0$, the FGT index is the poverty headcount ratio; (ii) if $\alpha = 1$, the index is the poverty gap index (FGT-1); and (iii) if $\alpha = 2$, the index is the severity of poverty index (FGT-2). The parameter α reflects the sensitivity of the index to poverty. For all $\alpha > 0$, the FGT is strictly decreasing- when one's living standards increases, the FGT will decrease (less poor). For $\alpha > 1$, the FGT will be greater when the degree of poverty increases (when a poor becomes poorer as her/his living standards decrease). The FGT measures can also be disaggregated at different levels such as different population subgroups (e.g., urban/rural, female/male). Poverty gap index and squared poverty gap index are known as complementary to headcount ratio. For instance, the headcount ratio can be low but the intensity of poverty for those who are poor can be high, or the headcount ratio can be high, but intensity of poverty can be low meaning that a majority of those who are poor are just below the poverty line.

In the literature, there are also various papers focusing on (monetary) unidimensional poverty measures (or known as welfare measures) such as Sen (for 'Sen index' see 1976, 1979), Watts (for 'Watts index' see 1968), Kakwani (1980a, 1980b, 1984), Chakravarty (1983), Clark et al. (1981), Atkinson (1987, 1992), Foster (1983), Seidl (1988), Foster and Shorrocks (1988, 1991), Zheng (1997), Shorrocks (1995), Ravallion (1996), Kumar et al. (1996). Ravallion (1992), Rio Group (2006) provide a comprehensive review on poverty lines (e.g., absolute, relative, subjective). Anand and Harris (1994) show how the welfare of a person/household may substantially change due to the type of indicator (e.g., household income per capita, household total expenditure per capita, household food expenditure per capita, household calorie intake per capita, household inverse food-share). As the main focus of this thesis is not monetary poverty, we don't provide any other technical details for other measures of monetary poverty.

Culture of poverty, cycle of deprivation, underclass

The use of cultural terms to explain a social phenomenon is not a new practice in the literature. In this regard, Max Weber's work *The Protestant Ethic and the Spirit of Capitalism* (1905) can be considered as one of the master pieces in modern sociology, pointing out how cultural values have played central role in the life of North-western European protestants. Weber showed that delaying immediate gratifications of doing or having something for future rewards, as one of the doctrines of Protestantism, contributed significantly to increase economic prosperity of North-western European protestants. Based on his observations and his collected statistics, Weber argued that in North-Western Europe, the Protestant ethic had oriented people to adopt a secular mentality and notably it stimulated people to create their own firms, to engage in trade and to save for the accumulation of wealth. The 'Protestant work ethic', therefore, led to the emergence of capitalism. Weber's marvellous work unavoidably spread the idea of working hard, saving money, and investing as one of the fundamental steps for becoming rich and for thus avoiding poverty. After the 1950s, the modernization theory (see Moore 1966, Parsons 1967) tried to explain in cultural terms why some countries are poor compared to others, and what the poor countries have to do for becoming advanced (industrial) countries[30]. Notably, fatalism and inability to delay present gratifications are considered as the most important brakes

[30] In the original work, instead of country, the word 'nation' is used. But, as each nation does not have their own countries, we prefer to use the term country.

inhibiting development. The examples of some East and Southeast Asian countries known as 'Asian miracles' have been also subject to similar studies to explain how the Asian values contributed to economic development. In opposition to this idea, Kerbo states (2006) that for instance, despite Burma, Laos, Cambodia, Thailand have similar cultural heritages from their past, only Thailand did remarkable efforts to reduce poverty. Unfortunately, the other three countries remain very poor.

In the 1980s, the World Bank focused on cultural explanations in their development related studies and financed several projects on this issue. For instance, in the book titled *Underdevelopment Is a State of Mind. The Latin American Case,* Harrison (1985) has argued that the main driver of underdevelopment and poverty in Latin America was due to Spanish Catholics values assimilated by Latin American countries[31]. In next paragraphs, we will see some of the most important studies made in the field of culture and poverty. It is important to keep in mind that there is neither agreement nor scientific proof in the literature on the idea that culture[32] is the main driver of poverty. Nonetheless the notion of culture had substantial impact on poverty discourse notably from the 1960s. Rich and powerful people including politicians, have used cultural matters to convince public that poor are poor because the poor do not merit to be rich as they don't work hard, or because their life style is different than those of rich people.

During the period of welfare regime[33] dominance in the northern societies, when the anthropologist Oscar Lewis came up with his culture of poverty framework in 1959[34] the politicians did not feel ashamed for blaming the poor. For many of them, the poor are responsible for their own poverty because of their typical behaviours, which led them to live in poverty. In opposite, Lewis thought that the framework of culture of poverty might help reducing bad regards of well-off people towards the life and behaviours of the disadvantaged poor people, as he mentioned "the concept of the culture of poverty may help to correct misapprehensions that have ascribed some behaviour patterns of ethnic, national or regional groups as distinctive characteristics" (Lewis 1966, 19). He has particularly emphasized that not all poor people live in culture of poverty; there are some preconditions should be met. In his words, these conditions are "the setting is a cash economy, with wage labour[35] and production for profit and with a persistently high rate of unemployment and underemployment, at low wages, for unskilled labour. The society fails to provide social, political and economic organization, on either a voluntary basis or by government imposition, for the low-income population. There is a bilateral kinship system centred on the nuclear pro genitive family, as

[31] See also Harrison and Huntington's edited book *Culture Matters: How Values Shape Human Progress* for different angles on culture and development, published in 2001.

[32] According to Linton (1968, 78), the principal way of transmission of acquired behaviors between people is from parents to their children in the sense that members of new generation employ the habits of their parents and transmit these habits to their own children with some modifications due to their own experience. He defines *culture* in general sense as "the total social heredity of mankind", but also in a more specific sense as "a particular strain of social heredity". Therefore, for him, *culture* "as a whole, is composed of a vast number of cultures each of which is characteristic of a certain group of individuals".

[33] See Titmuss (1969 [1958], 1968, 1974) for the concept of welfare (i.e., The Welfare State, Welfare Society).

[34] Gans (1995, 24) mentions that earlier some other initial works are done on the culture of poverty by other scholars but with the works of Lewis the concept has become very known notably among public.

[35] The word 'labour' (in British English) is written as 'labor' in American English.

distinguished from the unilateral extended kinship system of lineage and clan. The dominant class asserts a set of values that prizes thrift and the accumulation of wealth and property, stresses the possibility of upward mobility and explains low economic status as the result of individual personal inadequacy and inferiority" (1966, 21). Therefore, when these conditions are met, Lewis considers that the person lives (his lifestyle) in a certain culture of poverty. It is very important to note that regarding him, these conditions can be more likely developed in capitalistic systems characterized by strong individualistic behaviours, and class stratification; more specifically, he mentions "free-enterprise, pre-welfare-state stage of capitalism", but also "colonial societies" (1966, 24).

For Lewis, the culture of poverty is a sort of both an adaptation and a reaction of the poor to their disadvantage position in a stratified, individualistic, and capitalist society. He argues that poverty influence in a negative way the participation of the poor (i.e., lower-class) to their society detracting them from the national culture and letting creating them their own sub-culture. He came up with the idea that the sub-culture of the poor has some similarities regarding "family structure, the nature of kinship ties, the quality of husband-wife and parent-child relations, time orientation, spending patterns, value systems, and the sense of community" in different parts of the world such as amongst lower-class families in London, Puerto Rico, Mexico City slums, Mexican villages, and the Afro-American population in the U.S (1959, 2). Thus, he argued that no matter where they live, the poor (in a less prosperous country) might have more in common with the poor (in a prosperous country) than with their well-off fellow citizen. He specifically focused on four aspects for describing the culture of poverty: the link between the sub-culture of the poor and the rest of society, the life style of communities where the poor live (he calls it the *slum community*), the structure of family (i.e., how the family functions, its characteristics), and the behaviours of individuals (see 1966). As we can see, his main aim was to observe both family relationship and individual behaviours. And he did not only focus on the poor, but he also searched for the context of country as a whole, considering carefully the social class system. In next paragraphs, we will see in more details his studies.

Through his anthropological studies, Lewis (1944, 1950, 1958, 1959, 1961, 1963 [1951], 1964, 1966a, 1967 [1965, 1966b], 1968) focused on the culture of poverty of middle, and lower-class poor families. He attempted to describe their life from different aspects such as their habits, emotions, hardships, residential conditions, working life, family relationship, friendship, social contacts in his books Five Families: Mexican Case Studies in the Culture of Poverty (1959), The Children of Sanchez (1961), Pedro Martinez: A Mexican Peasant and His Family (1964), La Vida: A Puerto Rican Family in the Culture of Poverty-San Juan and New York (1967 [1965, 1966b]). In his book *La vida*, he mentions that his main aim is "to give a voice to people who are rarely heard, and to provide the reader with an inside view of a style of life which is common in many of the deprived and marginal groups in our society…indeed, one of the major objectives of this volume is to bridge the gap in communication between the poor and the middle class personnel…lead to a more sympathetic view of the poor and their problems…provide a more rational basis for constructive social action" (1967, xii). In his narratives, one striking feature is that he leaves the reader to interpret. He successively achieves to not show his prejudices, personal judgments and evaluations (see chapter 3 in Valentine 1968 for an excellent review on Lewis' works on culture of poverty).

Lewis had done research in Mexico since 1943 by interesting firstly in the urbanisation of peasants (i.e., peasants who had leave the village Azteca for immigrating to Mexico City), and

later in the life of all residents in *vecindades* in Mexico City. In his book (1959) *Five Families: Mexican Case Studies in the Culture of Poverty*, he has attempted to portray the culture of poverty in contemporary Mexico, and he has given insights into the common features of lower-class in general by providing some apercus on the daily life of five Mexican families (four of the families are from the lower income group) based on his observations over five ordinary days of these families in Mexico City. Lewis clearly affirmed that he found very important to know the life of peasants and urban dwellers in developing countries, which was constituting almost eighty per cent of the world population at that time. With this motivation, Lewis was very interested in psychology of these population. He wrote that "what happens to the people of these countries will affect, directly or indirectly, our own lives. Yet we know surprisingly little about them. While we have a great deal of information on the geography, history, economics, politics and even the customs of many of these countries, we know little about the psychology of the people, particularly of the lower-classes, their problems, how they think and feel, what they worry about, argue over, anticipate, or enjoy" (1959, 1) Nevertheless, he did not give a clear definition of neither poverty, either culture of poverty. He was just interested in observing whether these families have or not similar behaviours and living standards as a common sub-culture in contemporary Mexico. He argued that "there are striking similarities in terms of family structure, personal relationship, timetable, money, hierarchy of values, conception of community among the poor areas of London, Glasgow, Paris, Harlem, and Mexico" (1961, 30). Further, he inquired in more details the life style of each of the five members of a low-income family *Sánchez* (1961) living in slum of Tepito (*vecindad* Casa Grande) in Mexico City, describing *vecindades* (i.e., dwelling conditions, social life, economic situation), but in particular the children's souvenirs, experiences, observations, opinions on family and proletariat life. Moreover, in 1967, by his fieldwork in Puerto Rico, he wanted to check for the validity of his theory. For that, he observed and compared 100 lower-class families living in the slums of Greater San Juan, Puerto Rico, and another 100 Puerto Rican lower-class families living in New York, U.S. Through his studies on the culture of poverty, he employed a family study technique rather than traditional anthropological methods. He justifies his approach in that way: "Unlike earlier anthropological studies, the major focus of this study is the family rather than the community or the individual…whole family studies bridge the gap between the conceptual extremes of culture at one pole and the individual at the other; we need both culture and personality as they are interrelated in real life" (1959, 17)[36].

In his work, *The children of Sánchez* (1961), he listened each family member's telling his own story with their own words, perspectives, and he saved these narratives by a magnetophone. Lewis, notably, has emphasized that there were very few in-depth poverty studies investigating psychology of poor people in developing countries. This is why, in this book, he wanted to take in consideration this psychological aspect. He mentions that despite problems and complexities

[36] Lewis summarizes his technique by 9 steps as follows: "In summary, the major steps involved in producing a well-rounded family study are as follows: (1) census-type data are gathered on a large number of families selected on the basis of the major variables of interest to the study; (2) from this sample a smaller group of families are selected for more intensive study; (3) interviews are conducted with each family member to record their life story and to question them on a wide range of topics; (4) a week or more of consecutive days are reconstructed on the basis of intensive interrogation; (5) complete days in the life of the family are observed and recorded; (6) recorded interviews are transcribed from the tapes; (7) typed data are translated, edited and organized; (8) re-interviewing is done to fill in gaps in the data; (9) he final version of the autobiographies and days are edited for publication" (1967, xxiv).

in their life, the *Sánchez* brought us to their life and made us understand their possibilities, but also their wasted talents. During his interviews, Lewis tried to collect systematically information on their "souvenirs, dreams, hopes, fears, joys, pains, works, relationships with friends, family, patron, sexual life, conception of justice, of religion, of politics, knowledge on geography, and history; in brief, their worldview." (1961, 25). In introduction, Lewis starts with this sentence "This book tells the story of a poor family in Mexico: Jesús Sánchez, the father, fifty-two years old, and his four children, Manuel, thirty-two years old; Roberto, twenty-nine years old; Consuelo, twenty-seven years old, and Marta, twenty-five years old. My goal is to offer to reader a detailed vision of a family life, and what does mean grow up in a one-piece apartment inside a dilapidated building, in a centre of one of the biggest cities of the Latin America during the economic and social evolution" (1961, 14)[37]. Later, he writes that "certainly, the life of the poor is not monotone. Stories in this book expose a universe of violence and death, pain and deprivation, infidelity and broken families, delinquency, corruption, police brutality and cruelty of poor to poor." (1961, 15).

In earlier pages of the book, Lewis describes delicately without personal judgments the neighbourhood (*vecindad*) and Casa Grande[38] where the Sánchez lives. Thanks to his descriptions, reader can easily get the picture of poor living standards of Casa Grande and vecindad[39] (e.g., one-piece apartments without window, common toilets, no accessible water inside building), and some living habits (e.g., street vendors, streets in night). Despite multiple difficulties, habitants are reported to be satisfied and proud of living there. Lewis explains that all basic necessities can be found easily in the neighbourhood. Habitants leave rarely the vecindad, they are almost foreigners to the rest of Mexico City. The population is composed of poor traders, artisans, and labourers. They have generally low wages, for instance Jesús Sánchez was having an average level income approximately one dollar per day. He could barely afford the expenses, he was needing a second job such as selling lottery tickets, pigeons. Some families could afford a domestic coming every day for doing household tasks such as laundry or cooking.

In fact, through his narratives, in some way, Lewis has pointed out multidimensional aspects of poverty, and he showed that the economic and material deprivation were not the only negative situations that the poor were facing off. Their life was encountering many difficulties in multiple aspects such as fragilized family relationships, get stuck in low living standards, degraded social life, feeling of hopeless for their future and limited opportunities for both present and future. To handle hardships, the poor created their own environment, (strong bonds) communities featured by their own particular habits, activities, actions, reflexions, rules, and fatalism. For instance, the social participation of the poor adults to the institutions was quite weak. Welshman (2006, 90) writes that "people with the culture of poverty, it was alleged, did not belong to trade unions, were not members of political parties, were not participants in the welfare system, and did not make use of banks." In some way, community sense was strong, for instance, they were creating long standing friendship, they were going to same school, they

[37] Lewis met Jesús Sánchez and his four children in October 1956 during his fieldwork in Casa Grande. The father, Jesús, was living there for more than twenty years, but meanwhile his children have left and come again for several times to their one-piece apartment in Casa Grande. Thus, Casa Grande is a very important place for them. Lewis mentions that Casa Grande is a kind of symbol of stability in the life of the Sánchez (1961, 15).

[38] Casa Grande is composed of one-hundred and fifty-seven one-piece apartments without windows.

[39] However, Lewis mentions that some families with relatively higher income have good quality furniture in their apartment and they are waiting for an opportunity for move out.

were dancing in the yards, they were getting married between them. But mutual help was limited when it comes to basic necessities. Similarly, Coates and Silburn (1970) in their investigation of poverty in Nottingham, have found that most of interviewed people has reported feeling hopelessness, powerlessness, not participating in the major institutions such as trade unions, and accepting their poor living conditions as their destiny.

As we have seen, Lewis has attempted to describe the culture of poverty. He reminds that in anthropology, the term culture means a lifestyle tradition including beliefs, customs retransmitted from one generation to another. Based on his studies, he observed that the poor were having a certain lifestyle stable and persistent passing from one generation to another within a family, as an intergenerational reproduction of the same kind of life and behaviours over years. For instance, when a person was born in a disadvantaged family it was extremely difficult for her/his to get out of this vicious circle, to build a new life and have a brighter vision of the future. In his book, he mentions that "children have usually absorbed the basic values and attitudes of their subculture and are not psychologically geared to take full advantage of changing conditions or increased opportunities which may occur in their lifetime" (1967, xli). Based on his studies in Mexico (1961, xxvi), he lists some *universal characteristics* (e.g., lower life expectancy, low level of education and literacy, relatively higher level of death rate), *economic traits* (e.g., unemployment, underemployment, low wages, unskilled occupations, tackle for surviving), *social and psychological characteristics* (e.g., high incidence of alcoholism, mother-centered families, physical violence in raising children); related to that for the complete list including other traits, and other living characteristics, Leeds (1971, 239-241) provides a complete list. Concerning social and psychological consequences of the culture of poverty he mentions that it blocks the poor participating fully to national culture, for instance, "on the level of the individual, major characteristics are a strong feeling of marginality, of helplessness, of dependency, and of inferiority…weak ego structure…little ability to defer gratification and to plan for the future…resignation and fatalism…" (Lewis 1966b, xlvii). In the end, the culture of poverty becomes the sub-culture of the poor, which can be seen as the reaction of the poor to poverty, and a solution for their problems because they do not have an easy access to the institutions, they are deprived of real opportunities, sources, means, and they are socially excluded from the society. For instance, as they know that they cannot get a credit from banks, the poor organize their own debt system. The poor were very critical towards elites, and institutions, for instance, they do not like police, they mistrust government because they believe that institutions do not stand by them. Growing up in such sub-culture with some particular psychological characteristics inhibit them to get the normal behaviour for escaping from poverty. From this perspective, the culture of poverty becomes self-perpetuating (see Valentine 1971).

Lewis was conscious that labour class in developed countries was not having the same conditions as in developing countries. For that reason, he highlights that "the culture of poverty is not the synonym of labour class, proletariat or peasantry…. The culture of poverty can be observed among the most disadvantaged people regarding socio-economic status, the most disadvantaged labourers, small farmers, plantation workers, small artisans and traders which is habitually called *Lumpenproletariat"* (1961, 29-30). But he also mentions that some of the items of the culture of poverty do exist in different forms among middle and upper class in Mexico. For instance, alcohol is a social pleasure for middle class but for the poor alcohol is consumed for forgetting problems or having self-confidence to face with hardships. However, Lewis did not express concisely what he was understanding by saying poverty. He just

mentioned some situations of the poor in Mexico such as "…most of the poor have a very low educational level, they do not belong neither labour union, nor political parties, they do not benefit of medical, maternity help, either old pensions offered by social security, they do not go to banks, hospitals, big stores, museums, galleries (of paintings), and aerodromes." (1961, 31). He also provided some economic, social, psychological characteristics of the culture of poverty such as underemployment, unemployment, low wages, precarious work, child labour, alcoholism, violence, living in the areas with high density of population (see Lewis 1961, 33-34 for complete list). In fact, the two notions *poverty* and *culture of poverty* were a kind of overlapped.

From empirical perspective, Lewis did not provide solid social indicators designing his theory, he did not attempt to show any cross-sectional or dynamics of the impact of poverty on culture of poverty. On the one hand, he did not prove statistically whether psychology (i.e., behaviour) of the poor cause or has an impact on their poverty, or whether their poverty cause or has an impact on their behaviours. On the other hand, the sample was relatively small which can cause bias in generalization of the results. Eventually, his results were open to tricky political debates on poverty. Even if he mentioned that not all poor people were having the same sub-culture[40] and not in every society the culture of poverty appears[41]; however, his works stimulated a new regard to poor people. In particular, politicians exploited this approach to manipulate and convince public that the poor have pathologic culture, and they reproduce the same mistakes constantly as living in and reproducing a stigmatised hazardous culture for whole society. Lewis (1966b, li-lii) suggested to eliminate the culture of poverty in the world; however, he mentioned that "it is much more difficult to eliminate the culture of poverty than to eliminate poverty per se," and "the elimination of physical poverty per se may not be enough to eliminate the culture of poverty." Based on his observations, he also advised amelioration of the poor's material conditions, and he predicted worsen social conditions for Mexico because of high inequality.

As mentioned by Leacock (1971a), major works in the field of culture of poverty are massively influenced by Lewis's statements on the existence of specific self-perpetuating, self-destructing, fatalism and immediate gratification-based *subculture* among the poor. However, besides Lewis's culture of poverty, there has also been other works dealing with more specific issues such as the approaches of *lower-class culture*, and the *culturally deprived or disadvantaged* (see Valentine 1968, 1969; Banfield 1968; Clark 1965; Leacock (ed.) 1971b)[42]. Similar to Lewis, Walter Miller's *independent lower-class culture* (1958, 1959) has emphasized that lower-class status and poor living conditions led the poor to have their own culture- e.g., the poor have generally different behaviours regarding situations and opportunities. Miller has argued that stable jobs and relatively good income won't be enough for eliminating poverty.

[40] He wrote that (1968, 57) "my rough guess would be that only about 20 per cent of the population below the poverty line (between six and ten million people) in the United States have characteristics which would justify classifying their way of life as that of a culture of poverty".

[41] He wrote that (1968, 54) "in effect, we find that in primitive societies and in caste societies, the culture of poverty does not develop. In socialist, fascist and in highly developed capitalist societies with a welfare state, the culture of poverty tends to decline. I suspect that the culture pf poverty flourishes in, and in generic to, the early free-enterprise stage of capitalism and that it is also endemic in colonialism."

[42] Gans (1995, 25) emphasizes that Lewis' culture of poverty concept has stimulated the emergence of the concept of lower-class culture and its use among the journalists in the 1970s for mentioning the undeserving poor.

Regarding the life of lower-class American Negroes, another interesting study by Rainwater (1966) expresses that the poor prefer to live with their own culture (i.e., norms, behaviours) because they think they will never reach to higher living standards despite they do also have aspirations for better living conditions. One of the studies carried out by the New York State Commission Against Discrimination resulted that the children of Negro families had in fact more aspiration compared to those of White families from the same income level, but less real opportunities to accomplish their ideas (cited in Harrington 2012 [1962], 77). From this perspective, Gans (1970, 152; see also Gans 1995, 1968) has argued that it is very important to take in consideration both aspirations and behavioural norms in the culture of poverty studies, because the poor are not necessarily poor because of their bad behaviours but they can be just discouraged for pursuing their own aspirations for having a good life. Gans (1970, 153) has also argued that some behaviours and aspirations of the poor are not constant over time, "they are situational, developing as responses to situations which people face". H.G. Lewis (see his writings in the report financed by the National Institute of Mental Health, cited in Seligman 1968, 51) mentioned that the poor are not generally different than others in terms of value, but their problems are significantly different. For instance, the poor's children are often absent at school because they do not have enough food bringing with them for the day and/or they do not have (appropriate) clothes to wear for the school.

Moreover, Harrington (2012 [1962]), in his book *The Other America*, has asserted that poverty in the U.S. is a specific *culture* that is significantly different than the life of the mainstream society and this specific culture generates *multi-problem* families. But most importantly this culture corresponds to a *hidden* (new) poverty affecting around 50 million people[43]. Harrington (2012 [1962], 162) described the situation of the poor as "who lack education and skill, who have bad health poor housing, low levels of aspiration, and high levels of mental distress" including more specifically "unskilled workers, the migrant farmworkers, the aged, the minorities, and all other who live in the economic underworld of American life" (2012 [1962], 2)[44]. For instance, he mentions that "…beauty and myths are perennial masks of poverty. The traveller comes to the Appalachians in the lovely season. He sees the hills, the streams, the foliage-but not the poor. Or perhaps he looks at a run-down mountain house and, remembering Rousseau rather than seeing with his eyes, decides that "those people" are truly fortunate to be living the way are and that they are lucky to be exempt from the strains and tensions of the middle class. The only problem is that "those people", the quaint inhabitants of those hills, are undereducated, under privileged, lack medical care, and are in the process of being forced from the land into a life in the cities, where they are misfits" (2012 [1962], 4). Concerning the invisibility of poverty, he yields some examples such as poor people living in city centre but socially isolated, or properly dressed poor people lacking access to adequate housing, healthy nutrition, and health services, or the elderly people living in loneliness, or poor people ruled out of politics (e.g., no being member of unions, political parties). To show the magnitude and

[43] Gans (1995, 26) mentions that Michael Harrington (2012 [1962], 1984, 1981), American active socialist, is the first to used Lewis' concept to explain poverty in the United States. According to Gans, Harrington's way of using this concept is more economic related and contains less focus on the meaning of culture in sense of inter-generational transmission in explaining poverty in the U.S.

[44] Harrington (2012 [1962], 1-2) emphasizes that poverty in the U.S. is not the same as in many developing countries; however, "if these people are not starving, they are hungry, and sometimes fat with hunger, for that is what cheap foods do. They are without adequate housing and education and medical care".

worrisome part of American poverty, he mentions that the poor "…are not simply neglected and forgotten as in the old rhetoric of reform; what is much worse, they are not seen" (2012 [1962], 6).

Notably, Harrington (2012 [1962], 12-13). considered poverty as a result of: (i) transformation of American cities mainly due to technological changes in the market; and (ii) psychologic difficulties, as a malaise, adaptation problem to American society. The poor have been seen as having particular traits, attitudes involving them with poverty. He highlighted the structural changes regarding the requirements of labour market in terms of technological and educational levels. Greater American society has become more technology based higher educational and skills level were demanded in the market. This was a major disadvantage for the new poor lacking access to higher education, various opportunities, being left alone in ghettos[45] with a multitude types of deprivation. Older generation were used to be employed based on their lower level education and skills as the industry at that time was functional with such level; for instance a young person might prefer to start working as a labourer instead of going to university and after he or she might go up the stairs thanks to her/his work experience through years. But the new contemporary poor were clearly threatened by the arrival of machines, or in other words by new technology. Harrington mentioned inter-generational risk of such poverty in the sense that if the parents of the new poor "were penalized a generation ago for their lack of schooling, their children will suffer all the more." For that reason, he has expressed the situation of the new poor as living in the culture of poverty. Harrington (2012 [1962], 16-17). was also very interested in qualitative works to understand specific aspects of the Other Americans' way of living. He argued that "there is, in short, a language of the poor, a psychology of the poor, a worldview of the poor. To be impoverished is be an internal alien, to grow up in a culture that is radically different from the one that dominates the society…They need an American Dickens to record the smell and texture and quality of their lives."

Lewis and Harrington's ideas had significantly influenced the creation of War on Poverty programme[46] (during the presidency of Lyndon Johnson, announced by him on January 8, 1964) in the U.S (see Brauer 1982 for the War on Poverty). This programme has aimed at expanding social service programmes (notably by hiring more social workers and adopting a new welfare system) for reducing poverty in the U.S. As Gladwin (1967, 26) mentions this governmental programme was "designed very explicitly to correct the social, occupational, and psychological deficits of people born and raised to a life of poverty." Meanwhile, Seligman (1968, 217) has argued that these anti-poverty programmes were useless because the typical poverty traits pass from generation to generation becoming "ineradicable"; he calls this type of inter-generational

[45] Glasgow (1968, 43) mentions that "the ghetto is patently a northern invention employing the tool of discrimination to keep the Negro apart from the rest of society".

[46] A number of acts came into force such as The Economic Opportunity Act (1964), Food Stamp Act (1964), Elementary and Secondary Education Act (1965), Social Security Act (1965). Also see The Great Society programme introduced by the President Johnson in 1964-1965, aiming at reducing poverty and racial discrimination. The War on Poverty is generally perceived as the continuation of former President Roosevelt's New Deal programme (1933-1936) which was implemented to fight Great Depression period's effects in the U.S. There are some critics as well on the efficiency of the Economic Opportunity Act for instance, see Levitan (1969).

sub-culture of poverty as "a social syndrome"[47]. Kerbo (2006) points out that the War on Poverty was not efficient because poverty reduction in the 1960s and 1970s was mainly driven by the fact that the poor got more cash thanks to the welfare system. This reduction was not realized because hired social workers taught well the poor on how to work harder, and behave better for finding and keeping jobs, or in turn the poor learnt well how to behave correctly and succeed to change their bad behaviours to good behaviours.

Valentine in 1968 has pointed out in his book the link between culture and poverty. Basically, he perceives poverty from a comparative perspective as a relative deprivation (see Townsend 1979). He argues that the conditions implying poverty status depend on societies and cultures, which change as well over time due to various factors. He specifically mentions that the source of poverty is inequality. Not everyone has advantaged position in society, there are always some people who are more prosperous than others. For instance, he makes clear that "the poor are deprived in comparison with the comfortable, the affluent, and the opulent" (1968, 13). To highlight his proposition, he provides some examples from the U.S. such as in terms of "occupations, education, and political power….Desired resources in each of these areas of life are quite unevenly distributed among the several strata of the American class system." (1968, 14). Related to inequality, he criticizes the U.S. population in the sense that the question of equality comes to play very often in debates and in public perception; however, at the same time, poverty appears dramatically in different forms especially if we look at the situation of underclass[48], socially excluded people due to their race, ethnicity, and cultural values, and also due to hierarchical social class system such as almost all people of lower-class are poor and some people from working class find themselves disadvantaged in American society. From this perspective, Valentine is convinced that this contradiction lead to conceding of poverty as a social problem in the U.S. He clearly expressed in his book *Culture and Poverty* (1968) his worries concerning the emanation of the idea that poverty is the consequence of poor people's own subculture (i.e., culture of poverty- see Lewis above). And, related to latter he is interested in showing how prejudice, misconception on the poor and their lifestyle have emerged in public opinion (for that, he suggests particularly see the papers of the 1966 meeting of the American Anthropological Association).

Through his review and discussions on culture and poverty (Valentine 1968, 141-144), he proposed three models on life styles (as a sub-culture) of the lower-class (as a sub-society): model 1 as the self-perpetuating sub-society with a defective, unhealthy subculture; model 2 as the externally oppressed sub-society with an imposed, exploited sub-culture, and model 3 as the heterogeneous sub-society with variable, adaptive sub-cultures. Regarding first model, he mentions that the poverty sub-culture makes difficult to get out from poverty mainly due to the social relationships adopted by the poor and its psychosocial consequences. In his second model, he highlights the use of power of the higher strata to define the position of the poor in society in the sense that the higher strata would like to protect their advantages and because of that, they are not interested in redistribution of sources, and in creation of opportunities for the poor. And, Valentine draws attention on a positive aspect of sub-culture of the poor in his third

[47] Seligman (1968, 16) pointed out that "the poor had become a hidden subculture of America, a culture that perpetuated itself in endless cycles, a culture beyond the reach of the contemporary welfare state, a culture whose inhabitants were lonely, insecure, fatalistic, pleasureless".
[48] See other coming paragraphs below for more explanations on that.

model in the sense that these people achieve to adapt themselves to their conditions (for more patterns information on low class people and their subculture, see Valentine 1968 chapter 6).

Not very far from the concept of culture of poverty, but with a special focus on difficult family situations, the concept of *problem families* (Blacker (ed.) 1952, Spencer 1970) has emerged in early 1940s. This concept has attired a few attentions from scholars because it was perceived that analysing separately some families from the rest of other families was not making sense; but also taking some families as a homogeneous group was another problematic issue (see Welshman 2006, 119). However, Rutter and Madge (1976) were convinced that it was important to examine various problems and cycles of disadvantages of families. It was recognized, in the literature, that the disadvantages and its inter-generational characteristics were manifesting in different forms due to various situations. There was no unique type of problem families, nor a stereotype of transmission and continuity of disadvantage cycle between generations. For instance, Brown and Madge (1982) found some correlation between income level and occupational group across generations, and "health, criminality, and emotional experiences in childhood could lead to poor parenting behaviour in adulthood" (cited in Welshman 2006, 124). Some studies showed also examples of discontinuity of disadvantage situations, but also falling into disadvantage situations despite they weren't born in such families (see Welshman 2006, 120). Similarly, Atkinson et al. (1983) have investigated the income situation of families from inter-generational perspective; however, his results did not show any cycle of disadvantage. Moreover, for having a comprehensive look on 'problem families', Philp and Timms (1957), Schlesinger (1963), Jordan (1974) provide an extensive overview on this topic.

Rutter ad Madge's work fundamentally rests on Sir Keith Joseph's (the Secretary of State for Social Services in U.K) speech in 1972 during the pre-school playgroups association meeting, where he queried why the problems of deprivation were still existing and continuing persistently despite the U.K was known to be in a quite prosperous, full employment period and the improvements in social services after the Second World War. He defined deprivation as the circumstances, the environment of the person making difficult him or her to explore and take advantage of her/his physical, and intellectual skills and abilities, and to encounter opportunities to progress in her/his life. Deprivation entails people to poverty preventing them having access to good quality of education, health services, and to emotional smell/imbalance, fragile physiological state such as depression, despair, disorder. He clearly perceives deprivation as an interactive multiform fact existing in each stratum of society (not only among the poor). For that reason, Sir Keith Joseph suggested new deprivation studies in a dynamic perspective to examine life cycles of people and to know more about their disadvantaged periods and the evolution of their deprivation through years. In his speech, he said that "it seems perhaps that much *deprivation* and *maladjustment* persist from generation to generation through what I have called a '*cycle of deprivation*'. People who were themselves deprived in one or more ways in childhood become in turn the parents of another generation of deprived children…ways and extent to which children in working-class families, particularly those with semi-skilled or unskilled fathers, are disadvantaged from birth" (Rutter and Madge 1976, 5). In this perspective, Rutter and Madge (1976) focused on cycles of disadvantaged families and on the nature of transmitted deprivation from one generation to another within the family. Their study is based on the concept of deprivation which they define as "varied aspects of a person's environment" (1976, 1) including economic status, housing, academic performance and educational

attainment, occupational status, crime and delinquency, psychological imbalance, parenting behaviour, family problems, and ethnic minorities.

During the last half of the 20th century (notably since the late 1970s except Myrdal's earlier works- see below), the American scholars' view on poverty had begun to change and to include a more social perspective. This especially led to emerge the concept of *underclass* (coined by Myrdal), which can be seen as a specific label attributed to a particular group of people (notably poor and vulnerable) emerging in inner cities in American society. In other words, the *Underclass*[49] designates a group of people (in U.S. context, young and minorities were mainly part of them) who lives in poverty and vulnerability because of their physical appearance, ethnic, origin, race, or other specific demographic or social characteristics. In 1977, The *Time* magazine qualified the American underclass as "a large group of people who are more intractable, more socially alien and more hostile than almost anyone had imagined. They are the unreachables" and it "produces a highly disproportionate number of the nation's juvenile delinquents, school dropouts, drug addicts and welfare mothers, and much of the adult crime, family disruption, urban decay and demand for social expenditures" with an estimation of at least 7 to 8 million Americans (August 29, 1977, 14-17)[50]. Gans (1995, 31-32) claims that from that publication titled "The American Underclass: Destitute and Desperate in the Land of Plenty", the initial look to underclass in behavioural terms has become a specific label notably for the Black and Hispanic poor families. In the literature, various studies have attempted to show different causes of the emergence of underclass such as economic problem, dynamics of social situation in cities, cultural matter, or social welfare programmes.

Classically, the underclass studies have focused on the *American ghettos* such as Afro-American, Hispanic population or other minorities (Glasgow 1980, Auletta 1983 [1982], Myrdal 1963, Clark and Nathan 1982, Cottingham (ed.) 1982c, Wilson 1985, 1987, 1989, (ed.) 1993, Ricketts and Sawhill 1988, Simon 1993, Katz 1986, 1989, 1995, 2013, Katz (ed.) 1993, Jencks 1989, 1992, Jencks and Peterson (eds.) 1991, Kornblum 1991, Smith (ed.) 1992, Gans 1971), and on the *urban underclass* which has played a substantial role in poverty debates in the U. S. (Katz 1995). For instance, as Welshman (2006, 128) mentions, the *underclass* is used in different terms by Right-hand and Left-hand politicians in the U.S. Right-hand politicians have agreed that underclass represents "unemployed and unemployable people who faced lives of misery in the inner-city, characterized by violent crime and illegitimacy". On the left side, underclass was recognized as representing "a group of undereducated and unskilled workers apparently left behind by profound shifts in technology and in the economy as a whole." Katz (1993, 5) summarized in six items the main concerns in underclass debate: distinguishing individual vs. structural poverty, determining the link between cultural factors and poverty, studying the social and family environment of child, finding the environmental effects (e.g., neighbourhood) on poverty emergence, questioning the efficiency of institutions for combatting poverty and examining how public policies are defined and how they contribute eliminating or increasing the level of poverty in American society.

[49] In many texts, the *Underclass* is employed for meaning the American poverty (for instance, see Auletta 1983 [1982]).

[50] Vol 110. No.9. In addition to that, as mentioned by Gans (1995, 31), previously in the June 17, 1974 issue of Time magazine despite the cover story was about the rise of black middle class, on pages 26 and 27 in a little box entitled "The Underclass: Enduring Dilemma", poor black families were referred to "troubled underclasses".

In underclass studies, Charles Murray's controversial book *Losing Ground* (1984) can be considered as the most important one. Murray is the first person who introduced the term *underclass* in the early 1960s through his American studies. In his book *Losing Ground*, he discusses the effectiveness of the welfare regime, voluntary poverty, social policies in the U.S. between 1950 and 1980. According to Murray, the welfare state regime is not efficient unlike many people do agree (see Garland 2014 for 'welfare state')[51]. He argues that the functioning of welfare regimes creates, in long-term, a certain dependence among the disadvantaged people to social assistance, charities, or any sort of private help, despite the welfare regime appears very attractive in short-term to policy makers for solving poverty problems. But also, to him, welfare dependency causes break-up of traditional family structure, encourages disadvantaged people to not work and even increases criminality. Morris (1994, 80) refers this situation as "socialisation into a counter-culture". Murray's main ideas are: (i) in any cases, the eligibility criterion refrains some people reach to social help; (ii) social transfers (as a result of social assistance) make more attractive to being eligible to social programmes; (iii) as long as the change does not come from the person, social programs do damage more the society rather than make it more prosperous. He explains that welfare regime is not accurate for abolishing poverty in a sustainable way. In that sense, he considers that poverty relieve programs can stimulate more poverty.

For supporting his idea of inefficiency of welfare regime, Murray showed with numbers that from the second half of the 1960s, the situation for the poor and other disadvantaged people become worse despite the federal welfare programmes including those on employment, education, family, poverty, security. For instance, he expressed (i) "*latent poverty"* concerning those who were below the official poverty line, plus those who were above it only because of welfare payments; (ii) an increasing unemployment rate (i.e., his theory of discouraged worker): (iii) out-of-wedlock births (see Murray 1984, Welshman 2006). According to his results, "*latent poverty*" had decreased in the 1950s, but started to increase again after 1972, to reach 22 per cent by 1980" (cited in Welshman 2006, 134-135). This result was due to two main reasons: (i) family problems; (ii) economically inactive young men (in the labour market). Notably, he mentions that the results are more intense for black families. Murray, therefore, associated the emergence of underclass with two problems: black single mothers, and exclusion of black youth population from labour market[52]. Concerning unemployment situation, Murray found that (i) in the second half of the 1950s and in the late 1960s, black youth unemployment rate had increased; (ii) from the 1970s, the difference between the labour participation rate of white and those of black had dramatically manifested. Regarding out-of-wedlock births, the percentage had augmented from 17 in 1950 to 48 in 1980 for black single women, while it had augmented from 2 in 1950 to 11 in 1980 for white single women. As we can see, the difference is quite remarkable. In addition, he showed that the rate was much higher for teenage black women, for instance, in 1980, 82 per cent of all births among black women aged 15-19 were out-of-wedlock. Through these findings, Murray wanted to point out that it is important to measure and evaluate the impact of social policies on people's behaviours, in particular, focusing on people receiving social welfare. For instance, he has argued that some social assistances

[51] With the coming of Reagan to the presidency in 1981, in the U.S., debates on war on poverty and social welfare programs had a new direction influenced by the Republican Party approach.

[52] . As mentioned by Morris (1994, 80), these two populations can be considered as outsiders compared to mainstream society.

(benefits and services) may be positive encouraging people to sort their hardship out, but some of them can be negative (i.e., increasing delinquency or deviance) or receiving social help and other allocations may become a normal process.

Despite its large success, *Losing Ground* had also received some critics notably regarding its contained statistics. For instance, in her book, Alice O'Connor (2001) critics heavily *Losing Ground* based on historical data on U.S. poverty. First of all, she argues that the poor had benefited positively from social transfers. Secondly, she states that empirical analysis shows no statistically significant link between social allowance and out-of-wedlock births. Related to this, she wrote "in fact, Murray had used the wrong measure altogether, by pointing to the illegitimacy ratio (nonmarital births to all live births), which had risen, rather than the illegitimacy rate (nonmarital births to the number of women of childbearing age), which had declined" (2001, 248). Regarding O'Connor, Murray should have used 'rate' instead of 'ratio'. Thirdly, she highlights that the U.S. welfare programme on elderly helped to reduce dramatically eldest poverty, and the Aid to Families with Dependent Children (AFDC)[53] had started to decline from 1969[54]. She mentions that Murray did not choose correct indicators, there was something wrong with the numbers he collected. According to O'Connor, in opposition to Murray, "the vast majority of poor people did not choose welfare over work" (2001, 248). Fourthly, she critics the latent poverty concept of Murray by saying that he has used pre-transfer poverty measures developed by the Institute for Research on Poverty (IRP). IRP researchers were used these measures for evaluating the impact of social welfare programmes- whether the programmes reach well to its target population or not, especially, for showing the help of these programmes to keep low the U.S. official poverty rates. However, Murray used these measures to show that (i) the aid programmes such as AFDC had an aversive effect on the poor to not work or to not enough search jobs in labour market; (ii) the poor had become dependent to social programmes; (iii) the poor were not able to obtain real opportunities for getting through hardship and misery because of their dependency to social aids. On the other hand, Katz (1993) evokes that first beneficiaries of the AFDC were a small group of widows and any specific critics were done against this programme until the beneficiary group had likely become unmarried and black women[55].

Prior to *Losing Ground*, in *Challenge to Affluence* (1963), the 1974 Nobel Prize Winner Swedish/multinational economist Gunnar Myrdal has focused on the population of unemployed and unemployable people in America, which he called the *underclass*[56]. Regarding this approach, the emergence of underclass is due to labour market, economy failure to respond

[53] The AFDC is originally called Aid to Dependent Children (ADC), which is entered into force in 1935. The ADC was the first social assistance programme offered by the federal to women with children, within the initiative of Social Security.

[54] Notably, the establishment of *workfare* programs introduced by James Charles Evers in 1968, and promotion of these programs by the president Nixon in 1969 had influenced the diminishment of social welfare programs in the U.S.

[55] In this perspective, Katz (1993) underlines inconsistent and massive discriminatory public opinion on how unmarried and black mothers are counted as undeserving poor in American society.

[56] Gans (1995, 184) mentions that "when Gunnar Myrdal formulated the term 'underclass', he meant it to apply to people *being excluded from the economy*".

sufficiently the demand and to not offer enough secure jobs[57]. In opposition to Murray's approach which relates the underclass issue with the cultural traits, bad and deviant attitudes, behaviours of some group of people in American society (especially the black people), Myrdal's approach emphasizes the role of structural inequalities to generate disadvantaged groups excluding them from the mainstream society. Myrdal argues that technological and economic change had most affected the less qualified people (in terms of educational attainment and job market requirements), and this change had pushed these people into the most disadvantaged situation in American society. In fact, the technological revolution at that time affected significantly the working organization, notably manual labourers. This modification in the labour market due to technological revolution, therefore, stimulated the polarization in the U.S. society. Myrdal used the term *underclass* "to describe the effects of technological change on the American workforce" (Welshman 2006, 128). Myrdal explains that "an underclass of unemployed, unemployables, and underemployed, more and more hopelessly divorced from the nation at large and without a share in its life, its ambitions and its achievements" (1963, 19). Myrdal also points out that underclass family members enter to a vicious circle, and for children it becomes more and more difficult to get out of underclass and to create for themselves the necessary conditions for having a successful and pleasant life. The underclass lifestyle of the parents become a present and a future lifestyle of the children.

As Myrdal refers to economic and technological change as the source of underclass, from his perspective, the underclass is a structural problem rather than a cultural or a behavioural one. However, in the 1960s, the underclass issue, specifically the economic restructuring and long-term joblessness have been linked by some scholars to cultural behaviour (see Rainwater 1966, Liebow 1967, Hannerz 1969). For instance, Hannerz (1969, 177) mentioned the "ghetto-specific complex" which describes black man living in ghettos as "toughness, consuming a substantial amount of alcohol, having broken social relationships with his family, spending time outside his family in an informal way, being suspicious toward others, having trouble relationships with women." In this approach, scholars used the term 'lower-class culture', rather than underclass. They were interested in whether these groups have their own sub-culture, or whether their behaviour is just a sort of adaptation to their situation.

It is also important to note that Myrdal is also interested in black poverty (see Jencks 1992 for the cultural patterns of black poverty) in his work *American Dilemma* (1944) by asking *Why the Negro people are so poor*? The main aim of American Dilemma was to explore American people's perspectives on race. He concluded that the problem of poverty trap of black people was due to white people's behaviours to black people as the black race was considered as an inferior race. In other words, he argued that black people were systematically rejected from white community. Glasgow (1980, 1) explained the black's detachment from mainstream White society in terms of the behaviour of institutions (e.g., police), labour market conditions, and welfare regime features, by qualifying such situation as "the *new camouflaged racism*". Glasgow (1980, 5) detected the two major problems mining more and more the gap between the Blacks and Whites, leading the young Blacks (he calls also the Inner-city youth) to the permanent underclass status: (i) high poverty rate among the Blacks (three times higher than the Whites according to the U.S. Bureau of the Census's figures in 1959 and 1974): (ii)

[57] As evoked by Gans (1995, 28-29), Harrington's use of culture of poverty in his book The Other America (2012 [1962]) and Myrdal's use of underclass are quite in same direction in terms of "nature, causes, and effects of poverty" in the U.S.

structural unemployment among the Blacks (between 1959 and 1974, unemployment rate was decreased from 28.3 million to 16.3 million among the Whites but it was only decreased from 9.9 million to 7.5 million among the Blacks). In this regard, Glasgow (1980, 9-11) defined underclass as "a relatively new population in industrial society" including massively the long-term poor living perpetually at the subsistence level; however, they are not necessarily discriminated on the basis of their race but rather they are kept outside of the core of society due to their "social profile" (e.g., chronic unemployment, failure of schools, living in ghettos, distant relationship with the institutions). He also qualified underclass by its *lack of mobility* compared to the lower-class in the sense that underclass status passes from one generation to another whilst among the young lower-class people the likelihood of doing better than their parents appears higher. Similarly, Seligman (1968, 42) pointed out the "pathologic isolation" preventing the Negroes to move up regarding many socio-economic aspects and hence pushing them living in a certain poverty sub-culture in ghettos.

Furthermore, in terms of *negro poverty*, Moynihan's *The Negro Family: The Case for National Action* (1965)[58], Du Bois's *The Philadelphia Negro* (1899), Chicago Commission on Race Relations' *The Negro in Chicago* (1922) can be considered as some of the substantial works done in this area, reflecting the ideas on the relationship between cultural pathology and poverty. On the one hand, Moynihan's report, assistant secretary of labour for the president Lyndon Johnson, argued that out-of-wedlock births, female headed household type along with vicious circle of permanent assistance to social welfare were the main drivers of Afro-American families. On the other hand, DuBois found that Philadelphia Negroes were regularly marginalized from the urban industrial economy, and in other sectors as well. In 1960, the Department of Labour reported that 4 per cent of Negro employees were professional, technical and kindred workers vs.11.3 per cent for the White employees; 2.7 per cent of Negro employees were managers, officials and proprietors vs. 14.6 per cent for the White employees (cited in Harrington 2012 [1962], 73).

Wilson (1982, 113) emphasized that the majority of poor black families were headed by females (e.g., 74 per cent by 1978), and since 1960, black people have encountered increasing discriminations in labour market in terms of occupational group. There was a clear distinction between white's occupations and black's ones. For instance, based on the U.S. Department of Commerce data, from 1960 to 1978, on the one hand, the percentage of employed black workers decreased 13.4 per cent in service and labour, and 5.0 per cent in farmworker jobs; on the other hand, it is increased by 10.7 per cent in professional and technical occupations, and by 9.6 per cent in clerical jobs. Wilson drawn attention to the fact that underclass black community was composed of lower-class workers with low level income, or often below the poverty line, the long-term unemployed and discouraged people. This share of population was quite become dependent in a permanent way to social welfare. Concerning unemployment, based on the U.S. Monthly Labour Review, Wilson shows that compared to white men, the unemployment rate of black and other men substantially increased for 16-24 age interval (e.g., for 20-24, in 1955 it was 12.4 per cent and in 1978 it become 20.0 per cent). It was the same trend also for black and other women compared to white (e.g., for 20-24, the unemployment rate was 13.0 per cent in 1955 and it became 21.3 per cent in 1978). For white men between 20 and 24 yrs. old, the unemployment rate remained same as 3.0 per cent between 1955 and 1978. For white women between 20 and 24 yrs. old, the rate was 5.1 per cent in 1955 and it slightly shifted to 4.5 per

[58] See Rainwater and Yancey (1967) the full text of the report.

cent in 1978. In addition, Monthly Labour Review data showed that among black people aged between 16 and 19 living in poor family with a monthly family income under $5,000, 41.0 per cent of them was not enrolled at school, and among those living in a poor family with a monthly family income between $5,000 and $9,999, 33.6 per cent of them was not enrolled.

Moreover, in the U.S. literature, there are also other well-known authors for their precious work in this field. Auletta[59] (1983 [1982], xvi) argues that there are 4 groups of the underclass: (i) "the hostile" group such as street criminals, drop-outs, drug addicts; (ii) "the passive poor" who benefit permanently social welfare assistance, for instance, chronically assisted mothers; (iii) "the hustlers" who are relatively less dangerous than "the hostile" group because they are generally less involved in (violent) crimes however they do work in an informal sector; (iv) "the traumatized" people such as drifters, drunks, mental patients, homeless. Based on his work[60], he concluded that underclass is the result of behavioral or income problem. He also mentioned that there was no consensus at that time on the nature and extent of underclass. It is clear that Auletta does not mention a geographical perspective such as ghetto, urban lower-class to define underclass people; instead, he integrates different groups into the same category. Lemann (1986) points out that underclass appears because of social situation in American cities, and of cultural characteristics in southern blacks' lives. Cottingham (1982a) has emphasized the intergenerational character of poverty and disadvantage position of underclass people who are generally lack of good education, qualified skills, social activities, economic power which led them to being employed in precarious works with unstable working conditions and low wages. This situation makes them apart from well-paid, technological skills required prestigious jobs. Cottingham portrays underclass as "a permanently impoverished social stratum", affected severely by limited mobility in terms of occupational group, social life, and property (1982a, 179). Because of that, underclass people remained far from the American mainstream in terms of economic, social, and political power. The major population of underclass is the urban lower-class living in ghettos which he describes as "nutritional, health, neurological, or organic afflictions associated with densely populated, poverty-ridden, inner-city, urban environments" (Cottingham 1982b, 1-3).

As mentioned by Reid (1982, 83-84), the environment of ghettos suffers incredibly negative aspects such as pollution, cancer-causing factors, toxic chemicals. She mentions that minorities and the poor are more likely die from cancer, for instance, "while 41 per cent of U.S. cancer patients survive for 5 years after the discovery of cancer, only 30 per cent of black Americans survive cancer for 5 years after diagnosis." She provides a deep insight into the potential threats from "asbestos, chemical dumps, water, lead, and electrical transformers" on the life of minorities and the poor (1982, 85). In terms of infrastructure and facilities, the ghettos of the cities are often deprived of sufficient public services (e.g., educational, sanitation, transport, health services), but also voluntary organizations. Cottingham (1982a, 195) mentions that "government services are reduced first in the poorest inner-city neighborhoods, whose populations are less effectively and politically organized to protest such actions". Moreover, Cottingham has also drawn attention to the impact of industry and technology on underclass people's fragilized family and personal lives. Clark (1965), Clark and Nathan (1982)

[59] In addition to his book, the journalist, Ken Auletta, wrote previously three articles for the magazine New Yorker in 1981. Gans (1995, 33) highlights that four years after the famous underclass cover in Time Magazine, Auletta's works were strengthened the use of the term underclass as a specific label.
[60] Case study of the Wildcat Skills Training Center in New York.

highlighted in a similar way the situation of ghettos in the U.S, adding that underclass people have also limited mobility. Clark and Nathan (1982, 50) express mobility problem as "…. People in the urban underclass are not geographically and occupationally mobile in a manner that enable them to respond to new opportunities…". Based on their quantitative studies in the 1980s, they also agreed that because of disadvantaged and vicious living conditions, black families' poverty problem has become an intergenerational issue. Ricketts and Sawhill (1988) also used a quantitative approach based on the 1980 U.S. census to describe underclass in American society. They came up with the following definition of underclass: "people whose behavior departs from these norms [attending school, delaying parenthood, adult men and women in regular work, and being law-abiding] and in the process creates significant social costs" (316-25), as a sort of deviance.

In particular, in the 1980s, William Julius Wilson's works oriented underclass studies in a new direction compared to Myrdal's social welfare focus, but also, he pioneered urban poverty studies. He introduced the term *ghetto poor* (1991, 6), which can be considered as a type of (spatial) poverty suffered by disadvantaged people living in ghettos. After Auletta's prominent work (1983 [1982]), Wilson's *The truly disadvantaged* (1987) is generally recognized as the first book analyzing delicately and systematically the American underclass population living in ghettos, notably why the underclass population was growing and why this was happening. He introduced the concept of *weak labour-force attachment* (1987, 12-13). Regarding him, underclass problem has emerged particularly in ghetto neighbors of large cities (metropolitan centers) because the poor and vulnerable people living in these areas were not possessing the necessary skills to work in new financial, services sectors and suburban manufacturing jobs which were appeared in the 1970s. In fact, at that time, three phenomena were happening: a massive migration of poor black people from South to Northern cities, middle and working-class families were started to move from inner-city neighborhoods to suburbs, and the American economy was restructuring and educational requirements for employment were taking new forms. This situation, in particular, created a mismatch between the job skills of migrated black people and the demand of northern employers. It became very difficult for southern blacks to find new employment opportunities in Northern cities (see also Kasarda 1985, 1989). Dismissed from the mainstream society, low skilled inner-city working class (especially the blacks) stayed alone in their ghettos, and they have massively lost their secure manufacturing jobs. Affected by high unemployment and undereducation, people in ghettos found themselves disadvantaged and isolated from the rest of society, accompanied by joblessness, dysfunctional institutions, increasing number of crimes, and teen pregnancy (see Tienda and Stier 1991, Duncan and Hoffman 1991, Anderson 1991 for some case studies). Wilson mentions that "the central predicament of inner-city ghetto resident's joblessness reinforced by a growing social isolation in impoverished neighborhoods, as reflected, for example, in the rapidly decreasing access to job information network systems" (1991, 9). In same line with Wilson, Massey and Denton (1993, 2) highlighted the strong impact of segregation on persistent poverty which opened a road to "disadvantaged environment that become progressively isolated-geographically, socially, and economically-from the rest of society." As a result, poverty was dramatically increased in inner-city neighborhoods such as above 40 per cent, especially in large older industrial cities with high segregation of Hispanic and black population.

Wilson has also made clear that his approach was different than Lewis's culture of poverty. Wilson argues that Lewis could not distinguish concisely structural poverty and the way that poor people try to cope with poverty, vulnerable living conditions, and consistent hopeless for

their future in which they find themselves. Wilson emphasizes that as poor people find themselves excluded from the rest of society, they develop their own social milieu, behaviors and culture as a reaction to life, to society. For him, this phenomenon was not a culture of poverty, but it was a sort of social isolation. He mentions that "the underclass exists mainly because of large scale and harmful changes in the labour market and its resulting spatial concentration as well as the isolation of such areas from the more affluent parts of the black community" (1991, 5). In this sense, we can also think that on one hand social isolation is the way that the disadvantaged underclass people show themselves and prove that they exist in society, on the other hand it can be seen as a part of structural poverty. Poor living conditions may lead to fragilized emotions, instability, social isolation, and violent behaviors. For instance, he mentions that "a social context that includes poor schools, inadequate job information networks, and a lack of legitimate employment opportunities not only gives rise to weak labour-force attachment but increases the probability that individuals will be constrained to seek income derived from illegal or deviant activities" (1991, 10). The poor people in ghettos are clearly in disadvantaged position regarding labour market and unfortunately their social environment in ghettos and the view of others towards their social milieu and their behaviors make worse their vulnerable living conditions. For Wilson (1987, 145-146), therefore, the behavior of the underclass should be seen as "a symptom of class and racial inequality".

While the concept of underclass was preoccupying in the U.S. debates, on the other side of the Atlantic, in Britain the underclass was less discussed despite the terms such as 'the lumpen', 'the rabble', 'the vagrants', or 'the dangerous' were often used for mentioning some poor and notably potential threat population to the British society (see Wilson 2006, 104). But, notably from the 1980s, the underclass started to be explicitly used in British poverty studies. In 1989, The Sunday Times invited Murray to England whether this country was producing an American type of underclass. Following this one, Murray expressed his opinions in 1990 in his book *The Emerging British Underclass* by evoking the existence of similar characteristics of underclass between British and American societies[61], such as unemployment, out-of-wedlock births, and crime. Later in 1993, Murray is invited again by the Sunday Times to England to observe if there were any changes since 1989 in British society on this issue. Murray combined his observations from 1989 and 1993 in one book titled *Underclass: The Crisis Deepens* published by the Sunday Times in 1994. As symptoms of underclass, Murray has focused on three subjects: crime, illegitimacy, and economic inactivity among working-aged men. Concerning crime, based on data, he mentions that the property crime more likely occurs in England compared to U.S. For instance, "the risk of being burgled in England is more than twice that in the U.S." (1994, 2). But the violent crime rate remains still higher in the U.S. When it comes to illegitimacy, the figures are quite striking. Murray mentions that in 1987, 23.3. per cent of English births were illegitimate; and by 1992, the figure had increased to 31.2 per cent. And, for economic inactivity, among the working-aged men (16-64 yrs.), based on 1991 census, the unemployment rate was 11.0 per cent, and the economically inactive rate reached 13.3 per cent.

[61] As Welshman (2006, 162) mentions Murray convinced that Mayhew's (2008, [1852]) *undeserving poor* theory was in fact a sort of underclass. In underclass framework, the term 'underserving poor' was related to two factors: race, and deviant behavior. This means that in American society, some people due to their race and to their eccentrical, or their opposite behaviors to mainstream/core society norms, are considered poor (see also Gans 1990).

Murray qualified this trend alarming. In addition, he also expressed his worries about the future of lower-classes families. He predicted that "society would be split between the New Victorian and the New Rabble. The first would edge back towards traditional morality, while the second would come to resemble the American underclass" (cited in Welshman 2006, 165; see Murray 1994, 15).

Similar to Myrdal, Richard Titmuss (1965, 363) has argued that technological change was the primary cause of the emergence of a certain underclass in British society, and he described fragilized underclass as "the solidifying of a permanent underclass of deprived citizens, uneducated, unattached and alternating between apathetic resignation and frustrated violence." As we have seen in the U.S. debates, underclass was more related to racial segregation, but in British literature, underclass is used to express ethnic migrant workers' disadvantaged situation. From employment perspective, Giddens (1973) mentioned that migrant workers were having less paid jobs such as unskilled manual workers, but also the rate of unemployment was high, and the education level was low among migrants. Welshman (2006, 160) states that Townsend (1979) reflected some forms of underclass in his famous poverty survey in Britain, such as the situation of disabled, chronic sick, long-term unemployed people, and one-parent families, especially regarding their employment and wage conditions. It was found that they were generally less paid and living in difficult conditions. Similar to Townsend, Gallie (1988, 1994) has found that the British economy was generating secondary sector jobs characterized by poor wages, and short-term contract. Another idea of underclass appeared in Britain was Dahrendorf's (1987) vision which was emphasized the family breakdown besides unemployment, low wages, under education, poor living standards of the underclass. Besides all different approaches to underclass, in Britain and in European Union, from mid-1990s, rather the term social exclusion has started to be used by researchers in poverty studies. Similar to Britain, in France as well, the similar case of underclass was appeared in industrial and urban areas. Interestingly, Westergaard and Resler (1975, 356) argued that despite ethnic minorities were having some difficulties in British labour market, in Germany and Switzerland the situation was seeming more precarious (e.g., less paid jobs) for migrant labourers.

As we can see, among several authors, it was generally agreed that (i) welfare dependency might be bad for the poor and disadvantaged people if its impact is not deeply analyzed; (ii) welfare programs may produce lawlessness, discouraged, fragilized, socially excluded people (e.g., drug users, out-of-wedlock births, unemployed, inactive, welfare dependent people) becoming threat and deviance in the eyes of others in their society. Studies have showed that from many perspectives the underclass suffer from hardships and they do not have access to better living conditions (such as good education, good environment), repeating from one generation to another in a vicious life circle. However, it is interesting to note that among underclass studies, generally, being poor is not clearly and concisely defined. Poverty was often based on the U.S official monetary poverty line, or it is calculated in terms of being disadvantaged in a broad sense, or even it was not explicitly mentioned. We witness mainly four issues: (i) how poverty can be transformed into a persistent and intergenerational underclass problem; (ii) existence of specific vulnerable groups suffering from misery, hopeless, hard living conditions, racial/ethnical segregation; (iii) a strong need of well-designed evaluated programs of social welfare programs- because welfare programs may result in harmful impacts on the poor and disadvantaged people's life in long term; (iv) poverty conceptualization should also include social exclusion, social isolation, family aspects. The underclass approach opened room to further research such as urban poverty, ghetto poverty (see

Jargowsky and Bane 1991), vulnerable disadvantaged groups, and stimulated discussions for finding solutions to combat poverty without relying only on social welfare programs.

Working poor

The concept of *working poor* mainly refers to active labourers who have a low wage and poor living conditions at the same time despite the fact that they are selling their workforce in the labour market. This issue is more and more elaborated by governments and organisations, notably by the European Union with their 'make work pay' policies and the promotion of full-employment policies in the context of the European Employment Strategy and the Europe 2020 mid-term strategy of Europe. To be considered as working poor, literally there are two main conditions to simultaneously fulfil: (i) being employed; and (ii) living with poor standards (e.g., living under monetary or multidimensional poverty line). When these two primarily necessary conditions are satisfied, based on different criteria such as low wage, less working hours, different working poor groups can be identified. Among different sorts of criteria, low wage is the most frequently used one (Andreß and Lohmann (eds.) 2008, Nolan and Marx 2000, Pena-Casas and Latta 2004, Strengmann-Kuhn 2002). Working poor can be considered as a specific form of poverty or as a part of decomposed analysis of overall poverty once the entire poor population is identified. One thing is clear, working poor do not cover the whole poverty problem but it is a more specific issue instead.

Various type of working poor studies have been done by researchers with different research axes such as investigating working conditions or/and poor living standards of badly paid workers, analysing the link between specific welfare programmes (e.g., child support) and the workers living under the poverty line, the impact evaluation of welfare regimes and the situation of working poor in a country, the assessment of anti-working poor social policies[62] (i.e., mainly there are three approaches: minimum wages, social transfers, pro-employment programmes), different types of working poor by different topics such as household size, household composition, gender, age interval, region, the working poor analysis at the dynamic level focusing on their entry and their exit from labour market over a given period, investigating labour market conditions (i.e., labour market protections, labour costs, social security contributions) with the national economic situation such as the impact of economic shocks or different types of economic problems on labour market and working poor, international comparisons of working poor and its specific forms across countries. Regarding statistical information on working poor, in general, static (cross-sectional) descriptive statistics on (i) incidence of poverty or at risk of poverty, *working poor* (or called *in-work poverty*) by gender, country, different mechanisms of working poor such as the proportion of working poor in total working population, or the proportion of low paid/wage workers among the total working poor population; (ii) distribution of working poor by different variables such as age categories, educational attainment, level of annual earnings, full-time vs. part-time work, type of contract (i.e., permanent vs. temporary), status in employment, occupational group, industrial activity, household composition[63], number of earners in household; (iii) multivariate regression analysis

[62] According to Crettaz (2011, 104), the most efficient anti-working poverty social policies are: minimum wages, tax credits for workers, cash transfers towards families, provision and cost of childcare services.

[63] Different types of household composition such as one adult with no children, one adult with children, two adults with no children, two adults with children, three or more adults with no children, three or more adults with children.

on the probability of being in-work poverty (i.e., expressed by odds ratios) by different factors such as economic, socio-demographic items, are provided in research papers.

In order to compute the number of working poor in a society, first we need to know the definitions of work and poverty thresholds. Generally, in the literature, (i) poor living standards are indirectly measured via monetary poverty lines, compared to multidimensional poverty thresholds; (ii) there is no consensus on the definition of work (i.e., being employed) and national work definitions vary across countries; (iii) work definitions are generally based on arbitrary thresholds regarding number of hours per week, or number of months per year spent in the labour market; and (iv) informal sector is rarely captured in analysis (see Crettaz 2011, Crettaz and Bonoli 2011). Based on the works of Crettaz (2011, 22), Pena-Casas and Latta (2004), we will see some examples of work and poverty threshold definitions across countries. In European Union, regarding Eurostat regulation, a person is considered having worked if he or she employed at least 15 hours in her/his most frequent activity status in last year. As to poverty threshold, there are two definitions: (i) low-income threshold: a household is considered poor if it has less than 60% of the median equivalized household income (i.e., it corresponds to the relative monetary poverty); (ii) at risk of poverty: individuals living in a household with an equivalized disposable income below sixty percent of the median. In the U.S, for work definition, there are three different sources: the U.S. Census Bureau (USCB onwards) criteria, the U.S. Bureau of Labour Statistics (USBLS onwards) criteria, and other criterion used by scholars in American literature. On the one hand, the USCB defines work as total hours worked by family members greater than or equal to 1750 hours (44 weeks). On the other hand, the USBLS defines work as individuals who have spent at least six months (27 weeks) of the year in the labour market (working or searching for a job). In addition, researchers in general consider work as if adult works, on average, at least half time (approximately 1000 hours). Concerning poverty threshold, the USCB and USBLS use Federal poverty line criterion (absolute monetary poverty), but several researchers use less than 125%-150%-200% of Federal poverty line criterion (absolute monetary poverty line)[64].

Crettaz (2011, 22-23) investigated different types of working poor and how they differ across welfare regimes (see Esping-Andersen's typology (1990): social-democratic, liberal, conservative corporatist welfare regimes, Bonoli 1997). He used four categories of welfare regime (2011, 68), namely, *liberal*, *conservative corporatist*, *social-democratic*, *Mediterranean*, and he argues that (2011, 25) "there is no such thing as 'the working poor', but different groups of poor workers who experience very different situations", and "this social problem can take different shapes in different countries" (see also Crettaz and Bonoli 2011). He distinguishes (2011, 192-193) various groups within working poor based on three working poor mechanisms: low labour force attachment, low earnings per unit of time, and high needs due to household size and composition (notably the number of children per working-age adult). Crettaz[65] explains these three mechanisms (2011, 36, see also Crettaz and Bonoli 2011, 48-49),

[64] Moreover, in Australia, the Social Policy Research Centre considers all active individuals as worked regardless of the number of hours worked, and as to poverty line they use Henderson absolute poverty line (absolute monetary poverty line).

[65] One of the interesting technical points in his analysis is: instead of arbitrary work thresholds, he used International Labour Organization's (ILO) employment definition that is *those who work at least one hour during a reference week are considered employed*. Hence, this larger definition of employment allows to take in consideration (2011, 22) "various groups of disadvantaged workers who are in different

which according to him increase the likelihood of being working poor : low labour force attachment covers underemployed (i.e., individuals who would like to work more), intermittent workers and people who cannot or who do not want to work more hours because of child or children in their household; low earnings per unit corresponds to being badly paid corresponding to two-thirds of the median wage or half median earnings; the concept of high needs is related to the situation of large number of dependent children in the household, or breakup or divorce situations as single parents. He examined how these working poor mechanisms have a bearing on working households in terms of age, gender, education, household type, and immigrant or minority status. Some of his results include that (2011, 194) youth people are more likely being unemployed or badly paid, most young parents are more likely exposed to working poor because of child/children's needs (e.g., especially single parent households), women are more likely to be badly paid than men, low skilled workers are more likely to be low paid and unemployed.

As for the link between welfare regime and working poor, Crettaz (2011) shows the importance of each of these three mechanisms mentioned above regarding different types of welfare regimes. Due to that, according to him, there are two results: (i) there is no one type of working poor; and (ii) the structure of working poor population varies across countries. To demonstrate his ideas, he provides some statistics of the year 2000, from US (in liberal cluster), Spain (in Southern European (Mediterranean) cluster), Germany (in conservative corporatist cluster), and Sweden (in social-democratic cluster) based on different data sources. For instance, concerning low-wage rate, the per centage of low paid workers[66] reaches 24.7 in U.S, 16.2 in Spain, 12.9 in Germany, and 6.1 in Sweden. Notably, the proportion of low-paid workers (i.e., "share of workers with full-time year-round earnings below 50 per cent of the median" (2011, 173)) among the working poor[67] is more than fifty per cent in Germany and Sweden, around fifty per cent in US, and a little higher than twenty per cent in Spain. Concerning low labour force attachment measured in terms of "the ratio of the volume of work performed by the head of household and his or her spouse (if any) to their full work potential, namely a full-time job for each partner." (2011, 174)[68], the mean and median labour force attachment is found superior for the non-poor compared to poor working class in three countries except Sweden where working poor class people work the most. However, in US, the median labour force attachment is around seventy per cent, and in Germany and in Spain it is around sixty per cent[69]. Concerning high number of children relative to the number of working-age adults, in terms of mean child-to-adult ratio[70], the highest level appears in US with 0.8 for working poor, a slightly higher than 0.5 for Spain, and for Germany and Sweden the number is quite near (almost 0.4 for Germany, between 0.3 and 0.4 for Sweden). Crettaz (2011, 196) mentions that all three mechanisms play an important role in US labour market with labour force participation rate is lower compared to

situations that require different policy interventions,…., so that the no poor adult who participates in the labor market is left out of the anlaysis, whatever his or her degree of labour force attachment."

[66] (2011, 156). The source is OECD website, labour statistics and country statistical profiles, as of June 2009.

[67] (2011, 173). Source is author's own calculations based on Luxembourg Income Study (LIS)- 2000 for US, Germany, Spain, and 1995 for Sweden.

[68] Source is author's own calculations based on Luxembourg Income Study (LIS)- 2000 for US, Germany, Spain, and 1995 for Sweden.

[69] Crettaz (2011, 175) mentions that « Sweden and the US are the countries in which low-income active persons work the most."

[70] (2011, 175). Source is author's own calculations based on 2000 Luxembourg Income Study (LIS).

low-income rate and the child-to-adult ratio. In Germany, low labour force participation and low income, in Spain, the child-to-adult ratio and low labour force participation, and in Sweden, low income appear to be the most important factors to be working poor. Based on his results (for more results, see his book Crettaz 2011), he suggests that each country must find its own combination of policies for better social policies for abolishing working poor.

Furthermore, based on his cluster analysis of welfare regimes, Crettaz (2011, 196-197) puts in evidence that for combatting working poor, the liberal cluster focuses on a 'make work pay' strategy (MWP) by establishing minimum wages and tax credits for workers; social-democratic cluster aims at keeping high employment rates (with low unemployment rate) and keeping low income and earnings inequality; conservative corporatist cluster puts priority on keeping high the quality of jobs in the market with employment protection regulations and collective bargaining; and southern European cluster is characterized by social transfers, income transfers for the people who are not in the labour force, and strict regulation of labour market. He showed that for reducing working poor, "the minimum wages, social transfers, and employment maximizing strategies" (2011, 200) are mainly used. Related to that, the minimum wages strategy can work out for combatting low pay, the employment maximizing strategy for combatting low labour force attachment, and social transfers for combatting low pay, low labour force attachment and also providing support to family needs related to child (e.g., supplement benefits, family cash benefits, incentives to work).

In line with Crettaz's working poor concept, Crettaz and Bonoli (2011, 48) sheds light on three mechanisms leading to working poverty (i.e., low hourly earnings, low labour force attachment, a large number of dependent children in the household). They elaborate these three mechanisms with an application to OECD countries. For computing low earnings, they take in consideration "the number of hours usually worked and the number of weeks spent in the labour market over the income reference period." These two factors are used for adjusting annual earnings of people working part time or did not work regularly over the reference period, to obtain a hypothetical "full-time year-round earnings." Concerning low labour force attachment, it is obtained by "dividing the amount of work performed by household members in full-time equivalents by the number of adults[71]."- it covers head of household and their spouse (if any) aged 18-65. And, concerning a large number of dependent children in the household, they "divide the number of children by the number of working-age adults[72]." As in the work of Crettaz (2001), Crettaz and Bonoli (2011) mention also that (i) a working poor will have at least one of the three mechanisms mentioned above, workers who do not correspond to any of these three mechanisms are more unlikely to be in risk of poverty; (ii) none of these mechanisms has to lead to working poor status; and (iii) when a person faces more than one mechanism, the likelihood of being working poor increase. By the comparative analysis, they mainly show that (2011, 66) household size and composition are the most important factors to working poor status. For that reason, authors recommend improving family policies such as family cash benefits, parental leave schemes, childcare services.

In a comparative perspective, Goerne (2011) focuses on Europe (23 countries in consideration) and analyses the in-work-poverty (IWP onwards) patterns with the cross-national differences

[71] Crettaz and Bonoli (2011, 50) yield an example for this: "a couple with children where one parent works full time and the other one has a 50 per cent job will have a labour force attachment of 0.75 (1.5 full-time equivalents/2 adults = 0.75).

[72] For instance, the ratio for a family of four (two parents and two children) is 2/2=1.

by using the 2007 wave of the European statistics on income and living conditions (EU-SILC). He found three mechanisms of IWP (2011, 15): "job quality (remuneration levels), job quantity (labour market participation of household members) and household expenditure (cost of dependents)." Concerning the basis of IWP, classical European monetary poverty line (i.e., sixty per cent of median equivalized household annual income adjusted by OECD equivalence scale) to measure poverty, and the most frequent activity status[73] as 'working' to measure active people in labour market are used. As Crettaz (2011), Goerne (2011, 22-23) points out as well three clusters: Nordic, Southern European, liberal and conservative-continental welfare regimes. In general, as studies show (see Lohmann and Marx 2008), Nordic countries are known for high participation of women in labour market, low wage inequalities, and various welfare programmes, for that reason, author mentions that it is expected to have none of the three IWP mechanisms in Nordic countries. In Southern European counties, for instance in Spain, Greece, and Italy, the situation is quite opposite such as high wage inequalities, low female labour participation, and less welfare and public programmes (e.g., childcare). However, Goerne mentions that the IWP mechanisms are more complex in liberal, and conservative-continental welfare regimes[74]. The results show that the three Nordic countries Denmark, Finland, and Sweden have the best and the four Mediterranean countries Greece, Spain, Italy and Portugal have the worst performance regarding IWP. However, among the continental regimes, much more variation occurs for instance, Belgium, Netherlands, Austria, France and Germany are in the middle or lower half of the distribution. The two liberal countries, UK and Ireland, are found in the middle of the series. And, interestingly, there is no unique tendency amongst Eastern European countries such as Czech Republic has the lowest level of IWP in opposition to Poland that is at the top end. In addition, Goerne (2011, 29) searched for the determinants of being working poor based on individual, household, and employment variables amongst 5 countries UK, Spain, France, Sweden, and Poland. He finds that low wage is the primary problem, but also workers in single earner households (especially in the UK), and workers in households with many dependents (Especially in Poland and Spain) have a higher risk of being working poor. Furthermore, single parents are also one of the highest risk group of IWP in all countries.

In the existing literature, there are also working poor studies focusing on the situation of specific groups and disparities such as in terms of gender, immigrants. For instance, Alvarez-Mirada (2011), by using the 2007 wave of Statistics on income and living conditions (SILC) survey of European Union, explore the IWP and poverty situations of immigrants in the European Union (EU). As a result of her analysis, she comes up with the typical immigrant profile[75] in the EU, and she found that there are three best predictors of IWP among employees living in the EU (but born outside the EU): "low wages, part-time work, and incomplete labour force participation of household members." (2011, 272). Concerning gender issue, Pena-Casas and

[73] It is defined as "working for more than 6 of the past 12 months." (2011, 17).

[74] Goerne explains that (2011, 20) in UK low-wage employment occurs relatively high but the female labour force participation is high as well. In conservative-continental countries, the social welfare programmes such as public spending is high but there are high amount of wage inequalities and low wages, and low female labour participation.

[75] "This is a man who is middle aged, who did not go further than secondary education in his school years and does not hold citizenship of the country of residence; who lives in a single person household or, if in a larger household, has dependent children and not all members of working age are regularly in work; who earns a relatively low wage at a job of low occupational status, in a small workplace, for less than 30 hours of work per week." (2011, 272).

Ghailani (2011, 202-204) mention that, despite Eurostat's findings show no any striking differences between women and men regarding IWP and in the risk of poverty (i.e., regarding own calculation of the authors based on EU-SILC 2007, the gender ratio of poverty is 1.1 for Spain, 1.1. for France, 0.9 for Poland, 0.9 for Sweden, 1.1. for UK, and the gender ratio of IWP is 0.7 for Spain, 0.9 for France, 0.8 for Poland, 0.8 for Sweden, and 1.0 for UK[76]), there are considerable disadvantaged situations that the women face in the labour market. These negative situations for women can be summarized as having low access to high quality job and opportunities (Ghailani 2009, OECD 2008), being less paid than men (OECD 2008), doing part-time jobs (Bettio and Verashchagina 2009), struggling more as to balancing work and family life in terms of time-use, and resulting in low labour force participation (Lewis 2009, Lewis et al. 2008, Plantenga and Remery 2005), being one of the vulnerable groups as lone parents (Brady and Kall 2007, Fagan et al. 2006) with a high risk of poverty, and being dependent on financial support of social welfare programmes, or on their spouses (Pasqua 2008).

In opposition to paradox results of Eurostat on IWP, Pena-Casas and Ghailani (2011, 206) proposes another IWP measurement framework based on individual net incomes instead of basing only on household level income. They call this process "individualization of in-work poverty risks[77]", and they suggest seeing the results of the process as "'latent poverty', reflecting the potential income situation of individuals in case of rupture of their household attachments." We can say that this method puts in evidence the vulnerability of women regarding their personal income. Authors compute a net individualized disposable income based on net *individual* income components deduced from taxes ("earnings, income from self-employment, private pension schemes, or social protection benefits such as unemployment, pensions, survivor, sickness/disability, education, or minimum income guarantee" (2011, 207)) and net *common* income components[78] deduced from taxes ("revenues from property or renting, capital, inter-households transfers, family allowances and housing benefits" (2011, 207)). Later, based on the individualized income, authors compute IWP at the individual level (by using 2007 EU-SILC Spain, France, Poland, Sweden, UK). Authors find that women face two to three times higher IWP, and various precarious situations in labour market compared to men. Especially, women have to more cope with time-use between family and work life, they are in general work in lower quality jobs, and they are earning less than male partner in their household. Authors suggest also to focus on longitudinal data analysis encompassing individual trajectories, which will put in evidence potential vulnerability of individuals to external shocks, and alterations in their household life. This dynamic perspective will help to uncover and examine various types of precariousness[79] (see Paugam 2000, Wresinski 1987), the extent of

[76] For all calculations, authors used all adults of working age as reference population (2011, 203).

[77] Authors suggest looking at Meulders and O'Dorchai (2010) and Meulders et al. (2009) because their methodology is based on these two works.

[78] Family allowances and housing benefits are equally divided among household's adults, and some of other common income sources are imputed based on the rules of taxation (2011, 207).

[79] The term precariousness (in French, 'précarité') comes from the French literature, which is used for indicating various worst conditions of employment and being vulnerable due to labour market conditions. The Wresinski (1987) Report defines precariousness as "the absence of one of more forms of security, notably that of employment, which allow people and families to meet their professional, familial and social obligations and to enjoy their fundamental rights. The insecurity which results may be more or less extended and can have consequences of varying severity and nature. It leads to serious poverty when it affects several aspects of existence, when it becomes persistent, and when it

poverty or in risk of poverty, and transitions between different positions such as prosperity, in risk of poverty, IWP over time (see European Union's '*flexicurity* approach').

In the context of dynamic analysis of IWP, Gutiérrez et al. (2011, 175) focus on IWP and examine poverty dynamics in European Union. As we have seen through previous examples, cross-sectional data analysis gives information at macro level on a country situation at one point of time. However, dynamic analysis provides information at micro level such as move out or move into IWP by following individuals through years in a longitudinal data set[80]. As the authors mention, in general, poverty dynamics analysis focuses on mobility and persistence (or called chronic) of poverty. On the one hand, the term *mobility* is used for move-in and move-out poverty (i.e., transitions between poverty and non-poverty), on the other hand the term *persistence* is used for mentioning staying in-poverty situations over years. There are two main basic statistics of *poverty mobility*: the entry rate vs. exit rate. The entry rates correspond to the per centage of people who were not poor at a reference year, but they become poor the following year. The exit rates correspond to the per centage of people who were poor at a reference year, but they become non-poor the following year. The authors show the transition situations (i.e., in and out of IWP) by using EU-SILC 2005-2007 data sets for Spain, France, Poland and UK. The year 2005 is taken as the reference year and the working poor in 2005 are followed in the following two years. Three main types of movements are found (Gutiérrez et al. 2011, 186): "one of persistence of IWP; one of exit; and one of recurrence[81]." They also found that the variables occupational group and status in employment have a more impact on entry to IWP compared to household composition and social transfers variables.

In American context, Charles Murray (1987, 3) sheds light on the issue of working poor through the question "Can any American who is willing to work hard make a decent living?" He mentions that since the 1960s the answer to that question has become *yes* due to some dysfunctions in economy causing barriers and leading some people living in structural poverty. Murray speaks about the dynamic of poverty in U.S. based on Duncan et al. (1984)'s poverty studies called *Years of Poverty, Years of Plenty*, which employed Panel Study of Income Dynamics (PSID onwards). PSID is a longitudinal data base started to be collected in 1968 with a sample of around 5,000 interviewed American families. They found that in America mostly poverty is not a chronic issue. Instead, it is a kind of temporary (short term) poverty characterized by move in and move out of poverty. But also, many people are found vulnerable to economic shocks leading them to poverty. Murray was curious about the results; then, he also used PSID for his own computations of working poor. His main research question was "Given that a person is poor but in a household that is participating in the labor market, what happens over the next ten years?" (1987, 6, and see his article for more information on technical

compromises the opportunity to reassume responsibilities and to reclaim rights by oneself, in the foreseeable future" (Gordon and Spicker (eds.) 1999, 107).

[80] For instance, based on Eurostat's data base, authors show that IWP risk rate in per centage was 11 for Spain in 1997, 7 for France in 1997, 11 for Poland in 2000, and 5 for UK in 1997. For Spain, France, and Poland these numbers stayed approximately same in ten years, but for UK the rate is jumped to 9 in 2008 (Gutiérrez et al. 2011, 177).

[81] One of persistence of IWP means they remained in the same IWP situation. One of exit of IWP has three forms: (i) whether they continue to work, and they escape of poverty, (ii) whether they are no longer employed, and they are still living under poverty line; or (iii) whether they are no longer employed but they are no longer poor as well. One of recurrence means they moved out IWP in 2006 but they move into IWP in 2007 (see 2011, 186).

issues regarding his computation). Murray found different results compared to Duncan et al. He found that in 1970, 95 per cent of working age adult population were not poor. Among the 5 per cent, 2.7 per cent are qualified as living in working poor households. His conclusion was «it is extremely rare for a person to get into the labor market, stick with it, and remain poor. It is quite common for the rare person who is both working and poor not only to escape poverty but to reach the median income range or higher within a few years and remain there securely" (1987, 17).

In working poor sub-section, we are specifically focused on living conditions of economically active people, with an emphasize on their degrading social and working conditions. In next sub-section we will see the vulnerability approach, which is generally examined in the literature by employing sophisticated forecasting methods in order to predict the risk of being exposed to negative hazards such as disasters, crisis, the risk of falling into poverty or the probability of moving out of poverty. Thus, vulnerability approach is linked with dynamic analysis taking in consideration time aspect.

Vulnerability, poverty dynamics, vulnerability to poverty

The World Development Report 2000/2001 (World Bank 2000, 139) mentions vulnerability as a measure of "the resilience against a shock-the likelihood that a shock will result in a decline in well-being", where the resilience can be seen as "the manifestation of positive adaptation despite significant life adversity" (Luthar 2003, xxix), or as the extent to which a person (or a household or a family, and so on) resists against harmful experiences[82]. In another perspective, Ranci (2010, 16) defines vulnerability as "a high degree of exposure to serious damage". Thus, vulnerability indicates us how units such as a household, individual, community, region, country, or a group of people are ready for coping with a possible damage (e.g., drought, increasing food prices, illness, natural disasters), or in other words, how they are in disadvantaged position regarding the possible damage given their resources and capabilities. Chambers (2006, see also Chambers 1989) distinguishes two sides of vulnerability: external vs. internal. External side includes risks, shocks that faced by households, individuals, groups of people. Internal side is related to incapacity or capacity deficit of people to cope with damages.

Poverty is more related to deprivation, or capabilities deficit; however, vulnerability deals with insecurity, shocks, stress, risks, and how to deal with these factors in living over. Vulnerability does not primarily aim at describing the evolution of poverty over time, or of other negative social phenomenon such as social exclusion; instead, it is a more specific concern focusing on being exposure to a hazard (i.e., probability that the damage happens). In this sense, vulnerability corresponds rather to a risk analysis such as the risk of falling into poverty or of the deterioration of well-being. Castel (1995a) mentions that the probability of a hazard happens do not necessarily depends on social class, but it depends on various situations in which the person finds him or herself. Castel especially has focused on precarious living conditions of wage-workers since the 1970s, in France. He describes the vulnerable situation of wage-workers through two main axes: labour market, and social bonds. Regarding labour market mechanism, fragile wage-workers have generally short-term contracts, or worthless, small jobs.

[82] The World Bank's Social Protection strategy has recommended the Social Risk Management (SRM) to analyze vulnerable households to poverty (see Alwang et al. 2001, which includes a comprehensive literature of vulnerability from different disciplines).

In this group, Castel also integrates unemployed and notably long-term unemployed people. In terms of social bonds, he mentions degradation of family relationship.

According to Klasen and Povel (2013, 17-18), vulnerability has three main features: first, it does not affect only economic power (i.e., income, consumption patterns, purchasing power) but affects also well-being in a negative sense. Vulnerability renders more fragile the person against adverse shocks, for instance, becoming not able to "prevent, mitigate, and cope with these shocks." Second, vulnerability can be "a cause of long-term deprivation". In this regard, vulnerability plays an important role in reducing long-term poverty. Third, "vulnerability shifts the attention from ex-post poverty outcomes to ex-ante poverty risks by trying to identify the (types of) households with highest future poverty risks." Households below the poverty line are vulnerable to different negative shocks such as chronic poverty, however, it does not mean that all households above the poverty line can never be vulnerable to poverty. It is very important to investigate upward poverty tendencies for households below the poverty line.

Vulnerability studies can be realized at different level/unit. For instance, as we can see from the vulnerability definition of Klasen and Povel (2013) above, they worked at household level in their article. Household vulnerability studies (see Baffoe and Matsuda 2017, Notenbaert et al. 2013), generally, deals with the position of household with respect to hazard, adverse events, and their capabilities of coping with these negative situations, hardships (i.e., household's risk profile or risk attitudes). Some studies are done at country level scoping economic vulnerability of country (see Briguglio 1995, 1997; Briguglio and Galea 2003, Briguglio et al. 2009), or its environmental vulnerability (e.g., disaster management) against earthquakes, typhoons, hurricanes, droughts, floods, climate change, and so on (see Guillaumont 2009). Furthermore, some other studies are realized at regional, or community level for instance, local vulnerability index in South Africa (Naudé et al. 2008), geographic poverty traps (Jalan and Ravallion 2002), community resilience in Mississippi in terms of economic development and social capital[83] (Sherrieb et al. 2010), community resilience and social cohesion in rural Canada against disasters[84] (Townshend et al. 2015).

Although research topics of interest in vulnerability can be quite wide given there are various types of adverse events and shocks can be happening, the literature is relatively focused on two issues: health, and weather shocks (Klasen and Povel 2013, 19). Concerning health shocks, studies of Gertler and Gruber (2002), Skoufias and Quisumbing (2005), Beegle et al. (2006), Grimm (2010), and concerning weather shocks, with a particular attention to droughts and floods, studies of Fafchamps et al. (1998), Dercon an Krishnan (2000), and Kazianga and Udry (2006) give insights to different aspects of vulnerability. In the literature (see Klasen and Povel 2013, 20), health shocks-except the epidemics, are called *idiosyncratic* events, and weather shocks as disasters for instance are known as *covariate* shocks. Generally, weather shocks concern a group of people such as community, country, region; nevertheless, health shocks are quite known as individual events. In empirical perspective, OECD's 'Pressure-State-Response' (P-S-R) and UNDP's 'Human vulnerability and Risk Index' aim to measure people's environmental vulnerability (see UNDP 2004). On the one hand, the P-S-R model defines factors causing environmental problem (*Pressure*), variables presenting the actual situation of

[83] See Norris et al. (2008) for the model of adaptive capacities for economic development and social capital, and the social vulnerability index- SOVI (Cutter et al. 2003).

[84] Buckner's Index of Cohesion (Buckner 1988) and the Index of Perceived Community Resilience (Kulig et al. 2013) are used.

environment (*State*), and indicators of how people cope with the environmental problems (*Response*). On the other hand, UNDP's index computes the mortality rate due to a natural disaster.

Recent years, as in the case of social exclusion, the question of social vulnerability has also won an important place in poverty and well-being related debates in Europe. Ranci (2010) explains delicately the nature, tendency and extent of social vulnerability in Europe. We will shortly see some of his arguments in next two paragraphs. He starts by emphasizing the actual persistence of social insecurity and instability in Europe, particularly in western Europe societies, despite the social welfare programs (i.e., social protection) occurred from the second half of the twentieth century. To explain this new wave, he mentions Esping-Andersen's (1999) concept of the failure of '*work, family, welfare'* institutions for providing social security and well-being to citizens, and Castel's (1995b) '*progressive erosion'* arguments on the emergence of new social order and fragilization of individualistic capitalist systems affecting notably children, elderly, and middle-class populations. Ranci (2010, 4-5) gives insights into the 'progressive erosion' through three main aspects. First, '*the weakening of the labour market to function as the principal mechanism of social integration*' that leads to increasingly job insecurity, long-term unemployment, and permanent labour-market exclusion. Second, '*the gradual weakening of kinship support networks*' related to new emerging household types such as single-person, and single-parent households, and to progressive decrease of traditional male breadwinner families (see Crompton 1999). Third, '*the great rigidity of welfare systems*' signaling the limited coverage of welfare systems compared to past. Current welfare allocations do not really respond to the needs of mostly disadvantaged people. In fact, current social protection systems in Europe are more likely to meet with the needs of who are 'fully integrated in the labour market'. As a result, new types of social risk have been emerging in European societies (we bear in mind that in industrial societies, the traditional social risk was mainly the loss of a permanent job).

Ranci (2010, 6) mentions four types of new social risk in Europe. First risk is the integrated poverty as developed by Paugam (1997) for describing temporarily relative poverty, which primarily causes income instability, and makes uncertain household's financial conditions. Second risk is the housing deprivation. Increasing housing costs and modifications in housing has been creating difficulties for some part of the population for getting access to dwelling, notably youth population and people living in metropolitan areas find themselves in relatively disadvantaged position in housing market. Third risk is the increasing number of insecure and temporary jobs in labour market. This situation touches especially workers with low educational attainment and unqualified skills. This type of jobs means lower wage, which makes not easy to fully participate in social activities; but also, as these jobs do not contribute in a positive way in terms of working experience and occupational skill, people performing these jobs risk also exclusion from the labour market and becoming distant from high qualified and secure jobs. For instance, youth people becoming attached to insecure and low qualified jobs might lose their autonomy and get hopeless for their future. And fourth risk is the difficulty of reconciliating work and family life, especially regarding childcare. In one way, it is important arising female participation in labour market in terms of women empowerment. But, on the other side, this implies the necessity of better social programs designed and applied by governments for amelioration of childcare services to helping notably single-mothers, and single-parents. But also, for other types of households, when two parents work at the same time, it is becoming quite complex to manage time sharing between child activities, household

obligations, and work. It may also create an effect on gender inequalities and demographic changes in society, and even push women to not be active in labour market in order to stay at home and take care of children.

Ranci and Migliavacca (2010, 219) analyzed social vulnerability from a multidimensional perspective in Europe. We will see how they conceptualized social vulnerability based on which dimensions, and indicators, and which main results they obtained. Authors used 5 dimensions that are: economic vulnerability, housing risk, work vulnerability, difficulty of reconciliating work and childcare, and presence of dependent persons in the family. Regarding economic vulnerability, they took into account the probability of falling below the income poverty threshold (i.e., monetary poverty line). Their results show that 5 per cent of households were permanently poor (i.e., being under the poverty line at least 4 times in a period of 5 years), 7.8 per cent of them faced recurrent poverty (i.e., being under the poverty line two or three times in a period of 5 years), ad 9.8 per cent experienced transitory poverty (i.e., being under the poverty line once in a period of 5 years). Regarding housing risks, authors emphasize the importance of material living bases such as home in daily life of people to "develop stable familial and social relationship". They mention that being deprived in living standards is not something rarely happens in Europe. Serious material deprivation and vulnerability vis à vis housing conditions are quite show itself; for instance, they found that 5.5 per cent of households face severe housing deprivation, and 10 per cent of them live in poor living standards or live-in overcrowded dwelling. The situation regarding work status shows also a number of types of vulnerability of households such as 9.3 per cent of households are without any active member working regularly, but also 9.1 per cent of households has an active member working in precarious[85] conditions or an unemployed, with an active member working regularly. As we can see, in that dimension, the tendency is to focus on unemployed, irregular (temporary) workers, or workers in precarious employment (see Rodgers and Rodgers 1989; Paugam 1991a, 1996, 2000; Burgess and Campbell 1998; Tucker 2002, Vosko et al. 2003, Cingolani 2005; Vosko (ed.) 2006, Fudge and Owens (eds.) 2006, ILO 2016 for 'precarious employment'). Authors also show that some households experience difficulties of reconciling their work and childcare. This issue has been on European debates for long time, notably due to considerable increase of female employment[86]. For instance, their results (2010, 223-225) show that 7.4 per cent of households is found in difficulties for balancing childcare and work responsibilities, and 49 per cent of households prefer to look after their children by themselves that means that at least one of the parents does not work. In latter case, the risk of being in poverty multiples by three times and the average per capita income reduces by 29 per cent (based on ECHP 2001 calculations). And, concerning last dimension, authors mention that households with dependent person(s) deal with both providing necessary care services to dependent family member(s) and earning necessary income for affording care services. They become concerned with public assistance services and employment conditions of the family members. Their results show that fifty per cent of the families with a dependent member do not need to reduce significantly their consumption (including holidays, inviting friends for dinner, meat or fish on a weekly basis, ability to afford new furniture), or they are not in severe financial deterioration because of care costs. And, in total, including all households with dependent members, 7.2 per cent of them experience severe material deprivation.

[85] Authors suggest also looking at the works of Edgar et al. (2002).
[86] Authors suggest see Crompton (1999), Esping-Andersen (ed.) (2002), Jacobs and Gerson (2004).

Moreover, authors particularly highlight that dependency plays a substantial role in deterioration of life standards such as the risk of falling into poverty increases by more than 200 per cent for households with a dependent member (2010, 225)[87]. Through the analysis based on these 5 dimensions, authors extended their analysis to the types of households most exposed to social vulnerability. They found that (2010, 234) especially the male breadwinner households are most likely fall into vulnerability or live in multidimensional vulnerability. In contrast, dual-income households with small children appear less vulnerable. But also, lone-parent households and single pensioners living alone (i.e., to note that households with adults of different generations are quite vulnerable regarding housing and work) are the second and third most vulnerable groups respectively. Authors also point out that a single working people and working couples without children are the less exposed to vulnerability.

After having seen the definition of vulnerability and its different types such as ecological, economic, and social; now we will focus on how is poverty linked with vulnerability? In the literature, vulnerability is related to poverty especially in respects to poverty dynamics including transient vs. chronic poverty, and poverty traps that keep people in poverty in a persistent way over their course-life (see Modigliani 1986). Here, it is worth to mention that conceptually poverty dynamics come from Rowntree's life-cycle model. After him, for long time, this issue was not priority of poverty studies. But, recently, poverty studies took a second level orienting from static to dynamic perspective. Poverty dynamics focus on the evolution of poverty status of the unit (e.g., individual, household) over time (see Bane and Ellwood 1986, Gaiha 1988, 1989, Hulme and Shepherd 2003, Budowski and Suter 2002, Cellini et al. 2008, Addison et al. (eds.) 2009a, Carter and Barrett 2006, Lawson et al. 2008). The temporal aspect of poverty becomes important notably for distinguishing short-term vs. long-term poverty. The former is generally used to mention being poor for a certain number of months; whilst the latter known as structural poverty is rather a permanent lack of resources and capabilities. Regarding time perspective, most of poverty studies deal with poverty trends focusing on the comparison of headcount ratio across the country at different times (e.g., the evolution of headcount ratio of poverty over years in the country), which is known as *conventional static poverty analysis*. But, this type of analysis does not inform on chronic poverty, move in to and/or move out of poverty status of individuals/households over time. The main argument behind this type of analysis is that, as mentioned by Addison et al. (2009b, 6), not every citizen in the country achieves to improve her/his well-being; so that, to know more about on "ups and downs of welfare status" of individuals/households, we need a dynamic perspective. This is why, in recent years, poverty dynamics studies have gained importance in the literature.

Addison et al. (2009b, 5-6) emphasize two other positive features of dynamic studies: first concerns depth, severity of poverty and disadvantaged groups; for this, they give an example: "hypothetically two different countries might have the same scores for the headcount, depth, and severity of poverty. Apparently, poverty in both of these countries is similar. However, in the first country poverty is largely transitorily and is a phenomenon that many of its population experience but only for short durations. In the other, most of the population are non-poor but a minority are trapped in poverty for most or all of their lives. In the former country policies need to help those experiencing short spells of poverty-unemployment insurance and benefits, reskilling, microcredit, temporary social safety nets, health services. In the latter, deeper structural problems must be addressed- inclusion of the poor in access to health and education

[87] Authors suggest see Grammenos (2003).

services, asset redistribution, tackling social exclusion, and regional infrastructural development." Second, they suggest that "past experiences shape the agency (motivation, preferences, and understandings) of people". But there are two main shortcomings: first is dynamic studies are often focus on monetary perspective such assets, and income (or consumption, expenditure), rather than multidimensional perspective (nevertheless, it is widely accepted that poverty is beyond monetary issue). Second is, in a large variety of citizens profiles over time, we will be needing a more synthetic examination due to size of information obtained, which in turn will induce some loss of information. Dynamic analysis appears richer in terms of providing a comprehensive information on individual/household status, but this process does not much help policy makers to create rapid and effective solutions for reducing poverty in the country.

On the one hand, *transient poverty* means facing poverty for a certain period of time, being poor from time to time. In this case, move in to and move out of poverty over time are the subject of interest (see Gaiha and Deolalikar 1993, Jalan and Ravallion 1998, Baulch and Hoddinott 2000 for examples). On the other hand, *chronic poverty* tackles persistent poverty situation- staying in poverty for a long-time period (see Hulme and Shepherd 2003; Chronic Poverty Research Centre (CPRC) 2005, 2009; Hulme and McKay (2005) for examples, see Foster 2009 for adaptation of the FGT[88] poverty measures into the measurement of chronic poverty, and see Osmani (2009) for the application of Sen's capabilities framework to measure chronic poverty), which can affect an entire life course or even transmits across generations (i.e., *inter-generational transmission of poverty*). More specifically, Haughton and Khandker (2009, 214) distinguish chronic poor from persistent poor in the sense that the former "those whose average consumption per capita over time is at or below the poverty line", whilst the latter are "those, among the chronically poor, who *never emerge from poverty*, not even for a year or two". Poverty dynamics studies might help finding solutions to move out from chronic poverty, and to prevent people to fall into chronic poverty (e.g., preventing transitorily poor before their poor living conditions become persistent over time).

Chronic Poverty Research Centre (2004, 5) proposes 5 groups: always poor, usually poor, fluctuating poor, occasionally poor, and never poor. The first two groups include in *chronically poor*, third and fourth form *transitorily poor*, and the last group correspond to *non-poor*. Regarding their schema (see Figure 2.1 in Shepherd 2013, 9), the always poor are always living under the poverty line over their life-course, the usually poor are generally situate under the poverty line but rarely they achieve to be above the poverty line, the fluctuating poor come and go between under and above the poverty line but mostly their mean score is under the poverty line, the occasionally poor are also fluctuating between under and above the poverty line but their mean score is above the poverty line (where the mean score is the mean score of poverty, which can be in terms of household expenditure, income, consumption, material deprivation,

[88] The FGT measures are created by James Foster, Joel Greer, and Erik Thorbecke in 1984 (see their article entitled '*A class of decomposable poverty measures'*). The FGT measures did not initially include time perspective. Foster (2009) included the duration of time passed in poverty into the FGT poverty measures. This new measure can be called as 'duration-adjusted FGT measure'. And he applied it to Argentinian panel data sets. But one negative point is that she is generally applied in monetary perspective such as income to measuring poverty.

nutritional threshold, or assets, or multidimensional poverty)[89]. And the never poor have always a mean score above the poverty line during their life-curse.

In the literature, some studies specifically focused on 'the occasionally poor' such as Hübinger (1996) used the term 'precarious prosperity' for people who are above the poverty line but risks at the same time falling below the poverty line or they do transit between above and below the poverty line[90]. Budowski et al. (2010, 201) qualify this group as a socio-structural category emerged from empirical research, referring to "a critical dynamic or uncertain situation (precarious) and a certain level of material well-being or income (prosperity) that allows for opportunities of agency and for setting priorities, yet lacking financial reserves to buffer unexpected shocks or invest rapidly in upcoming opportunities"[91]. It is interesting to see that in their work Budowski et al. (2010, 214) define prosperity as "a level of socially desirable material well-being". Herewith, 'precarious prosperity' covers the vulnerable people in terms of material well-being. But the term 'prosperity' has a broader meaning; it does not only include material aspects of life but does include also emotional and social aspects, which can vary over time.

Shepherd (2013, 8) yields three areas of chronic poverty: long-term poverty, life course poverty, and intergenerational poverty. In first case, poverty persists for a considerable long period of time, and notably it becomes quite impossible to move out from poverty if the external conditions (might affect the living) stay unchanged. In second case, a person (or a household or a family) stays poor in her/his entire life. And the third case corresponds to the transmission of poverty from one generation to another (i.e., from parents to children). For instance, Günther and Klasen (2009) applied a dynamic perspective to measure inter-generational multidimensional poverty including education, health, and nutrition, based on Sen's capabilities in Vietnam. They found that education plays a crucial role across generations. Children of older generations with low education have also low education. Quisumbing (2009) shows how intergenerational transfer of assets can affect the poverty status of households in a dynamic perspective. Notably, Bird (2013) provides a comprehensive picture of persistent poverty across

[89] In the literature, the term "trajectory" is generally used to describe the life course of people (e.g., or household, or family) by showing their upwards and downwards regarding poverty line.

[90] It is important to note that the term 'precarious prosperity' is different than the term 'precarious employment'. For 'precarious employment', see sub-section 'Working poor'. However, the word 'precarious' means a pragmatic situation in the sense that the person faces both some advantages and some disadvantages at the same time. For instance, in 'precarious employment' context, Paugam (1991a, 2000) emphasizes social disqualification of employed people in certain perspective. In their article, Budowski et al. (2010) link the term 'precarious' to social inequalities in comparative perspective. In practice, the term 'precariousness' is often used in labour market context such as similarly to working process-employees having precarious employment conditions including unstable work, short-term contracts (see Hübinger 1996, Portes 1985, Wresinski 1987). In a broader context, Villeneuve (1984) measured precariousness in France based on the survey of disadvantaged situations by developing 47 variables across several dimensions including labour, health, finance, discontent.

[91] Budowski et al. (2010) criticized social exclusion, underclass, and culture of poverty approaches in the sense that they lead to dualization process in society (poor vs. non-poor, excluded vs. non-excluded, marginal vs. main (core) stream). The 'precarious prosperous' concept focuses on potentially poor (i.e., being above the poverty line at a certain time but facing poverty spells over certain periods). This concept is an important step to analyze poverty degrees but still she does not avoid separating the society into the two main categories: above vs. below the poverty line. For evaluating someone as 'precarious prosperous', we still need a relative benchmark which in certain way will conduct to distinguish different levels in a society in terms of prosperity and precariousness.

generations in the U.S. This type of study aims at unveiling how poverty affect the future well-being of children, such as difficult economic conditions of household, poor neighborhood, weak educational attainment, less opportunities due to poor district (see underclass studies in the U.S, Lewis's culture of poverty, see also Boggess et al. (eds.) 1999). In other words, persistent poverty in many ways makes children vulnerable to their futures; for instance, school dropouts, child labour, health problems in childhood constraint children's ability to catch up opportunities, and as a result these negative factors translate in increasing the likelihood of living in poverty in their adulthood[92]. Moreover, the resilience discourse in vulnerability studies shows (i) how children resist to poverty, cope with hardship; (ii) the extent to which children experience vulnerability (e.g., irreversible vs. reversible poverty effects); and (iii) children's resilience capacity given their environment[93], social interactions, personality (e.g., self-esteem), cognitive development (e.g., child development trajectories), and other various psychological and economic factors. Boyden and Cooper (2009) draw attention to genetic and environmental factors as the most important ones to explain children's poverty.

Barrientos (2013, 89) recalls three main approaches to measure chronic poverty: *duration* approach, *shortfall* approach, and *predicted consumption* approach. In first approach, chronically poor people correspond to those who are persistently below the poverty line. In second approach, there are two components of income or consumption: constant vs. fluctuating. So that, for chronically poor people, the constant part of their income or consumption is permanently at or below the poverty line (and the fluctuant part corresponds to transient poverty phases). And the third approach consists of forecasting the future income or consumption level of units based on their current income or consumption. In that case, if the estimation shows that the unit will probably suffer of falling below the poverty line in her/his income or consumption; then, he or she is considered potentially chronically poor. Basically, it is considered that if for instance a household's current income or consumption is at or below the poverty line, it is highly probably that this household will face again being at or below the poverty line in future[94]. Barrientos (2013, 90) points out that among these three approaches, only the third one includes explicitly the notion of vulnerability through the forecasting process (i.e., vulnerability as expected poverty-see below for the formulas); and therefore, the first and second approaches are the ex post, but the third approach is an ex ante approach to define and measure chronic poverty.

In the literature, there are also various studies proposing different methods to measure chronic poverty. For instance, Carter and Barrett (2006) propose an asset-based approach by which they define two different groups of poor: structurally poor vs. stochastically poor. The structurally poor are in a position of not being able to overcome poverty (i.e., staying persistently under the poverty line) because of their insufficient assets. On the other side, stochastically poor are potentially capable of living above the poverty line but it happens that at the moment of the

[92] For the case of U.S, Bird (2013, 61-62) describes five theories found in the literature to explain intergenerational poverty: "the economic resources model, the family structure model, the correlated disadvantaged model, the welfare culture model, and the social isolation model."

[93] For instance, Yacub (2002) mentions that children are more likely being irreversibly affected by poverty in utero and in early infancy. Similarly, Bird (2013, 78) recalls that "there crucial periods in an individual's life-course appear to be foetal development, early childhood, and youth".

[94] Barrientos (2013, 89-90) suggest looking at Baulch and Hoddinott (2000), and Baulch and Masset (2003) for first approach, Ravallion (1988) and Jalan and Ravallion (2001) for second approach, and Pritchett et al. (2000), and Chaudhuri (2002) for third approach to chronic poverty.

survey they can be below the poverty line (e.g., stochastic variations in their assets returns). From more social perspective, Barrientos (2013, 104) proposes to use Sen's capabilities framework for studying vulnerabilities in the sense that he refers the latter as "the deficits in commodities and/or difficulties in converting these commodities into functionings". He emphasizes the importance of freedom with an example of "deficits in voice and participation by women in household decision will make them especially vulnerable." He also points out that capabilities are a good basis for understanding the inter relation among different vulnerabilities, such as a one type of vulnerability make vulnerable the individual/household in realizing other beings and doings. For instance, he mentions that for elderly people, their capabilities for converting their commodities to functionings are limited due to their age, reduction in assets and life-course capacity.

Vulnerability to poverty studies try to explain the possible negative impact(s) of single, successive, or simultaneously shocks (e.g., economic, ecologic, health), crises (e.g., financial), or other negative events (e.g., unexpected events such as sudden changes in household structure, becoming unemployed, natural disasters, conflicts, war) on well-being/welfare of people; more specifically, how these impacts stimulate poverty trap (i.e., vulnerability generate poverty), or make persist poverty among those already poor or in a in disadvantaged position (see the World Bank's framework for analyzing vulnerability to poverty (2009, 244), and Sirven 2007, Fujii 2016 for a comprehensive review). In World Bank's framework (2009), inspired by Dercon (2001, 17), several examples of risk given various types of assets, incomes, and well-being/capabilities are provided. Notably, as mentioned by Barrientos (2013, 87) "the livelihoods of poorer people are highly vulnerable in large measure because they have fewer buffers or because the range and effectiveness of the buffers available to them provide inadequate protection" (also see Scoones 1998, Ellis 2000 for livelihoods approach to poverty measurement); so that, on one hand some studies focus on the quantity and quality of buffers available to poor, disadvantaged people (see Chambers and Conway 1991), and on the other hand some others explore how units (e.g., household, family, individual) try to reduce their exposure to shocks, stresses, crises and strengthen their capabilities.

Several studies in the literature show that units (e.g., household, individual) having encountered uninsured shocks, those with fewer assets, entitlements, and buffers are more likely in risk of poverty. For instance, assets (e.g., human capital, social capital, financial capital) help the poor to escape from poverty trap and cope more easily with shocks, unexpected negative events. Addison et al. (2009b, 16) mention that "assets bring the production dimension into poverty measurement, adding to the income, consumption, and human development dimensions." (see Hulme et al. 2001, Carter and Ikegami 2009 for application of assets-based approach to poverty dynamics). As for coping with vulnerability, studies point out survival strategies, possible protective entitlements, social policies against hazards (see World Bank's Risk and Vulnerability Assessments), various ways and behaviors (e.g., diminishing health expenditure, children dropping out and starting to work (see Dehejia and Gatti 2002, Guarcello et al. 2010 for child labour examples), performing informal employment) experienced and adopted by people for escaping from 'vulnerability to poverty', and from insecure situations to protect them.

Mostly, vulnerability studies are based on probabilistic models and (managing) risk analysis by using panel data allowing to observe the duration of poverty, specifically to see inter-temporal variability in well-being status of individuals over consecutive years (e.g., how poverty status

changes for a household between certain periods of time). For instance, the inter-temporal analysis of poverty known as the poverty dynamics (or the duration of poverty or longitudinal poverty studies) in the literature requires panel datasets (i.e., the same individuals are followed for certain periods of time). However, as Barrientos (2013, 91) emphasized, the limited amount of available longitudinal data sets in developing countries is quite a huge gap in this field (but also see Lawson et al. 2003, Hulme and McKay 2008 for more information on available panel datasets in developing countries). In same line, Bird (2013, 76) mentions that in developing countries, panel data following children from their childhood to their adulthood are not very available. Besides longitudinal analysis, cross-sectional data can also be used for some specific purposes such a time-series analysis. As shocks are the unexpected events, the issue of uncertainty is at the core of vulnerability studies, for that reason, sophisticated statistical modelling take stage as an important tool in empirical analysis. But of course, complementary to quantitative studies, prominent qualitative studies (e.g., family histories, biographies, life-history in- depth interviews) contribute as to describing vulnerable situation of the poor such as their daily struggle, vicious life circle, recurring hardships, utter destitution, despair against shocks (see Narayan and Petesch 2007; Narayan et al. 2000, 2009).

Concerning measurement of vulnerability, Klasen and Povel (2013, 24) provide the main points to take in consideration: first, time horizon (as vulnerability is an ex ante issue, it is important to know the period of interest relevant to vulnerability), second, subjective time preference (to know for the unit such as household, which event is the primary matter of vulnerability), third, benchmark (such as poverty line, or current level of well-being) for evaluating the vulnerability of current situation, fourth, risk attitudes of the unit (e.g., household). They mention that "risk-neutral households behave in a different way than risk-averse ones do." As aforementioned, poverty and vulnerability are two disentangled concepts. Vulnerability is a more dynamic concept while poverty is usually analyzed statically. Even chronic, and transient poverty studies are done over several years; in poverty dynamic studies, the analysis is backward looking. In opposite, vulnerability studies are forward looking with the aim of identifying those who are expected to be poor in the future (Dercon 2005). Klasen and Povel (2013, 23) defines vulnerability as "the future threat of poverty", or similarly, vulnerability to poverty can be seen as an "ex-ante assessment of poverty, based on current or past information" (Barrientos 2013, 85). In this light, units who are expected to be poor in their future are considered to be vulnerable to poverty. Dercon (2001, 27) clearly distinguishes poverty and vulnerability in the sense that vulnerability measures "exposure to poverty rather than the poverty outcome itself". In addition, it is important to note that vulnerability to poverty studies can be important for policy makers to anticipate the impact of poverty reduction policies in future. In following paragraphs, we will directly focus on vulnerability as expected poverty (Ligon and Schechter 2004, Gaiha and Imai 2009), and vulnerability to poverty (Calvo and Dercon 2005). For more broad knowledge on vulnerability, Alwang et al. (2001), Dercon (2001) can be consulted.

Concerning *vulnerability as expected poverty*, there are two main ways of calculating: the risk of falling below the poverty line in the future, or the risk of experiencing poverty at least once in the next few years. The works of Ravallion (1988), Holzmann and Jorgensen (1999) play important role in the former approach, for instance, Holzmann and Jorgensen (1999, 6) define vulnerability as "the risk of economic units (such as individuals, households, and communities) to fall below the poverty line (i.e., having insufficient consumption and access to basic services) or, for those already below the poverty line, to remain in or to fall further into poverty". In this approach, poverty lines and probability thresholds of vulnerability are generally set arbitrarily.

For instance, as a benchmark for assessing vulnerability, households having at least fifty per cent of probability of falling below the poverty line in the future are considered vulnerable (Chaudhuri et al. 2002). On the other hand, Rajadel (2002) defines the *vulnerability line* as the observed headcount ratio of poverty.

Regarding the risk of experience poverty at least once in the next few years, there is a more focus on probability, for instance, Pritchett et al. (2000) give empirical evidence for households experienced episodes of poverty. Moreover, Christiaensen and Subbarao (2005), Günther and Harttgen (2009) yield different applications of vulnerability as expected poverty. Chaudhuri et al. (2002) formularized actual degree of vulnerability as expected poverty (VEP) problem such as:

$$VEP_{ht} = pr(C_{h,t+1} \leq z)$$

where h is used for household, t for present time, pr for probability, C for consumption, $t+1$ for future time, and z for poverty line.

By this formula, the goal is to find the probability that a household's future consumption ($C_{h,t+1}$) falls below the poverty line in the future.

However, this formula has two drawbacks: it does not consider risk sensitivity- degree of vulnerability of households, neither severity of expected poverty- falling very hard from poverty line does not mean same hardship as falling a little bit from the poverty line.

Hoddinott and Quisumbing (2003) reformulates the formula of Chaudhuri et al. (2002) based on Foster-Greer-Thorbecke (1984) measures of poverty as follows:

$$VEP_{ht} = \sum_{i=1}^{N_h} p_{hi} \times I_{hi}\left(C_{hi,t+1} \leq z\right) \times \left(\frac{z - C_{hi,t+1}}{z}\right)^{\alpha}$$

where h is used for household, $\sum_{i=1}^{N_h} p_{hi}$ for the sum of probabilities of all possible future scenarios faced by household, the function $I_{hi}(.)$ is an indicator as:

$$I_{hi} = \begin{cases} 1, & C_{hi,t+1} \leq z \\ 0, & otherwise \end{cases}$$

and the parameter α is used for indicating the depth of poverty such as: (i) if $\alpha = 0$, VEP_{ht} is equal expected poverty headcount; (ii) if $\alpha = 1$ (risk neutrality), VEP_{ht} is equal expected poverty gap; and (iii) if $\alpha > 1$, VEP_{ht} is equal expected severity of poverty. Meanwhile, it is important to mention that this given formula does not include perfectly the risk attitudes of household (see Klasen and Povel 2013, 29). It is important to note that subjectivity of setting poverty lines is a problematic question as well in these computations.

Moreover, Kamanou and Morduch (2002) employ consumption patterns and household characteristics to identify vulnerability of households to poverty in terms of their expected expenditures. To estimate future expenditures, authors observed household characteristics and household's past consumption trends. In this type of analysis, assumptions play crucial role (e.g., in a non-homogenous society, household's risk averse and risk exposure will be various, which will make more sophisticated statistical modelling and predictions). Furthermore, in this type of vulnerability studies, poverty is seen as a purely monetary phenomenon based on

consumption patterns. For that reason, these studies can be criticized because in recent years it is widely recognized that poverty is a multidimensional issue rather purely economic one.

Concerning *vulnerability to poverty*, Calvo and Dercon (2005) introduced axiomatic approach to measure vulnerability to poverty. They focus on probability of the state of the world and state-specific degree of deprivation. They propose the following formula:

Vulnerability to poverty for a household = $1 - (\sum_{i=1}^{N_h} p_{hi} \times x_{hi}^{\alpha})$

where h is for household, i for the state of the world to occur, p_{hi} for the probability of the state of the world to occur, x_{hi} for state-specific degree of deprivation, and α changes between 0 and 1. The parameter α designs the risk aversion, when it approaches to 1 the less risk aversion is assumed.

The state-specific degree of deprivation is equal:

$$x_{hi} = \frac{\tilde{y}_{hi}}{z} \times \tilde{y}_{hi}$$

It is called a censored outcome because when y_{hi} is superior than the poverty line z, the related data is censored, which means that they do not have any impact on vulnerability measure. Concerning axioms, Klasen and Povel (2013, 31) mention two most important axioms of this formula: (i) "the axiom of risk sensitivity implies that while keeping the expected outcome constant an increase in uncertainty faced by a household should be reflected in a higher degree of vulnerability"; (ii) "the focus axiom is supposed to ensure that the vulnerability measure is exclusively sensitive to negative future states of the world. Instead, positive future outcomes must not be reflected in the measure."

Regarding the link between multidimensional poverty and vulnerability, Abraham and Kumar (2008, 77) computed the vulnerability to poverty of households in India, by ranking 15 major Indian states in terms of multidimensional poverty and vulnerability to multidimensional poverty in 1993-1994 and 1999-2000. They define poverty as a "status of an entity at a point of time"; whilst vulnerability to poverty is the likelihood of an entity falling into poverty, given the current status of the household". Authors notably emphasize the importance of such study to define preventive policies against poverty. In their study, vulnerable households correspond to whose achievements in one or more dimensions are above the deprivation threshold but at the same time they are not very far from the poverty line. Among those above the poverty line, households having higher deprivation levels are considered more vulnerable to being poor. In this regard, Qizilbash (2002) provides an interesting work focusing on who is above the poverty line but vulnerable to poverty in South African provinces, by using fuzzy logic (for instance, see Miceli 1998 for fuzzy logic- the Cerioli-Zani function[95]). In his work, the degree of belonging or membership to the set of poor, measured by fuzzy logic, corresponds to the extent of vulnerability to poverty. Inspired by Qizilbash's work, Abraham and Kumar (2008) applied

[95] The Cerioli-Zani function (see Cerioli and Zani 1990) defines the membership value of the set of poor $\mu_i = \frac{(X_h - X_i)}{(X_h - X_l)}$ where X_h is the highest level of achievement in the series, X_i is the achievement of individual and X_l is the lowest achievement in the series. The membership value of an individual (μ_i) with the highest achievement in a dimension, will be equal zero. In opposite, the membership value of an individual with the lowest achievement in a dimension, will be equal one. The higher the membership value, the more vulnerable the person.

fuzzy logic to compute vulnerability to multidimensional poverty based on 6 dimensions that they call 'basic capabilities' (consumption, education, sanitation, access to water, source of energy for cooking and dwelling). They identified vulnerable people as those having a membership value greater than 0.7 (an arbitrarily selected threshold) in a dimension (every dimension has an equal weight). Moreover, Calvo (2008) computed vulnerability to multidimensional poverty in Peru, for the period of 1998-2002, based on Calvo and Dercon's (2005) vulnerability modelling.

Vulnerability can be concerned with specific target groups (i.e., vulnerable groups) such as women, children, different types of households that are affected by negative shocks. For instance, during the United Nation's World Summit on Social Development in Copenhagen, in 1995, it is mentioned that "women bear a disproportionate burden of poverty and children growing up in poverty are often permanently disadvantaged. Older people, people with disabilities, indigenous people, refugees and internally displaced persons are also particularly vulnerable to poverty. Furthermore, poverty in its various forms represents a barrier to communication and access to services, as well as a major health risk, and people living in poverty are particularly vulnerable to the consequences of disasters and conflicts" (see United Nations 1995, cited in Halleröd 2000, 166).

A brief introduction to vulnerability and its relation to poverty has given in previous paragraphs in this sub-section. As the goal of this thesis do not concern with vulnerability, we don't enter to more specific details. However, various papers investigate advantages and shortcomings of vulnerability to poverty and of vulnerability as expected outcome (Klasen and Povel 2013, Klasen et al. 2013).

As we have seen all approaches to poverty evoked in our literature review are based on objective measures except subjective monetary poverty lines and social necessities in relative deprivation. But, in next sub-section, we will explore subjective poverty approach based on subjective well-being. It is important to note that in the literature, the interest in subjective indicators increases, notably combination of objective and subjective indicators is highly recommended.

Subjective poverty

In the literature, some studies have attempted to compare poverty as a concept to well-being and quality of life (Sen 1993, Anand et al. 2005, Rojas 2006, 2015). Originally, poverty is recognized as a more objective concept focusing on achievements and deprivations in terms of doing, being, and having in a socio-economic perspective; on the other hand, well-being and quality of life come mainly from the discipline of psychology (see Møller and Schlemmer 1983, Diener 1984, Nussbaum and Sen (eds.)1993, Kahneman et al. 1999, Cobb 2000, Diener and Suh 2000, Hagerty et al. 2002, Noll 2004, Diener et al. 2010, Dodge et al. 2012, for well-being and quality of life). Well-being studies are generally interested in mental and physical health of individuals or in social psychology[96] at societal level by employing subjective approach, for instance, Diener et al. (2002, 63) define subjective well-being as "a person's cognitive and

[96] Mead (1934) argues that the person belongs to a social arrangement and the main role of *social psychology* is to find out the impact of social groups on the determination of the personal behavior and experience. Societal experience (including behavioral process and diverse mechanisms) is therefore dependent on both personal experiences and the operational process between personal and societal experiences.

affective evaluations of his or her life". Whilst quality of life covers a larger set than subjective well-being including both objective and subjective approaches focusing on multiple aspects of life such as education, health, economic activity, leisure, social relationships, values, opinions, feelings. Sen (1993, 31) argues that quality of life can be assessed in terms of "the capability to achieve valuable functionings". Besides, many quantitative studies of well-being and quality of life have proposed various typology of well-being reflecting it in different categories such as material, financial, social, political in terms of domain, sphere or dimension (see Campbell 1976, 1981, Campbell et al. 1976, Diener 1984, Griffin 1986, Nussbaum and Sen (eds.) 1993, Hagerty et al. 2001, Phillips 2006, OECD 2013). Researchers arguing that poverty is a purely financial, monetary or material phenomenon have tendency to perceive quality of life as a larger set compared to poverty. However, poverty is an older concept than well-being and quality of life in sociology as we know that pioneer quantitative sociological studies (notably from the eighteenth century in Britain) carried out comprehensive social surveys and quantitative social analysis including a number of living items such as nutrition, housing, economic activity, health to examine poverty; and especially from the 1990s, poverty has been recognized as a multidimensional fact including more than monetary/financial aspect, which is promoted by international agencies (e.g., the United Nations, World Bank). In this regard, quality of life approach cannot substitute or encompass multidimensional poverty studies; rather they are more parallel works. Quality of life and poverty are two different concepts in the sense that quality of life is particularly important to investigate subjective well-being of people such as overall life, satisfaction regarding different items of life, happiness. Poverty includes both economic and social aspects; however, in terms of subjective well-being, the quality-of-life studies are more fruitful. Some studies define poverty as a low level of well-being (Kakwani 2006), wellbeing deprivation (Rojas 2014) or they prefer to use interchangeably both poverty and well-being like "considering that individual wellbeing/poverty manifests itself in multiple dimensions" (Ruggeri et al. 2006, 10). Rojas (2014, 320) mentions that "because the concept of poverty is closely associated to the notion of wellbeing, different conceptions of wellbeing will clearly lead to different understandings of poverty" and "any study of the concept of poverty must necessarily address the related issue of how people's wellbeing is conceived"[97]. Notwithstanding, in many other studies, poverty is not always linked to nor directly based on well-being (e.g., basic needs, absolute poverty, underclass). We can say that in the literature, well-being, quality of life and poverty are often defined in similar way and there is no unique definition of each of them, so then such context makes difficult to draw a strict line between these three phenomena.

[97] Rojas (2014, 320) mentions the two main approaches to define wellbeing: *presumption* and *imputation* traditions. The first one, despite accepting that wellbeing is a personal experience, lets define and assess wellbeing based on third-persons such as the United Nations, the World Bank, researcher, institution's criteria. In second approach, the third-person is considered as the most adequate person/institution to define and evaluate wellbeing of others. Here, in these two approaches, the risk is making the normative judgement, so that the third-person defines and measure the wellbeing of others based on her/his personal judgment without relying on any scientific proof for justifying her/his choice. Rojas (2014) adds that the *subjective* wellbeing approach is most recently emerged as opposite to these two older approaches, taking in consideration personal experiences as the only criteria for defining and measuring wellbeing; in other words, the person's own evaluation, perception, satisfaction level will define her/his level of well-being on her/his life as whole or on a specific domain/aspect/dimension of life.

In methodological terms, well-being, quality of life and multidimensional poverty studies are all considering several indicators to reflect multiple aspects of life. In terms of indicators[98], we can do following statements: (i) if objective indicators are used in well-being analysis it is called objective well-being analysis; (ii) if subjective indicators are used in well-being analysis, it is called subjective well-being analysis. In both analysis, domains/dimensions/aspects of life may be the same however the difference comes from how the questions are asked to people. If the question is asked for receiving personal feelings, evaluation, perception of the individual on an event or on her/his life as a whole, then well-being of the person regarding that event/life is considered analyzed from subjective angle. On the other hand, quality of life analysis does generally include combination of both objective and subjective indicators. Concerning multidimensional poverty studies, general tendency is to include only objective indicators. At this point, we emphasize that employing subjective indicators in multidimensional poverty analysis is a flourishing topic, which is known as subjective poverty. Depending on employed definitions we can say that (i) objective well-being and multidimensional poverty analysis can be based on similar objective indicators; (ii) subjective well-being and subjective poverty analysis can be based on similar subjective indicators; (iii) quality of life and multidimensional poverty analysis can be based on a similar combination of objective and subjective indicators if multidimensional poverty analysis includes also subjective indicators. But we also note that generally instead of saying objective well-being, the term well-being is directly used if only objective indicators are employed.

We can consider the works of van Praag (1968), Goedhart et al. (1977), van Praag et al. (1980), Hagenaars and van Praag (1985), Blaylock and Smallwood (1986), Kapteyn et al. (1988), Hagenaars and de Vos (1988), de Vos and Garner (1991) as pioneers, and the works of Pradhan and Ravallion (2000), Ravallion and Lokshin (2002), Ferrer-i-Carbonell (2002), Ferrer-i-Carbonell and van Praag (2001), Carletto and Zezza (2006), Rojas (2006, 2007, 2008, 2009), Rojas and Jiménez (2008), Ravallion (2012a) as some of most recent poverty studies from the 2000s for employing subjective well-being[99] tools. In previous sub-sections, we saw that generally objective poverty lines (e.g., absolute monetary poverty line such as 1USD per day), which are arbitrarily set by experts or institutions, are considered in poverty studies. But, alternatively, subjective approach brings another level in construction of poverty lines by referring to (i) whether household/individual find sufficient or is satisfied with her total income, consumption, expenditure, living standards, social environment; or (ii) what is the minimum income level that the household/individual needs in their opinion for making ends meet (e.g., Leyden approach, see van Praag et al. 1982); or (iii) whether individual/household perceives herself as poor (i.e., self-perceived poverty for instance see Rojas and Jiménez 2008). In other words, subjective approach takes into account household/individual's own evaluation, perception, opinion on her monetary and/or non-monetary situation in different items of living conditions. For instance, Rojas (2008) distinguishes *experienced* poverty, *experienced economic* poverty and income poverty in Mexico by employing a subjective well-being

[98] Subjective indicators are based on personal opinions, perceptions, evaluations, and emotions. For instance, the indicator of educational attainment based on the years of schooling is an objective indicator whilst the indicator of educational attainment based on the personal perception of utility of educational attainment is a subjective indicator. Maggino (2016, 222) defines *objective indicator* as "based upon explicit criteria, shared by external observers" whilst *subjective indicator* is "based on subjective evaluations and implicit criteria which can vary from individual to individual".

[99] Generally denoted as SWB in the literature.

approach in terms of life-satisfaction[100]. He defines experienced poverty as a "low life satisfaction" and experienced economic poverty as a "low economic satisfaction" (2008, 1078). He argues that income poverty is not an accurate indicator for capturing poverty that people experience in their life, for instance, he shows that greater income does not always correspond to greater well-being. He also adds that these three different approaches lead to different headcount ratios of poverty because not always the same people are identified as poor in each of these three concepts.

Monetary subjective poverty lines take in consideration personal or household's perception and/or evaluation about her own financial situation such as whether what the individual/household possess as material or monetarily correspond to what they really need or to what they wish to have (e.g., income level) for leading a good life (see van Praag and Ferrer-i-Carbonell 2008, Corazzini et al. 2010, Castilla 2010, Rojas 2007). On the other hand, non-monetary subjective poverty lines try to capture if social conditions of individual/household's life meet her expectations in various components of life such as education, health facilities, drinking water (see Mack and Lansley 1985 as a pioneer work for defining social necessities based on subjective approach, Pradhan and Ravallion 2000). Subjective poverty lines can be based on one or multiple questions. One of the most popular practice is to estimate poverty level of an individual/a household by simply asking directly them one question in form of how they in general appraise their life at a specific time; for instance, 'Overall, how satisfied (content, happy) are you with your life?' (Ravallion and Lokshin 2002). Possible answers to this question bear typically in an ordinal metric such as Likert scale (Likert 1932), or gradual ascendant/descendent order satisfaction categories (e.g., from very satisfied, satisfied, neither satisfied nor dissatisfied, dissatisfied, to very dissatisfied; or, from very unsatisfied, unsatisfied, neither unsatisfied nor satisfied, satisfied to very satisfied (Ravallion 2012a), or in Cantril's ladder of life scale (e. g. from the poorest to the richest in a number of step ladders, see Cantril 1965, Riffault 1991). Rojas (2008, 1082) used the question "Taking everything in your life into consideration, how satisfied are you with your life?" with seven options of response in an ordinal scale (i.e., extremely unsatisfied, very unsatisfied, unsatisfied, neither satisfied nor unsatisfied, satisfied, very satisfied, extremely satisfied). The satisfaction question is also adapted for various components of life[101] such as housing conditions, financial situation; for instance, 'how satisfied are you with your housing conditions?", "how satisfied are you with your household's financial situation?" (Rojas 2008, 1082). In the literature, there are also studies focusing on the link between subjective well-being and monetary poverty, on the link between subjective well-being and basic needs, or on the link between subjective well-being and multidimensional poverty by employing traditional regression analysis (see Kingdon and

[100] See Campbell 1981, Diener et al. 1985, Cummins 1996, Rojas 2007, van Praag et al. 2003 for life satisfaction. Maggino (2016, 215) mentions that subjective well-being includes a *cognitive* and an *affective* component. Regarding *cognitive* aspect, the person evaluates her/his life situation as a whole or a specific domain/dimension/aspect of her/his life based on satisfaction term. On the other hand, *affective* component reflects emotions of the person in positive (i.e., pleasant affects such as being happy) or negative way (i.e., unpleasant affects such as feel anxious, stressed) on a specific event or an aspect of life (e.g., working conditions).

[101] Rojas employs the domains of life approach for mentioning different aspects of life (see also Cummins 1996).

Knight 2006, Posel and Casale 2011, Stevenson and Wolfers 2013), which are in particular useful for controlling the validity of objective indicators of poverty[102].

Computation of subjective poverty by national statistical offices is not a common practice. However, in 2012, the Statistics South Africa published a report of subjective poverty measured at national and provincial levels in South Africa based on Living Conditions Survey (LCS onwards) 2008/2009[103]. In this report, subjective poverty is defined as "an individual's assessment of his or her own welfare, utility or happiness" (Statistics South Africa 2012b, 8). Subjective poverty measurement process in this report includes two parts: financial subjective poverty, multidimensional subjective poverty. Concerning the financial aspect, the following four questions in the survey are used (Statistics South Africa 2012b, 15) : "*self-perceived wealth* question (SPWQ)" (i.e., "Would you say your household is at present (1) wealthy; or (2) very comfortable; or (3) reasonably comfortable; or (4) just getting alone; or (5) poor; or (6) very poor"), "*minimum income* question (MIQ)[104]" (i.e., "Which net household income per month in Rand would be the absolute minimum for your household? That is to say, that you would not be able to make ends meet if you earned less"), "*income evaluation* question (IEQ)[105]" (i.e., "Is the total monthly income of your household higher, lower or more or less the same as this figure (i.e., the figure given in Q24.5): (1) much higher; (2) higher; (3) more or less the same; (4) lower; (5) much lower"), "*economic ladder* question (ELQ)" (i.e., "Please imagine a 9-step ladder where on the bottom, the 1st step, stand the poorest people, and on the highest step, the 9th, stand the rich. On which step would you consider you and your household to be?"). Concerning the multidimensional aspect, the following five questions in the survey are used (Statistics South Africa 2012b, 17): "*food consumption* question" (i.e., "Over the past month, was your household's standard of food consumption: (1) less than adequate; (2) just adequate; (3) more than adequate; (4) not applicable"), "*housing* question" (i.e., "Over the past month, was your household's standard of housing: (1) less than adequate; (2) just adequate; (3) more than adequate; (4) not applicable"), "*clothing* question" (i.e., "Over the past month, was your household's standard of clothing (1) less than adequate; (2) just adequate; (3) more than adequate; (4) not applicable"), "*health care* question" (i.e., "Over the past month, was your household's standard of health care: (1) less than adequate; (2) just adequate; (3) more than adequate; (4) not applicable"), "*children's schooling* question" (i.e., "Over the past month, was

[102] The literature of subjective well-being recommends taking in consideration personal perception, evaluation in various aspects of life for assessing well-being, rather than let experts, policy makers, governments, institutions to define well-being level of individuals based on objective indicators and thresholds. However, it is important to note that poor people have tendency to adapt themselves to their hardship, disadvantages and to being satisfied in their life despite violation of human rights and multiple deprivations. Adaptation may lead to ignorance of injustice, human rights violation and to not search for better life conditions. In this perspective, it is very important to test the validity of subjective indicators with objective indicators. For instance, a child labour who dropped her/his elementary school may feel satisfied with her/his educational attainment; but in fact, this situation reflects violation of her/his educational right.

[103] It is important to note that the Living Conditions Survey (LCS) 2008/2009 was the first survey prepared for measuring poverty in South Africa, which is conducted between September 2008 and August 2009, and is suitable for producing both objective and subjective indicators for a multidimensional analysis. The subjective poverty report is prepared in collaboration with the School of Development Studies at the University of KwaZulu-Natal (Statistics South Africa 2012).

[104] As a concept, developed by Goedhart et al. (1977).

[105] As a concept, developed by van Praag (1968, 1971).

your household's children's schooling: (1) less than adequate; (2) just adequate; (3) more than adequate; (4) not applicable").

We, therefore, finish the literature review by the approach of subjective poverty mentioned above.

9.2 Comparative summary of poverty concepts

As we have seen previously through the literature review, different concepts of poverty have been developed by scholars in socio-economic perspective since the eighteenth century. Among these concepts, some of them are remained at theoretical level and some of them are operationalised for empirical studies. To determine which concept of poverty is better than other is a complex and difficult task. Depending on purpose of research, each concept may have its own advantages and disadvantages. In chapter 2, we expressed some critics on the approaches of absolute poverty, basic needs, relative deprivation; but also, why the capabilities approach is important for ensuring development and reducing poverty. After presented the other approaches on poverty from the literature, we will mention the advantages and disadvantages of these approaches in the following paragraphs.

In section 9.1 we saw the culture of poverty and underclass concepts, which are originally emerged in the U.S literature. These two concepts are often studied by qualitative research. A lack of empirical studies based on comprehensive social surveys is a significant gap in these two approaches. Oscar Lewis' narratives for describing culture of poverty are certainly masterpiece (life stories are delicately written with participant observation far from personal judgments); but, unfortunately American politicians (i) used this approach to blame the poor, minorities, and the most disadvantaged groups for being lazy, unskilled, not adequate for industrialized life; and (ii) tried to explain poverty to public based on behavioural characteristics of the poor, rather than focalising on structural factors. The aim of Lewis was to portray living conditions of poor people under different circumstances for showing us how these people do live on their daily basis including their expectations, fears, opinions, habits, and how they try to cope with their vulnerability. If instead of the term 'culture', Lewis was used 'living conditions' or 'living situation', the critics towards the poor could be different? probably not because it is much easier to manipulate public by showing different sides of poor compared than those of upper, or upper-middle class, instead of trying to understand why the poor may adopt deviant behaviour, or marginal behaviour, or why they are frustrated or socially excluded. Underclass studies showed in particular the life situation of unemployed, single-mother, Hispanic, Afro-American people, and urban poverty. As in the case of culture of poverty, underclass studies are also used by politicians and elitists to explain poverty in behavioural terms instead of showing racial discrimination, unequal public services, geographic segregation, labour market dysfunctions in American society. In a similar way, Germani examined in major Latin American cities the poor and the disadvantaged people staying out of mainstream whom he called the marginals, by employing a considerable number of descriptive statistics. It is important to note that these three approaches lack a concise definition of poverty. Rather, poverty is tried to be explained based on living conditions of some disadvantaged groups. However, different sub-analysis done in these three approaches through qualitative methods and descriptive statistics, can be an important guide today for highlighting multidimensional aspect of poverty and for improving the scoop of poverty analysis in terms of different social groups.

In a similar line with relative deprivation approach, social exclusion studies have focused on the risk of being in a position of not being able to afford properly social necessities incorporating social, economic, political participation in society. Debates on whether social exclusion is the same phenomena as poverty, or whether poverty means social exclusion, or whether social exclusion means poverty, or whether social exclusion is a dimension of multidimensional poverty, or whether poverty and social exclusion can be both combined are emerged in the literature. Notably, social exclusion as a concept and tool have become very popular in European context since the late 1990s to conceptualize and measure poverty. For instance, Eurostat considers material deprivation and low work intensity as the indicators of social exclusion and combine them with the relative monetary threshold for measuring poverty and social exclusion in European Union. The fact that social exclusion is a multidimensional phenomenon, determining the link between poverty and social exclusion becomes complex. All poor people won't necessarily feel socially excluded or a very prosperous person may find herself socially excluded as well. As we have mentioned in relative deprivation, people are not obligated to get, or to be happy or to be satisfied with the same materials, habits, social, political, economic, cultural activities. Even if some items are considered/voted by a majority of society as the most important items for living a decent life, the idea of not being able to have access to these activities corresponds to social exclusion remains problematic in terms of variety of living styles and real liberty. Furthermore, investigating why some people feel socially excluded or is there any specific phenomenon in society that makes some people socially excluded appear other topics in examining poverty.

In a more specific matter, the approaches of working poor and social disqualification focus on living conditions of a particular group of people. On the one hand, working poor as a concept aims at reflecting precarious, insecure, unstable working conditions of economically active people who suffer of deprivation in some aspects of life. On the other hand, the situation of assisted people (i.e., dependency in social welfare) in labour market and their social bonds are explored by studies of social disqualification. Working poor studies are mainly based on quantitative data and distinguish the poor in terms of monetary threshold such as income level. In disqualification studies, the position of assisted people regarding monetary poverty threshold is also often examined. However, the link with multidimensional poverty is a quite new area in working poor and social disqualification studies. Working poor, assisted people can be seen as sub-groups of poverty but they cannot be taken as a whole definition of poverty. In a multidimensional perspective, a person may be poor in some aspects of life but not necessarily in her working conditions, or not all multidimensionally poor people would become dependent vis à vis social welfare programmes.

Most recently, there is a growing interest in vulnerability and subjective poverty studies. The former approach does not provide a specific definition of poverty but instead the risk of falling into poverty is analysed thanks to sophisticated forecasting models. Vulnerability to poverty literature is generally based on monetary poverty thresholds such as income level or consumption patterns; the risk of being multidimensionally poor is quite a new practice in this field. The latter approach involves a subjective perspective to poverty analysis based on subjective well-being. In opposition to many quantitative poverty studies that are based on objective indicators, subjective poverty studies require subjective indicators considering personal evaluation, perception, feelings, satisfaction vis à vis different items of life in a multidimensional perspective.

Through previous paragraphs, we provided a short summary of each approach to poverty that we saw in the literature review. In the table below, we provide main advantage(s) and disadvantage(s) of each poverty concept described in the literature review.

Table 18. Main advantage and disadvantage of each poverty approach described in the literature review

Concept	**Main advantage(s)**	**Main disadvantage(s)**
Subsistence	-Focus on nutrition, which allows take priority hunger and starvation in poverty analysis.	-Not multidimensional.
Absolute poverty	-Focus on nutrition, shelter, clothes, which are very important items for identifying the poorest, notably in case of developing countries.	-Limited number of items, which does not allow to capture the various forms of poverty. -Interpretation in absolute terms emerges problematic.
Basic needs	-Focus on nutrition, shelter, living standards, clothes, which are very important items for identifying the poorest, notably in case of developing countries.	-Limited number of items, which does not allow to capture the various forms of poverty. -The notion of "basic" appears problematic (weak scientific arguments for justifying 'basic' approach)
Assistance	-Focus on dependency vis-à-vis social welfare, which allows to observe how the assisted people risk losing her/his real liberty and become dependent to help.	-Longitudinal surveys are important for capturing life trajectory of assisted people; however, the use of panel data on social welfare dependency is quite limited in this area in developing countries. -This approach focuses on a specific group analysis, which does not reflect whole poverty. Not all poor people are dependent on social welfare. Notably, in developing countries, social welfare programs are limited and the access to these programs may be problematic due to governance issues (e.g., corruption).
Monetary poverty	-Focus on monetary side of poverty that reflects income and consumption patterns, which is one of the important aspects from material perspective in life.	-One-dimensional analysis. - Access to various items of life such as family relationships, friends, being confident, having a respect and a dignity in society cannot be measured in terms of money. There are many items such as democracy, freedom, family that we cannot buy. Such approach reduces the value of life to money.
Relative deprivation	-A special focus on social participation, which was an important step in the literature for understanding poverty as a multi-dimensional phenomenon beyond the monetary perspective.	-In practice, this approach is considered as a monetary poverty through relative monetary poverty lines and/or as material deprivation. -The content of such approach is more suitable for developed countries.

		-To define social necessities in a society, relative deprivation studies do not generally evoke violation of human rights, cultural aspect, adaptation to poverty, context of non-democratic environment.
Culture of poverty	-For understanding daily life of poor and disadvantaged people in many perspectives, poverty from generation to generation, Lewis' narratives are masterpiece.	-Most of studies in this area are qualitative. Hence, target group remains small, which does not allow generalization of the theory. -The notion of 'culture' allows politicians and elitists to manipulate public for blaming the poor for their behaviours and living style. -There is no specific definition of poverty.
Underclass	-For understanding urban poverty, geographic segregation, racial discrimination, difficulties of minority groups, this approach brings a substantial contribution to the literature. -Descriptive statistics provided by some authors for showing underclass are important as case studies for showing sub-group analysis of poverty.	-Most of studies in this area are qualitative. Hence, target group remains small, which does not allow generalization of the theory. -There is no specific definition of poverty. -Politicians and elitists can easily use this approach based on behavioural characteristics of poor for blaming them to be poor, rather than focusing on structural terms to explain the determinants of poverty.
Social exclusion	-As in the relative deprivation approach, taking in consideration social items is the most important feature of this approach to poverty studies. -Some panel data analysis in this area brings time aspect to poverty studies.	-As social exclusion is also a multidimensional fact, the link with multidimensional poverty emerges complex. -The use of subjective indicators is quite limited. In general, material deprivation, low work intensity, and relative monetary poverty are used to explain social exclusion; however, feeling and perception of social exclusion are important features. Not all materially or monetarily poor people may feel socially excluded. Social exclusion appears as an indirect measure of poverty. -Definition of social exclusion is not based on any sociological theory.
Social disqualification	-Focus on the extent of assistance to social welfare studies by considering labour market status and social bonds. -There are robust sociological explanations behind the elaboration of the concept.	-To examine social disqualification of the disadvantaged people, the poor are generally defined based on monetary poverty line. The multidimensional poverty approach is needed for future research in this area.
Marginalidad	-Focus on outsiders of mainstream society in terms of urban poverty	- Target group is limited to the poor living in urban areas.

	-Descriptive statistics in various socio-economic aspects are provided. -In particular, this approach contributed to the literature of poverty in Latin America.	- A specific definition of poverty is not provided.
Working poor	-Focus on working conditions of economically active population.	-Poverty is generally defined in terms of monetary poverty thresholds. Elaboration on multidimensional aspect is needed for further research in this area.
Capabilities	-Conceptualization of poverty in terms of real liberty and capabilities. Instead of monetary thresholds and material deprivation, poverty is explained in terms of life situation of the person in various aspects.	-The link between real liberty and indicators is often invisible in the studies. -Achievements are taken as proxies of capabilities. However, capabilities reflect more than simple deprivation.
Multidimensional poverty	-Emphasize the importance of non-monetary perspective to poverty. This approach takes in consideration human rights, freedom, social needs, economic needs.	-In general, measurement of multidimensional poverty is based on capabilities approach. -In methodological terms, a binary classification of society as poor vs. non-poor is obtained in such analysis. However, multidimensional aspect requires discovering different forms of poverty rather than simple poverty ratio statistics.
Vulnerability to poverty	-Take in consideration probabilistic models for predicting the risk of falling into poverty.	-In general, vulnerability to monetary poverty or to material poverty are evoked in this area. -In some cases, instead of advanced robust statistical models, arbitrary thresholds are used for determining vulnerability level.
Poverty dynamics	-Take in consideration time aspect, which is important for following individuals over a certain period of time (e.g., fall into, move out of poverty).	-Following the same individuals require panel data, which is generally unavailable in developing countries. -In a multidimensional perspective, given many items of poverty, biographic analysis becomes a relatively difficult task.
Subjective poverty	-Subjective well-being is considered for measuring poverty, which can be an important process for testing the validity of objective indicators. -Emotions, perceptions, subjective evaluations of individuals on their life situation in many aspects are obtained in this area.	-Subjective analysis can hide some poverty problems in society. Due to adaptation, or inaccessibility to information, a disadvantaged person may adapt and become satisfied despite her/his hardships and deficit of freedom. -In a non-democratic environment, it may happen that citizens don't feel comfortable to express their satisfaction level regarding various items such as regarding public health services, transports, trust in legal system, trust in governance.

9.3 Statistical background: data mining, probability theory, classification trees

This section aims at giving insight into the statistical method used in the analysis. First, we will start off by explaining briefly the data mining and knowledge discovery, then we will continue with the introduction to statistical terms including some basic principles of probability theory, and we will finish by explaining in detail the classification trees method we used in the analysis.

Data mining and knowledge discovery

Knowledge Discovery in Databases (KDD onwards) is an interdisciplinary (computer science and statistics) process, aiming at exploring, and analyzing large (complex) data sets based on pattern recognition approach by employing data mining tools[106] (see Duda et al. 2001[1973], Piatetsky-Shapiro an Frawley (eds.) 1991, Ripley 1996, Bishop 2006 [1995], Flach 2012, Russell and Norvig 2009 [1994], Mitchell 1997, Kelleher et al. 2015, Langley 1996, Maimon and Rokach (ed.) 2010 [2005], Hastie et al. 2009 [2001], Michie et al. 1994, Rokach and Maimon 2008). Frawley et al. (1991, 3) define *knowledge discovery* as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from the data". Since the late 1980s onward, the statistics literature had seen a wide variety of new data analysis tools including innovative data mining techniques. The main reason behind of this new wave was the need of more advanced data analysis methods for especially large data sets such as advanced database systems, the Web, cleaned and integrated data warehouse, or other types of data repositories. Han et al. (2012 [2000], 5) have described the situation at that time as "the world is data rich but information poor" by adding that "the fast-growing, tremendous amount of data, collected and stored in large and numerous data repositories, has far exceeded our human ability for comprehension without powerful tools." In the 1990s, among the computer science community, it was very clear that traditional statistical techniques were not enough for filling the gap between mass data generation and data exploration and interpretation. For instance, in 1989, it was estimated around five million of databases in the world (see Frawley et al. 1991). Silberchatz et al. (1990) have recognized data mining as one of the most important research areas for the 1990s.

Data mining tools provide numerous algorithms for exploring and analyzing large-scale data sets or data bases to obtain *interesting* and *useful* patterns; for instance, Witten et al. (2017 [2000], 4) specify data mining as "the process of discovering patterns in data". By definition, a database contains many data sets, which can be either based on one or multiple sources (i.e., collected data sets). If database is well organized then it can be very helpful to store efficiently, or to modify easily mass information. Data mining tools are used in different application areas such as medicine for health diagnosis (Blum 1982), finance (Kovalerchuk and Vityaev 2005), agriculture (Michalski and Chilausky 1980), marketing and sales (Fernandez-Aguirre et al. 2008), environment (Gibert et al. 2008), insurance, physics and chemistry (Buchaman and Mitchell 1978). However, it is important to note that data mining tools can also be used by users/researchers for just understanding or discovering the structure of the data rather than making predictions.

[106] In the literature, on the one hand, data mining is directly referred to Knowledge Discovery from Data (i.e., KDD) by some researchers; on the other hand, it is considered as an important specific tool for KDD. Here, we follow the second approach. Data mining is also known as "knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, or data dredging" (Han et al. 2012 [2000], 6).

Data mining is linked with various domains such as statistics, data warehouse, machine learning (see Figure 60, Han et al. 2010 [2012, 2000], 23).

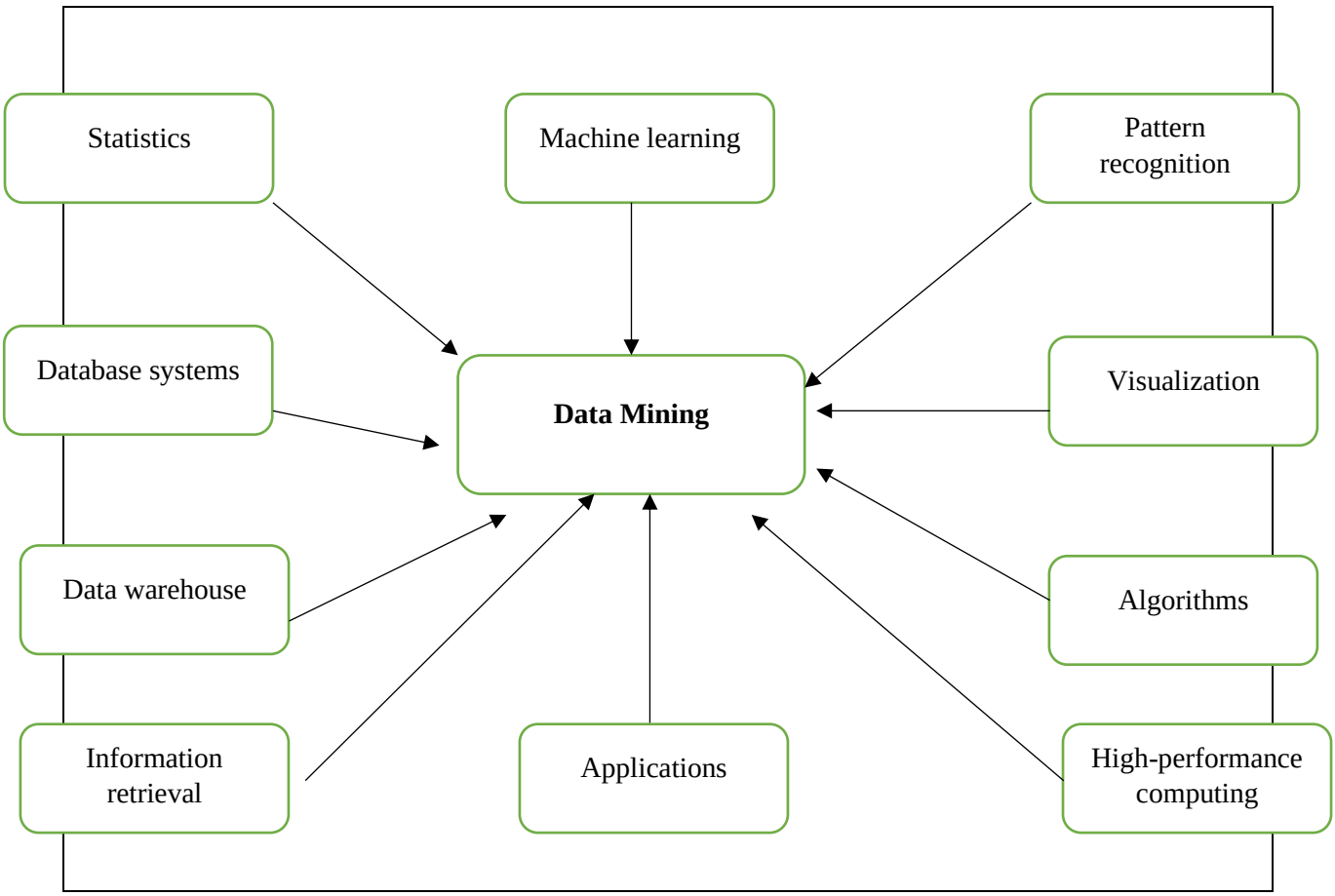


Figure 1. Data mining and its relationship with other domains. Source: Han et al. 2010 [2012, 2000], 23.

But we will especially be interesting in how data mining is associated with statistics, machine learning, and pattern recognition. We note that statistical modelling plays an important role in data mining process. There are two ways for building statistical models: (i) based upon the discovered patterns; or (ii) before the *pattern identification* process, researchers/users can construct the statistical modelling a priori and based on that the algorithm then will detect the related patterns in data. Statistical modelling indicates the variables that will be used and their role in the prediction/forecasting; for instance, the respondent variable (predicted), independent variables (predictors) in the model.

On the one hand, the *exploration* phase includes discovering the (hidden[107]) patterns in data; on the other hand, the *analysis* phase yields the prediction (i.e., estimation, forecasting) based on the patterns obtained from the exploration phase. Especially, data mining helps to seek "nontrivial predictions" (Witten et al. 2017 [2000], 6). A *pattern* corresponds to the prediction path of a phenomenon based on several variables. In technical terms, for instance, Frawley et al. (1991, 3) define a *pattern* as: "given a set of facts (data) *F*, a language *L*, and some measure of certainty *C*, we define a *pattern* as a statement *S* in *L* that describes relationships among a

[107] Hidden patterns mean discovering/uncovering previously unknown pattern in data.

subset F_S of F with a certainty c, such that S is simpler (in some sense) than the enumeration of all facts in F_S." The obtained pattern, namely, *discovered knowledge* can be considered as the most striking information among many others, which reflect relationships between facts[108]. Numerous patterns are possible for predicting a phenomenon. These patterns are discovered thanks to the data mining algorithms. The patterns are unknown before the data process- for that reason, we qualify the patterns as *hidden*. Herewith, the data mining process can be seen as the data analytics, and the patterns can be considered as *insights* to decision making process. As Kelleher et al. (2015, 1) mention data analytics extract insights from data, and the insights are used for taking decisions. Different data mining techniques have their own languages that are operationalized by employing different probabilistic methods, yielding in the end the patterns in specific visualization forms. Herewith, it becomes very important to understand the three points that are: form of algorithm, graphical representation of obtained patterns, and uncertainty of results.

The main points in data analysis consist of interpretation of results such as pattern evaluation and visual presentation of discovered knowledge from data. According to Witten et al. (2017 [2000], 6), there are mainly two extreme patterns: (i) "*black box patterns*": discovered patterns are incomprehensible; (ii) "transparent *box patterns*": there is a structure in the pattern. They contribute clearly and effectively to decision making. These are also called "*structural patterns*". Witten et al. (2017 [2000], 6) emphasize that these two types of patterns both can make accurate predictions. On the other hand, Han et al. (2012 [2000], 8), in particular, highlights that authenticity of data mining process is to find "*interesting"* patterns for extracting the *new* knowledge on the object of study. For them, there are four conditions for considering a pattern as *interesting;* if the extracted knowledge "(i) can be easily understood by humans; (ii) valid on new or test data with some degree of certainty; (iii) potentially useful; and (iv) novel" (2012 [2000], 21). But they also add that a pattern can be interesting if it "validates a hypothesis that the user sought to confirm." Regarding Frawley et al. (1991, 4), the only *interesting* patterns can be considered as *knowledge*; and an interesting pattern should be "novel, useful, and nontrivial to compute". Especially, Frawley et al. point out the utility of a pattern. Not all discovered patterns will be useful for the user/researcher. It is very important that the discovered pattern brings a new or a useful information to the user/researcher in the framework of research project/target, to be considered as an *interesting* discovered knowledge. In this regard, it is worth to mention the *domain knowledge* in the process of knowledge discovery. When deciding on variables to be used in the analysis, it is crucial to have a good knowledge of the domain of application for obtaining interesting and useful patterns. To explain it, we take the example given by Frawley et al. (1991). For a study focusing on pregnant patients, it will be no sense to find out that the variable 'sex' discovered as one of the relevant variable in pattern(s), because simply it won't bring any new information to a doctor, hospital administrator, scientific community, or to other related institutions. And, in terms of 'non-triviality', the obtained patterns should not provide user/researcher information based on simple calculations such as the average, the maximum or minimum values; instead, the discovered information should be the result of a complicated calculation process (computational complexity) that user/researcher needs specific algorithms to obtain it. For instance, two-way relationships among variables can be obtained by using simple tabulation commands on the

[108] Given that *facts (F)* correspond to the variables, *language (L)* is the algorithm of the selected data mining technique, and *certainty (C)* is shown by the employed probabilistic calculus for extracting information in data.

computer; however, more complicated relationships between several variables at the same time in large-scale data sets/bases can only be obtained accurately via sophisticated data mining algorithms.

Furthermore, Han et al. (2010 [2012, 2000], 21-22) summarize some objective and subjective measures for identifying an *interesting* pattern. The objective measures include first the support and confidence measures of the association rules of the form $X \Rightarrow Y$, second, the accuracy and coverage of classification rules. Association rules (i.e., the relational conditions) are automatically constructed by the data mining algorithm for creating the patterns form the data. The support function is $P(X \cup Y)$ that designates the probability of both X and Y are included in the operation. The confidence function is $P(Y|X)$ that corresponds to the conditional probability of an operation containing X also includes Y. A threshold can be defined for assessing the interestingness of the pattern; for instance, all the association rules that do not satisfy the conditional probability of 50% would be considered as the rules that do not emerge interesting patterns. Classification rules are also automatically constructed by the classifier (i.e., data mining algorithm) for revealing the classification patterns from data. On the one part, the accuracy measure will inform us about the percentage of *instances* classified correctly by the classification rule. On the other part, the coverage measure shows the percentage of data to which a classification rule involves. The subjective measures are more arbitrary in the sense that they are based on the user/researcher's own criterion for evaluating whether the pattern is interesting or not. Han et al. (2010 [2012, 2000], 22) mention 2 types of subjective measures: unexpected patterns, actionable patterns. Unexpected patterns correspond to the patterns showing the opposite or an alternative to what the researcher/user is expecting. Actionable patterns can be considered as the patterns including a strategic information that the researcher/user can use it for a specific concern. In addition to objective and subjective methods, there are also two other issues: first is the completeness of the data mining algorithm that deals with whether the algorithm generated or not all possible interesting patterns. Second is whether the data mining algorithm is able or not to generate *only* interesting patterns.

Frawley et al. (1991, 5) designate four features to evaluate the knowledge discovery process carried by user/researcher: (i) *high-level language*: "discovered knowledge is represented in a high-level language. It need not be directly used by humans, but its expression should be understandable by human users"; (ii) *accuracy*: "discoveries accurately portray the content of the database. The extent to which this portrayal is imperfect is expressed by measures of certainty"; (iii) *interesting patterns*: "discovered knowledge is interesting according to user-defined biases. In particular, being interesting implies that patterns are novel and potentially useful, and the discovery process in nontrivial"; (iv) *efficiency*: "the discovery process is efficient. Running times for large-sized databases are predictable and acceptable". Concerning 'high-level language' and 'efficiency', these two criteria are generally carried out by computer scientists or computational statisticians. Nevertheless, applied statisticians or other scholars from different disciplines such as sociology, economy, environmental science, will be, in general, more concerned with accuracy and interestingness of obtained patterns (results).

Maimon and Rokach (2010 [2005], 5-6) determine the taxonomy of data mining methods (see Figure 2 below).

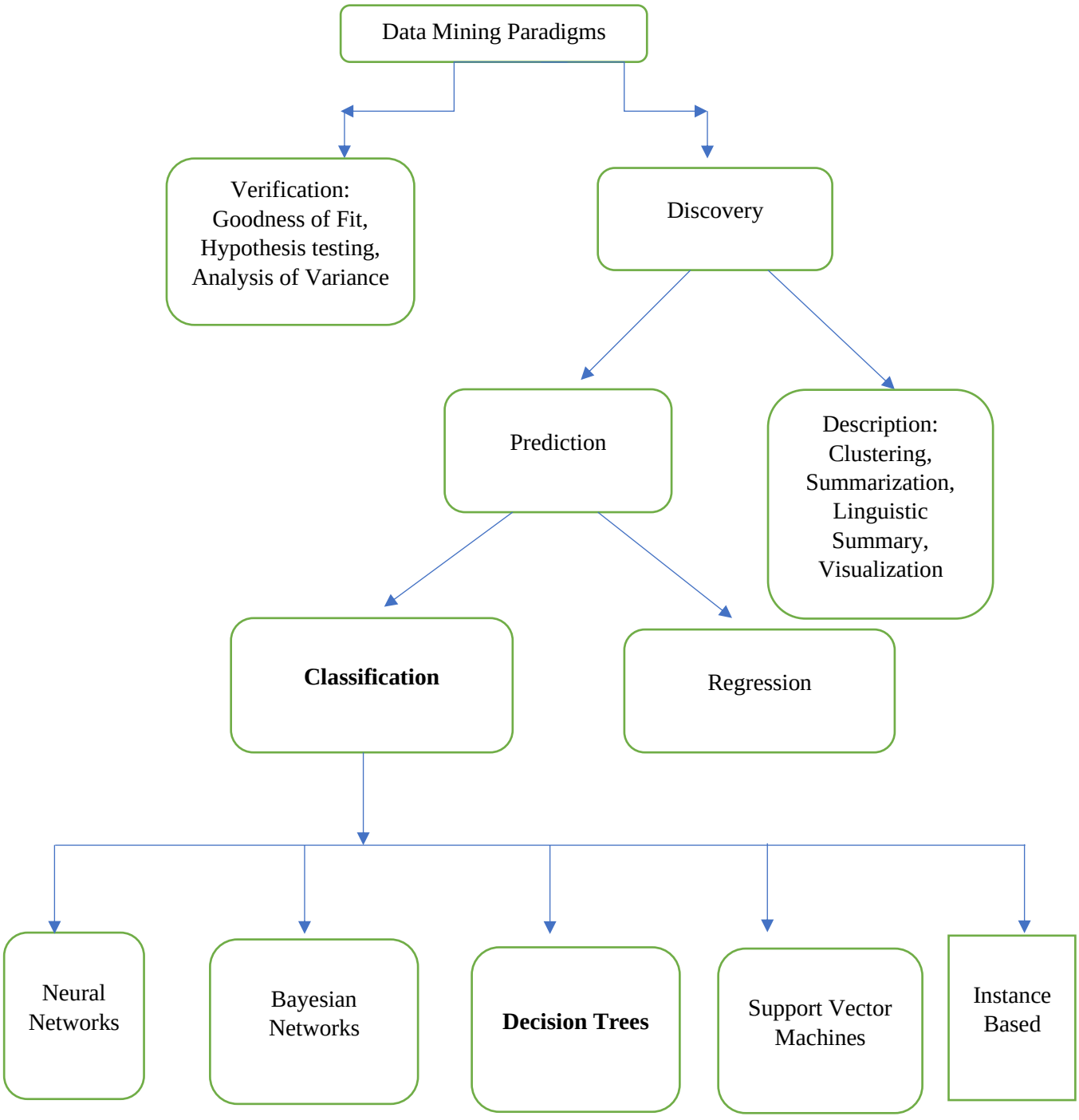


Figure 2. Taxonomy of data mining (from Maimon and Rokach, 2010 [2005], 5-6)

As Maimon and Rokach (2010, 5) mentions "there are two main types of data mining: *verification-oriented* (the system verifies the user's hypothesis) and *discovery-oriented* (the system finds new rules and patterns autonomously)." On the one hand, the verification-based data mining system's goal is to test a hypothesis proposed in the literature or by a researcher or simply by an expert. On the other hand, discovery-based data mining system's goal is to identify the hidden patterns in the data, which means that in opposition to the verification-based system, there is no hypothesis to test. The data analysis results help to build a model in discovery-based techniques; however, in verification-based models, the data analysis results help to verify if the hypothesis (null or alternative depending on the test) is rejected or not. Because of this

difference between these two types of data mining systems, the discovery-based ones are called as *evidence-based* tools. Among the evidence-based tools, there are two groups: descriptive tools vs. predictive tools. Descriptive methods show the properties of the data such as descriptive statistics, distribution of the variables, different visualizations (e.g., pie charts, bar charts multidimensional tables). Whereas predictive methods apply the induction in the data.

In the field of *machine learning*[109] (see Mitchell 1997, Langley 1996, Hastie et al. 2009 [2001], Russell and Norvig 2009 [1994]), there are four domains: supervised learning, unsupervised learning, semi-supervised learning, active learning (but we will only focus on supervised vs. unsupervised learning techniques). Machine learning techniques help to ameliorate the efficiency of the computers during the data mining process. Especially, when dealing with the large data sets, it is important that the discovery algorithm is performed automatically for capturing the complex patterns. If the machine learning process is better performed, this will create positive results on decision making process. This field comes originally from computer science, but it is also very related to statistics. For instance, as Witten et al. (2017 [2000], 31) mention "statistical tests are used to validate machine learning models and to evaluate machine learning algorithms."

Unsupervised methods do not include a response variable (i.e., target variable, predicted variable, class label attribute), the goal is simply to discover hidden clusters, groups among *instances* based on several variables included in the model- the aim is not to try to explain or predict a variable by other variables. At the end of the process, we can discover different classes as a function of the discovered clusters. There are no a priori labelled classes in unsupervised methods. For instance, suppose that there are 26 countries and 6 variables such as Gini index, monetary poverty headcount ratio, human development index, democracy index, annual economic growth rate, and unemployment rate. We would like to discover groups of countries based on these 6 variables. Cluster analysis, as an unsupervised method, will provide information on groups among 26 countries. Based on cluster results, we can form different classes of countries such as a specific typology. But it is important to note that cluster results do not inform us neither the impact of one variable on another nor the prediction results. Cluster results cannot be inferred for large population, the generalization is not possible. Hence, the results concern only the sample.

In opposition to unsupervised methods, supervised methods include a target response variable with *a priori defined* class labels. For instance, a response variable has three different categories. Each of these three categories is considered as a class (and labelled with a name) such as very satisfied, middle satisfied, or not satisfied at all. At the end of the data mining process, for each instance, a class is attributed based on the prediction. The response variable is predicted by other variables (i.e., predictors). Predictors are used to explain a phenomenon, which is indicated by the target response variable. Supervised machine learning techniques captive and model automatically the relationships between a set of predictors (i.e., descriptive features) and a target response variable[110] in the training data based on the data mining

[109] Kelleher et al. (2010, 3) define *machine learning* as "An automated process that extracts patterns from data." Witten et al. (2017 [2000], 4-6) describe machine learning as "mining knowledge from data".

[110] In the literature, a variable is also called as a *feature* or an *attribute*.

algorithm. Then, these relationships (can be considered as patterns[111]) are used to predict a class[112] for each observation in the test data. We will see later in more details in decision trees process section, but it is important to mention that the initial data is divided into two subsets: training data vs. test data. Training data is used for getting the patterns (i.e., prediction model). Test data is used for the prediction based on the patterns obtained from the training data. In other words, the patterns are obtained via the training instances, and then, the patterns are used to make predictions for the instances of the test data. Figure 3 depicts this process.

The classification and regression methods (see Figure 2 above) are both predictive tools; however, there is a difference between the two in terms of how they predict the response variable. Regression methods try to find the impact of predictors on the response variable in terms of ceteris paribus. Typically, if a predictor increases one unit[113], how the response variable will change. Classification methods try to predict a class of the response variable to each instance in the data and by using different types of visual tools such as tree or neural network we could see the patterns including interactions between the predictors (i.e., the prediction paths). However, both methods aim to bring out the hidden relationship/pattern in data. As we can see from Figure 2, there are five methods to classification: neural networks, Bayesian networks, decision trees, support vector machines, instance-based classification.

[111] Each pattern can be seen as a discovered model. Therefore, all patterns can be seen as the predicted models.

[112] Response variable is a categorical variable having different classes. For that reason, a class is predicted for each instance in the data. In case the response variable is quantitative, a number/value or a range of values is predicted for each instance in data.

[113] In statistical language, the term unit is used for two different things: (i) a variable is associated with a *measurement unit* such as km or kg. If the variable 'distance from school to house' is measured in terms of kilometers, then the unit of this variable is km; (ii) the *statistical unit* of the analysis means the analysis is done for who? The analysis can concern all units existing in the data such as individuals, or more specifically a group of units such as a specific community or some households.

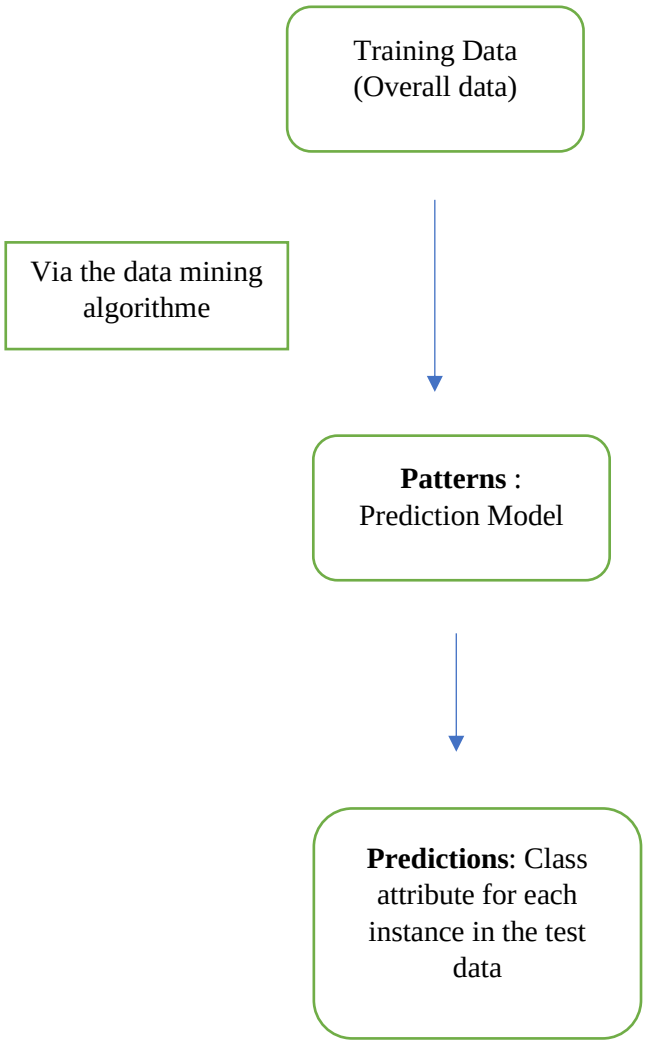


Figure 3. The process of prediction in supervised machine learning techniques.

Introduction to statistical terms: some basic principles of probability theory

In this sub-section, we will very briefly see some basic principles of probability theory, which will be useful for understanding the decision trees mechanism.

Probability theory

Probability is the branch of mathematics dealing with uncertainty, risk evaluation, and decision-making process[114]. We cannot always be sure about the occurrence of an event in the future such as we cannot know in advance the exact number of people will be admitting in a hospital. Probability calculations help us to understand the likely occurrence of an event and to take an optimal strategic decision under uncertain conditions. Despite its broad application in academic works, it still remains an abstract issue for many researchers in social and human sciences. Probably, the main problem is to find an easier way of connecting mathematics and social theories. However, even we don't necessarily notice, probability appears in many aspects in our life such as weather forecasting, insurance, lottery. At school, one of the most given examples of probabilities is tossing a coin. If we suppose that the coinage is straightforward, the probability of having a head is 50% as so for a tail. This way of looking at probability calculus

[114] See Ore (1960) for the invention of probability theory.

is considered as a *mechanic* approach (Crilly 2008, 124). But there is also another approach called *relative frequency*. In this second approach, an experiment such as tossing a coin is repeated for a number of times to look at the end whether head or tail is most obtained through all experiments. Another example is weather forecasting news that we follow almost every day. They are based on probability calculations. For instance, what is the probability of snowing for tomorrow? Basically, there are two options: it going to snow tomorrow, or it is not going to snow tomorrow. Nevertheless, in order to find out which option is more likely happen, we need more sophisticated calculations based on large historical data and advanced probabilistic models including various parameters for forecasting.

Crilly has briefly provided the history of probability in his book on *50 ideas of mathematics* (2008, 125; see also Rusell and Norvig 2010). The origin of mathematical modelling of probability comes from the discussions on gambling games between Blaise Pascal, Pierre de Fermat, and Antoine Gombaud (Chevalier de Méré) in the 17th century. Especially, in the 1650s, Pascal and Huygens have introduced the fundamentals of probability. In 1785, Condorcet has applied the probability theory in analyzing voting and jury systems. In 1812, Laplace's analytical theory of probabilities was published. In the early 1900s, Keynes had become known for his works on probabilities and its application in economics. And, in 1933, Kolmogorov pioneered the probability theory-based the axiomatic set theory. Crilly (2008, 125) has also mentioned an interesting discussion passed between Pascal, de Fermat, and de Méré. De Méré had asked: whether 'when we throw a dice four times, to obtain at least one 6', or 'when we throw a couple of dice 24 times, to obtain a 6-6', has a highest probability? So, which one is the best/rentable to put money on? At these days, the most common way was to put money on the '6-6' because there was more throw. However, probability calculus showed that it was not true because the option of '6' has in fact a higher probability. We can show it as follows (this following example is given on page 125 in the book of Crilly (2018)):

On the one hand, if we throw a dice once, the probability of not obtaining a '6' is 5/6. Herewith, if we throw a dice 4 times, as these events are independent (i.e., throwing a dice does not affect throwing a dice for second time), the probability of not obtaining a '6' will be $\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}\times\frac{5}{6}=\left(\frac{5}{6}\right)^4$. As a result, the probability of obtaining once a '6' at the end of 4 times throwing, will be equal:

$$1-\left(\frac{5}{6}\right)^4=0.517746.$$

given that the sum of all probabilities is equal 1.

On the other hand, if we throw two dices at the same time, the probability of not obtaining a '6-6' is 35/36. At the end of 24 times throw, the probability will be equal $\left(\frac{35}{36}\right)^{24}$. Hence, the probability of obtaining a least once a '6-6' will be equal $1-\left(\frac{35}{36}\right)^{24}=0.491404.$

given that the sum of all probabilities is equal 1.

Mosteller et al. (1961) mention that there are two main opposite approaches for interpreting and using probability. On the one hand, regarding the *objective* approach, probability can only be computed for the events that are *repeated* in the same conditions such as throwing a dice in the same conditions for several times. An objectivist does not interest in one-time events. Another

example is an electronical engineer working in a factory can check for the quality of televisions produced in the factory under the same conditions. The ratio of good quality televisions to the all produced televisions can be considered as the probability of producing good quality televisions given the producing conditions of the firm. Hence, probability can be seen as a conditional repeated number. In opposition, regarding the *subjective* approach, probability is seen as a measure of personal belief on a premise depending on personal knowledge and experience. In this perspective, a subjectivist will defend that there will be different personal degrees of belief on a likelihood of an event happens, for instance, expert's opinions (i.e., their own subjective probabilities). However, if the number of observations (i.e., sample) is high it is possible that subjectivists and objectivists obtain the same result(s). Moreover, Newbold et al. (2013, 101) report that there are three definitions of probability: classical probability, relative frequency probability, subjective probability. In line with Mosteller et al. (1961), we have already referred to both subjective and objective approaches. Regarding objective approach, as Newbold et al. (2013) evoke, we can also distinguish classical probability vs. relative frequency. The major difference between classical probability and relative frequency is that the latter concerns only a specific population; meanwhile, the former is a more general approach.

Probability calculations are generally realized for forecasting future events (i.e., predictive analytics). Prediction is based on the results of the repeated events as mentioned in paragraph above. For instance, calculating the probability of obtaining good quality televisions in the factory will help the firm to take optimal strategic decisions for different purposes such as satisfaction of clients, efficiency of production, evaluation of production process. Another example is forecasting weather. Based on the historical data of weather results in the past, meteorologists may predict weather of coming days. For decision makers in different areas, probability calculus becomes very important as the quality of probability models makes tremendous difference in decision process. If the probability model is good quality (i.e., explains adequately the phenomenon and based on cleaned and reliable data), then the prediction results will be more robust in terms of validity and reliability. Based on the fundamental properties of probability theory, different probability models can be developed addressing at specific issues (e.g., spread of an epidemic, get a disease, development of immunity to a disease). The probability models help to answer to various questions such as how long the disease may continue, what is the probability of ending the epidemic in city, or what is the probability of being vulnerable to the disease.

As Kelleher et al. (2015, 541) explain, it is very important to know the basic notions of probability (e.g., experiment, event, outcome, domain, sample space, experiment sample, random variable, domain of interest, variable, statistical model) to understand later the predictive analytics (e.g., decision trees). We start by explaining what does mean a random experiment. An **experiment** is a random process focusing on a phenomenon or phenomena (e.g., producing a television, throwing a dice) and leading to the **events** given that there is no a priori certitude on which event will occur. The events related to the experiment are considered its subsets. The design of an experiment is based on the construction of random variable(s) that are representing the experiment (generally, the names of random variables begin with an uppercase letter). In this sense, an **event** can be seen as the behavior of a random variable (e.g., throwing one dice) or the combination of random variables' behaviors (e.g., throwing two dices at the same time). For instance, Kelleher et al. (2010, 543) designate an event as an assignment of values to a random variable. A random variable is associated with the set of possible values that it can take on, which is called **domain**. Each event will be resulted with a value or a

combination of values. For each participant, the experiment has one result called **outcome.** So that the participant's outcome will be the set of her/his result(s) through associated event(s) (or an outcome can also be seen as the subset of the domain of the random variable). The set of outcomes correspond to all values obtained for all participants in the experiment. Moreover, the set of all possible outcomes is called the **sample space**, which are *mutually exclusive and exhaustive*. However, it is possible that the set of outcomes obtained from the experiment does not correspond totally to the sample space.

In light of this information, for measuring the likelihood (i.e., probability) of an outcome, we need an **experiment sample** and **random variable**(s). The **experiment sample** is the ensemble of individuals participated to the experiment. The **set of random variables** construct the **domain of interest,** which should be aligned with the nature of experiment. There are mainly 2 types of variables: **qualitative** vs. **quantitative (numeric)**. A **qualitative variable** takes *categories*, or called *modalities*, whilst a quantitative variable takes *values*. There are four types of **qualitative** variables: **binary** (a set of two values specifically coded as 0 or 1. Typically 0 for *no*, and 1 for *yes.* In machine leaning language, the terms 'true' is employed for 1 and 'false' is employed for 0), **dichotomous** (having only two categories such as male vs. female), **ordinal** (having ordered categories and the distances between the categories is unknown; for instance, satisfaction variable (Likert scale) with 5 categories: like, like somewhat, neutral, dislike somewhat, dislike. Value allows ordering but does not permit arithmetic operations), **nominal** (having several categories such as socio-professional groups. Each category can be labelled with a finite value, but these values cannot be ordered, and they do not allow arithmetic operations). On the other hand, a **quantitative (numeric)** variable can be **discrete** or **continuous**. For instance, a discrete random variable is the number of cars in a parking space, and a continuous random variable is the distance in km from school to home. A **discrete** variable takes on a finite or countably infinite number of values, whilst a **continuous** variable takes on an infinite number of values (i.e., any value in an interval, for instance fraction). Qualitative variables can be coded with (finite) discrete numbers in data, but these numbers do not have any mathematical function, it won't make any sense to make any algebraic calculus with them because the numbers are kind of synonym such as male is coded 1 and female is coded 2 in data. But, in the case of quantitative variables, the numbers are 'true numeric values' that allow arithmetic operations such as taking an average, median, summing, multiplying. In addition, even if the use of them is quite limited, there is also date (e.g., 22/08/2018), and textual variables (e.g., name of participants to the survey, such as Juliette). In machine learning, this typology of variables is often reduced to two categories: *continuous*, and *categorical*. The group *continuous* includes quantitative (numeric) and date type variables. The group *categorical* includes categorical, binary, dichotomous, nominal, ordinal, and textual type variables. Related to that, it is worth to mention that in machine learning language, the term *variable* can be replaced by the term *data, attribute,* or by the term *(descriptive) feature* (see Kelleher et al. 2015). In applied statistics or social indicators language, the term *variable* can be replaced by the term *indicator*. But the term *indicator* is more specific, qualified as indicating a nature or an extent of a specific phenomenon (e.g., child labour indicator). Moreover, there are **raw** vs. **derived** variables. Raw variables come directly from raw data sources. Derived variables do not normally exist in raw data source and they are computed/created based on raw variables. For instance, the variable age (as a numeric discrete variable) exists in the original data set as a raw variable, and the researcher/user can create another variable, derived variable, such as child (aged between 5 and 17) based on the raw variable age. Exceptionally, in some

national survey data sets, additional to raw variables some derived variables (computed directly by the national statistical offices for facilitating some computational issues) may exist as well. Or, the national survey data sets including some derived variables may be called as the derived national survey data sets, and the raw data (coming from the field) is not generally available to public and customers.

A **variable** is a quantity constructed for measuring a phenomenon, which can be seen as the *behavior or configuration*[115] of the object of study. The variable has changing values (reflects different behaviors) as a function of the question in study. In probability theory, variables are called **random variables**. The random variable(s) and the technique of calculus (e.g., probability) determined by the researcher correspond to the **statistical model** for measuring the object of study. According to the actions/preferences of participants to the experiment, the value/modality the variable takes may vary among participants[116]. Each obtained value/a combination of values/modality/a combination of modalities will be corresponding to an event in the experiment. Thus, researcher will do prediction (called *prediction task*) based on the observed behaviors of the participants on the object of study and based on the employed probabilistic technique.

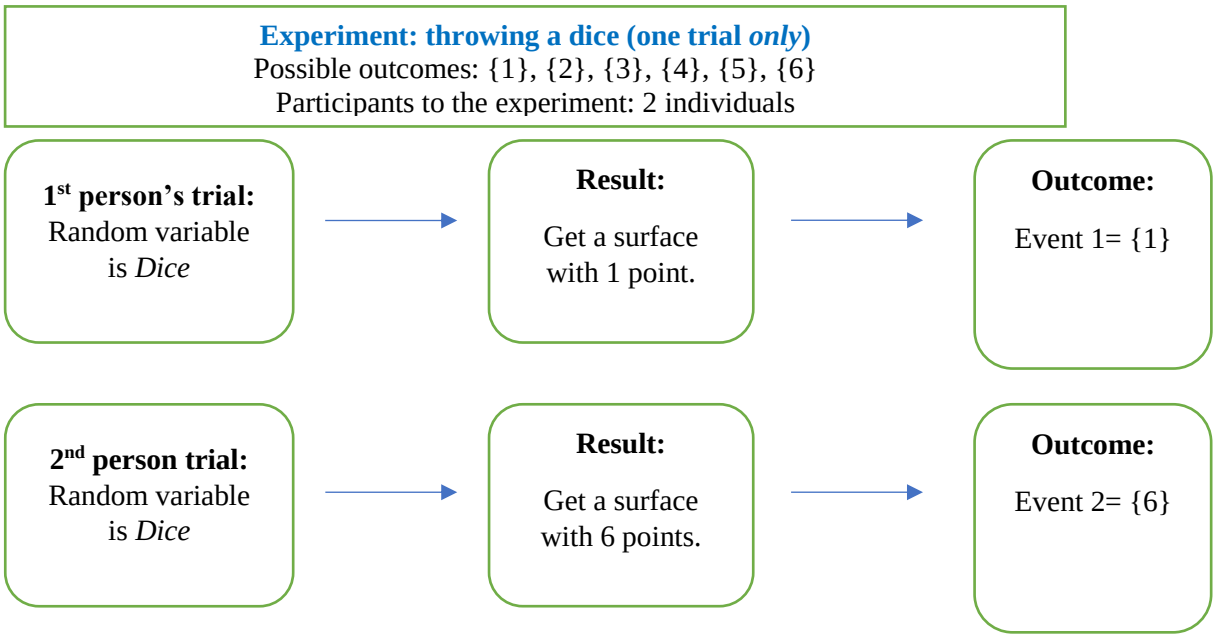


Figure 4. An example of the basic notions of probability

From Figure 4, we can see that the experiment is about throwing a dice with only one trial for each participant. The experiment is realized on the basis of a random process. Here, the random variable is *Dice* that represents the experiment. Depending on each participant's behavior, the random variable will take a value among six possible outcomes for each participant. Herewith, we can write that the domain of each *Dice* is {1, 2, 3, 4, 5, 6}. First participant gets the surface

[115] For instance, the behavior of a die or a configuration of poverty.
[116] Such as 'answering to survey question' or 'doing an action according to the framework of the experiment'.

with one point. Second participant gets the surface with six points. The results of two participants are independent because one's action does not depend on those of other. The results show that there are two events happened in this experiment.

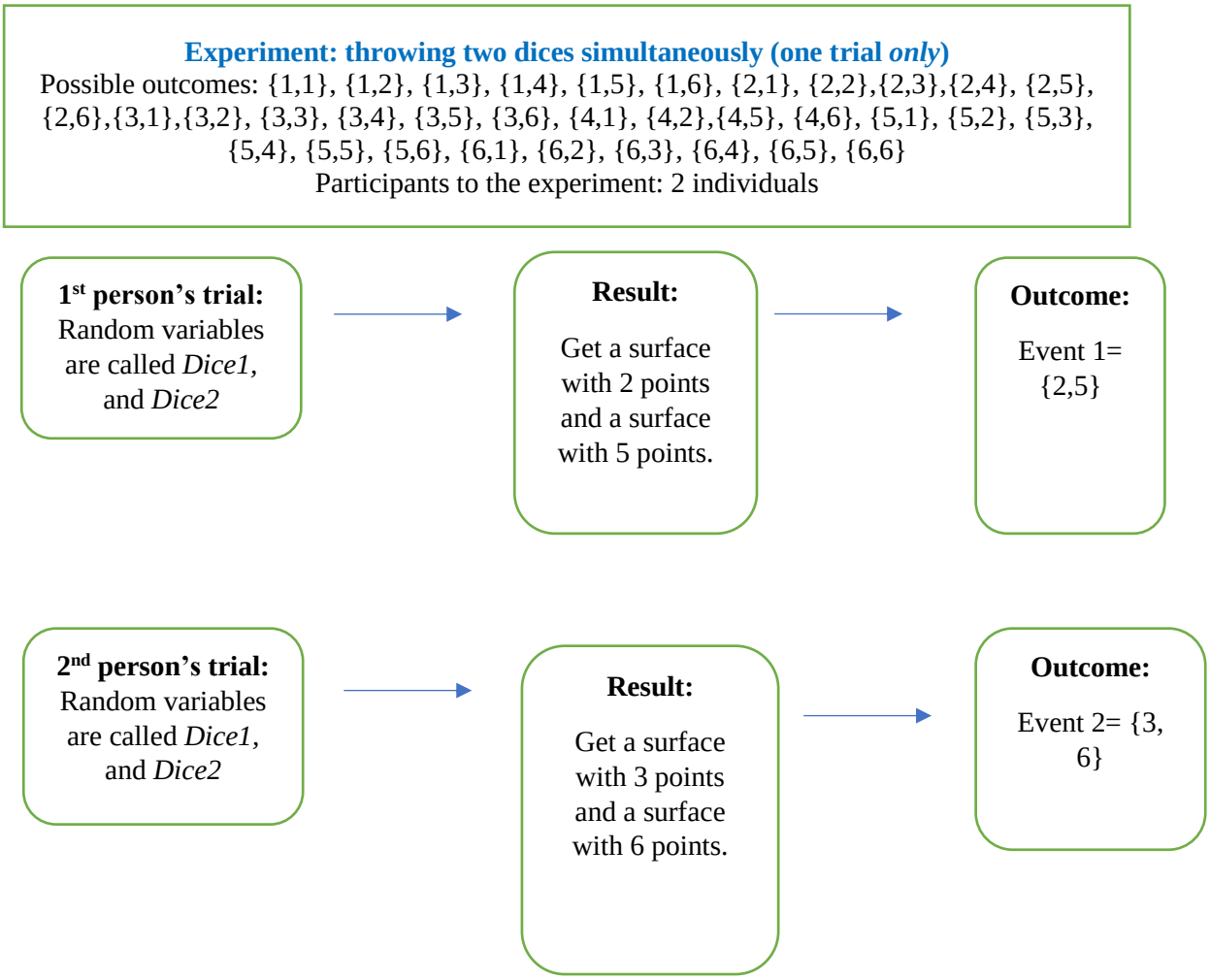


Figure 5. An example of basic notions of probability

From Figure 5, we can see that the experiment is about throwing two dices simultaneously with once trial possible for each participant. The experiment is realized based on random process. Two random variables called *Dice1*, *Dice2* represent the experiment. Depending on each participant's behavior and experiment's conditions, the random variables will take a value among six possible outcomes for each participant. The results of two participants are independent because one's action does not depend on those of other. The results show that there are two events happened in this experiment. Each event as a subset of the experiment is associated with a set of values.

It is crucial to define the accurate variable for better capturing the phenomenon. When a phenomenon is intangible such as poverty, social exclusion, it is more difficult to define the accurate variable(s) for measuring them. If a phenomenon is multidimensional (i.e., having different aspects simultaneously), more than one variable will be necessary for explaining the phenomenon. In addition, randomness is very important because the experiment conditions should not be defined a priori by researcher for obtaining a particular result such as any specific

result the researcher would like to obtain. The reliability of predicted results will be increasing when the researcher has a large sample and if he or she obtains similar results through several experiments.

For getting clear on probability calculus, we will see four fictitious examples as follows.

First example is: a researcher would like to compute the probability of being cancer in a community that he or she selected randomly. For this, first of all, he or she needs an enough sufficient and random[117] data including a large number of individuals[118] considered as the observations in the data, and a random variable according to her/his framework. The data is the experiment sample. Researcher selects randomly 100 individuals in the community and asks to each of them whether they are diagnosed with a cancer or not[119]. In this case, the experiment concerns only one phenomenon that is diagnosed having cancer or not, and as a result the domain of interest will be "Diagnosed with a cancer or not" (named as *Diagnosed cancer*). Two events are possible: have been diagnosed with a cancer or does not have been diagnosed with a cancer. If someone says 'yes' this means that that person has diagnosed with a cancer; but if someone says 'no' this means that that person has not diagnosed with a cancer. The scale of the variable is nominal with two possible categories of answer: yes or no. Hence, the type of outcomes will be yes or no. We suppose that an example of collected data for 10 persons is as follows:

Table 19. Example of data for 10 persons

Id (Identification number)	**Diagnosed cancer**
1	Yes
2	No
3	No
4	No
5	Yes
6	No
7	No
8	Yes

[117] Researcher will collect the data. The unit of her/his analysis will be individual. So, he or she has to gather a random data which means that the sample he or she will be collected should be based on a random process. Researcher should not select specifically some individuals because of their characteristics in order to avoid the subjectivism. The data should be collected objectively without having a pre-determined intention regarding a desired result.

[118] For a good prediction, at least 100 observations are suggested.

[119] In the statistics literature, there are many important documents concerning sampling issues (i.e., selection of sample and its techniques), collecting the data (i.e., how to collect the data and the time period for collecting the data), survey design (i.e., construction of questionnaires, time period defined in the questions such as the reference period for seeking a job). But here the purpose of our examples is just showing simple calculations of probability, so we do not enter into details about sampling issues specifically.

9	No
10	No

We note that the first variable Id in the first column is a unique identifier- to identify the individuals in the data, it is not used in the analysis. Of 10 people, 3 people says have been diagnosed with a cancer in the past, and the rest 7 people says not have been diagnosed with a cancer in the past. The statistical model is the simplest here: computing probability of diagnosed with a cancer by using one random variable. The probability computations will be done taking in consideration 100 people of the sample but just for example, concerning 10 people above, the probability of diagnosed with a cancer is 3 of 10 that is equal to 0.3. A probability takes a value between 0 and 1, with 0 means 0% possibility that the phenomenon happens, and 1 means 100% possibility that the phenomenon happens. For instance, 0.1 means a low probability, but 0.9 means a high probability.

Second example is a more sophisticated model than the first one. Supposing that researcher this time wants to develop her/his research question and he or she wants to predict the impact of a selected number of variables on the probability of being cancer; for instance, if air pollution becomes higher, does the probability of being cancer increase, decrease or not change. In this regard, the thematic of study is still being cancer vs. not being cancer. But the object of study extends including other variables for predicting their impact on being cancer. Distinctly to the first example, the experiment here includes more than one phenomenon- so, there are more than one variable. Researcher uses a cross-sectional data with the same individuals as mentioned in first example. The selected variables construct the domain of interest[120] that includes characteristics of individuals in terms of their socio-demographic, health, living situations. There are 7 random variables as follows: age, sex, having a chronic illness or not, degree of air pollution where the person spends most of her/his time, feeling often to be in stress or not, blood pressure, diagnosed as having cancer or not at the moment of gathering the data. Each individual in the data will have a value/modality for each variable. As a function of the responses, researcher will have the set of outcomes (i.e., the set of answers).

We give an example of 10 individuals:

Table 20. A fictitious example of a sample collected by a researcher for a cancer analysis

Id	**Age**	**Sex**	**Chronic illness**	**Air pollution**	**Feel stress**	**Blood pressure**	**Diagnosed with a cancer**
1	23	F	Yes	Low	No	High	Yes
2	79	M	No	High	Yes	Ideal	No
3	18	M	No	Low	Yes	Ideal	No
4	26	F	Yes	High	Yes	Low	No
5	34	M	No	High	Yes	Ideal	Yes
6	42	M	No	High	Yes	Ideal	No

[120] The set of possible outcomes regarding each variable is the domain of interest for each variable.

7	92	F	No	Medium	No	Low	No
8	85	F	No	Low	Yes	Pre-high	Yes
9	65	F	No	Medium	Yes	High	No
10	45	F	Yes	High	No	Low	No

We suppose that the researcher has a sample of 100 individuals, but we just focus on the case of 10 individuals. Table 21 above shows the distribution of variables for 10 individuals. The variable age is a quantitative variable having a value. Other variables are categorical variables having a modality. For instance, having a chronic illness is an event. There are in total 10 events regarding age, 2 events regarding sex (being female or being male), 2 events regarding chronic illness, 3 events regarding air pollution, 2 events regarding feel stress, 4 events regarding blood pressure, and 2 events regarding diagnosis cancer. Furthermore, there are also other subsets as a combination of values and/or modalities/categories such as having age of 23 and being female.

Third example is supposing that researcher would like to compute the probability of obtaining the 6 dots face and the 2 dots face when he or she throws two dices at the same time. Here, the experiment is throwing 2 dices at the same time. Researcher extends the sample and applies the experiment to 100 people under the same conditions (e.g., same dices). In that case, at the end of the experiment, researcher constructs a data base including 100 individuals at the row and 2 random variables at the columns. The variables called as the domain of interest are Dice 1 and Dice 2. Dice 1 represents the event of throwing the first dice. Dice 2 represents the event of throwing the second dice. Each individual participating to the experiment will have a result for the first dice and a result for the second dice. The results for each variable will be appeared in the columns. Therefore, based on the results regarding Dice 1 and Dice 2, researcher will be able to find the probability of obtaining the 6 dots face on one dice and the 2 dots face on other dice at the same time. The experiment is applied 100 times as there are 100 individuals in the data. An example (as shown by Kelleher et al. 2010, 542[121]):

Table 21. Example of throwing two dices

ID	**Dice 1**	**Dice 2**
1	3	4
2	1	5
3	6	5
4	3	3
5	1	1

In this example, there are many subsets that we call *events*. The events are: Dice1= {3}, Dice2={4}, Dice1={3} and Dice2={4}, Dice1={1}, Dice2={5}, Dice1={1} and Dice2={5}, Dice1={6}, Dice2={5}, Dice1={6} and Dice2={5}, Dice1={3}, Dice2={3}, Dice1={3} and Dice2={3}, Dice1={1}, Dice2={1}, Dice1={1} and Dice2={1}.

[121] In the example, authors use the figures for representing dices. But here instead we use the numbers to show which side of dice is obtained.

Fourth example is we suppose that there is a nominal variable corresponding to the event 'Traveling in Paris' with four different categories, and we suppose that only one modality among the four is possible for each interviewed person in the sample. The four modalities of this variable are: travelling Paris in a snowy day (E_1) or travelling Paris in a rainy day (E_2) or travelling Paris in a cloudy day (E_3) or travelling Paris in a sunny day (E_4). Modalities may also be associated with a discrete number such as 1, 2, 3, and 4 respectively. But also, each modality is associated to an event, hence the variable 'travelling in Paris' is a set of four possible events (i.e., The event 'traveling Paris is composed of four subsets). These four events are *mutually exclusive and exhaustive*. This means that among these four possibilities, only one can happen for each individual. Each event *E* is associated with a probability between 0 and 1. According to the properties of the probability mass function, in a case of travelling Paris, we obtain:

$P(E_1) + P(E_2) + P(E_3) + P(E_4) = 1.$

Herewith, based on the basic axiom of probability theory, we can formularize the *probability model* as follows:

$$0 \leq P(E) \leq 1 \; for \; every \; E \; and \; \sum_{E \epsilon \Omega} P(E) = 1 \; .$$

where the Greek letter Ω (uppercase omega) refers to the sample space (space of all possible events) given that each event of the sample belongs to the sample ($E \epsilon \Omega$), $P(E)$ refers to the probability that an event in the sample space realizes. We can see that the probability of an event $P(E)$ is between 0 and 1, and the sum of probabilities equals to 1.

After introduced the basic notions of probability, we continue with probability calculations. Related to this, we note that in the terminology of predictive analytic such as CART, generally, instead of variable, the term *feature* is often employed. Herewith, in order to find out the likelihood of a feature taking any particular value/modality (or any particular combination of values/ and/or modalities/categories), the **probability function** is used. The probability function, generally, denoted as P(E) where the notation P means *probability* and the notation E means *event*. In other words, probability function $P()$ yields the probability of an event- P defines a **probability distribution** for the random variable. We can call it as the probability that the event happens (see also equation above). For instance, if the event is getting ill, P(ill=true) indicates the probability of the ILL variable taking the value *true*. It is important to note that, in machine learning language, most frequently the Boolean language yes or no is used. Here, yes means that the event happens. In statistical literature, quantitative variables (continuous, discrete) are said taking values and qualitative variables (categorical, dichotomous, binary, nominal, ordinal) are said taking modalities/categories. But, in machine learning language, generally the term value is used for both qualitative and quantitative variables.

The probability distribution shows all possible values of the variable and the probability associated to each value. First, we will focus on **probability distribution for a discrete random variable** $P(x) = P(X = x), for \; all \; values \; of \; x$, representing the probability that the random variable X takes the value x. The probability distribution function, called **probability mass function,** yields the probabilities of the random discrete variable takes on for all the possible outcomes. Distribution of probabilities can be shown graphically, or by table, or algebraically. The probability function, for an event, takes a value between 0 and 1. It

corresponds to the relative frequency of that event in the dataset. The sum of the probabilities over the set of events must equal one. Herewith, there are two fundamental properties of probability distribution for discrete random variables:

Suppose X be a discrete random variable with probability distribution $P(x)$. Then,

1. $0 \leq P(x) \leq 1$ for any value of X,

2. the individual probabilities of all possible values of x sum to 1, $\sum_x P(x) = 1$. Based on this property, we can say that the events are *mutually exclusive and collectively exhaustive*.

We will explain it through an example of probability of travelling outside of the country in a year (see Table 23 below).

Table 22. An example of probability distribution of a discrete random variable

Probability of travelling outside of the country in a year				
	Travel= Number of travelling			
	0	**1**	**2**	**3**
Probability distribution	0.1	0.5	0.3	0.1
Cumulative probability distribution	0.1	0.6	0.9	1

In this example, the random variable *Travel* represents travelling outside of the country in a year. Regarding its distribution, it takes four possible values revealing four different events in terms of number of travelling. The row ***probability distribution*** shows the probability value of an event. We can see that the probability of travelling at least once in a year outside of the country is 50%. The probability of travelling two times in a year outside of the country is 30%. Besides, the probability of not travelling or the probability of travelling three times in a year outside of the country is 10%. For instance, the probability of not travelling outside the country can be shown as P(Travel=0) =0.1. Or, the probability of travelling at least once in a year outside the country can be denoted as P(Travel=1) =0.5. We can say that the highest probability is likely to be travelling once outside of the country in a year.

In addition, based on probability distribution, we can also compute the probability of more than one event such as the probability of travelling once and two times a year outside of the country, which is obtained by summing the probability of travelling once a year outside of the country and the probability of travelling two times a year outside of the country: P (Travel=1 or Travel=2) = P(Travel=1) + P(Travel=2) = 0.5 + 0.3 = 0.8.

The row ***cumulative probability distribution*** $F(x_0)$ shows the probability that the random variable is less than or equal to a particular value, $F(x_0) = P(X \leq x_0)$. For instance, the probability of travelling at most two times a year outside of the country is P(Travel<=2) = 0.9. This number is the sum of probabilities of no travelling (0.1), once year travelling (0.5), and two times a year travelling (0.3). This can be shown as:

$$F(x_0) = \sum_{x \leq X_0} P(x).$$

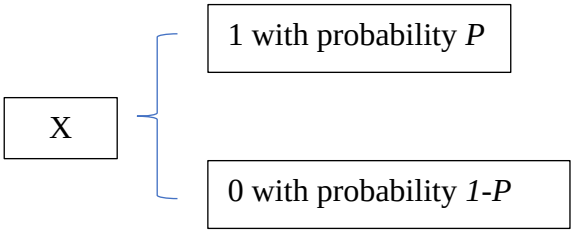
Based on the fundamental properties of discrete random variables, there are two fundamental properties of cumulative probability distributions for discrete random variables:

Given X s the discrete random variable and $F(x_0)$ is the cumulative probability distribution of X, then

1. $0 \leq F(x_0) \leq 1$ for each x_0, and

2. if $x_0 < x_1$, then $F(x_0) \leq F(x_1)$.

In the case of discrete probability distribution, **Bernoulli distribution**[122] is very important. The Bernoulli distribution concerns the probability distribution of a *binary* random variable. A binary random variable takes two values: 0 or 1. Consider a random experiment with a random variable X yielding two possible *mutually exclusive and collectively exhaustive* outcomes, namely 'success' and 'failure':



where P corresponds to the probability of success, and *1-P* corresponds to the probability of failure. Then, the probability distribution of this random variable is:

$P(X = 0) = (1 - P)$ and $P(X = 1) = P$.

We showed probability distribution of a discrete random variable. But we know that a random variable might be continuous as well. A **continuous random variable** can take on a *continuum* of values. As discrete random variable, continuous random variable has also probability and cumulative distribution functions. The probability is summarized by the **probability density function,** which is shown by a graphic where "the area under the probability density function between any two points is the probability that the random variable falls between those two points" (Stock and Watson 2007 2nd ed., 21). Suppose X is a continuous random variable taking a range of values x. The probability density function, $f(x)$, of the random variable fulfills three following properties:

1. $f(x) > 0$ for all values of x,

2. The area under the probability density function is equal to 1.0.

3. Suppose a and b are two possible values of the random variable X, with $a < b$. Then, the probability that X takes a value between a and b will be equal to the area under the probability density function between the points a and b. This probability is shown as:

[122] Bernouilli distribution is showed by Swiss mathematician and scientist Jacob Bernouilli at the 17th century.

$P(a \leq x \leq b) = \int_a^b f(x)dx$ (see Figure 6)

4. The cumulative distribution function, $F(x_0)$, corresponds to the area under the probability density function from x_m to x_0, where x_m is the minimum value of X. This can be shown as:

$F(x_0) = \int_{x_m}^{x_0} f(x)dx$.

There are different types of distribution such as uniform, normal, exponential, but we don't enter into detail here.

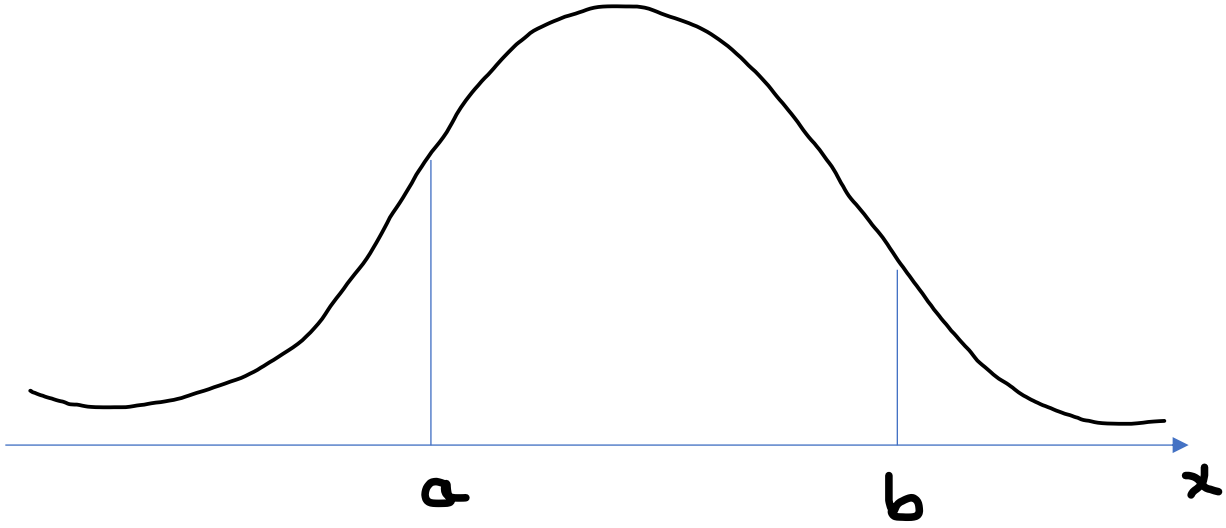


Figure 6. Probability density function

Figure 6 plots the probability density function of a normal distribution. We generated randomly 100 observations taking a value in the intervalley of [-4, 4], in the R statistical software. In *x* axis, the values of random variable exist and in *y* axis the density of each value is defined.

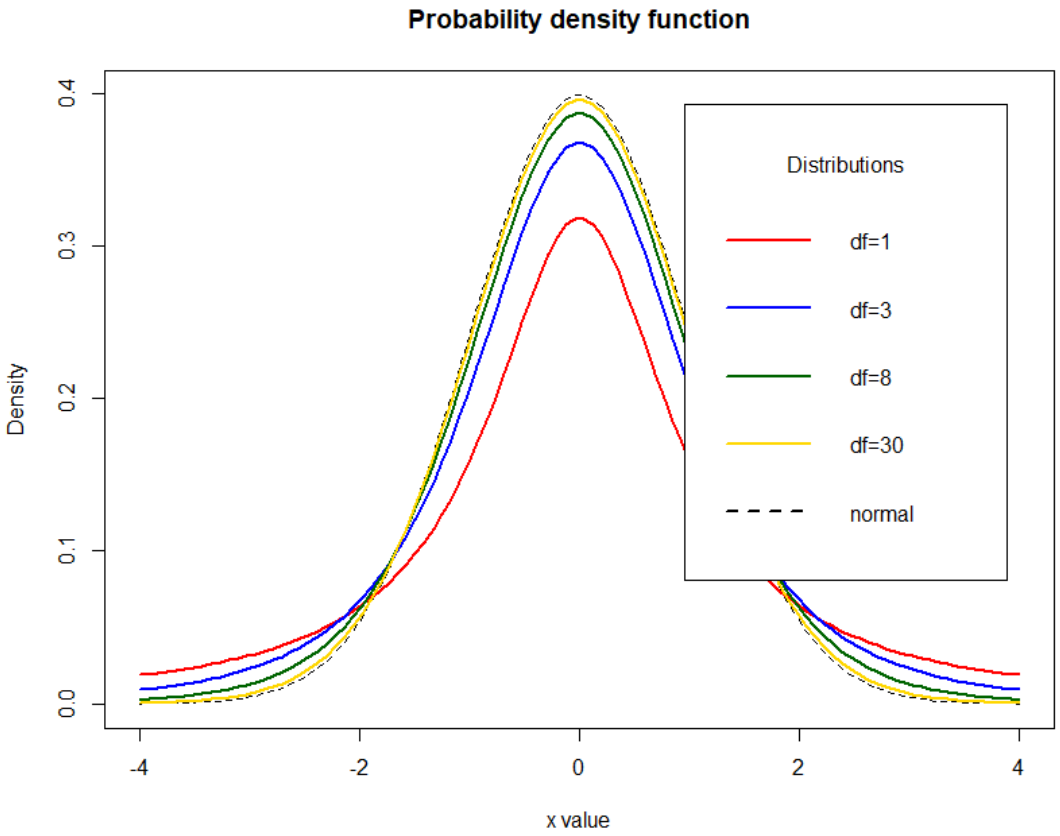


Figure 7. The probability density function of a normal distribution based on randomly generated 100 observations

We continue with the **probability mass function** through some examples for explaining the calculations of (simple) unconditional probability (prior probability), joint probability, marginal probability, conditional probability (posterior probability) respectively. This basic probability calculus is the fundamental for all other more sophisticated probability computations used in predictive analytics.

First of all, we will give the formula of classical probability. Classical probability is "the proportion of times that an event will occur, assuming that all outcomes in the sample space are equally likely to occur." (Newbold et al. (2013, 101). We obtain it by "dividing the number of outcomes in the sample space that satisfy the event by the total number of outcomes in the sample space". (Newbold et al. (2013, 101).

For an event X, its probability to be happened is:

$$P(X) = \frac{N_X}{N} \qquad \textit{Equation 1}$$

where N represents the total number of outcomes in the sample space, and N_X is the number of outcomes that satisfy the event X. (Newbold et al. (2013, 101).

Related to the classical probability, there are also the concepts of permutations and combinations. In this thesis, we don't use these two concepts in predictive analytics that we will carry on further. For that reason, we don't explain them.

Now, we will focus on the *relative frequency*[123] that covers probability computations regarding a particular population. The computation process is the same as those of classical probability. Suppose that there is a small data set composed of 20 randomly selected individuals, as showed in the table below, which is generally called as a *training set*.

Table 23. An example of classification and probabilities

ID number	**Age interval**	**Origin**	**Hobby**	**Season**	**Travelling**
Ind1	18-24	Europe	Museum	Winter	Paris
Ind2	18-24	Asia	Sport	Summer	Buenos Aires
Ind3	40-50	Asia	Art gallery	Winter	Buenos Aires
Ind4	51-60	Asia	Painting	Spring	London
Ind5	18-24	North America	Museum	Summer	Paris
Ind6	25-30	South America	Museum	Summer	Paris
Ind7	25-30	Europe	Shopping	Summer	London
Ind8	25-30	Europe	Shopping	Spring	London
Ind9	18-24	Asia	Shopping	Autumn	London
Ind10	18-24	South America	Art gallery	Winter	Buenos Aires
Ind11	51-60	Asia	Shopping	Winter	Tokyo
Ind12	40-50	Asia	Museum	Winter	Paris
Ind13	31-39	Asia	Museum	Spring	Paris
Ind14	18-24	Europe	Painting	Summer	Tokyo
Ind15	18-24	North America	Shopping	Summer	Tokyo
Ind16	25-30	Asia	Art gallery	Autumn	Buenos Aires
Ind17	18-24	North America	Shopping	Winter	London

[123] Newbold et al. (2013, 106) mention that "the relative frequency probability is the limit of the proportion of times that event A occurs in a large number of trials, n, $P(A) = n_A/n$ where A is the event, n_A is the number of A outcomes, and n is the total number of trials or outcomes." General probability approach is used for explaining theoretically the concept pf probability. However, the relative frequency of probability (which we use in this thesis) is the same process as the general probability, but it is applied in a specific data focusing on a specific population. As Newbold et al. (2013, 106) mention "…the relative frequency approach often can be obtained using existing data sources." The term n designates the population concerned in the data. The probability is the limit when n becomes larger.

Ind18	51-60	Asia	Sport	Spring	Buenos Aires
Ind19	25-30	Europe	Art gallery	Summer	Tokyo
Ind20	25-30	Europe	Museum	Autumn	Paris

As we can see from Table 24 above, there are 5 variables (age interval, origin, hobby, season, travelling). Variables have a number of categories: age interval (18-24, 25-30, 31-39, 40-50, and 51-60), origin (Asia, Europe, North America, South America), hobby (art gallery, museum, shopping, sport, painting), season (autumn, winter, spring, summer), travelling (Buenos Aires, London, Paris, Tokyo). Based on this training set, first of all, prior probabilities will be computed. The prior probability is the relative frequency of an event such as the probability of travelling to Buenos Aires, or to London, or to Paris, or to Tokyo (it concerns only one event and there is no other conditions).

Here, we have categorical variables. As in the case of discrete random variables, **probability distribution of a categorical random variable** is yielded by **probability mass function.** For instance, the probability of travelling to Buenos Aires is obtained as: the frequency of travelling to Buenos Aires divided by the total number of individuals in the training set. In the dataset, first, we count the number of rows where the variable *Travelling* takes the modality of Buenos Aires and then we divide this number by the total number of rows in the dataset. 2nd, 3rd, 10th, 16th, and 18th individuals in the dataset wish to visit Buenos Aires. In this light, the probability of travelling to Buenos Aires corresponds to:

$$\text{P (Travelling=Buenos Aires)} = \frac{|\{\text{Ind2,Ind3,Ind10,Ind16,Ind18}\}|}{|\{\text{Ind1,Ind2,Ind3,...,Ind20}\}|} = \frac{5}{20} = 0.25.$$

where (i) Ind2, for instance, refers to the instance in the dataset with a unique ID number of 2, and so on; (ii) the numerator is the number of times that the event happens in the data set-which is indicated by ID number of individuals fulfilling the event; (iii) the denominator is the total number of instances in the data set- which is indicated by ID numbers of all persons in the dataset from the 1st individual until the last person respectively such as Ind1,.,Ind20. This type of probability is called (**simple) unconditional probability**.

After having seen the basic probability computation, before to process other types probability calculus, we recall the fundamental probability postulates and the probability rules.

The probability postulates are the necessary conditions that the probability calculus must take in consideration. We assume that there are i number of events appeared in a random experiment. We denote X for an event in the sample space S and O for the outcomes of the events. For each event E, its probability $P(E)$ must fulfill following the two postulates:

1. $0 \leq P(E) \leq 1$ meaning that for an event in the sample space, its probability must have a value between 0 and 1. 0 means that the event will not happen, 1 means that the event certainly be happened.

2. $P(S) = 1$ meaning that the sum of all probabilities in the sample space should be equal to 1.

There are two main probability rules: (i) complement rule; (ii) addition rule of probabilities. These rules are necessary to known for computing probabilities of compound events.

If A is an event in the sample space, and as we know that the sum of all probabilities in the sample space is equal 1, the probability of the complement of the A will be equal $1 - P(A)$. For instance, when we throw once a die, the probability of obtaining the surface with 2 points (A) is $P(A) = 1/6$. Herewith, based on the complement rule, the probability of not obtaining the surface with 2 points is $1 - P(A) = 5/6$.

When there are two events such as A and B *mutually exclusive* (see the set theory in Mosteller et al. 1961), and see Figure 8), the probability of the two events happen (i.e., we denote the union as $P(A \cup B)$) simultaneously is equal $P(A \cup B) = P(A) + P(B)$. But suppose that two events are not mutually exclusive, so there is an intersection between A and B (see Figure 9).

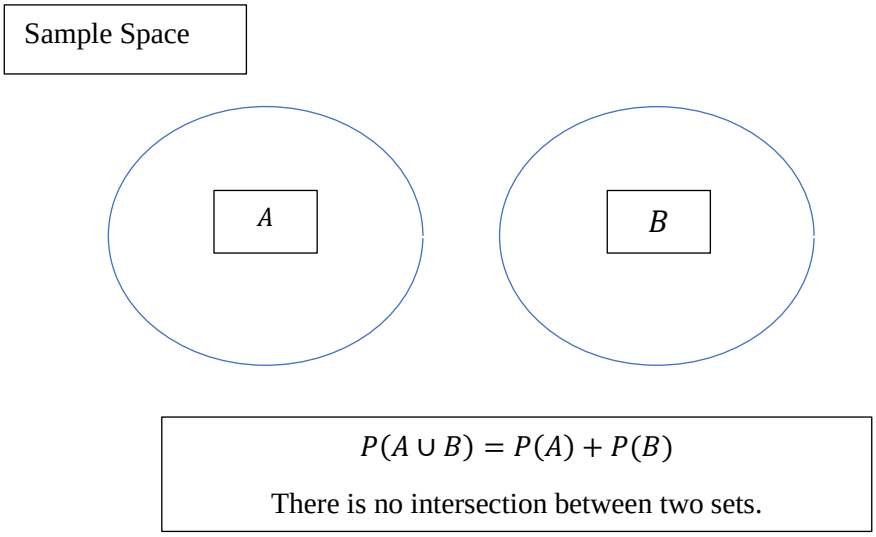


Figure 8. Union of two events in the case of no intersection between them by Venn diagram

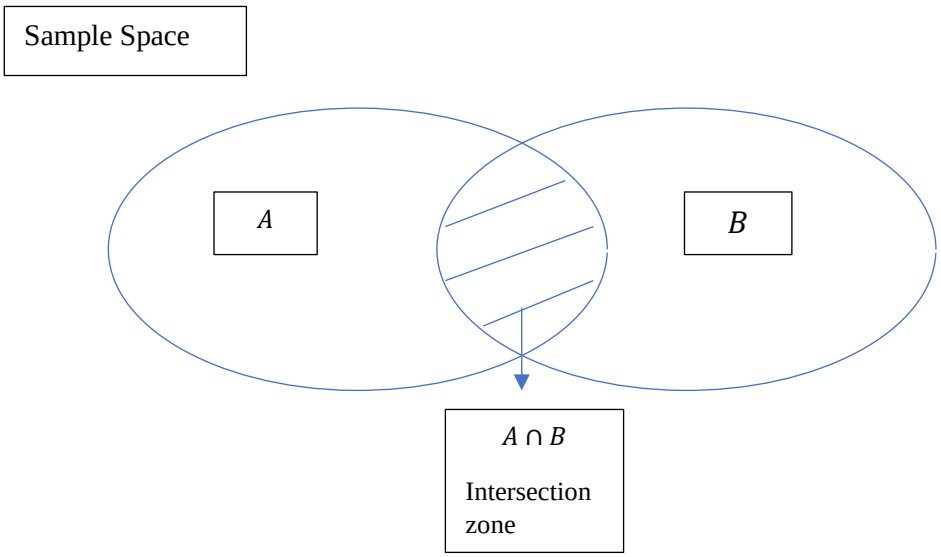


Figure 9. Intersection of two events shown by Venn diagram

The addition rule of probabilities takes in consideration the intersection between two events in the sample space. Let X and Y be two events and there are some common outcomes between them. In that case, the probability of their union will be:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

where the term $P(A \cap B)$ means the probability of intersection between two events.

We can also be interested in probability of more than one event happens at the same time (i.e., simultaneous occurrence of two or more events). This type of probability is called **joint probability**. As Stock and Watson mentions (2007, 29) "the joint probability distribution of two discrete random variables, say X and Y, is the probability that the random variables simultaneously take on certain values, say x and y. The probabilities of all possible (x, y) combinations sum to 1." Generally, the joint probability distribution is shown as $P(X = x, Y = y)$ where X , Y are the random variables, and the lowercase x, y are the possible values that the random variables take on. In terms of set theory, joint probability corresponds to the intersection of at least two events (the intersection of 'bivariate events') denoted as $X \cap Y$. The intersection $X \cap Y$ occurs if and only if both the two events X and Y happened at the same time, which embodies the set of outcomes common to the X and Y. So, the joint probability is used for expressing the probability of the intersection of X and Y, $P(x, y) = P(X = x \cap Y = y)$. This is why, joint probability is also known as *intersection probability* in the literature.

The joint probability distributions of discrete random variables, given X and Y are two discrete random variables with joint probability distribution $P(x, y)$, has two fundamental properties:

1. $0 \leq P(x, y) \leq 1$ for any pair of values x and y,

2. the joint probabilities over all possible pairs of values should sum 1.

For instance, there are two events that we are interested in such as being travelling to Tokyo and like to visit art galleries as hobby. When there are two events are the object of interest, we can call it as *joint event*. In such case, we have to find the number instance where these two events happen at the same time. Herewith, the joint probability corresponds to the relative frequency of the joint event in the dataset. In our example above, we see that the 19th individual in the dataset fulfills this condition. However, 11th, 14th, and 15th individuals in the dataset would like to also travel to Tokyo but their hobbies are not the visiting art galleries, so they are not included in the computation of this joint probability. We can compute the joint probability of travelling to Tokyo and like to visit art galleries as follows (two events are separated by a comma in the function):

P (Travelling=Tokyo, Hobby=Art gallery) = $\frac{|\{\text{Ind19}\}|}{|\{\text{Ind1,Ind2,Ind3,…,Ind20}\}|} = \frac{1}{20} = 0.05.$

where (i) Ind19, for instance, refers to the instance in the dataset with a unique ID number of 19, and so on; (ii) the numerator is the number of times that the joint event happens in the dataset- which is indicated by ID number of individual fulfilling the joint event; (iii) the denominator is the total number of instances in the data set- which is indicated by ID numbers of all persons in the data set from the 1st individual until the last person respectively such as Ind1, Ind20.

But of course, it is possible that the intersection of two events might be an empty set (i.e., no common outcomes between the considered events, shown as $X \cap Y = \{\ \}$. For instance, if two events X and Y have no common outcomes, they are called *mutually exclusive*.

So, as we have seen the intersection probabilities are called joint probabilities $P(E_1 \cap E_2)$, which focus on the intersection of more than one event. Concerning, single events, the probabilities for individual events such as $P(E_1)\ or\ P(E_2)$, are called **marginal probabilities**. Marginal probabilities are traditionally put at the margin of the table. We obtain them by summing the probability values in corresponding row or column. To show marginal probabilities, we will extract a subset of data mentioned in Table 24 above. The subset data is shown in Table 25 below:

Table 24. The effective, joint and marginal distribution of a pair of discrete random variables X, Y (extracted from Travelling data- see Table 24)

	Buenos Aires (B=1)	London (B=2)	Paris (B=3)	Tokyo (B=4)	**Total**
Asia (A=1)	4 (4/20)	2 (2/20)	2 (2/20)	1 (1/20)	**9 (9/20)**
Europe (A=2)	0 (0/20)	2 (2/20)	2 (2/20)	2 (2/20)	**6 (6/20)**
North America (A=3)	0 (0/20)	1 (1/20)	1 (1/20)	1 (1/20)	**3 (3/20)**
South America (A=4)	1 (1/20)	0 (0/20)	1 (1/20)	0 (0/20)	**2 (2/20)**
Total	**5 (5/20)**	**5 (5/20)**	**6 (6/20)**	**4 (4/20)**	**20 (20/20)**

In Table 25 above, rows represent the origin of tourists in terms of some regions in the world, columns represent the city to be visiting by tourists. The table shows the distribution of tourists as a function of their origin (the random variable A having four different categories) and the city that they would like to visit (the random variable B having four different categories). The full joint distribution is a 4×4 table and the total number of tourists (N) is 20. The joint probabilities (of A and B) appear in parenthesis in the $4 \times 4\ square$. We note that the joint probabilities sum to 1, conforming to the fundamental axioms of probability theory. The marginal probabilities appear in parenthesis along the right and bottom margins. The marginal distribution yields the probabilities of the values of the random variables in the subset without being conditioned to the values of the other random variables. This situation is the inverse of conditional distribution.

For instance, on the one hand, the joint probability of being European and travelling to Tokyo is:

$P(A = 2\ \cap\ B = 4) = 2/20 = 0.1$.

On the other hand, the **marginal probability** of being European is obtained by summing the probabilities of being European for each possible value of the Travelling random variable. In

other words, we obtain the marginal probability for the event Europe by summing the corresponding mutually exclusive joint probabilities:

$$P(A = 2) = P(A = 2 \cap B = 1) + P(A = 2 \cap B = 2) + P(A = 2 \cap B = 3) + P(A = 2 \cap B = 4).$$

thus,

$$P(A = 2) = \frac{0}{20} + \frac{2}{20} + \frac{2}{20} + \frac{2}{20} = \frac{6}{20} = 0.3 .$$

we note that $P(A = 2)$ can also be shown as $P(A_2)$.

In formal terms, the marginal probability distribution of the random variables X, and Y are shown as follows:

$$P(x) = \sum_{x} P(x, y)$$

$$P(y) = \sum_{y} P(x, y)$$

Now, we will explain **posterior probability** that is also known as **conditional probability,** which is denoted as $P(X = x_i \mid Y = y_i)$ for each possible i, j pair. In this type of probability, there are two terms: (i) an event that we want to know its probability to be happened; (ii) event(s) that that we know have already happened. So, **conditional probability** corresponds to probability of an event happens *given that* one or more other events are known to have happened. For instance, we would like to find the probability of travelling to Buenos Aires *given that* the traveler likes to visit art galleries during her/his travels as a hobby. So, to find it, we have to divide the number of individuals travelled to Buenos Aires and like to visit art galleries during travel as hobby by the number of individuals like to visit art galleries during travel as hobby, as follows:

$$\text{P (Travelling=Buenos Aires | Hobby= art gallery)} = \frac{|\{\text{Ind3,Ind10,Ind16}\}|}{|\{\text{Ind3,Ind10,Ind16,Ind19}\}|} = \frac{3}{4} = 0.75.$$

where (i) the vertical bar symbol " | " means the expression of *given that*, and the term appearing after the vertical bar represents the condition to be necessarily fulfilled that is called as *given event*; (ii) Ind3, for instance, refers to the instance in the dataset with a unique ID number of 3, and so on; (iii) the numerator is the number of times that the event of travelling to Buenos Aires and liking to visit art galleries during travel as hobby happens in the dataset- which is indicated by ID number of individual fulfilling that event; (iv) the denominator is the number of times that the event of like to visit art galleries during travel as hobby happens in the data set- which is indicated by ID numbers of individual fulfilling that given event in the data set. The numerator designates an intersectional event of 2 events (i.e., intersection of 'event' and 'given event') that are (1) travelling to Buenos Aires; and (2) liking to visit art galleries during travel as hobby respectively. And the denominator expresses the given event. Comparing to two other examples of prior probabilities above, in the computation of conditional probability, the denominator is *not* the whole data set, but it corresponds to the population of the given event.

Above, we computed conditional probability given a one variable, but we can also compute conditional probability given the combination of variables (i.e., by taking in consideration more

events at the same time as given events). For instance, what is the probability of travelling to Buenos Aires given season is summer and age-interval is 18-24? To obtain it, we divide the number of 18-24 yrs. aged individuals travelled to Buenos Aires during summer by the individuals 18-24 yrs. aged and travel in summer. We can formulate it as:

P (Travelling=Buenos Aires | Season=summer and Age interval=18-24) = $\frac{|\{\text{Ind2}\}|}{|\{\text{Ind2,Ind5,Ind14,Ind15}\}|}$

= $\frac{1}{4}$ = 0.25.

where (i) Ind2, for instance, refers to the instance in the dataset with a unique ID number of 2, and so on; (ii) the numerator is the number of times that the event of travelling to Buenos Aires during summer and being between 18 and 24 yrs. old happens in the dataset; (iii) the denominator is the number of times that the event of travelling in summer and having an age between 18 and 24 yrs. happens in the data set. The numerator designates an intersectional event of 3 events (i.e., intersection of 'event' and 'two given events'). And the denominator expresses the intersection of two given events.

Based on the example above, we can write the general formula of conditional probability. Let X and Y be the two events. The conditional probability of the event X given that the event Y has occurred (i.e., denoted as $P(X|Y)$) can be found as:

$P(X = x| Y = y) = \frac{P(x,y)}{P(Y=y)}$ *if and only if* $P(Y = y) > 0$. *Equation 2*

From the equation 2 we can write that

$P(x, y) = P(X = x \mid Y = y)P(Y = y)$ *Equation 3*

The equation 3 shows the multiplication rule of probabilities or known as the **product rule**. We can read this equation as follows: for X and Y to be true, Y must be true, and also X must be true given Y.

From predictive data analytics perspective, we can think that for example the feature *travelling* is the predicted target variable and other variables are descriptive features. We predict that there are four types of class in the training set as a function of the responses on travelling (i.e., travelling to Buenos Aires, London, Paris, or Tokyo). It is possible to find the probability of most likely classification given a feature value. For instance, the probability of most likely travelled class given the season is summer can be found through:

P (class=Buenos Aires | season=summer), corresponding to 'the probability that the class is travelling Buenos Aires given that the season is summer', equals to:

= $\frac{|\{\text{Ind2}\}|}{|\{\text{Ind2,Ind5,Ind6,Ind7,Ind14,Ind15,Ind19}\}|}$ = $\frac{1}{7}$ = 0.14.

P (class=London | season=summer), corresponding to 'the probability that the class is travelling to London given that the season is summer', equals to:

= $\frac{|\{\text{Ind7}\}|}{|\{\text{Ind2,Ind5,Ind6,Ind7,Ind14,Ind15,Ind19}\}|}$ = $\frac{1}{7}$ = 0.14.

P (class=Paris | season=summer), corresponding to 'the probability that the class is travelling to Paris given that the season is summer', equals to:

$$= \frac{|\{Ind5, Ind6\}|}{|\{Ind2, Ind5, Ind6, Ind7, Ind14, Ind15, Ind19\}|} = \frac{2}{7} = 0.28.$$

P (class=Tokyo | season=summer), corresponding to 'the probability that the class is travelling to Tokyo given that the season is summer', equals to:

$$= \frac{|\{Ind14, Ind15, Ind19\}|}{|\{Ind2, Ind5, Ind6, Ind7, Ind14, Ind15, Ind19\}|} = \frac{3}{7} = 0.42.$$

As we can see from these conditional probability results, there is a one dominant class for travelling during summer based on the training set, which is the city Tokyo. The probability of travelling Buenos Aires, and London during summer is the same. Hence, the most likely classification is travelling Tokyo given the season is summer. We can also compare these posterior probability results with the prior probability. In such case, the prior probability is 'the probability of travelling to Buenos Aires, or to London, or to Paris, or to Tokyo' as follows:

P (travelling=Buenos Aires) = 5/20 = 0.25

P (travelling=London) = 5/20= 0.25.

P (travelling=Paris) = 6/20 = 0.30.

P(travelling=Tokyo) = 4/20=0.20.

Based on these simple unconditional probability results, the most likely travelled city is Paris among the 4 cities. But we note the concept of prior probability and posterior probability are different. Therefore, the results are different as well.

Moreover, in machine learning language (e.g., computer programming language), the Boolean random variable is generally used, which only takes two possible values: *true* or *false* (see George Boole on algebraic system of logic)- the domain is {*true*, *false*}[124]. When, the value is *false*, the synonym ¬ is used. In addition, the intersection term is generally expressed with the synonym ∧ . As an example, we suppose that there are three random variables: *Weather* with four possible categories (sunny, rain, cloudy, snow), *Cinema* with two possible categories (yes, no), *School* with two possible categories (school, no school). Suppose that the probability that the child goes to cinema, given that the weather is cloudy, is 0.2. We can write it as follows:

$P(Cinema = yes \mid Weather = cloudy) = 0.2$.

Another example, the probability that the child goes to cinema, given that the weather is not cloudy, is 0.8, can be written as:

$P(Cinema = yes \mid \neg\, cloudy) = 0.8$.

If we use an example with two conditions like the probability that the child goes to cinema, given that the weather is not cloudy and there is no school at that day, is 0.3, can be written as:

$P(Cinema = yes \mid \neg\, cloudy \wedge no\ school) = 0.3$.

Statistical independence

[124] The values are always lowercase (see Russell and Norvig 2009 [1994]).

Let X and Y be the jointly distributed random variables. They will be referred *independent* if and only if their joint probability distribution is the product of their marginal probability distributions:

$$P(x \cap y) = P(x)P(x)$$ *Equation 4*

Or it can be shown also as:

$$P(x, y) = P(x)P(y)$$

Equation 5

And k random variables are independent if and only if

$$P(x_1, x_2, \dots, x_k) = P(x_1)P(x_2) \dots P(x_k)$$

If the random variables X and Y are independent, the conditional probability distribution of X, given Y will be the same as the marginal probability distribution of X:

$$P(x_i|y_j) = P(x_i) \ \ if \ P(y_j) > 0$$

$$P(y_j|x_i) = P(y_j) \ if \ P(x_i) > 0$$

Theorem of Bayes

Theorem of Bayes, which is developed by Thomas Bayes (1702-1761), is based on conditional probability (originally it is published in 1763, but it is reprinted in 1958-see Bayes 1958). Notably, many of works in the field of decision theory in the post second world war area, are developed based on Bayes' original works. We will explain the Bayes' theorem through a fictitious example. Suppose that cholera epidemy is spread in a tropical city of Southern America during the 19th century. Researchers are particularly interested in how identifying an ill person at the early stage of cholera. They think that the fever is the most important indicator. However, they know that it does mean that every person having fever has a cholera. To analyze it, they collect information in a village. In their sample, there are 120 people. Among them, 40 of them has already cholera, 65 of them declared that they had fever, and the rest 15 people had no cholera neither fever during the reference week of collecting the data. Besides, among 40 cholera patients, only 5 of them do not have fever; and, among people having fever, only 35 of them have cholera. We visualize the situation with a Venn diagram as follows (where 15 people do not appear because they do not belong neither cholera set either fever set).

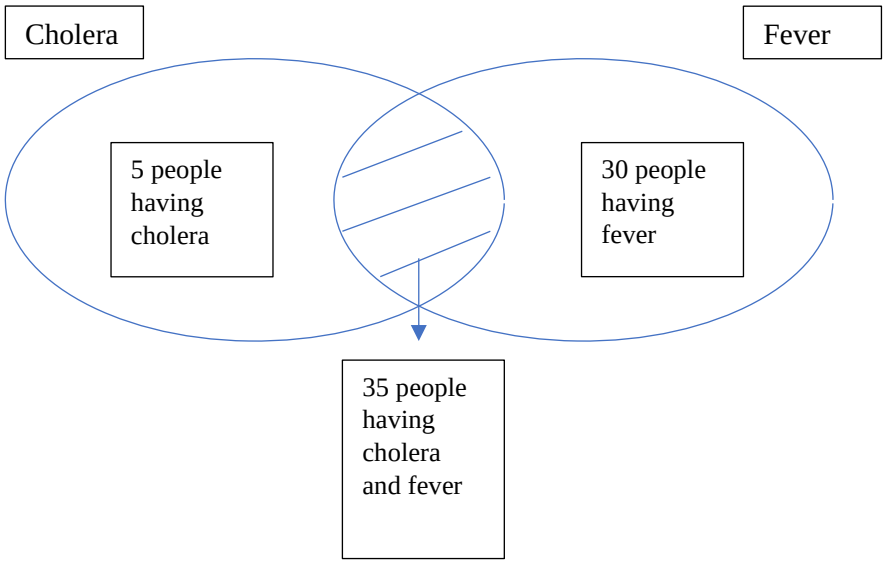


Figure 10. Example of Bayes theorem with the set illustration

We denote the number of people having cholera as *Cholera*, the number of people having fever as *Fever*, and the number of people having both cholera and fever as *X*. In total, there are 35 people having cholera and fever at the same time (*X*=35), 40 people having cholera (*Cholera*=40), and 120 people in the sample (*n*=120). Hence,

The probability of having cholera will be:

$$P(Cholera) = \frac{cholera}{n} = \frac{40}{120} = 0.33. \qquad \textit{Equation 6}$$

The probability of having fever will be:

$$P(Fever) = \frac{fever}{n} = \frac{65}{120} = 0.54. \qquad \textit{Equation 7}$$

The probability of having both cholera and fever at the same time will be:

$$P(Cholera, Fever) = P(x) = \frac{x}{n} = \frac{35}{120} = 0.29. \qquad \textit{Equation 8}$$

Researchers can also obtain two conditional probabilities:

(i) the probability of having fever given that the person has cholera:

$$P(Fever\ |Cholera) = \left(\frac{x}{cholera}\right) = \frac{35}{40} = 0.875. \qquad \textit{Equation 9}$$

(ii) the probability of having cholera given that the person has fever:

$$P(Cholera|\ Fever) = \left(\frac{x}{fever}\right) = \frac{35}{65} = 0.538. \qquad \textit{Equation 10}$$

As we can see, both the meanings and the results of these two conditional probabilities are different. But if we know one of them, the **formula of Bayes** then will help us to find the value

of other conditional probability. Before showing the formula of Bayes, we start by computing the joint probability of having cholera and fever at the same time.

We can find the joint probability of having both cholera and fever at the same time, by using the equation 1 and the equation 4, as follows:

$$P(Cholera, Fever) = P(Fever \mid Cholera) \times P(Cholera) = \left(\frac{x}{cholera}\right) \times \left(\frac{cholera}{n}\right) = \left(\frac{x}{n}\right) = \frac{35}{40} \times \frac{40}{120} = \frac{35}{120} = 0.29.$$

Equation 11

We can also obtain this joint probability by using the equation 2 and the equation 5, as follows:

$$P(Cholera, Fever) = P(Cholera \mid Fever) \times P(Fever) = \left(\frac{x}{fever}\right) \times \left(\frac{fever}{n}\right) = \left(\frac{x}{n}\right) = \frac{35}{65} \times \frac{65}{120} = \frac{35}{120} = 0.29.$$ *Equation 12*

As we can see, the results of the equations 10 and 11 are the same. As the results are the same, we can equalize the two formulas of the joint probability:

$$P\,(Fever, Cholera) = P(Fever \mid Cholera) \times P(Cholera) = P(Cholera \mid Fever) \times P(Fever)$$ *Equation 13*

Then, based on equation 8, we can obtain the **Bayes formula**:

$$P(Cholera \mid Fever) = \frac{P(Cholera)}{P(Fever)} \times P(Fever \mid Cholera)$$ *Equation 14*

As we have seen the Bayes formula (see equation 13) gives the inverse relation between the two conditional probabilities. For instance, if researchers know already the probability of having fever given the patient has cholera, they can also find the probability of having cholera given that the patient has fever by using the Bayes formula. Particularly, the Bayes theorem is known as a useful tool for updating the beliefs (i.e., subjective approach-based probabilities) as a function of the proves (see Crilly 2008, 131).

After the example above, we can also provide the formal development of Bayes' theorem. Suppose that there are two events based on two random variables A and B. Based on the equation 12, we can write that:

$$P(A \cap B) = P\,(A \mid B)\; P\,(B) = P(B \mid A)\; P\,(A)$$

Equation 15

Then, Bayes' theorem states that

$$P(B \mid A) = [P(A \mid B)\; P\,(B)] \;/\; P(A)$$

Equation 16

and

$$P(A \mid B) = [P(B \mid A)\; P(A)] \;/\; P(B)$$

Equation 17

In another example, we suppose that there are two random variables satisfaction of clients (Satisfaction), price of goods (Price) determined in a market research study. The variable Satisfaction has four categories: very satisfied, middle satisfied, few satisfied, not at all satisfied. The variable Price has three categories: high price, middle price, low price. The table below shows the probability distributions :

Table 25. The effective, joint and marginal distribution of a pair of discrete random variables Price, Satisfaction

	Very satisfied (VS)	Middle satisfied (MS)	Few satisfied (FS)	Not at all satisfied (NS)	**Total**
High price (HP)	12 (12/108)	6 (6/108)	8 (8/108)	12 (12/108)	**38 (38/108)**
Middle price (MP)	10 (10/108)	8 (8/108)	6 (6/108)	12 (12/108)	**36 (36/108)**
Low price (LP)	8 (8/108)	4 (4/108)	12 (12/108)	10 (10/108)	**34 (34/108)**
Total	**30 (30/108)**	**18 (18/108)**	**26 (26/108)**	**34 (34/108)**	**108 (108/108)**

In the table above, the joint probabilities appear in parenthesis and the marginal probabilities appear in parenthesis along the right and bottom margins. The marginal probabilities are as follows: P (VS)=0.277, P(MS)=0.166, P(FS)=0.240, P(NS)=0.314, P(HP)=0.351, P(MP)=0.333, P(LP)=0.314. We want to find the probability that the consumers buy middle price products *given that* they are very satisfied, $P(MP \mid VS)$.

Regarding Bayes' theorem, we can write that:

$$P(MP \mid VS) = [P(VS \mid MP)\, P(MP)] \,/\, P(VS)\,.$$

Equation 18

Based on equation 14, we can find that:

$$P(VS \mid MP) = \frac{P(VS \cap MP)}{P(MP)} = \frac{0.0925}{0.333} = 0.277\,.$$

Herewith, $P(MP \mid VS) = \frac{[0.277 \times 0.333]}{0.277} = 0.333\,.$

Of course, $P(MP \mid VS)$ could be obtained by using the equation 14:

$$P(MP \mid VS) = \frac{P(MP \cap VS)}{P(VS)} = \frac{0.0925}{0.277} = 0.333\,.$$

But also, the equation 17 can be developed as:

$$P\,(MP \mid VS) = [P(VS \mid MP)\, P(MP)] \,/\, [P(VS \mid HP)\, P(HP) + P(VS \mid MP)\, P(MP) + P(VS \mid LP)\, P(LP)]\,.$$

Equation 19

where

$$P(VS) = P(VS \cap HP) + P(VS \cap MP) + P(VS \cap LP) = P(VS \mid HP)\,P(HP) + P(VS \mid MP)\,P(MP) + P(VS \mid LP)\,P(LP)\,.$$

Equation 20

The conditional probabilities $P(VS \mid HP), P(VS \mid MP), P(VS \mid LP)$ can be find based on equation 14:

$$P(VS \mid HP) = \frac{P(VS \cap HP)}{P(HP)} = \frac{0.111}{0.351} = 0.316\,.$$

$$P(VS \mid MP) = \frac{P(VS \cap MP)}{P(MP)} = \frac{0.0925}{0.333} = 0.277\,.$$

$$P(VS \mid LP) = \frac{P(VS \cap LP)}{P(LP)} = \frac{0.074}{0.314} = 0.235\,.$$

So, we can obtain $P(MP \mid VS)$ by using the equation 18:

$$P(MP \mid VS) = \frac{0.277 \times 0.333}{[(0.316 \times 0.351) + (0.277 \times 0.333) + (0.235 \times 0.314)]} = \frac{0.0922}{0.110 + 0.0922 + 0.073} = \frac{0.0922}{0.2752} = 0.33\,.$$

Statistical method: classification trees

In the following sub-sections, we will explain the classification trees method we used in the thesis by focusing on mechanisms of decision trees, entropy function, information gain, gini index of diversity, misclassification, pruning tree, predictive accuracy of classifier, independent test set, k-fold cross-validation process, n-fold cross-validation process, leave-one-out cross-validation, stratified cross-validation, evaluation measures, confusion matrix, missing values, advantages of classification trees, and statistical software available for applying classification trees, respectively.

Mechanism of decision trees

Decision trees is a top-down, predictive, heuristic, non-parametric classification[125] method allowing to *partition recursively*[126] the training data set[127] into smaller as homogeneous as

[125] In the literature, there are also bottom-up methodologies. However, Rokach and Maimon (2010, 151) mention that there is a 'clear preference' for the top-down approach. In this sentence, the approach *top-down* means that the data mining process (i.e., the algorithm of a decision tree inducer) is based on discovering iteratively the paths hidden in the data (see Appendix 9.3- section of data mining and discovery of knowledge for the meaning of hidden path). The analysis becomes more complex each time a new hidden path is identified. It is important to mention that the terms *top-down* and its opposite *bottom-up* are also used for specifying the type of data analysis whether the theory is based on the results or the results allow to build a theory: (i) the *top-down* approach designates that a researcher builds a statistical model based on an already determined theory- the data analysis results aim to show different characteristics of the theory depending on research questions; (ii) but, the *bottom-up* approach is used for mentioning that a researcher conceptualizes a phenomenon or constructs a theory based on the results obtained from the data analysis. We can say that for instance a tree-based clinical research by a set of collected variables through the patients' information regarding a specific illness may help to conceptualize a theory in terms of pattern of the illness (i.e., which variable(s) contribute to the likelihood of the illness).

[126] Rokach and Maimon (2010, 151) label the *recursive portioning* as "divide" and "conquer".

[127] Data set can be called as the *instance space*.

possible (data) subsets/samples. This method is employed in the fields of data mining, machine learning, pattern recognition, and artificial intelligence (see Loh 2014 for historical perspective, and see Moon and Lee 2017, Yannakoulia et al. 2016, Tyrovolas et al. 2012, Kuhn et al. 2014, Chen et al. 2011, Xiaogang et al. 2011, Zhang and Ye 2008, Pino-Mejias et al. 2008, Lewis 2000, Yohannes and Hoddinott 1999, Moisen 2008, De'ath 2002, Marshall 2001, De'ath and Fabricius 2000, Vayssieres et al. 2000, Hess et al. 1999, Yohannes and Webb 1999, Zhang and Singer 1999, Baker 1993 for different applications of decision trees). Han et al. (2012 [2000], 385) define classification as "a form of data analysis that extracts models describing data classes. A classifier, or classification model, predicts categorical labels (classes)". The first decision tree algorithm (more specifically, *regression tree*), namely the *Automatic Interaction Detection* (AID onwards), is proposed by Morgan and Sonquist in 1963 (cited in Loh 2014). Later, in 1972, Messenger and Mandell have built a new classification algorithm called *Theta Automatic Interaction Detection* (THAID) based on the AID (cited in Loh 2014). Nevertheless, decision trees have been significantly promoted and developed by the two groups of researchers approximately at the same time, during the 1980s. On the one hand, in 1984, four Californian statisticians, Breiman, Friedman, Olshen, and Stone (Breiman et al.) have published the book *Classification and Regression Trees* (CART) that proposes a binary decision tree algorithm. On the other hand, between the 1970s and early 1980s, Australian researcher in machine learning, Quinlan (1986, 1987, 1988, 1992) has developed a decision tree algorithm (i.e., ID3- *Iterative Dichotomiser*) for producing classification trees. Quinlan later has developed another algorithm called C4.5. The two algorithms CART and C4.5 are considered as the fundamental of decision tree induction. Each decision tree methodology has its own specific algorithm, namely the *decision tree inducer* or *decision tree induction* that allows to construct decision tree(s) from data set (Rokach and Maimon 2010)[128]. Decision trees' '*inductive inference*' algorithm is information-based allowing to "determine which descriptive features provide the most information about a target feature and make predictions by sequentially testing the features in order of their informativeness" (Kelleher et al. 2015, 120; see also information theory Shannon and Weaver 1949)[129]. Due to its 'inductive inference algorithm', the decision trees method is also known as a *learning* method.

The method of decision trees has two main goals:

[128] As Han et al. (2012 [2000], 347) have mentioned the ID3, C4.5, and CART algorithms are more suitable for relatively small data sets, but still national micro data sets can be fit in this category. Especially for the "very large real-world databases" generating "very large training sets of millions of tuples", the efficiency of these algorithms becomes problematic in terms of memory-the training data exceed memory (i.e., the *scalability* of these algorithms). In the literature, other methods such as AVC-set ("Attribute-Value, Classlabel"), BOAT ("Bootstrapped Optimistic Algorithm for Tree construction") are proposed; but we won't use them in this research (for more information, see Han et al. (2012 [2000]).

[129] Claude Shannon (1916-2001) is considered as the pioneer of information theory. He worked for AT&T Bell Labs, where he developed information theory based on his works on encoding messages for telephone communication.

- first is to show the patterns of the target response variable[130] as a function of several descriptive (predictor) variables, based on extracted information from the training set[131]. Especially, for large data sets, it is very difficult to discover combinations of variables simultaneously taken into account for explaining a response variable[132] by using traditional statistical methods such as regression, descriptive statistics. For instance, when we do a regression analysis, the impact of an independent variable on the target response variable, ceteris paribus, is obtained[133]. An interaction term composed of two variables can be also integrated into the model. But compared to decision trees, we cannot discover other available most important interactions existing in the data set. In regression, we confine ourselves to an interaction term pre-determined by the researcher and whether it is statistically significant or not. In other words, regression models define in an ex-ante approach the variables to be used for explaining the phenomenon. To discover in an ex-post approach which variable(s) are the most influential[134] including interactions, regression techniques remain not sufficient. For that reason, the decision trees method emerges as an efficient *data compression* process for uncovering hidden patterns, and (complex) interactions between variables in the data. This process makes easier data analysis when the data is complex and big sized including substantial number of variables.
- Second, decision trees yield the prediction of a response variable based on a classification process. A pattern corresponds to a *prediction path* of a phenomenon based on a number of descriptive variables. In decision trees, therefore, a pattern corresponds to a prediction path (or also called model) of the response variable based on one predictor or on a combination of predictors. This second process, therefore, helps to categorize/classify future data sets.

In this thesis, we will use the 'CART' methodology (see Breiman et al. 1984). The CART handles both with quantitative and qualitative types of target response variables. When the response variable is qualitative, it is called *classification* trees; when the response variable is quantitative, it is called *regression* trees. However, we will work only with qualitative type of

[130] Response variable represents a phenomenon. For instance, a binary variable such as being child labour vs. not being child labour corresponds to a child labour phenomenon. And, in machine learning language, it is called '*query instance*'. Moreover, machine learning language employs often the term 'instance' for describing the data set (as a row in data). For response variable, the term *output*, and for descriptive variable the terms *input*, or *explanatory* are also used in the literature.

[131] The initial data set is generally divided into two samples: training set vs. test set. First of all, the computations are done in the training set, secondly, the prediction of classification is done based on the test set. We will describe in more detail the notions of training and test sets in next sub-sections. For moment, in following examples, we just assume that the given data is the training data for facilitating the comprehension of the decision trees mechanism including splitting rules and attribution of classes.

[132] We use the two terms 'target response variable' and 'response variable' interchangeably throughout the document.

[133] In regression analysis, we interpret only the linear impact of the statistically significant variable on the dependent variable, by assuming other independent variables constant. If and only if the researcher/user determines a priori an interaction term in the model, then its significance can be found at the end of computations. This means that, in opposition to decision trees, there is no a posteriori discovery of interactions among the variables.

[134] To find out the most influential variables, first step is that the decision trees algorithm takes in consideration all independent variables in the model at the same time. Then, at second stage, based on some criteria (see next paragraphs), the most important ones are selected to do classification.

target response variable. So that we will focus on classification trees and we won't explain the technical details of regression trees[135]. Before proceeding in detail, we start off with a general introduction to the mechanism of classification trees. Then, we will address more specific technical topics in building classification trees, such as entropy function, information gain, Gini index of diversity, pruning trees, training and test sets, cross-validation process predictive accuracy of classifier, confusion matrix, missing values, advantages of decision trees, respectively.

Introduction to classification trees

In classification trees[136], the target response variable[137] (called also as *output*, or *categorical dependent variable*) is qualitative and descriptive variables (called also as *predictors*, or *inputs)* can be both qualitative and quantitative. But often, classification trees are applied in cases where the *target response variable* takes only *two distinct and mutually exclusive modalities/categories* (e.g., *Boolean classification*). Suppose that a marketing manager of a cosmetics department store wants to classify the client satisfaction on eye make-up products as a function of four descriptive variables. The target variable is defined as the client satisfaction at three levels such as very satisfied, medium satisfied, few satisfied. As descriptive variables, price, brand, type of product, and place made (categorical variables) are selected. By employing classification trees method, the marketing manager expects to find out the distinguished patterns separately for each satisfaction level. Suppose that the variable *brand* is identified as the most important one which best distinguishes three classes of client satisfaction. In addition to *brand,* price or another descriptive variable may bring some additional information to differentiate classes. For instance, it can be found that if the brand is A, clients are very satisfied; if the brand is B and price is between 15 and 25 euros, clients are middle satisfied; and if the brand is C and price is between 5 and 10 euros, clients are few satisfied. Such results can be helpful for the marketing manager to design new sale campaigns, launching events, beauty product posters, store organization in the future. Classification trees method makes enable decision makers to (i) find out models (or functions) that describe and classify the data by concepts (i.e., discovering relevant models distinguishing in most efficient way each concept or class from the others); (ii) predict the response of the population of their interest (i.e., target population) regarding an event or a phenomenon, in terms of discovered patterns (e.g., behavioral characteristics of this population). Discovered patterns are considered as the results of the learning process from the dataset regarding a particular phenomenon, to make better diagnosis and assessment of the phenomenon for the future (e.g., diagnose medical cases, assess credit risk of loan applications, detect frauds, performance prediction, marketing strategy). In this line, this thesis will aim at discovering the most informativeness paths of different configurations of poverty, as an alternative method for policy makers in predicting multidimensional nature of poverty.

---

[135] But as explanatory variables (to explain target response variable), both qualitative and quantitative variables can be used in classification trees.

[136] Recall that the decision trees approach is one of the classification methods in data mining, including two sub-approaches: *classification* trees (the target response variable is qualitative), and *regression* trees (the target response variable is quantitative). Even if the latter is called *regression*, its aim is still classification as in the other methods of classification existing in the literature. In this thesis, we will use the CART methodology in classification tree approach.

[137] We remind that in machine learning language, a variable is also called as an *attribute* or a *feature*.

We will come further on the technical details concerning the prediction process. But first, we start off by a fictitious example of twenty individuals with five variables: distance, homework, outdoor activity, family activity, football match respectively. Assuming that we asked five questions to randomly selected twelve students at a high school about going to football match, which will be held on Saturday. First question is: What is the distance in kilometers (km onwards) from home to the football game place? Second question is: Is there any homework to do for Monday? (Response options: Yes or No). Third question is: Is there any outdoor activity planned by the student at the same day of the football game? (Response options: Painting, Tennis, Theatre, Swimming). Fourth question is: Is there any family activity planned at the same day of the football game? (Response options: Yes or No). And the main (classification) question is: Will you come to the football game on Saturday? (Response options: Yes or No). The answers of students are recorded and shown in the table below:

Table 26. An example of sample for showing a classification tree

Student id[138]	**Distance (km)**	**Homework**	**Outdoor activity**	**Family activity**	**Football match classification**
St1	10	Yes	Tennis	No	Don't Come
St2	11	No	Swimming	Yes	Don't come
St3	1	No	Theatre	No	Come
St4	0.6	No	Theatre	Yes	Come
St5	25	Yes	Theatre	Yes	Don't come
St6	12	Yes	Swimming	Yes	Don't come
St7	7	Yes	Tennis	No	Come
St8	5.5	Yes	Tennis	No	Don't Come
St9	4	Yes	Painting	No	Don't Come
St10	9	No	Tennis	Yes	Come
St11	11	Yes	Theatre	No	Come
St12	8	Yes	Painting	No	Don't Come

From this example, we will try to determine the rules for (i) coming or; (ii) not coming to the football match on Saturday (7 students don't come and 5 students come to the match: the categorical dependent variable as our target response variable here is composed of two distinct and mutually exclusive groups), as a function of student's answers to the four questions, in the training set. The variable Distance has numerical values, the variables[139] Homework, Outdoor activity, and Family activity are categorical variables (where Homework and Family activity are dichotomous variables). The response variable Football match has two outcomes (i.e.,

[138] Student id is for distinguishing students in the sample. The sample should be unique (i.e., each student should be interviewed only once). This is why, for controlling the unique process, it is important to assess a unique id number to each student.

[139] In data mining and machine learning languages, variables are generally called *attributes* or *features* (see for instance, Witten et al. 2017 [2000]).

categories): come, or don't come to the match. As the response variable has only two categories, the classification of students will be binary. We note that some classification tree algorithms build only binary trees where each internal node creates only two branches (and a terminal node corresponds to one of the two categories of the response variable). But of course, there are also non-binary trees.

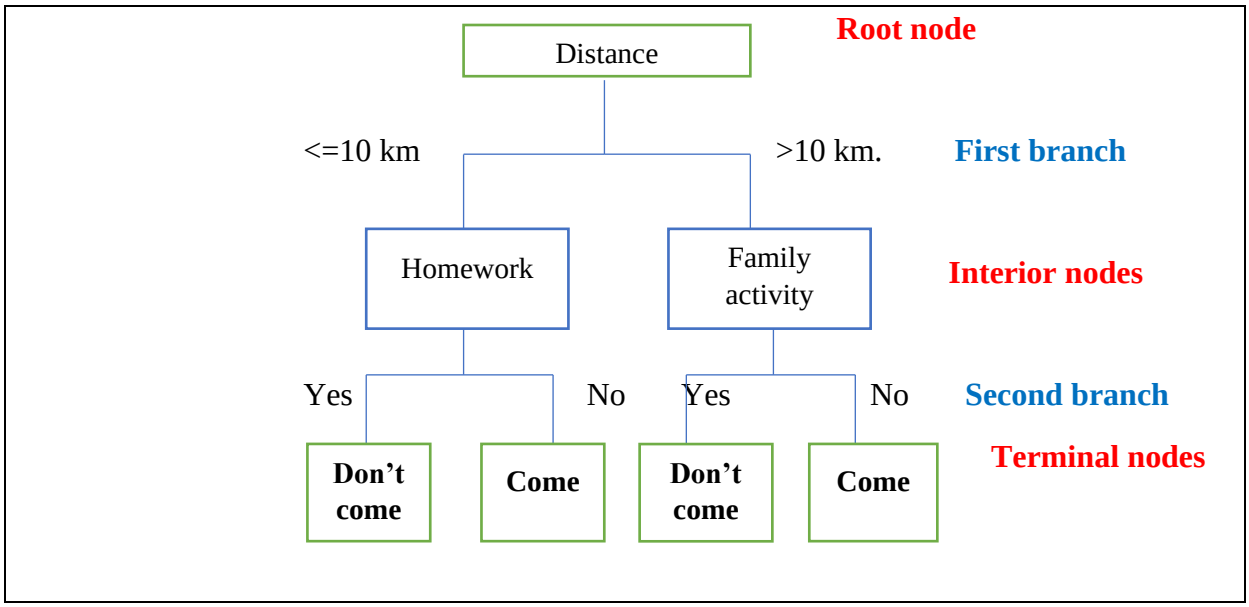


Figure 11. An example of a decision tree and a classification (don't come vs. come) based on the sample in Table 27

Figure 11 above shows the decision tree built based on training/learning set. The aim is to predict whether a student is likely to come or not come to the football match on Saturday, and the prediction paths of coming or not coming to the match. A decision tree visualizes the classification patterns in a *tree* format flowchart, which consists of nodes and branches[140]. How we will read the tree? We will use the top-down approach: reading the tree from the top node to the bottom *terminal/leaf* nodes. The topmost node is called *root node* including the entire sample[141] (all instances/observations of the data) representing the undivided data. The entire sample is initially heterogenous or called impure. For instance, if the target response variable has two categories such as being child labour vs. being not child labour-and every person in the data is labelled as one of these two modalities, then these two categories will be included in the root node (there is not yet a separation of the data into the groups, which will be coming as we descend into the branches of the three). There is always a variable/a combination of variables appears at the root node. We can also say that the root node is the first layer of the tree. Other nodes between the root and terminal, are called *interior* nodes. In Figure 11 above, the tree has three layers of nodes: one root node in the first layer, two interior nodes in the second layer, four terminal nodes in the third layer. Interior and root nodes are also known as *non-terminal* or *parent* node that splits into *daughter/child* nodes[142]. For instance, if the split is binary, then two daughter/child nodes will be generated from the parent node. We note that (binary) trees

[140] Related to the visualization of decision models, it is important to note that these models can also be represented by classification rules (i.e., *IF-THEN* rules), or neural networks. But also, there are various methods of classification in the literature such as Naïve Bayesian classification, rule-based classification, support vector machines, or k-nearest-neighbor classification (see Han et al. 2012 [2000]).
[141] All the sample is known as the entire learning set L.
[142] It is also called *offspring* node.

are often built upon the *binary* recursive partitioning process, where the term *recursive* means that each daughter/child node becomes parent node (or in other words, each parent node is split into two child nodes that in turn may be split again to build new additional child nodes) except the terminal nodes. The nodes other than root ones are called as *internal* or *test* nodes. Nodes are connected by *branches* (i.e., a node creates at least two branches).

The tree separates into branches according to the *splitting criterion* (or known as *attribute selection measures*[143] or *splitting rules*) appearing at the node. At each split, the data is partitioned into two mutually exclusive groups (i.e., right side/right-hand branch of the tree: right child node vs. left side/left-hand branch of the tree: left child node). The sum of the instances of the right and left sides of the branch should be equal to the number of instances at the node. In principle, at child/daughter nodes, we are supposed to obtain more homogenous groups compared to the root node. At a terminal node, if all remained instances are classified with the same category, then the node will be considered as the most (purest) homogeneous (impurity is equal zero). For best classification, it is important that instances with the same category regarding the target response variable are correctly predicted and regrouped together at the terminal node (i.e., homogeneous subsets). The splitting criterion consists of a test on a descriptive variable (i.e., through independent variables, the initial data set is split into fragments and the algorithm will try to find out which independent variable is the best split for obtaining homogeneous sub-sets), which is determined based on different techniques such as entropy function- we will see it in more detail later. In other words, a test at the node corresponds to the process of choosing the best descriptive variable for classification, namely the *goodness of split criteria* that compares different splits and decides which of the splits is the most optimal in terms of reduction of impurity[144] (i.e., maximization of reduction of heterogeneity in terms of the distribution of the outcome at the given node) for obtaining a homogenous sub-sample. The best descriptive variable(s) in terms of information efficiency to classify instances in the data, are generally located in the top layers of the tree. Each branch descending from the node corresponds to one of the possible modalities/categories or values of the variable at the node. If the split variable is qualitative (such as binary or categorical), the splitting criterion is about the modalities/categories of the split variable. But, if the split variable is quantitative, the splitting criterion includes the numerical values of that variable forming a range. The number of possible modalities/categories or values that a variable can take determine the number of downward branches from a non-terminal node.

The sequence of tests continues until a (large) saturated tree is obtained. The saturated tree means that the training/learning data is divided *recursively* into subsets/partitions according to splitting criterion until there is no more an adequate splitting criterion and each instance in the data is placed at one of the terminal nodes (i.e., at the terminal nodes, there is no more split). Rokach and Maimon (2010, 157) have mentioned five stopping rules for obtaining a saturated tree size of tree: "(i) all instances in the training set belong to a single value of target variable; (ii) the maximum tree depth has been reached; (iii) the number of cases in the terminal node is less than the minimum number of cases for parent nodes; (iv) if the node were split, the number of cases in one or more child nodes would be less than the minimum number of cases for child nodes; (v) the best splitting criterion is not greater than a certain threshold". Each terminal is assigned by a *predicted* class (predicted value/modality/category) of the target response

[143] The algorithms, such as ID3, C4.5, CART, use different attribute selection measures.
[144] Impurity is the indicator of heterogeneity.

variable. But, in some cases, terminal node may be a probability of the attributed class as well. It is possible that not every instance in the data is correctly assigned at the end. For instance, a non-child labour may be classified as a child labour. Some misclassifications can be considered less severe error than others. Misclassifying a child labour as a non-child labour may be considered a problem rather than misclassifying a non-child labour as a child labour[145]. We get the classification paths in the number of terminal nodes in the tree. Each instance/observation/unit in the training/learning set falls into one of the terminal nodes. It is possible to obtain more than one terminal node with the same class in the tree. But, if there is only one split with two terminal nodes, the tree will be known as a *stump*.

In Figure 11 above, there are seven nodes (Distance, Homework, Family activity, Don't come, Come, Don't come, Come) and six branches (<=10 km, >10 km, Yes, No, Yes, No). Remind that each node is affiliated with a variable (the label appearing in the rectangle at the node indicates which variable is at the node), and its branches are labeled with its values or modalities/categories (i.e., the labels on each branch indicate the values or modalities/categories that the variable at the node can take). In our case, the root variable is *Distance*. If the value of *Distance* is a maximum 10 km., the next variable to be considered is Homework; otherwise, it is the variable *Family activity*. Hence, the first splitting criterion based on the variable *Distance* is less or equal to 10 km vs. superior to 10 km. In the second layer of the tree, there is one left branch and one right branch. At the left branch, the variable *Homework* appears. In parallel, at the right branch, the variable *Family activity* appears. These two variables are the second split criterion in the tree. On the one hand, the splitting criterion of *Homework* (doing homework vs. not doing homework) creates another two branches at the left side of the tree. On the other hand, the splitting criterion of *Family activity* (participating do family activity vs. not participating to family activity) creates another two branches at the right side of the tree. If student's home is maximum 10km to the football game place and he or she has homework to do; then, decision is don't come to the match. But, if student's home is maximum 10km to the football game place and he or she has s homework to do; then, student does not come to the match on Saturday, (ii) if student's home is maximum 10km to the football game place and he or she does not have a homework to do; then, student come to the match on Saturday, (iii) if student's home is at least 11 km to the football game place and he or she has a family activity; then, student does not come to the match, (iv) if student's home is at least 11 km to the football game place and he or she does not have any family activity; then, student comes to the football match on Saturday. For instance, the third path can be show as follows:

$$Path = (Distance > 10km \cap Family\ activity = No).$$

Each path can be called as a resulting rule (i.e., all paths constitute a set of resulting rules). In classification trees, the discovered paths are the most important ones in terms of gathering information to predict the target response variable. The *decision tree learning* algorithm tests always the most important variable first. Russell and Norvig (2009 [1994], 700) describe this algorithm as "a greedy divide-and-conquer strategy". Herewith, the variable(s) that discriminate(s) the best the instance space into the categories/modalities of the target response

[145] Depending on statistical software/package employed by researcher, it is possible to integrate desired misclassification costs as a form of matrix into the splitting rules. Or, the user may prefer to keep equal all possible misclassifications as well (see statistical software sub-section).

variable is considered as the most important one(s)[146]. The goal is to obtain a relatively small tree (with a small number of tests) with a good rate of correct classification (i.e., much as possible homogeneous groups at the nodes). In our example above, the tree classifies students as if they can come or not to the match in weekend based on three descriptive variables. As we have seen from Figure 11, there are three splitting rules/criteria (i.e., splitting according to the attributes) that create subgroups/partitions in the training set. At terminal nodes, there are four predicted classes (i.e., *don't come* or *come* at the left terminal nodes vs. *don't come* or *come* at the right terminal nodes). In other words, we obtain four different paths (through the decision tree from the root node to the terminal nodes) that predict in a most efficient way (in terms of gathered information) the classification of coming or don't coming to the match in the data. The classification tree method as a classifier allows teachers or school managers to predict the response of students regarding with their participation status to the football match over weekend, based on four discovered patterns- behavioral characteristics of students. The prediction model is as follows:

if Distance > 10 km **and** Family activity=Yes **then**

OUTCOME= Don't come

if Distance >10 km **and** Family activity=No **then**

OUTCOME=Come

if Distance<=10km **and** Homework=Yes **then**

OUTCOME=Don't come

else

OUTCOME=Come

The prediction model above corresponds to the set of classification rules. We note that from a small data set, it is easy to draw *manually* the prediction model. But, in large data set it becomes very complicated. Concerning quantitative descriptive variable, the number of possible splits is equal the number of its distinctly observed values minus one. The form of split is at the node is in terms of 'values less than, or greater than', compared to the selected value. On the other part, for a qualitative descriptive variable with M categories/modalities the number of possible splits any combinations of categories/modalities is possible to form a split) is equal $2^{M-1}-1$ (i.e., for a qualitative variable with two categories, only one type of split is possible; but, for a qualitative variable with more than two categories, many types of splits according to the given formula is possible. For instance, for 3 categories, the possible number of splits is equal 3). To deal with this complexity, the data mining algorithms are very important because they make enable to deduce automatically the prediction model in large data sets. We can say that this is the main contribution of data mining to data analytics[147].

For large data, several prediction models are possible in the literature; but data mining algorithms selects the best model(s) in terms of best capturing the relationship between the

[146] It is important to note that predicted categories/modalities are also called as classes. We will use the term *class* in following paragraphs.

[147] In Figure 12, we provide the classification tree of our example. We note that it can be easy to convert classification trees to classification rules in a small data set.

target response variable and a set of predictors[148]. The way to select the best model depends on the structure of algorithm. Normally, data mining algorithms define classification rules[149] in sequence in the training data (e.g., if 1st classification rule = X then CLASS=Y else if 2nd classification rule=W then CLASS=U else if 3rd classification rule=V then CLASS=Y else CLASS=W. This is an example in a case of two outcomes (Y and U) with two attributes (X and W). The command "else if" indicates the order in an algorithm). Then, the ordered classification rules are applied in the test data for prediction. Prediction is not necessarily to be a hundred per cent correct. It is normal to have misclassification because of misprediction. Misprediction means that the classification rules extracted from the training data do not match exactly with those of the test data (e.g., overfitting). With a small example given above (see Table 27 and Figure 11), we assumed that the training data and test data are the same- we did not randomly divide the initial data into training vs. test data because our intention is to just show an example of induction of classification rules. Also, in our example, the prediction is a hundred per cent correct as we dealt with a small data (i.e., the predicted classes from Figure 11 match with the outcomes in Table 27). Meanwhile, we keep in mind that the data mining algorithm will automatically segment the data into two parts for prediction (i.e., learning vs. testing) and it will automatically generate the rules in order. Moreover, we draw attention on a difference between classification rules vs. association rules. Without having an intention of classifying instances in the data, it is also possible to identify the association rules that simply show the associations between attributes. Below, we show a basic decision tree algorithm summarized by Han et al. (2012 [2000], 333):

Algorithm: Generate_decision_tree. Generate a decision tree from the training tuples of data partition, D.

Input:

- Data partition, D, which is a set of training tuples ad their associated class labels;
- *attribute_list*, the set of candidate attributes:
- *Attribute_selection_method*, a procedure to determine the splitting criterion that "best" partitions the data tuples into individual cases. This criterion consists of a *splitting_attribute* and, possibly, either a *split-point* or *splitting subset.*

Output: A decision tree.

Method:

(1) create a node N;

(2) **if** tuples in D are all of the same class, C, **then**

(3) return N as a leaf node labeled with the class C;

(4) **if** *attribte_list* is empty **then**

(5) return N as a leaf node labeled with the majority class in D; //majority voting

[148] Some researchers call *response variable* as *target feature*, and they call *predictors* as *descriptive features* (see Kelleher et al. 2015, 5).

[149] Classification rules are also known as *splitting criterion*, *splitting rules* or *attribute selection measures*.

(6) apply **Attribute_selection_method** (*D*, *attribute_list)* to **find** the "best" splitting_criterion:

(7) label node *N* with *splitting_criterion;*

(8) **if** *splitting_attribute* is discrete_value **and**

multiway splits allowed **then** //not restricted to binary trees

(9) *attribute_list* ⟸ *attribute_list – splitting attribute; //* remove *splitting_attribute*

(10) **for each** outcome *j* of *splitting_criterion*

// partition the tuples and grow subtrees for each partition

(11) let D_j be the set of data tuples in *D* satisfying outcome *j*; // a partition

(12) **if** D_j is empty **then**

(13) attach a leaf labeled with the majority class in *D* to node *N*;

(14) **else** attach the node returned by **Generate_decision_tree** (D_j, *attribute_list*) to node *N*;

endfor

(15) return *N*;

Figure 12. Basic algorithm for inducing a decision tree from training tuples. Source: Han et al. (2012 [2000], 333)

As we have seen from the **Attribute_selection_method** illustrated in Figure 12, (see **Information Gain** and *Gini index* sub-sections in next pages[150]), it is very important to know how the splits are obtained (i.e., *splitting rules*). We note that if the instances in the training data are all of the same class, then the node *N* becomes a terminal (leaf) node and is labeled with that class (the algorithm finishes at 5th step-see Figure 12). But, whether the instances in the training data are not of the same class, then the step 6 comes into play and the algorithm needs to determine a splitting criterion (i.e., *Attribute_selection_method* command). Often, the splitting criterion is only one variable. However, there are also multivariate splitting criteria (see Rokach and Maimon, 2010). Now, we will see three different partitioning scenarios of splitting, as mentioned by Han et al. (2012 [2000], 334-335). In first scenario, suppose that the split variable *A* at the node *N* is a qualitative variable taking different categories, $a_1, a_2, \dots a_k$, and the outcomes of the test at node *N* appearing at the branches grown from the node *N*, are the *known* categories/modalities of the variable *A*. In second scenario, suppose that the split variable *A* at the node *N* is a quantitative continuous variable. Due to the quantitative scale of the split variable, two outcomes of type (i) $A \leq$ *split_point*; and (ii) $A >$ *split_point*, will be appear at the branches grown from the node *N*. In third scenario, suppose that the split variable *A* at the node *N* is a qualitative variable and a *binary tree* must be generated. The test at node *N* will take the form "$A \epsilon S_A$?", where the term S_A represents the splitting subset for *A*. The branches grown from the node *N* will take the *known* categories/modalities of the variable *A*. Generally, the left branch grown of the node *N* fulfills the split condition and therefore takes the category '*yes*'. And, intrinsically, the right branch takes the category '*no*'. If the descriptive variable *A* takes *k* different possible values/modalities, then there will be 2^k possible subsets.

[150] Gini index produces only a binary tree. However, information gain allows multiway splits (i.e., two or more branches generated from the node).

For instance, if the descriptive variable Color has three possible modalities, namely {red, blue, pink}, then the possible subsets will be: {red, blue, pink}, {}, {red, blue}, {red, pink}, {blue, pink}, {red}, {blue}, {pink}. Among these eight subsets, the power set, {red, blue, pink}, and the empty set, {}, cannot be considered as *splitting subsets* since intrinsically they do not represent a split[151]. So, in the case of binary split on descriptive variable at the node, we will have $(2^k - 2)/2$ possible configurations to build up two partitions of D. Figures 13 and 14 below illustrate these three scenarios. The algorithm until the step 14 mentioned above continues to partition recursively the data to form a classification tree until one of the following three conditions fulfills: all remaining sub samples have the same class (of target variable), or there are no remaining descriptive variables to be applied for partitioning the sample, or there are no samples for a give branch in the tree. In second case, the majority voting process is applied which means that the node N is converted to the terminal node and it is labelled with the most common class in D. In the last case, the terminal node is labelled with the *majority class* in data partition D (i.e., by a majority class vote process, the terminal node is assigned/labelled with one of the categories/modalities of the target response variable). For instance, if the target response variable has two categories, namely A and B, and there are more instances predicted as the class A at the terminal node X; hence, X will be labelled with the class A.

To illustrate a binary tree and how to read it, we can base on an example given by Izenman (2008, 282-283). Suppose that a researcher aims to classify a sample of individuals living in a city called BLUE into the four mutually exclusive groups/classes of the living regions in the city (e.g., c_1= north, c_2= south, c_3= west, c_4= east, $c_5 = periphery$), by two descriptive variables. As descriptive variables, she selects one quantitative variable (e.g., income-*INC*) and one qualitative variable (e.g., educational attainment-*EDUC* taking three possible outcomes: less than high school level, high school level, university degree). The classification tree diagram composed of a sequence of questions (i.e., splitting criteria) is given in Figure 15. We start reading the tree by the root node where the first question (i.e., splitting criterion) appears $INC \leq 4000$? If the answer is yes, we follow the left branch. Then, another splitting criterion appears at the left branch of the second layer in the tree, which is $EDUC = High\ school$? If the answer is yes, we follow the left branch and we reach the class c_1 with corresponding domain $C_1 = \{Educ \leq High\ school, Income \leq 4000\}$; if no, we follow the right branch and we reach the class c_2 with corresponding domain $C_2 = \{Educ > High\ school, Income \leq 4000\}$. On the other hand, if the answer to the question $INC \leq 4000$? at the root node is no, we follow the right branch. Then, another splitting criterion appears at the right branch of second layer of the tree, which is $INC \leq 6000$? If the answer is yes, we follow the left branch and another splitting criterion appears at the right branch of third layer of the tree, which is $EDUC = University$? If the answer is yes, we follow the left branch and we reach the class c_3 with corresponding domain $C_3 = \{Educ \leq University, 4000 < Income \leq 6000\}$; if no, we follow the right branch and we reach the class c_4 with corresponding domain $C_4 = \{Educ > University, 4000 < Income \leq 6000\}$. Moreover, if the answer to the question $INC \leq 6000$? in the second layer is no, then we follow the right branch of second layer of the tree and we

[151] In a binary split, the descriptive variable should be split into two subsets. If the power set such as {red, blue, pink}-the subset that includes all categories/values of the descriptive variable, is considered as the split variable then there will not be possible to have two subsets of this because all categories/values will be placed only in one part.

reach the class c_5 with corresponding domain $C_5 = \{Income > 6000\}$. The tree is composed of three layers, eight branches, four splitting criteria and five terminal nodes. Through four splitting criteria, the sample in partitioned into mutually exclusive five groups assigned by a class label: c_1, c_2, c_3, c_4, c_5, corresponding to the five terminal nodes (see Figure 15 where the five disjoint regions representing the five groups are separated by line segments. The terminal nodes in Figure 15 correspond to the regions in Figure 16).

The type of classification trees depend substantially on how the descriptive variables are selected as splitting criterion in the tree, but also on how to prune tree (we will see the pruning phase in next sub-sections). There are several methods for selecting the splitting criterion (e.g., information gain which is used in ID3 algorithm (see Quinlan 1986), Gini index, likelihood-ratio chi-squared statistics, DKM criterion, gain ratio, distance measure, binary criterion, orthogonal criterion, Kolmogorov-Smirnov criterion, AUC-splitting criteria): "(i) according to the origin of the measure: information theory, dependence, and distance; (ii) according to the measure structure: impurity-based criteria, normalized impurity-based criteria, and binary criteria." (Rokach and Maimon 2010, 152). In following pages, we will present a few of them but we will first start off by the entropy function, which is the base of any splitting criterion.

Entropy function, Information gain[152]

The goal of an entropy function is to minimize uncertainty, hence, to maximize the information gain in the process of choosing the most adequate variable among many others in the training set for defining the split criterion in the roof and interior nodes of the tree (Bramer 2007). The origin of this function bears in information theory especially regarding transmission of messages in telecommunication (see Shannon and Weaver 1949, Shannon 1948), which is often referred to Shannon's *entropy model*. Let concentrate on Shannon's entropy model, starting off with an example. Suppose that there are three different boxes of books including each of them 6 books, in a library. First box has 6 same books of 19th century, the second box has 4 same books of 16th century and 2 same books of 18th century, and the third box has 6 different books of 15th century on history of art. If we select randomly a book in the first box, we will have *zero uncertainty* for predicting the type of book because each of books in this box is equal. Herewith, the first box has *zero entropy*. In contrast, if we select randomly a book in the third box, we will have a *high level of uncertainty* for prediction because there are 6 different possible outcomes. In this case, the third box has *very high entropy*. And second box has entropy values between zero entropy and very high entropy.

Entropy is a measure of uncertainty of a random variable and its value is measured based on probabilities. A randomly selected outcome in the training set with a *high probability* to be happened will be associated with a *low entropy value*. In contrast, a randomly selected outcome in the training set with a *low probability* to be happened will be associated with a *high entropy value*. The **logarithm**, or **log** (onwards), function in mathematics will be used for obtaining entropy values from the probability values. The log function $log_b(a)$ is pronounced 'the log of a to the base 'b. A logarithmic function can have different bases such as 10 (e.g., $log_{10}100 = 2$). By definition, if $log_2(X) = Y$ with $X > 0$; this means that $2^Y = X$. For instance, $log_2(16) = 4$ because $2^4 = 16$, or $log_3(243) = 5$. If the **base is 2**, the logarithm is qualified **binary**. The $log_2 X$ is equal to 0 when $X = 1$; is negative when $X < 1$; and is positive when $X > 1$. When we plot **binary log** function of probabilities, $[-\infty, 0]$, by probability function

[152] This section is based on Kelleher et al. (2015), chapter 4, and Han et al. (2012 [2000]), chapter 8.

($log_2 P(x)$ in y axe ranging from -10 to 0, and $P(x)$ in x axe ranging from 0 to 1 in the graph), we can see that the *large negative numbers of log function* are associated with *low probability values*, and inverse. In absolute terms, this relation is convenient as a measure of entropy: (i) for low level of probability (high level of uncertainty), we will obtain high level of entropy; (ii) for high level of probability (low level of uncertainty), we will obtain low level of entropy (see Kelleher et al. 2015).

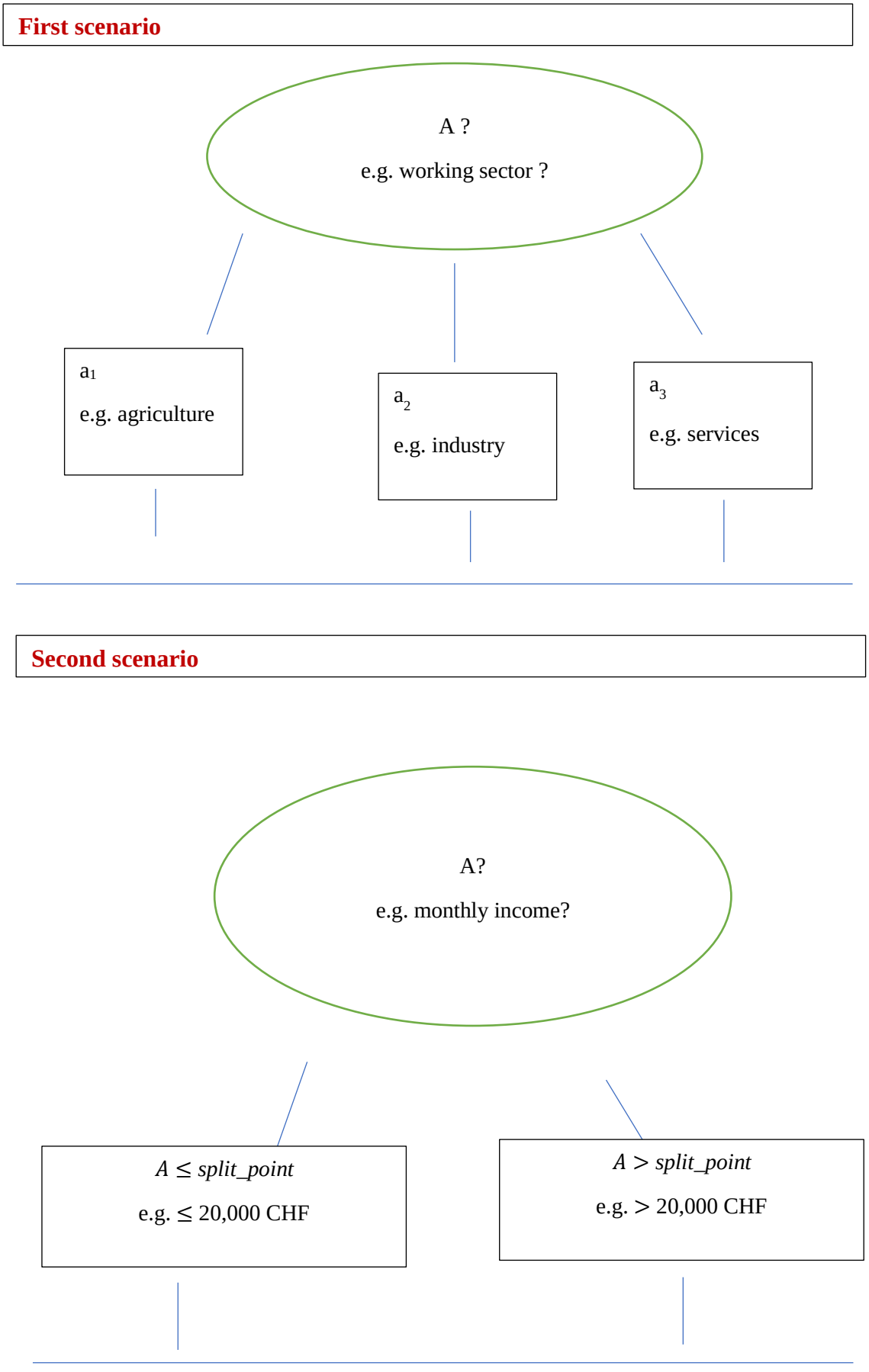


Figure 13. Splitting partioning scenarios. Source: Han et al. (2012 [2000], 334)

Third scenario

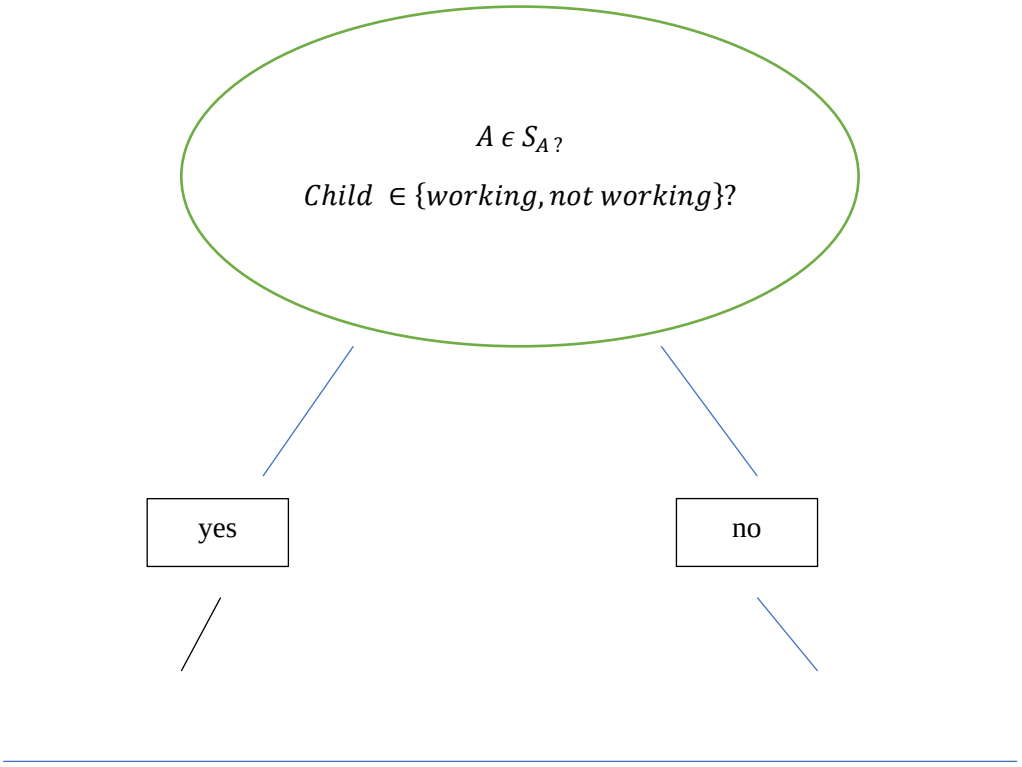


Figure 14. Splitting partitioning scenarios Cont'd. Source: Han et al. (2012 [2000], 334)

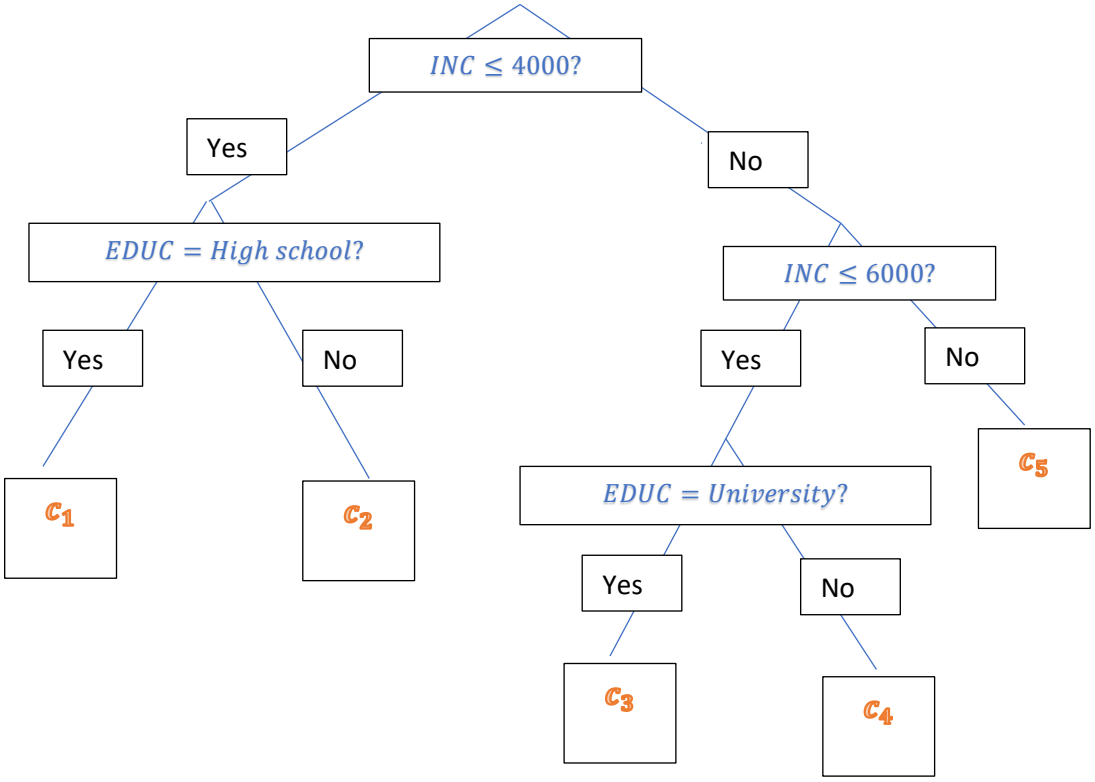


Figure 15. Classification tree example with five terminal nodes and four splitting criteria

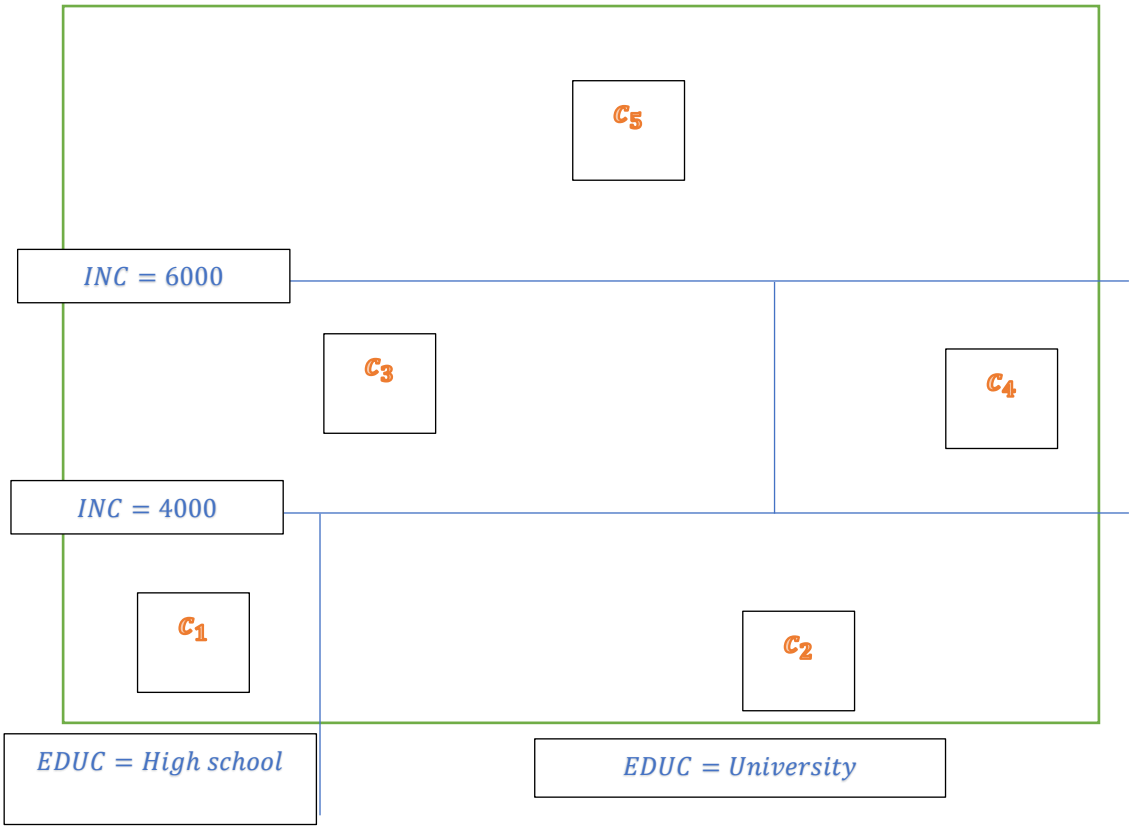


Figure 16. The partitioning of the sample into five groups, corresponding to the five terminal nodes

Based on the link between binary log function of probabilities and probability function, Shannon's entropy model (denoted as E, onwards) is defined as follows (see Barmer 2007)- this following equation is the fundamental of modern information theory:

$E\ (t) = -\sum_{i=1}^{K} P(X = i) \times\ log_b\ P(X = i)$ (with $P(X = i) \neq 0$) [153].

where, we suppose that there are K possible values of target response variable in the training set, $P(X = i)$ is the probability that the target variable takes the value i (i.e., the number of occurrences of the value i is divided by the total number of instances (N) in the training data- or it is simply called training set). As we know that a probability (i.e., $P(X = i)$) takes a value between 0 and 1 (i.e., with 0=the event is not possible, and 1=the event is certainly possible), on the one hand, E will be positive for the values of $P(X = i)$ greater than zero and less than 1. On the other hand, E will be zero if and only if all instances in the training data have the same class (i.e., $P(X = i)$ is equal to 1 and there is no other probability). The minus sign at the beginning of the equation allows to obtain positive numbers, as the log function yields some negative numbers (see paragraph above). Shannon's entropy model is measured in terms of 'bits' (i.e., called as binary digit) of information; this is why we will always use 2 as the base (i.e., $b = 2$). We note that the maximum amount of entropy for a set of two values/modalities is 1.00 bit, in the case of $b = 2$. For more than two variables, the entropy may be higher than 1.00 bit.

For instance, if there are two values/modalities with probabilities $P(X = 1) =$ 0.3 and $P(X = 2)$= 0.7 respectively, the we compute entropy as follows (we can also use the term p_i for mentioning probability):

$E = -\sum_{i=1}^{K=2} p_i\ log_2\ p_i$ = - [(0.3 log_2 0.3) + (0.7 log_2 0.7)] = - [(0.3 *-1.736) + (0.7 *-0.514)] = - [-0.52 - 0.359] = 0.879 bits[154].

We also note that E is maximized when each value/modality of the random variable has the same probability.

By using a traditional statistical calculator, it can become difficult to compute the logarithm base 2. However, the népérien logarithm function can help us to solve this problem. For all real numbers a strictly positive and different to 1, and for all real numbers x strictly positive, the népérien logarithm function can be written as:

$log_a(x) = \frac{\ln(x)}{\ln(a)}$.

The entropy value will be higher when heterogeneity of a set is higher. This is why, the entropy measure is used as a major indicator of impurity. When the set is very heterogenous (meaning that there are several values/modalities that the random variable takes on), the set is considered very impure. For instance, in a books box including 6 same books, the heterogeneity, and therefore the impurity is zero. We can compute the entropy as follows:

[153] We can also use the expression p_i for $P(X = i)$. And, for expressing entropy, the term $Info(D)$ is also used in the literature, for instance see Han et al. (2012 [2000], 337).

[154] The computation of log base 2 can be done on the internet by using the following web site : https://www.miniwebtool.com/log-base-2-calculator/. Accessed 5 August 2018.

$$E = -\left(\frac{6}{6}\right) log_2\left(\frac{6}{6}\right) = -1 \times log_2(1) = -1 \times 0 = \mathbf{0\ bits}.$$

As we can see, we obtained a zero value of entropy that shows the zero impurity in a set including all identical values (there is only one type of value/modality). If the set includes 6 different books, then as the entropy value will show the impurity will reach its maximum for this set, as follows:

$$E = -\left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right) - \left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right) - \left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right) - \left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right) - \left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right) - \left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right) = (-0.166 \times -2.59) \times 6 = \mathbf{2.579\ bits}.$$

We can conclude that the entropy score for this set is significantly increased to 2.579 bits when the set includes 6 different types of books, compared to the initial set with 6 same types of books.

In classification trees, the entropy function is used for selecting a split variable in root and interior nodes. To make clear this process, we will do entropy computations with a fictitious example below. We suppose that there is a competition among bachelor students offered by their university to go to Antarctica for a month with a team of researchers. University receives 30 applications, which constructs our training data set including three *descriptive* (characteristics) variables (Study year, Study field, Gender) and one *target response* (*classification)* variable (Class) covering three categories: strong physical capacity (SPORT), environmental activist spirit (ENVIN), strong management capacity in difficult times (MANAG) of students. Possible study fields are: economics, electronic engineering, geography, mathematics, music, plastic arts, sociology. Possible gender: female vs. male. Possible study year: first (1); second (2); third (3) year of bachelor. See the table below for the complete training data set.

Table 27. Training set for the computations of the information gain on split

Study year	Study field	Gender	Class (Target response variable)
1	Economics	F	SPORT
1	Sociology	F	MANAG
1	Sociology	M	MANAG
1	Plastic arts	M	SPORT
1	Music	M	SPORT
1	Plastic arts	F	MANAG
1	Electronic engineering	M	ENVIN
1	Sociology	F	ENVIN
1	Sociology	M	ENVIN
1	Music	M	ENVIN
2	Electronic engineering	M	ENVIN
2	Electronic engineering	M	SPORT

2	Music	F	SPORT
2	Sociology	F	ENVIN
2	Plastic arts	M	SPORT
2	Sociology	F	ENVIN
2	Sociology	M	MANAG
2	Mathematics	F	ENVIN
2	Economics	F	MANAG
2	Economics	F	MANAG
3	Mathematics	M	SPORT
3	Mathematics	F	SPORT
3	Geography	F	ENVIN
3	Geography	F	ENVIN
3	Geography	M	SPORT
3	Plastic arts	F	MANAG
3	Geography	M	MANAG
3	Sociology	M	ENVIN
3	Mathematics	F	ENVIN
3	Electronic engineering	F	SPORT

Through this example, the aim is to compute the entropy for classification based on the variable 'Study year' that has three subsets: first year (10 students), second year (10 students), and third year (10 students) of bachelor students. We note that the split variable is defined if its *average entropy* is less or equal than the average entropy of the previous training set (see Bramer 2007). Thus, first, we compute entropy for these three subsets (i.e., these three subsets alone can be seen as a different training set) as follows:

Given that $i = 3$ (i.e., 3 different types of classification respectively: (1) SPORT; (2) MANAG; (3) ENVIN),

In first subset (first year of bachelor students):

p_1=3/10= 0.3

p_2=3/10=0.3

p_3=4/10=0.4

Hence,

$E_1 = -\sum_{i=1}^{K=3} p_i \, log_2 \, p_i$ = - [(0.3 log_2 0.3) + (0.3 log_2 0.3) + 0.4 log_2 0.4] = - [(0.3 *-1.736) + (0.3 *-1.736) + (0.4 * -1.321)] = - [-0.520-0.520-0.528] = 1.568 bits.

For the second subset (second year of bachelor students):

p_1=3/10= 0.3

p_2=3/10=0.3

p_3=4/10=0.4

Hence,

$E_2 = -\sum_{i=1}^{K=3} p_i \log_2 p_i$ = - [(0.3 $\log_2$ 0.3) + (0.3 $\log_2$ 0.3) + 0.4 $\log_2$ 0.4] = - [(0.3 *-1.736) + (0.3 *-1.736) + (0.4 * -1.321)] = - [-0.520-0.520-0.528] = 1.568 bits.

And, for the third subset (third year of bachelor students):

p_1=4/10=0.4

p_2=2/10=0.2

p_3=4/10=0.4

Hence,

$E_3 = -\sum_{i=1}^{K=3} p_i \log_2 p_i$ = - [(0.4 $\log_2$ 0.4) + (0.2 $\log_2$ 0.2) + 0.4 $\log_2$ 0.4] = - [(0.4 *-1.321) + (0.2 *-2.321) + (0.4 * -1.321)] = - [-0.528-0.464-0.528] = 1.520 bits.

Then, we compute the *average* entropy of the three subsets (we just multiply each entropy by its weight-10/30 for each subset): $E_{subsets}$= (10/30) [E_1 + E_2 + E_3] = 0.33 [1.568 + 1.568 + 1.520] = 0.33 * 4.656 = **1.536** bits.

In order to find out the *information gain* by splitting on the variable Study year in the tree for class prediction, we do need also the entropy of the previous training set (before partitioning). In a classification tree, splits continue until the entropy obtained from the subsets reach to zero. The information gain allows us to see the expected reduction in entropy.

Given that there are three classifications (10 students in SPORT class, 10 students in MANAG class, and 10 students in ENVIN), the entropy of the main training set is calculated as:

$E_{initial}$= - (10/30) $\log_2(\frac{10}{30})$- (10/30) $\log_2(\frac{10}{30})$- (10/30) $\log_2\left(\frac{10}{30}\right) = (-0.33 \times -1.599) \times 3 =$ $\mathbf{1.583}\ \boldsymbol{bits}$.

As a result, the *information gain* is equal to the entropy of the training set minus the entropy of the subsets. This gives: $E_{initial} - E_{subsets}$ = 1.583 – 1.536 = **0. 047 bits**. In the beginning (i.e., before splitting) we have 1.583 bits information for classification. Then, after the split on the variable Study year, we obtain 0.047 bits information. For example, since the information gain from this split is not equal to zero, the classification tree can still continue to be splitting on another variable. We could also compute the information gain for the other two variables Study field, and Gender in same way as we did for the variable Study year above. The variable which will have the highest value of information gain will be selected as the splitting variable (called also splitters) at the node.

Now, we will see another example where we will focus on information gain by taking in consideration more than one descriptive variable to predict the classification in regards with the target variable. This example will guide us how to build a sequence of splits in what order in

the training data given the target variable values/modalities. The table below describes the training data set.

Table 28. Second example of computation of entropy, information gain, impurity

ID	Personal Income level (Income)	Obligatory educational attainment (Educ)	Work in a secure job (Job)	Classification in terms of housing location (House)
1	High	Yes	No	Periphery
2	Low	Yes	No	Periphery
3	Low	No	No	City Centre
4	High	Yes	Yes	City Centre
5	Low	Yes	Yes	Periphery
6	Low	Yes	No	City Centre

In this training data set, we have three descriptive variables (Income, Educ, Job) and one classification target response variable (House). The aim is look at which descriptive variable brings more information gain to discriminate between living in periphery (Periphery class) and living in city center (City Centre class). As we can see, Income variable has two modalities (High vs. Low), Educ variable has two modalities (Yes vs. No), Job variable has two modalities (Yes vs. No), and House variable has two modalities (Periphery vs. City Centre). The goal is to discriminate efficiently between the two modalities of classification target variable, based on the informativeness of descriptive variables. For this, we need to calculate information based on entropy function. We will follow the same steps that we did in previous example, but this time, we will measure entropy for all descriptive variables. We will follow the three equations provided by Kelleher et al. (2015, 130-131), corresponding to 'the entropy of the dataset D regarding the target response variable t', 'the entropy remaining after the dataset D is partitioned by the particular descriptive variable d', 'the information gain obtained by splitting the dataset D with the descriptive variable d'.

Equation of the entropy of the dataset D regarding the target response variable t :

$$E(t, D) = -\sum_{l \epsilon\, levels(t)} \big(P(t = l)\big) \times log_2 P(t = l)$$

Equation 21

where $levels(t)$ is the set of categories/modalities that the target response variable t takes on (e.g., l is one of the categories/modalities). Thus, $P(t = l)$ is the probability that the target response variable takes the category/modality l.

Equation of the entropy remaining after the dataset D is partitioned by the particular descriptive variable d :

$$rem\ (d, D) = \sum_{l\ \epsilon\ levels(d)} \frac{|D_{d=l}|}{|D|} \times E(t, D_{d=l})$$

Equation 22

On the left side of the equation 22, D represents the descriptive variable in general sense, d corresponds to the descriptive variable in computation, rem means the entropy *remaining* after the dataset is partitioned by a particular descriptive variable d. When the dataset D is partitioned by the descriptive variable d, a number of partitions emerge, $D_{d=l_1 \ldots} D_{d=l_k}$ where $l_1 \ldots l_k$ are the k levels/categories/modalities that the descriptive variable d takes on. On the right side of the equation, the notation l in sigma sign designates different levels/categories/modalities of the descriptive variable in computation d. First part of the formula in right side of the equation, is the weight of the level/category/modality of the descriptive variable in computation. Second part of the formula in right side of the equation is the entropy function of the partition $D_{d=l}$.

Equation of the information gain obtained by splitting the dataset D with the descriptive variable d :

$$IG(d, D) = E(t, D) - rem(d, D)$$

Equation 23

First, we compute the entropy for the descriptive variable *Income*, as follows:

$$rem\ (Income, D) = \left(\frac{|D_{income=high}|}{|D|} \times E\left(t, D_{Income=high}\right)\right) + \left(\frac{|D_{income=low}|}{|D|} \times E(t, D_{income=low})\right) = (2/6 \times (-\sum_{l\epsilon\{periphery,center\}} P(t = l) \times log_2 P(t = l)) + (4/6 \times (-\sum_{l\epsilon\{periphery,center\}} P(t = l) \times log_2 P(t = l))$$

$$rem(Income, D) = \left(\frac{2}{6} \times \left(-\left(\left(\frac{1}{2} \times log_2\left(\frac{1}{2}\right)\right) + \left(\frac{1}{2} \times log_2\left(\frac{1}{2}\right)\right)\right)\right)\right) + \left(\frac{4}{6} \times \left(-\left(\left(\frac{2}{4} \times log_2\left(\frac{2}{4}\right)\right) + \left(\frac{2}{4} \times log_2\left(\frac{2}{4}\right)\right)\right)\right)\right) = 0.333\big((-0.5 \times -1) + (-0.5 \times -1)\big) + 0.666\big((-0.5 \times -1) + (-0.5 \times -1)\big) = 0.333 \times (1) + 0.666(1) = \mathbf{0.999\ bits}.$$

In our dataset, there are 6 people. 4 of them have a low level of income, and 2 of them have a high level of income. This is why, there are two parties in the equation above. First part focuses on people having a high level of income (the probability of having a high income is 2 of 6). Among people having high level of income, 1 person is living in periphery and 1 person is living in center. Related to that, in the first part of the equation above, there are two sub-parties. The probability of living in periphery and having a high level of income is 1 of 2, and the probability of living in center and having a high level of income is 1 of 2, among people having a high level of income. Second part focuses on people having a low level of income (the probability of having a low income is 4 of 6). Among people having low level of income, 2 people are living in periphery and 2 people are living in center. Related to that, in the second part of the equation above, there are two sub-parties. The probability of living in periphery and

having a low level of income is 2 of 4, and the probability of living in center and having a low level of income is 2 of 4, among people having a low level of income.

Second, we compute the entropy for the descriptive variable *Educ*, as follows:

$$rem\ (Educ, D) = \left(\frac{|D_{Educ=yes}|}{|D|} \times E\left(t, D_{Educ=yes}\right)\right) + \left(\frac{|D_{Educ=no}|}{|D|} \times E(t, D_{Educ=no})\right) = (2/6 \times (-\sum_{l\epsilon\{periphery,center\}} P(t=l) \times log_2 P(t=l)) + (4/6 \times (-\sum_{l\epsilon\{periphery,center\}} P(t=l) \times log_2 P(t=l))$$

$$rem(Educ, D) = \left(\frac{5}{6} \times \left(-\left(\left(\frac{3}{5} \times log_2\left(\frac{3}{5}\right)\right) + \left(\frac{2}{5} \times log_2\left(\frac{2}{5}\right)\right)\right)\right)\right) + \left(\frac{1}{6} \times \left(-\left((0 \times log_2(0)\right) + \left(\frac{1}{1} \times log_2\left(\frac{1}{1}\right)\right)\right)\right)\right) = 0.833\big((-0.6 \times -0.736) + (-0.4 \times -1.321)\big) + 0.166\big((-0 \times inf.) + (-1 \times 0)\big) = 0.833 \times (0.441 + 0.528) + 0.166 \times (0-1) = 0.807 - 0.166 = \mathbf{0.641\ bits.}$$

In our dataset, there are 6 people. 5 of them have an obligatory level of education, and 1 of them does not have. This is why, there are two parties in the equation above. First part focuses on people having an obligatory level of education (the probability of having an obligatory level of education is 5 of 6). Among people having an obligatory level of education, 3 people are living in periphery and two people are living in center. Related to that, in the first part of the equation above, there are two sub-parties. The probability of living in periphery and having an obligatory level of education is 3 of 5, and the probability of living in center and having an obligatory level of education is 2 of 5, among people having an obligatory level of education. Second part focuses on people having not an obligatory level of education (the probability of having not an obligatory level of education 1 of 6). Among people having not an obligatory level of education, 0 person is living in periphery and 1 person is living in center. Related to that, in the second part of the equation above, there are two sub-parties. The probability of living in periphery and having not an obligatory level of education is 0, and the probability of living in center and having not an obligatory level of education is 1, among people having not an obligatory level of education.

Third, we compute the entropy for the descriptive variable *Job*, as follows:

$$rem\ (Job, D) = \left(\frac{|D_{Job=yes}|}{|D|} \times E\left(t, D_{Job=yes}\right)\right) + \left(\frac{|D_{Job=no}|}{|D|} \times E\left(t, D_{Job=no}\right)\right) = (2/6 \times (-\sum_{l\epsilon\{periphery,center\}} P(t=l) \times log_2 P(t=l)) + (4/6 \times (-\sum_{l\epsilon\{periphery,center\}} P(t=l) \times log_2 P(t=l))$$

$$rem(Job, D\ ) = \left(\frac{2}{6} \times \left(-\left(\left(\frac{1}{2} \times log_2\left(\frac{1}{2}\right)\right) + \left(\frac{1}{2} \times log_2\left(\frac{1}{2}\right)\right)\right)\right)\right) + (\frac{4}{6} \times \left(-\left((\frac{2}{4} \times log_2\left(\frac{2}{4}\right)\right) + \left(\frac{2}{4} \times log_2\left(\frac{2}{4}\right)\right)\right)\right)) = 0.333((-0.5 \times -1) + (-0.5 \times -1)) + 0.666((-0.5 \times$$

$-1) + (-0.5 \times -1)) = 0.333 \times (0.5 + 0.5) + 0.666 \times (0.5 + 0.5) = 0.333 + 0.666 =$ **0.999 bits.**

In our dataset, there are 6 people. 2 of them have a secure job, and 1 of them does not have. This is why, there are two parties in the equation above. First part focuses on people having a secure job (the probability of having a secure job is 2 of 6). Among people having a secure job, 1 person is living in periphery and 1 person is living in center. Related to that, in the first part of the equation above, there are two sub-parties. The probability of living in periphery and having a secure job is 1 of 2, and the probability of living in center and having a secure job is 1 of 2, among people having a secure job. Second part focuses on people having not a secure job (the probability of having not a secure job is 4 of 6). Among people having not a secure job, 2 people are living in periphery and 2 people are living in center. Related to that, in the second part of the equation above, there are two sub-parties. The probability of living in periphery and having not a secure job is 2 of 4, and the probability of living in center and having not a secure job is 2 of 4, among people having not a secure job.

Fourth, we compute the entropy for the whole training data set as follows:

$$E(t, D) = -\sum_{l \in \{periphery, center\}} P(t = l) \times log_2 P(t = l)$$
$$= -P(t = periphery) \times log_2 P(t = periphery) - P(t = center) \times log_2 P(t = center)$$

$$E(t, D) = -\left(\left(\frac{3}{6} \times log_2\left(\frac{3}{6}\right)\right) + \left(\frac{3}{6} \times log_2\left(\frac{3}{6}\right)\right)\right) = -(0.5 \times -1) - (0.5 \times -1) = \mathbf{1\ bit}.$$

At the end, we can compute the information gain (IG) for each descriptive variable, as follows:

$IG(\boldsymbol{Income}, D) = 1 - 0.999 = \mathbf{0.001\ bits}$.

$IG(\boldsymbol{Educ}, D) = 1 - 0.641 = \mathbf{0.359\ bits}$.

$IG(\boldsymbol{Job}, D) = 1 - 1 = \mathbf{0\ bits}$.

To split the training data into pure sets regarding the target variable, it is crucial to look at the information gain in bits of each descriptive variable. When the information gain in bits is higher, the split will be the purest. In other words, the information gain indicates how much would be gained by splitting the node by the given variable. For instance, the variable having an information gain score of 1 bit will be the best split variable for discriminating the target variable values/modalities (i.e., when the information gain score is equal to the entropy of the entire dataset). So, the aim is to look at for the descriptive variable bringing the highest information gain, which will be selected as the *splitting variable*. Based on the results we obtained in this example, we can see that the variable *Educ* has an information gain of 0.359 bits. On the other side, the two other variables have an information gain of 0 bits, which is usefulness for discriminating periphery vs. city center. Splitting on the variable *Educ* (or saying that 'partition on the variable *Educ*'), therefore, will be the 'best classification' that will provide the most useful information for better prediction of the target variable, *House*. Therefore, the variable *Educ* becomes the splitting variable at the root node of the classification tree. Branches will be grown for each category/output (yes vs. no) of the *Educ*. If the information gained by

the descriptive variable at the node is higher then, the amount of information needed to finish classification tree (arriving to the terminal nodes) will be less. Here, the gained information by splitting on the variable *Educ* is not insignificant but it is not perfect as well.

In the example below, we worked on a qualitative descriptive variable case, but if we have a quantitative variable instead, how the best split-point for that variable should be determined? For such variable A, first of all, its values should be sorted in increasing order. The average between each pair of adjacent values based on sorted series of values, will be considered as a possible split-point. For instance, a possible split-point between the two adjacent values a_i and a_{i+1}, in sorted series is:

$$\frac{a_i + a_{i+1}}{2}.$$

As we can see, if the variable A takes m values, then the number of possible splits is equal $m - 1$. Next step is to compute the information, $Info_A(D)$, for each possible split-point for A. In each split, there will be two partitions such as D_1 is the training set in D satisfying the condition $A \leq split_point$, and D_2 is the remaining training set in D satisfying the condition $A > split_point$. Given j is for mentioning the two partitions in the split, the information for each possible split_point for A, will be as follows:

$$Info_A(D) = \sum_{j=1}^{2} \frac{|D_j|}{|D|} \times Info(D_j)$$

In the end, the split-point yielding the minimum information for A, will be selected as the split-point at the node.

For instance, suppose that there are 6 people in the sample taking each of them a value for a random quantitative variable, namely X as follows:

$$X = \{2,4,4,6,8,10\}$$

These 6 people do also take a category for the target response variable $Class$, in the form of *Yes (Y)* or *No (N)*. Suppose that the distribution for the variable $Class$ is as follows:

$$Class = \{Y, Y, N, N, Y, Y\}$$

Th first person in the sample takes a value 2 for X and Y for $Class$, and so on.

We would like to compute the information for each possible split-point. The set is already ordered. The average between each pair of adjacent values in the series can be shown as follows (let name the average series as Y):

$$Y = \{3,4,5,7,9\}$$

As we can see above, given we have 6 values in the original series, we obtain 5 possible split-points. The next step is to compute the information for each possible split-point.

$SplitInfo_{\{3\}\epsilon Y}(D) = \left(\frac{1}{6} \times ((-1/1 \times log_2\left(\frac{1}{1}\right)) - (\frac{0}{1} \times log_2\left(\frac{0}{1}\right))\right) + (\frac{5}{6} \times (\left(-\frac{3}{5} \times log_2\left(\frac{3}{5}\right)) - (\frac{2}{5} \times log_2\left(\frac{2}{5}\right))\right) = \left(0.166 \times \left((-1 \times -0) - (0 \times \text{inf})\right)\right) + (0.833 \times \left((-0.6 \times -0.736\ ) - (0.4 \times -1.321)\right) = (0.166 \times 0) + \left(0.833 \times (0.441 + 0.528)\right) = 0 + 0.807 =$ **0.807** bits.

where the first partition (D_1) is for $X \leq 3$ that includes 1 person who takes a value at most 3 for X (among them, 1 person takes the class *Yes*), and the second partition (D_2) is for $X > 3$ that includes 5 persons who take a value superior than 3 for X (among them, 3 persons take the class *Yes* and 2 persons take the class *No*).

$SplitInfo_{\{4\}\epsilon Y}(D) = \frac{3}{6} \times \left(-\frac{2}{3} \times log_2\left(\frac{2}{3}\right) - \frac{1}{3} \times log_2\left(\frac{1}{3}\right)\right) + \frac{3}{6} \times \left(-\frac{1}{3} \times log_2\left(\frac{1}{3}\right) - \frac{2}{3} \times log_2\left(\frac{2}{3}\right)\right) = 0.5 \times (-0.666 \times -0.586 - 0.333 \times -1.586) + 0.5 \times (-0.333 \times -1.586 - 0.666 \times -0.586) = 0.5 \times (0.390 + 0.528) + 0.5 \times (0.528 + 0.390) = 0.459 + 0.459 = \mathbf{0.918}$ bits.

Where the first partition is for $X \leq 4$ that includes 3 persons who take a value at most 4 for X (among them, 2 persons take the class *Yes*, and 1 person takes the class *No*), and the second partition is for $X > 4$ that includes 3 persons who take a value superior than 4 for X (among them, 12 persons take the class *Yes*, and 1 person takes the class *No*).

$SplitInfo_{\{5\}\epsilon Y}(D) = \frac{3}{6} \times \left(-\frac{2}{3} \times log_2\left(\frac{2}{3}\right) - \frac{1}{3} \times log_2\left(\frac{1}{3}\right)\right) + \frac{3}{6} \times \left(-\frac{1}{3} \times log_2\left(\frac{1}{3}\right) - \frac{2}{3} \times log_2\left(\frac{2}{3}\right)\right) = 0.5 \times (-0.666 \times -0.586 - 0.333 \times -1.586) + 0.5 \times (-0.333 \times -1.586 - 0.666 \times -0.586) = 0.5 \times (0.390 + 0.528) + 0.5 \times (0.528 + 0.390) = 0.459 + 0.459 = \mathbf{0.918}$ bits.

$SplitInfo_{\{7\}\epsilon Y}(D) = \frac{4}{6} \times \left(-\frac{2}{4} \times log_2\left(\frac{2}{4}\right) - \frac{2}{4} \times log_2\left(\frac{2}{4}\right)\right) + \frac{2}{6} \times \left(-\frac{1}{2} \times log_2\left(\frac{1}{2}\right) - \frac{0}{2} \times log_2\left(\frac{0}{2}\right)\right) = 0.666 \times (-0.5 \times -1 - 0.5 \times -1) + 0.333 \times (-0.5 \times -1 - 0 \times inf) = 0.666 \times (0.5 + 0.5) + 0.333 \times (0.5 + 0) = 0.666 + 0.166 = \mathbf{0.832}$ bits.

$SplitInfo_{\{9\}\epsilon Y}(D) = \frac{5}{6} \times \left(-\frac{3}{5} \times log_2\left(\frac{3}{5}\right) - \frac{2}{5} \times log_2\left(\frac{2}{5}\right)\right) + \frac{1}{6} \times \left(-\frac{1}{1} \times log_2\left(\frac{1}{1}\right) - \frac{0}{1} \times log_2\left(\frac{0}{1}\right)\right) = 0.833 \times (-0.6 \times -0.736 - 0.4 \times -1.321) + 0.166 \times (-1 \times -0 - 0 \times -inf) = 0.833 \times (0.441 + 0.528) + 0.166 \times (0 + 0) = 0.807 + 0 = \mathbf{0.807}$ bits.

Before processing to computation of information gain for each possible split-point, we compute the expected information in D as follows:

$Info(D) = -\sum_{i=1}^{2} p_i \, log_2(p_i)$,

$Info(D) = -\left(\frac{4}{6} \times log_2\left(\frac{4}{6}\right) + \frac{2}{6} \times log_2\left(\frac{2}{6}\right)\right) = -((0.666 \times -0.586) + (0.333 \times -1.586)) = 0.390 + 0.528 = \mathbf{0.918}$ bits.

The information gain for each possible split-point will be:

$Gain(split - point = 3) = 0.918 - 0.807 =$ **0.111** bits.

$Gain(split - point = 4) = 0.918 - 0.918 = \mathbf{0}$ bits.

$Gain(split - point = 5) = 0.918 - 0.918 = \mathbf{0}$ bits.

$Gain(split - point = 7) = 0.918 - 0.832 = \mathbf{0.086}$ bits.

$Gain(split-point = 9) = 0.918 - 0.807 = \mathbf{0.111}$ bits.

Based on the information gain results above, the two split-points {3} and {9} yield the highest information gain for A, so one of them (randomly) will be selected as the split-point at the node.

Gini index of diversity[155]

As a measure of splitting criterion, the Gini index is used in CART methodology. It focuses on discrepancy between the probability distributions of descriptive variables given their modalities/values (i.e., it is computed for each descriptive variable in the classification model) and generates a *binary* split on descriptive variable A at the node N (i.e., $A \epsilon S_A$?, see 3rd scenario in Figure 14). Suppose that there are K number of classes and p_i is the probability that the i^{th} class (of target response variable) occurs in training data (or training samples or data partition) D, the Gini index of diversity can be written as:

$$Gini\ (D) = 1 - \sum_{i=1}^{K} p_i^2.$$

This index is also called as the measure of impurity in D. With an example below, we will show how this index can be computed in practice.

Based on the information in Table 29 above, we know that the initial (prior) probabilities of the three classes are: p(class=Sport) = 10 / 30 = 0.33, p (class=Envin) = 8/30= 0.266, p (class=Manag) = 12/30= 0.4.

The *Gini index initial value*, therefore, can be computed as follows:

$Gini_{initial} = Gini(D) = 1 - (0.33^2 + 0.266^2 + 0.4^2)$ = 1- (0.1089 + 0.070 + 0.16) = 1- 0.3389 = **0.6611**.

To find out which splitting criterion will be the best at the node in D, we have to compute the Gini index for each descriptive variable, *Study year*, *Study field*, *Gender*, respectively. First, we consider the variable Study year and its each of the possible splitting subsets, {1}, {2}, {3}, {1,2}, {1,3}, {2,3}.

The general formula indicates the weighted sum of the impurity of each partition. Given that the binary split on descriptive variable A partitions the training set D into two subsets D_1 and D_2, the Gini index of D will be equal:

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2)\ .$$

The Gini index values for possible splitting subsets of *Study year* can be computed as follows:

$$Gini_{Study\ year\ \epsilon\ \{1\}}\ (D) = \frac{|D_1|}{|D|} Gini\ (D_1) + \frac{|D_2|}{|D|} Gini(D_2) = \frac{10}{30} Gini(D_1) + \frac{20}{30} Gini(D_2) =$$
$$\frac{10}{30} \times \left(1 - \left(\frac{3}{10}\right)^2 - \left(\frac{3}{10}\right)^2 - \left(\frac{4}{10}\right)^2\right) + \frac{20}{30} \times \left(1 - \left(\frac{7}{20}\right)^2 - \left(\frac{5}{20}\right)^2 - \left(\frac{8}{20}\right)^2\right) = 0.333(1 - 0.09 -$$
$$0.09 - 0.16) + 0.666(1 - 0.1225 - 0.0625 - 0.16) = 0.217 + 0.436 = \mathbf{0.653}\ .$$

Concerning the subset {1} of descriptive variable *Study year*, in first partition (D_1), there are 10 persons who satisfy the condition "$study\ year\ \epsilon\ \{1\}$". The remaining 20 people would be placed in second partition (D_2). We note that as the split is the binary character, there will be

[155] The content and some notations in this section are based on Han et al. (2012 [2000], 341-343).

always two partitions. The split criterion "$study\ year\ \epsilon\ \{1\}$" will be at the node. The rule is units (i.e., person here) satisfying the split condition will go to left branch, and units who do not satisfy the split condition will go to right branch of the tree.

$Gini_{Study\ year\ \epsilon\ \{2\}}\ (D) = \frac{10}{30} Gini(D_1) + \frac{20}{30} Gini(D_2) = \frac{10}{30} \times \left(1 - \left(\frac{3}{10}\right)^2 - \left(\frac{3}{10}\right)^2 - \left(\frac{4}{10}\right)^2\right) + \frac{20}{30} \times \left(1 - \left(\frac{7}{20}\right)^2 - \left(\frac{7}{20}\right)^2 - \left(\frac{6}{20}\right)^2\right) = 0.333(1 - 0.09 - 0.09 - 0.16) + 0.666(1 - 0.1225 - 0.1225 - 0.09) = 0.217 + 0.442 = \mathbf{0.659}$.

$Gini_{Study\ year\ \epsilon\ \{3\}}\ (D) = \frac{10}{30} Gini(D_1) + \frac{20}{30} Gini(D_2) = \frac{10}{30} \times \left(1 - \left(\frac{4}{10}\right)^2 - \left(\frac{2}{10}\right)^2 - \left(\frac{4}{10}\right)^2\right) + \frac{20}{30} \times \left(1 - \left(\frac{6}{20}\right)^2 - \left(\frac{6}{20}\right)^2 - \left(\frac{8}{20}\right)^2\right) = 0.333(1 - 0.16 - 0.04 - 0.16) + 0.666(1 - 0.09 - 0.009 - 0.16) = 0.211 + 0.439 = \mathbf{0.650}$.

$Gini_{Study\ year\ \epsilon\ \{1,2\}}\ (D) = \frac{20}{30} Gini(D_1) + \frac{10}{30} Gini(D_2) = \frac{20}{30} \times \left(1 - \left(\frac{6}{20}\right)^2 - \left(\frac{6}{20}\right)^2 - \left(\frac{8}{20}\right)^2\right) + \frac{10}{30} \times \left(1 - \left(\frac{4}{10}\right)^2 - \left(\frac{2}{10}\right)^2 - \left(\frac{4}{10}\right)^2\right) = 0.666(1 - 0.09 - 0.09 - 0.16) + 0.333(1 - 0.16 - 0.04 - 0.16) = 0.439 + 0.213 = \mathbf{0.652}$.

$Gini_{Study\ year\ \epsilon\ \{1,3\}}\ (D) = \frac{20}{30} Gini(D_1) + \frac{10}{30} Gini(D_2) = \frac{20}{30} \times \left(1 - \left(\frac{7}{20}\right)^2 - \left(\frac{5}{20}\right)^2 - \left(\frac{8}{20}\right)^2\right) + \frac{10}{30} \times \left(1 - \left(\frac{3}{10}\right)^2 - \left(\frac{3}{10}\right)^2 - \left(\frac{4}{10}\right)^2\right) = 0.666(1 - 0.1225 - 0.0625 - 0.16) + 0.333(1 - 0.09 - 0.09 - 0.16) = 0436 + 0.219 = \mathbf{0.655}$.

$Gini_{Study\ year\ \epsilon\ \{2,3\}}\ (D) = \frac{20}{30} Gini(D_1) + \frac{10}{30} Gini(D_2) = \frac{20}{30} \times \left(1 - \left(\frac{7}{20}\right)^2 - \left(\frac{5}{20}\right)^2 - \left(\frac{8}{20}\right)^2\right) + \frac{10}{30} \times \left(1 - \left(\frac{3}{10}\right)^2 - \left(\frac{3}{10}\right)^2 - \left(\frac{4}{10}\right)^2\right) = 0.666(1 - 0.1225 - 0.0625 - 0.16) + 0.333(1 - 0.09 - 0.09 - 0.16) = 0.436 + 0.219 = \mathbf{0.655}$.

Based on the Gini index results above, we can say that the best binary split on the variable *Study year* is on $\{3\}$ as this subset minimized best the Gini index. Then, we compute the *reduction in impurity* of this split as follows:

$\triangle\ Gini(A) = Gini(D) - Gini_A(D) = 0.661 - 0.650 = \mathbf{0.011}$,

where A represents the descriptive variable, and D is used for the partition at the node N. The partition (training set) at the node N is partitioned into two subsets D_1 and D_2 as the split is binary.

We can compute the Gini index in same manner as above for the two other descriptive variables *Study field*, and *Gender*. At the end, the binary split yielding the maximum reduction in impurity in D will be chosen as the splitting criterion at the node, which yields the largest reduction in the value of the Gini index will be chosen as the *splitting variable* in the tree. The node will be labelled with the best criterion generating two branches. Units, therefore, satisfying the best split criterion will go to left branch, and units who do not satisfy the split criterion will go to right branch of the tree. For instance, if the best split is $Study\ year = \{2\}$, then the split in the tree will be visualized as in Figure 17.

The computations above have focused on qualitative type of variables. But, in the case of quantitative descriptive variable, how the split will be built? The method is the same as of information gain (see the information gain explained above). Recall that the descriptive variable that maximizes the reduction in impurity (or, equivalently, has the minimum Gini index value) will be selected as the splitting variable, and in each split, there will be two partitions such as D_1 is the training set in D satisfying the condition $A \leq split_point$, and D_2 is the remaining training set in D satisfying the condition $A > split_point$. This variable and its split-point yielding the minimum Gini index value together form the splitting criterion.

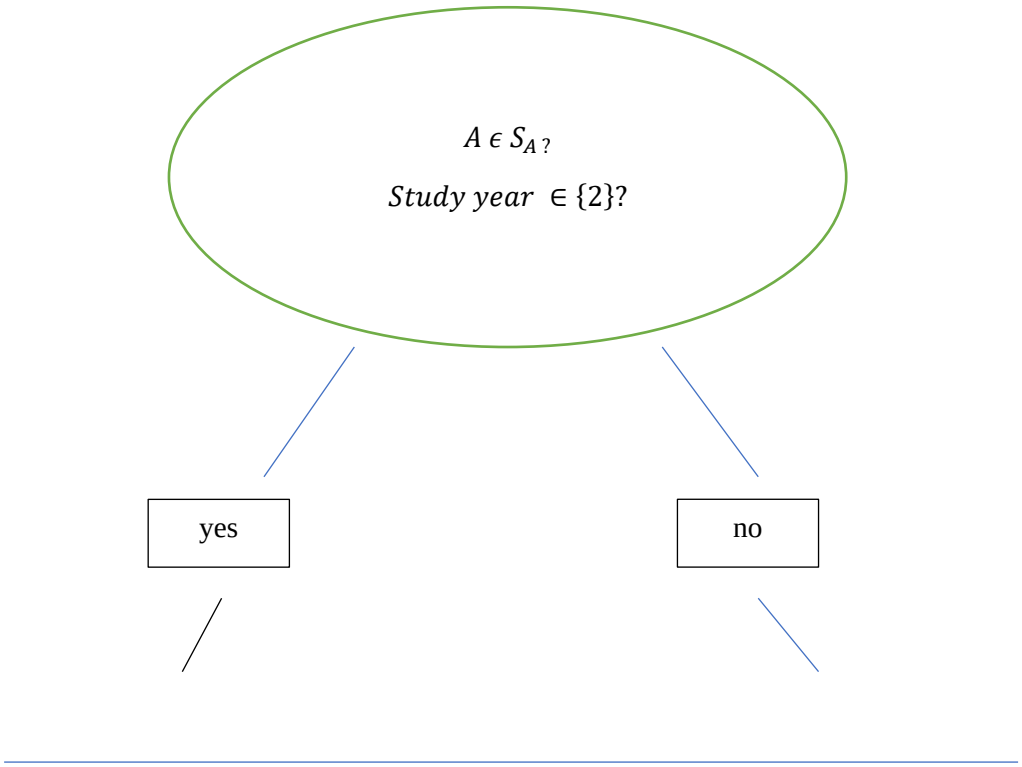


Figure 17. An example of visualization of the three after the split on descriptive variable (A) yielding the best split

We have seen two methods for determining the split criterion at the node of classification tree. The first method, namely the information gain is used in Quinlan's decision trees methodology. The second method, namely the Gini index is used in CART methodology to build decision trees. These two methods are commonly used for building decision trees. There are other methods as well such as gain ratio, the algorithm CHAID (Chi-squared automatic interaction detection), C-SEP, G-statistic, for attribute selection. The CART methodology provides multivariate splits as well (i.e., linear combination of descriptive variables). Each method has some bias; for instance, information gain and Gini index method are based concerning multivalued variables, and it is not very recommended to use a target response variable composed of a large number of categories/classes in Gini index method. But it is important to note that information gain and Gini index methods are recognized as efficient tools providing good results (see Han et al. 2012 [2000]).

Misclassification[156]

CART assigns a predicted class (one of the categories of the outcome/response variable) to each node with respect to the minimization rule of the misclassification costs. Nevertheless, all classification trees generate some errors. The resubstitution estimate $R^{re}(T)$ provides the misclassification rate in each terminal nodes of the tree; in other words, she measures the rate of subjects that do not correspond to the predicted class at the node. The resubstitution estimate is (for the formulas, see Izenman (2008, 294)):

$$R^{re}(T) = \sum_{l=1}^{L} r(r_l)p(r_l)$$

where $p(r_l)$ represents the proportion of all observations at node r_l, $r(r_l)$ designates the misclassification rate of observations at node r_l , $T = r_1, r_2, \dots, r_l$ the set of all terminal nodes with $l = 1,2, \dots, L$.

The general formula of the misclassification rate at the node is:

$$r(r) = 1 - max_k p(k|r)$$

The optimal result is a situation where the minimum as possible misclassification rate is obtained. Given that the class of terminal node (r) is k (determined by the majority class vote), greater observations at the node are classified as k lower the misclassification rate at the node. When the response variable has only two modalities, the general formula of misclassification rate simplifies to:

$$r(r) = 1 - \max(p, 1 - p)$$

It is important to note that resubstitution rate is considered relatively unreliable. Instead, the cross-validation rate, and 1-SE Rule are used to measure misclassification (Breiman et al. 1984).

We will use Edgar Anderson's Iris data[157] iris available in R to show how to compute the resubstitition error in practice. From Figure 18 below, we can observe that at the first terminal node (left side of the first lay of tree) the attributed class is setosa (there are 3 classes for response variable iris flower: setosa, versicolor and virginica) and in that node there are 0 versicolor and 0 virginica. This means that in that node there is no misclassification because all observations in that node are classified as setosa, which is the major class (class of node) in the node. Same reasoning for other terminal nodes can be done; for instance, in the second terminal node (right side of third lay of tree), the attributed class is virginica but there is a 1 iris flower classified as versicolor. So, in that node, the misclassification rate is 1/46, which can be shown as follows:

$$r(r_2) = 1 - \max(p, 1 - p) \; where \; p = \frac{45}{46} = 0.97826 \; and \; (1 - p) = 0.02174.$$

[156] The formulas and explanations in this sub-section are based on Izenman (2008, 294).

[157] Please refer to the link below for more information on the Iris data (accessed 22 September 2019): https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/iris.html

where p is the probability that an iris flower (subject) classifies as virginica (attributed class of node) and (1-p) corresponds to the probability that a iris flower classifies as setosa or versicolor (p is the probability of success and q=(1-p) is the probability of failure; a success is the event that the subject falls into the attributed class of terminal node (virginica),and failure is the event that the subject does not fall into the attributed class of terminal node).

Thus,

$$r(r_2) = 1 - \max(p, 1-p) = \min(p, 1-p) = \min(0.97826, 0.02174) = 0.02174.$$

and,

$$\min(0.97826, 0.02174) = \frac{1}{46} = 0.021.$$

This computation can be applied for the misclassification rate measure of each terminal node (in this example we have 6 terminal nodes $T = \{r_1, r_2, \dots, r_6\}$).

After computing the within-node misclassification rate, we obtain the global misclassification rate of the tree: $R^{re}(T) = (1+1)\ (50+1+45+47+1+2+1+3) = 2\ 150 = 0.0133$

$$R^{re}(T) = \frac{1+1}{50+1+45+47+1+2+1+3} = \frac{2}{150} = 0.0133.$$

where N (total population= total quantity of iris flowers) = 150. Consider now

$$R^{re}(T) = \sum_{r_1 \in T, k} [\min\{p(k|r_1), 1 - p(k|r_1)\} \times p(r_1)]$$

where $k = 1,2, \dots, K$. In this case, $K = 3$ because iris flower has 3 classes: setosa, versicolor and virginica. The number of terminal nodes in the tree: $r \in T; r = 1,2,, \dots,6$.

$$R^{re}(T = 6) = \sum_{r_1 \in T=6, k=3} \min\{p(k|r_1), 1 - p(k|r_1)\} \times p(r_1)$$

$$= \left[\left\{\min\left(\mathrm{p}(1|1), 1 - \mathrm{p}(1,1)\right)\right\} \times \mathrm{p}(\mathrm{r}_1) + \right\}\min\ (\mathrm{p}(3|2), 1 - \mathrm{p}(3|2))\} \times \mathrm{p}(\mathrm{r}_2) +, \dots, +\{\min\ (\mathrm{p}(2|6), 1 - \mathrm{p}(2|6))\} \times \mathrm{p}(\mathrm{r}_5) + \{\min\ (\mathrm{p}(3|6), 1 - \mathrm{p}(3|6))\} \times \mathrm{p}(\mathrm{r}_6)]$$

$$= \left[\left\{\min\left(\frac{50}{50}, 1 - \frac{50}{50}\right) \times \frac{50}{150}\right\} + \min\left(\frac{45}{46}, 1 - \frac{45}{46}\right) \times \frac{46}{150}\right\} + \left\{\min\left(\frac{47}{47}, 1 - \frac{47}{47}\right)\ \times \frac{47}{150}\right\} + \left\{\min\left(\frac{1}{1}, 1 - \frac{1}{1}\right) \times \frac{1}{150}\right\} + \left\{\min\left(\frac{2}{3}, 1 - \frac{2}{3}\right)\ \times \frac{3}{150}\right\} + \left\{\min\left(\frac{3}{3}, 1 - \frac{3}{3}\right) \times \frac{3}{150}\right\}\ ] = 0.013$$ that is equal to $R^{re}(T) = \frac{1+1}{N=150} = 0.013$.

The Re-substitution error rate measures the performance of the maximum sized tree on the training data set. The largest tree will always yield the lowest resubstitution error rate, because the maximum sized tree will always give the best fit to the training data. In general, it is not a good idea to let the tree keep growing until all terminal nodes are homogeneous, because it is very difficult to interpret the large trees.

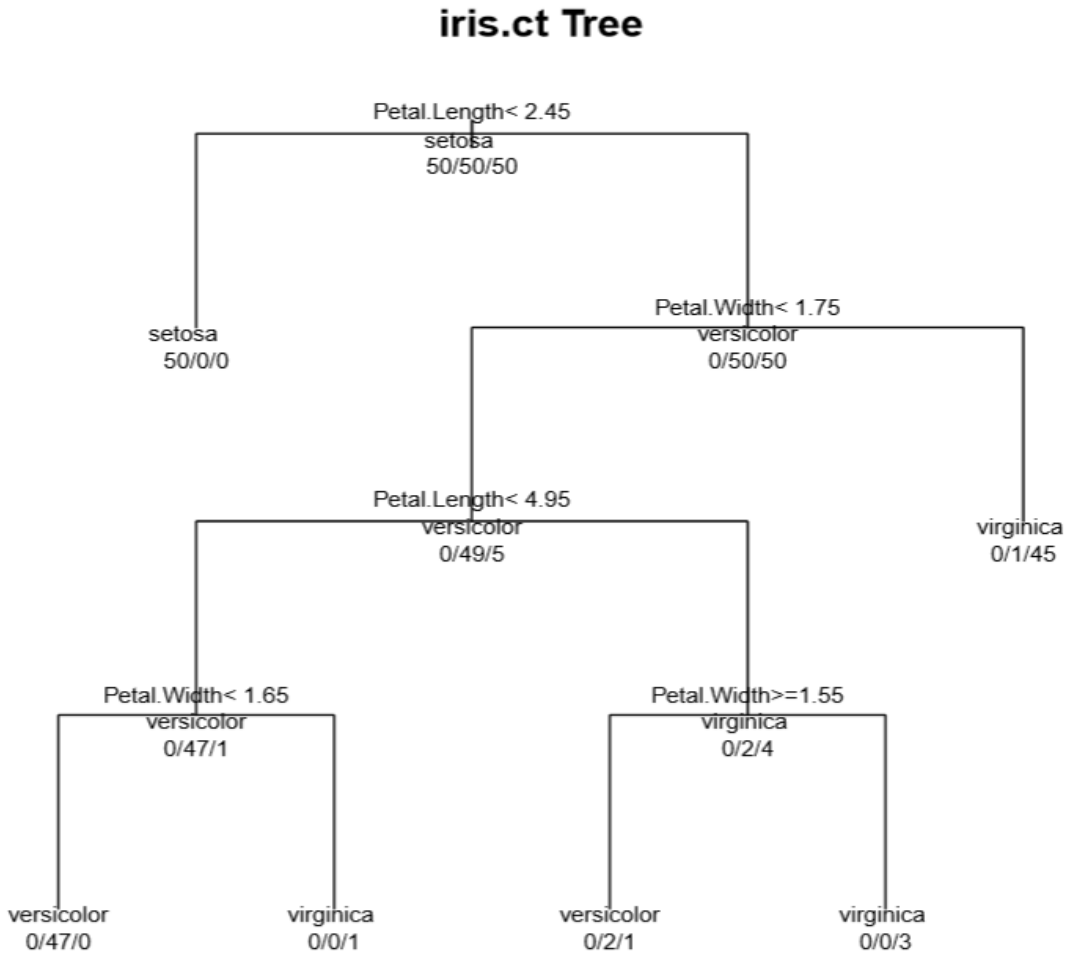


Figure 18. Iris data classification tree

Pruning trees[158]

The problem of *overfitting* such as outliers and noise in the training data, generally, emerges when a classification tree is built. To solve this problem, the bottom-up process of *pruning* (i.e., obtaining a sequence of subtrees from the initial big/saturated tree) is employed in order to remove the most problematic branches and leaves which do not bring substantial contribution in terms of information for prediction. A pruned tree is smaller compared to the non-pruned (large) tree in terms of size, with less but the most meaningful branches. The initial classification trees are generally very big and complex for interpretation and to get the comprehensible results. Herewith, pruning process makes also easier the interpretation and detection of patterns in the tree. As mentioned by Han et al. (2012 [2000], 344), there are two main approaches to prune the tree*: pre-pruning* and *post-pruning*. The *pre-pruning* approach consists of stopping the growt of the tree before the tree is built; for instance, the user/researcher can decide in which node the tree should stop to grow (i.e., user-defined criteria). However, this process in not very recommended due to difficulties for selecting the right threshold for stopping the growth of the tree. High thresholds may generate oversimplified tree, and in contrast low thresholds may produce insignificant reduction. The post-pruning approach is

[158] This section is based on Han et al. (2012 [2000], chapter 8).

more commonly used, which first let grow the tree (i.e., fully-grown tree), and then, to cut the tree at an optimal level (i.e., the tree will have less branches and nodes). By this operation, some branches will be removed and some interior nodes, therefore, will become terminal nodes (i.e., leaf). However, this in turn means losing some information in terms of split criterion and reduce the accuracy of prediction (i.e., increase of misclassification). As we have seen, it is a very delicate situation to decide where we have to cut the tree by being able to keep at the same time a relatively substantial amount of information and a good accuracy. This is why it becomes crucial to understand tree complexity.

We note that, in the literature up to now, only two methodologies, namely C4.5 and CART have employed a pruning process; otherwise, other methods just let the tree grow until no splitting is possible or the user defines a specific criterion to stop growing the tree (Rokach and Maimon, 2010, 151). For instance, the *cost complexity parameter* in the Bremain et al.'s CART methodology (1984) plays a crucial role in the post-pruning process, which is computed based on the number of leaves in the tree (i.e., terminal nodes), and the error rate of the three (i.e., the percentage of units misclassified in the tree). From the bottom of the tree, for each internal node, (i) the cost complexity of the subtree at the node N; and (ii) the cost complexity of the pruned subtree at the node N, are computed. If the cost complexity parameter of the pruned tree is smaller than those of non-pruned tree, then the tree will be pruned. The smallest tree with least cost complexity parameter is considered the optimal. In Quinlan's C4.5 method, the *pessimistic pruning* process is applied for pruning the tree. As Rokach and Maimon (2010, 151) mentioned finding an optimal (pruned) tree is NP-hard[159] (see also Hancock et al. 1996 for more details on NP-hard), and it is 'feasible only in small problems'.

Now, we will see in more detail the process of pruning, which is firstly relies on the estimation of the resubstitution or the error rate R(r) at each node r ∈ T. In the second step of the process, the operation of pruning from the bottom of tree toward the top of tree (root node) steps in, under the condition that the estimate of R(r) is minimized. However, as the resubstitution rate overestimates the performance of the tree on a test set, the complexity parameter is preferred.

For any node $r \in T$, we denote the complexity parameter $\alpha > 0$, then,

$$R_\alpha(r) = R^{re}(r) + \alpha$$

which can be written as

$$R_\alpha(T) = \sum_{l}^{L} R_\alpha(r_1) = R^{re}(T) + \alpha|\tilde{T}|$$

where $T = \{r_1, r_2, \dots, r_L\}, r \in T, \left|\tilde{T}\right| = L$ the number of terminal nodes in the subtree T, and $|\tilde{T}|$ corresponds to the penalty term that aims to punish the re-substitution error for generating very large trees.

If T represents all terminal nodes of the saturated tree (before pruning process), for each complexity parameter α, the minimization process is as follows:

[159] It is called *non-deterministic polynomial-time hardness*. The term comes from the field of computational complexity theory.

$$R_\alpha(T(\alpha)) = min_T R_\alpha(T)$$

The value of α is the size of tree. In other words, α determines from which branch of the tree, she can be pruned under the condition that the minimization of the misclassification rate is satisfied. If α is very small (e.g., α = 0.0001), the penalty term $\alpha|\tilde{T}|$ will be small also and then $R_\alpha(T)$ will be bigger. If we fix the complexity parameter to zero, the tree will grow to its maximum size with maximum terminal nodes T_{max} and in this case $R_{re}(T_{max})$ will be equal to zero. We can conclude that the maximum tree size minimizes the re-substitution rate. On the other hand, if we increase α, the pruned tree will have less nodes. But, if α is very large, we will have only the root node because the greater parameter tends to prune the entire tree.

Suppose that the node r in the saturated tree has two terminal child nodes r_L and r_R. Breiman et al.'s (1984) 4.2 proposition determines that $R^{re}(r) \geq R^{re}(r_L) + R^{re}(r_R)$ where r is the node in the tree, r_L denoted the left side node and r_R denotes the right side node (r splits into r_L and r_R). In this equation, when equality occurs then the tree's terminal nodes r_L and r_R are pruned. This process can continue until no further pruning is possible under this condition. Now, the important question is "When do we stop pruning?" Breiman et al. (1984) have proposed two estimation methods: independent test sample vs. cross-validation method. When dataset is very large, use of an independent test set is highly preferred, otherwise the cross-validation method can be chosen. We will apply the cross-validation procedure, the selection of the optimal tree based on this procedure will be directly shown in the case analysis with the rpart package codes in R.

Predictive accuracy of classifier

After having built a classification tree and get the results, it is important to evaluate the patterns obtained in the tree. To find out how accurately the tree classifier has estimated the behavior of units in the data according to the target variable, we can apply several tools. By saying 'accuracy', we question the performance of the decision tree classifier for class prediction. The decision tree classifier produces estimation based on learning and training phases. The main question will be how the classifier has achieved to realize correct predictions by learning from the training data set. The answer to that question will be emerged by analyzing the reliability of the classifier. The reliability checking process will be based on measuring the classifier's performance on a test set that consists of the data samples that were not used in the training data. We previously saw that the initial data is randomly fragmented into the two groups: training vs. test sets. All probability-based computations defined regarding the classifier's algorithm are done in the training data. The classification rules, therefore, discovered in the training data will be used to predict the target response variable in the test data. By comparing the predicted and the already existing classification values in the test data, the classifier's performance is evaluated for her capacity of producing accurate results. In case of a total accurate prediction, the predicted class should be fully matched with the actual class in the test data. To evaluate the predictive accuracy of classifier, we will focus on some essential tools provided in classification trees framework that are the training and test data sets, the k-fold cross-validation, the N-fold cross- validation, confusion matrix, re-substitution error, misclassification rate.

Independent test set: training (or called learning) and test sets

Data classification process consists of two steps: *learning* (*training* phase) and *classification*. The initial data set is partitioned into two parts: training vs. test sets. After that, firstly, in

learning process[160], the training data set is used to build the classification algorithm to get the classification rules that allow to predict the patterns of the target response variable as a function of regressors[161]. Secondly, in the test data, (i) the classification rules will be tested regarding their accuracy; (ii) in the literature, the classification process is also referred to the "learning of a mapping or function $y = f(x)$, that can predict the associated class label y of a given tuple X" (Han et al. 2012 [2000], 330). As we have previously mentioned, this mapping can be shown in three different formats that are classification rules, mathematical formulae, and classification trees that we focus on. As a function of the quality of predictions made, the discovered classification rules can be used for classifying other data sets. The optimal way of separating the initial data set into two distinct sets (training and test sets) is using the random division. Regarding Bramer (2007), the most usual random divisions are 1:1, 2:1, 70:30, 60:40 where for example, training set has randomly selected half of the initial data set and the test set has the other half; or, training set has randomly selected 70 per cent of the initial data set and the test set has the remaining 30 per cent of the initial set.

Suppose that there are N instances in the test set and X amount of them are correctly classified based on prediction from the training set; hence, the predictive accuracy of the classification trees in test set (i.e., the proportion of the correctly classified undiscovered cases by classifier) is X divided by N (i.e., $p = X/N$). We can even obtain a confidence level for the predictive accuracy, based on its standard error:

$$S = \sqrt{p\,(1-p)/N}$$

Hence, the confidence interval (CL) for the predictive accuracy is:

$$CL = p \pm Z_{CL} \times S$$

where, the term Z_{CL} can be taken as 1.96 for 95 per cent of confidence interval. For example, if there are 60 cases out of 100 cases of the test set are accurately predicted; hence, the predictive accuracy will be 60/100=0.6. The related standard error will be: $S = \sqrt{0.6(0.4)/100}$ = 0.048. According to this, CL= [0.6-1.96*0.048; 0.6+1.96*0.048] = [0.50; 0.69]. This means that the true predictive accuracy can take a value between 0.50 and 0.69 with a probability of 95 per cent. In parallel, the term of error rate can be used as well, that is equal to 0.4 when 60 cases out of 100 cases of the test set are accurately predictive.

k-fold cross-validation process

K-fold cross-validation is generally used when the initial data set contains relatively a small number of cases. The technique is to split the data set into *k* equal (or approximately equal size) and mutually exclusive parts (or subsets or '*folds*'). If the data set cannot be equally divided up, the last part will have fewer cases compared to the other *k-1* parts. Suppose that the data set can be equally segmented into *k* parts, $D_1, D_2, \ldots, D_k$. Training and testing process is used k times. For each k performance, we will use the term '*iteration*' to mention which partition of the data is used for test set. In iteration i, the partition D_i will be used as test set, and the remaining partitions will be used as training set (i.e., training the model). For instance, in first iteration,

[160] Learning from the training set.

[161] The training set is composed of training samples. In some books of machine learning, the term training samples is also used as training tuples. Throughout this text, we prefer to employ the term training samples.

I_1, the subsets (partitions) $D_{2,\ldots,}D_k$ will be used as first training set leading to the first results, which will later be tested on D_1; in second iteration, I_2, the subsets $D_1, D_3, \ldots, D_k$ will be used as second training set leading to the second results, which will later be tested on D_2, and so on. The predictive accuracy of classification trees is the overall number of correctly classified instances from the k iterations (i.e., overall classification accuracy), divided by the total number of samples in the initial data set.

N-fold cross-validation process

If there are N instances in the data set, the algorithm is that each of the instances in the data was considered as a test set while the other $N - 1$ cases were used as the training set. The difference compared to k-fold cross-validation process is in N-fold cross-validation instead of dividing up the data set we focus on each instance in the data. The predictive accuracy and its standard error are the same as in the k-fold cross-validation process. This method requires a large number of computations; for that reason, it is not very suitable for the big data sets.

Leave-one-out cross-validation

This type of cross-validation consists of one iteration compared to the k-fold cross-validation process. Only one sample is 'left out' for testing.

Stratified cross-validation

In case of stratified cross-validation, the samples (i.e., folds) are stratified in the sense that the distribution of classes in each fold is approximately the same as in the initial data set. The stratified 10-fold cross-validation is generally used and recommended for getting accurate predictions (e.g., low bias and low variance).

We also note that in addition to cross-validation process, there are three other methods, namely holdout, random sampling, and bootstrapping for partitioning the data into the training and test sets (See Han et al. 2012 [2000], chapter 8). The CART methodology employs cross-validation process. We don't enter into details concerning other methods of partitioning, such as holdout, random sampling, bootstrapping.

Evaluation measures and Confusion Matrix[162]

In this sub-section, we will see some evaluation measures, which will help us to find out whether the classification rules extracted from the training sample do accurate predictions on the test sample. With a good accuracy, the predicted and the known/actual values of the test data (or called validation data) should be matched. First of all, it is important to mention that there are two fundamental notions for understanding evaluation measures that are *positive samples (P)* vs. *negative samples (N)*[163]. In classification trees, as the target response variable is a categorical variable having at least two categories/modalities, the main class of interest among categories is called positive and other classes are called negative. Herewith, instances in the data having the main class of interest are called positive samples and instances in the data having other classes are called negative samples. For instance, suppose that we search for discovering the patterns of domestic workers. In that case, the variable being domestic worker vs. being not domestic worker (i.e., dichotomous variable) will be considered as the target response variables. The main class of interest in prediction will be 'being domestic worker'.

[162] This section is based on Han et al. (2012 [2000], chapter 8).

[163] We recall that the notion *sample* is replaced by *tuple* in some reference books and texts such as Han et al. (2012 [2000]).

Instances of being domestic worker will be called as positive instances (or being part of the positive sample). Instances of not being domestic worker will be called as negative instances (or as being part of the negative sample). Evaluation measures are summarized by a table in Han et al. (2012 [2000], 365) as follows:

Table 29. Evaluation measures of classification trees. Source: Han et al. (2012 [2000], 365)

Measure	**Formula**
Accuracy, recognition rate	$\frac{(TP+TN)}{(P+N)} = sensitivity \times \frac{P}{(P+N)} + specificity \times \frac{N}{(P+N)}$
Error rate, misclassification rate	$(FP+FN)/(P+N)$
Sensitivity, true positive rate, recall	$(\frac{TP}{P})$
Specificity, true negative rate	$(\frac{TN}{N})$
Precision	$(TP)/(TP+FP)$
F, F_1, F-score, harmonic mean of precision and recall	$(2 \times precision \times recall)/(precision + recall)$
F_β, where β is a non-negative real number	$\frac{((1+\beta^2) \times precision \times recall))}{(\beta^2 + recall)} \times precision$

As we can see in the table above, the terms TP, TN, FP, FN, P, N are the fundamental of classification trees evaluation measures. They refer respectively to: true positives, true negatives, false positives, false negatives, positive, and negative samples. True positives are the positive instances that are correctly predicted by the classifier. True negatives are the negative instances that are correctly predicted by the classifier. False positives are the negative instances that are not correctly predicted by the classifier (e.g., an instance is originally affiliated with the class 'not being domestic worker' but it is predicted by the classifier as 'being domestic worker'). False negatives are the positive instances that are predicted as negative by the classifier (e.g., an instance is originally affiliated with the class 'being domestic worker' but it is predicted by the classifier as 'not being domestic worker'). These terms are generally summarized by the confusion matrix (see Table 30).

By using a confusion matrix, it is possible to evaluate the overall predictive accuracy (i.e., the misclassification rate), quality of classifier. It informs both the percentage of correctly classified and of misclassified cases, for a binary classification problem (i.e., target response variable should be binary). For example, we suppose that 446 Japanese tourists will visit the city of Paris. As a function of the available information on Japanese tourists regarding their socio-economic, personal, and hobby information, the French travel agency did prediction for classifying Japanese tourists whether they prefer to visit Museum Louvre, or Eiffel Tower, or Museum d'Orsay at the first day of their visit to Paris. Table 31 below shows that 73 Japanese tourists prefer to visit Museum Louvre on their first day and this is correctly classified (52% of them) by classifier. However, among 141 Japanese tourists wishing to visit Museum Louvre on

their first day in Paris are misclassified by classifier such as 42 of them (30%) are classified as preferring to visit the Eiffel Tower, 26 of them (18%) are classified as preferring to visit the Museum d'Orsay. Regarding 160 Japanese tourists wishing to visit Eiffel Tower on their first day in Paris, 98 of them (61%) are correctly classified but 62 of them (the sum of 17 and 45) are misclassified. In total, the travel agency made a correct prediction about 37 % (i.e., (73 + 17 + 76) / 446).

Table 30. Confusion matrix of Japanese tourists' classification

	Classified as			
	Museum Louvre	**Eiffel Tower**	**Museum d'Orsay**	**Total**
Visiting Museum Louvre	*73*	42	26	141
Visiting Eiffel Tower	98	*17*	45	160
Visiting Museum d'Orsay	36	33	*76*	145
Total	**207**	**92**	**147**	**446**

If all tourists were correctly classified, the non-zero values should *only* be appeared on the *leading diagonal* from the top left to the bottom right in the confusion matrix (and all other cases should have zero).

Above we showed an example for multiple classes, but we can also build a confusion matrix in a similar manner for a binary classification problem. We can take the example of domestic worker; the predicted class and the actual class figures are reflected in the table below.

Table 31. Confusion matrix of domestic workers classification.

	Predicted class		Total	**Recognition (%)**
	Domestic worker=yes	**Domestic worker=no**		
Actual class				
Domestic worker=yes	68	34	**102**	66.7
Domestic worker=no	114	212	**326**	65
Total	**182**	**246**	**428**	65

Regarding the example of domestic workers classification, we can see from Table 32 that 68 persons are correctly predicted as domestic workers. Despite 34 persons are in fact domestic workers, they are not predicted as domestic worker by the classifier. On the other hand, 114 persons are not domestic workers, but they are predicted as domestic workers by the classifier

and the rest 212 no domestic workers are correctly classified by the classifier. The last column in the table shows the recognition rate of correct prediction in percentage. They are in total 102 domestic workers, and 68 of them are correctly predicted by the classifier. In this regard, when we divide 68 by 102 and multiply it by 100, then we find that around 67 per cent of the sample of domestic worker are correctly classified. But also, they are 326 non-domestic workers in the sample. Among non -domestic workers, 212 of them are correctly predicted, which in turn yields a recognition rate of 65 per cent. When we consider whole sample ($N = 428$), we find that in total 280 persons (68 TP + 212 TN) are correctly predicted by the classifier. The total recognition rate, therefore, corresponds to 65 per cent ($= ((68 + 212)/428) \times 100$).

In the case of a data set with only two classes (binary classification), the confusion matrix corresponds to:

Table 32. True and false positives and negatives

	Predicted class		Total
	Classified as +	**Classified as –**	
Actual class			
+	True positives-TP	False negatives (Type II Errors)	P
-	False positives- FP (Type I Errors)	True negatives	N
Total	P’	N’	P+N

The true positives and true negatives represent the correct classification. Hence, a hundred per cent correct classifier should have zero both in false positives and in false negatives.

After presented the confusion matrix, now we will see the predictive accuracy measures mentioned in Table 33. The formulas of these measures are based on the confusion matrix elements (see Table 30 for formulas). We start off by the **accuracy rate** of the classifier that corresponds to the percentage of correctly predicted samples in test data by the classifier. Secondly, there is an **error rate** or **misclassification rate** that equals to $1 - accuracy\ of\ the\ classifier$ in test set. In machine learning literature, if instead of test set, the error rate is computed in training set, then this error rate is known as **re-substitution error rate**. The accuracy rate is generally recognized as inefficient when **class imbalance problem** emerges in the data. Suppose that the target response variable is binary having two classes c_1, c_2. If the percentage of units having taken c_1 is significantly higher than the percentage of units having taken c_2 then, accuracy rate may not be acceptable due to unbalanced distribution of the target response variable in the data. To handle this problem, **sensitivity** and **specificity** measures come into play[164]. The **sensitivity rate**, refers as the *true positive rate*, is the percentage of correctly classified positive samples. The **specificity rate**, refers as the *true negative rate*, is the percentage of correctly classified negative samples. In addition, the **precision rate** shows the percentage of correctly classified positive samples compared to the

[164] But also, ensemble methods (i.e., “methods that generate many classifiers and aggregate their results” (Liaw and Wiener 2002, 18)) such as *bagging* (Breiman 1996), *boosting* (Shapire 1999), *random forests* (Breiman 2001) can be used to address this problem.

sum of correctly classified positive sampled and negative samples classified as positive samples. This rate is generally known as a measure of *exactness*. If it is equal to 1, this means that the classifier achieved to predict hundred per cent correctly all positive samples (i.e., the number of negative samples classified as positive samples will be zero). The measure, ***recall*** (i.e., sensitivity, true positive rate-see Table 30), is a measure of *completeness* that yields the rate of correctly classified positive samples among all positive samples. If the recall is equal to 1, this informs us that all positive samples are predicted as positive; but compared to the precision rate, the recall does not inform about the number of negative samples that are incorrectly predicted as positive. The $\boldsymbol{F}$ **measure** (or $F_1 score, or\ F-score$) and the $\boldsymbol{F_\beta}$ **measure** combine *precision* and *recall*. The $\boldsymbol{F}$ **measure** corresponds to the *harmonic mean* of precision and recall, whilst the $\boldsymbol{F_\beta}$ **measure** corresponds to the *weighted* measure of precision and recall, where β is a non-negative real number. Most frequently, the parameters $\beta = 2$ or $\beta = 0.5$ are used.

Furthermore, we note also that different classifiers (i.e., algorithms) can be evaluated in terms of accuracy by employing significance tests, or ROC curves.

Missing values[165]

It is possible that some instances in the data do not have the values/modalities for some variables. This situation is known as *missing values* in statistical literature. In case of a considerable amount of missing values, prediction problems may emerge. There are three possibilities to cope with missing values. First option is to drop observations having missing values in the data. But this process may reduce the reliability of the results in case there is a considerable amount of missing values. The second option is to replace missing values of a qualitative variable by its most frequently occurring category/modality in the training set, or to replace missing values of a quantitative variable by its average value in the training set. And third option is to create other additional nodes specifically for missing values (see Strobl 2014).

Concerning classification trees, if the observation has a missing value for the splitting variable at the node, one possible way is to drop this observation down the tree and construct the tree by using only the complete data as far as possible (see Clark and Pregibon 1990, cited in Izenman 2008, 312) or the second possible way is to drop down all observations with missing values to the same daughter node (Zhang and Singer 1999; cited in Izenman 2008, 312). Otherwise, if the variable with missing values is not selected for the split criterion in the tree, in this case the related observations will be dropped down directly to the terminal node according to its classes. In the framework of CART methodology, Breiman et al. (1984) have proposed *surrogate* splits method to employ a substitute variable that can best predict and replace the split variable with missing values. The surrogate variables are proxy variables for the main splitting variable. In case where an observation is missing for all surrogate splits then this observation will be dropped down in the tree and will be sending to the child node with the largest relative frequency. Moreover, for qualitative variable with categories/modalities, Kass (1980) has proposed to create an additional missing values category/modality, namely 'missing' or 'NA' for that variable (cited in Izenman 2008, 313).

[165] This section is based on Izenman (2008, 312-313 chapter 9).

Advantages of classification trees

Classification trees is a flexible and robust analytical method, which is able to deal with a large number of variables, non-linear relationships, high-order interactions, and missing values (De'ath and Fabricius 2000). In the process of construction of the trees, the distributional assumptions of any kind do not exist (i.e., non- parametric method). Thus, no variable should necessarily follow any kind of statistical distribution and classification trees do not require normality assumptions of the data. The classification trees method is very adequate for the interactive exploration, description and discovery of knowledge on patterns. Especially, it helps meaningfully to emerge the hidden patterns in the data. There is no restriction concerning the type of explanatory variables. As we mentioned in previous sub-section, classification trees can handle with missing values, except when a linear combination of variables is selected as a splitting rule. Besides the outliers, heteroscedasticity or collinearities which are some important topics when dealing with parametric models, are no problem for classification trees. Moreover, the transformation of explanatory variables to logarithms or squares or square roots has no particularly effect on the tree. The results are invariant with respect to monotone transformations of independent variables.

Another point that differs classification trees from other traditional statistical methods, is its ability to handle with the data set with high dimension. Classification trees algorithm selects automatically the most important variable(s) (i.e., classification trees does not require variables to be selected in advance) and then it puts these variables in the tree as a function of the split criteria, which are defined in accordance with the entropy function or Gini index. At the beginning even the data hold the important quantity of explanatory variables, classification trees can produce interesting and efficient results by using only a few important variables. Thus, from trees it is possible to detect the most influential variable(s) among all inputs on the estimation of the output. Feynman has mentioned that the bottom up approaches[166] like classification trees help to identify the contribution of each descriptive variable to the data set. For instance, the classical regression model requires the precise idea of the model to be constructed and to be estimated, the explanatory variables (predictors) and the interactions should be specified a priori in the model. This information is obtained in general from previous experiences (with similar data sets, literature on scientific theory). At this point, the advantage of classification trees emerges because it is possible that we don't have always the prior experience or scientific theory on concerned issue for statistical modelling. Or it is possible also that the previous results are not satisfactory and the analysis on the specific issue requires a new approach.

Statistical software

There are a number of statistical software packages to apply classification trees. The pioneering CART software is provided by the Salford Systems (Steinberg and Colla 1995[167]). As alternative, other statistical software such as SAS, SPSS, Statistica, S-Plus, and R offer specific packages (e.g., SAS Data mining, SPSS Classification trees, Rpart in R and S-Plus, the library

[166] Using the term 'bottom-up' is problematic because in the literature, some authors refer decision trees method as 'top-down', but some of them 'bottom-up'. The 'bottom-up' approach is employed to mention the contribution of each variable. But, when we read the tree from top to down for instance, we can say that decision trees method is a top-down approach. We suggest that depending on perspective, decision trees approach may be qualified as top-down, or bottom-up.

[167] Please refer to the link below for further information (accessed 1 July 2019):
http://docs.salford-systems.com/CART_Main.pdf

mda in R and S-Plus originally developed by MARS software). We will use the rpart (Recursive Partitioning and Regression Trees) package in R environment (see R Development Core Team)[168], which is developed by Therneau and Atkinson in 1997[169], updated in 2000, 2015, and in 2019[170]. The rpart is built upon the Breiman et al. (1984)'s CART (algorithm) methodology. Moreover, various sources yield useful and practical information on how to apply classification trees by using statistical software such as R and S-plus (see Kuhnert and Venables 2005, Maindonald and Braun 2010, Venables and Ripley 2002 [1994], Torgo 2016 [2010], Zhang and Singer 2010, James et al. 2013, Kabacoff 2015 [2011]). The detailed information on which options and commands are employed will be given in South Africa analysis (see section 5.1).

9.4 Case study of Turkey

Introduction

In case of Turkey analysis based on the SILC 2011 data, we will carry out the following analysis to measure multidimensional poverty, respectively: detecting multidimensional poverty patterns, dimensional poverty profiles, sub-group profiles, and some additional profiles such as elderly, unemployed. The list of variables is given in section 4.2. Depending on the purpose of the model, we will select indicators. The short names of variables (see the codes in parenthesis in first column of the table of selected indicators) will be displayed in the outputs including graphics. Before describing models and presenting results, we will yield a brief information on SILC and then the quality report of the variables.

SILC (Statistics on Income and Living Conditions-or known as EU-SILC, or the EU-SILC instrument)[171] is a comprehensive socio-economic survey notably for income, monetary poverty, multidimensional poverty, material deprivation, social exclusion, quality of life analysis, aims at gathering information at both household and individual level through household and personal data. This survey is designed by The European Union Statistics (Eurostat) under the EU-SILC project, is first launched in 2004 for the EU-15 (except Germany, the Netherlands, the United Kingdom) and Estonia, Norway and Iceland in collaboration with national statistical offices/institutes[172]. The EU-SILC is suitable for both cross-sectional and

[168] Please refer to the link below for further information (accessed 6 July 2019): http://www.r-project.org

[169] Please refer to the link below for further information (accessed 6 July 2019): https://www.mayo.edu/research/documents/biostat-61pdf/doc-10026699

[170] Please refer to the link below for further information (accessed 10 May 2020): https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf

[171] For further information, please refer to the link below (accessed 10 November 2019) : https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions

[172] For this research, the National Statistical Institute of Turkey (TUIK) has kindly provided me the Turkey-SILC 2011 data upon request from the author.

panel data analysis[173] and is convenient with the European Statistical System (ESS). In case of Turkey, the EU-SILC has been first launched in 2006[174].

In our analysis, we will employ several variables relating to geographic situation, individual characteristic, household composition, dwelling situation, material situation, educational attainment, environment, health, economic activity (labour), and financial situation. Labour, education, health information is collected for persons aged 15 and over whilst other variables are available at household level. The household level variables are the same for all members of the household. The Turkey-SILC 2011 questionnaires containing all questions at individual and household level asked to interviewees have provided us by the Turkey Statistical Institute; however, general information on questionnaires, potential derived variables based on these questionnaires, and classifications such as NACE, ISCO, NUTS is available at the website of the Eurostat[175].

Quality report of variables

As a first fundamental step, we created unique household (hid) and personal identifiers (id) in order to merge household and individual data together. More specifically, 3 data are used: personal register, personal, and household. In household data, there are 15,205 unique households; in personal register data there are 56,438 individuals including children; and in personal data based on the questions asked only for 15+ persons, there are 40,679 individuals. Through unique id number, personal register and personal data are merged together. After that, this merged personal data is merged with household data via unique household id number. This process generates at the end the final merged data including all household and individual information, consisting of 56,438 individuals and 15,205 households[176]. Based on the final merged data, all indicators (i.e., derived variables) defined in the methodology (section 4.2) are created and later derived data including both initial and derived variables is produced. Derived data and its subsets form the basis for our analysis. 28 variables are taken to construct the overall deprivation index, which means that all persons not having a full information of these variables were dropped from the derived data. The remaining missing values will be analyzed in last stage as a separate case. 28 variables, with their short names, are as follows: RO_D, HEATS, ENG_H, SAN1, SAN2, KIT, PWS, HWS, ROWW, INS, LIGHT, NOI, POL, CRI, TEL, MPHO, TV, COMPINT, WM, REF, DISHW, AIRC, CAR, EDU, MED, DEN, FS1.

Our analysis will cover persons aged 18 or over. In the data, there are 37,516 persons in this category, which constructs 66.5 per cent of the data. We remove in total 564 individuals aged 18 or over because they have missing values regarding at least one of the variables, HEATS,

[173] In the framework of the EU-SILC instrument, *cross-sectional data* is due to "a given time or a certain time period with variables on income, poverty, social exclusion and other living conditions", on the other hand *panel/longitudinal data* include "individual-level changes over time, observed periodically over a four-year period" (see https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions, accessed 10 November 2019). In our research, we use cross-sectional data.

[174]Please refer to the link below for further information (accessed 10 November 2019):
https://ec.europa.eu/eurostat/documents/203647/203704/EU-SILC+implementation+by+country.pdf

[175] https://ec.europa.eu/eurostat/web/income-and-living-conditions/methodology (accessed 10 November 2019). Please see also the link below for more information on classifications:
https://ec.europa.eu/eurostat/ramon/index.cfm?TargetUrl=DSP_PUB_WELC

[176] We note that among 15,205 households, there are 4 household composed of only people aged less than 18 yrs. old, and 63 households composed of a head whose age is less than 18.

EDU, ENG_H, MED, and DEN. We lose only 1.5 per cent of information, which can be considered quite a few amounts. Thus, the final data on which the analysis will be done includes 36,952 persons aged 18 or over across 14,848 households in the country. Additional to 28 variables mentioned in first paragraph, we created also several other variables for decomposition and supplementary analysis. There are in total 69 variables including PATH variable that corresponds to capabilities profile of each individual aged 18 or more in the data based on the main 28 variables that are used in building the overall deprivation index[177]. The variable PATH is a string variable reflecting the combination of 28 variables, as follows:

PATH=RO_D+HEATS+ENG_H+SAN1+SAN2+KIT+PWS+HWS+ROWW+INS+LIGHT+NOI+POL+CRI+TEL+MPHO+TV+COMPINT+WM+REF+DISHW+AIRC+CAR+EDU+MED+DEN+FS1

It is found 27,861 different capabilities profile (PATH) among 36, 952 individuals in the data. In other words, 75 % of the 18+ population in the derived data has her own singular capability profile/path. Such result is very complex to examine. So, we need to find out the most influential ones that will lead to carry out classification tree analysis in further sections.

We obtained the quality report of 69 variables in statistical software R via the package Hmisc[178]. We note also that descriptive statistics and analysis is based on the non-weighted data[179]. The short name of the variables corresponds to their codes mentioned in section 4.2.

Descriptive statistics outputs:

```
 TR2011derived[, 1:73]

 73  Variables      36952  Observations
---------------------------------------------------------------------------
hid
      n  missing distinct     Info     Mean      Gmd      .05      .10
  36952        0    14848        1     8157     5521    838.6   1645.0
    .25      .50      .75      .90      .95
 4044.5   8059.5  12285.0  14901.0  15765.0

lowest :     1     2     3     5     6, highest: 16560 16561 16562 16564 16565
---------------------------------------------------------------------------
id
```

[177] There are also household id (hid), personal id (id) numbers that are household, personal identifiers respectively; they are useful for checking whether there are duplicates in the data but also for merging household and personal data, and for identifying households in the data if it is necessary upon a specific condition. The variable PersonNR is the serial number for each individual in the data, which allows computing household size. The variable head is a binary variable taking the value 1 if the person is the head of the household, and it allows to construct household composition related variables such as educational attainment, age, gender of the head of the household. As these 4 variables are intermediate for checking the structure of the data, we don't provide summary statistics for them. However, they are explained in detail in the chapter of South Africa GHS 2011 analysis. Please, see South Africa GHS 2011 analysis for more information.

[178] See South Africa GHS 2011 analysis for more technical information on statistical software and packages.

[179] To obtain estimated numbers at national level typically for the national reports, the obtained figures can be multiplied by the appropriate weight provided in the data by the Turkey Institute of Statistics. But, in our analysis, we use the national sample, which is also an appropriate representative of the population in the country.

```
      n  missing distinct
  36952        0    36952

lowest : "    101" "    102" "    103" "    201" "    202"
highest: 1656501   1656502   1656503   87012     989602
---------------------------------------------------------------------------
PersonNR
      n  missing distinct     Info     Mean      Gmd      .05      .10
  36952        0    36952        1    27844    18707     2613     5528
    .25      .50      .75      .90      .95
  13805    27752    41837    50385    53150

lowest :     1     2     3     4     5, highest: 56434 56435 56436 56437 56438
---------------------------------------------------------------------------
head
      n  missing distinct     Info      Sum     Mean      Gmd
  36952        0        2     0.72    14768   0.3997   0.4799


---------------------------------------------------------------------------
PATH
      n  missing distinct
  36952        0    27861

lowest : 01111111111111111112111211111 01111111111111111112111212111 01111111111111111112111213111
01111111111111111114212230111 01111111111111112114112220222
highest: 14622111222221111111123112 14622111222222311111114113 14622111222223111111115113
14622111222222311511311O112 14622111222222311511311211 2
---------------------------------------------------------------------------
Age
      n  missing distinct     Info     Mean      Gmd      .05      .10
  36952        0       86        1    42.35    18.69       20       22
    .25      .50      .75      .90      .95
     29       40       54       66       74

lowest :  18  19  20  21  22, highest: 100 101 106 108 110
---------------------------------------------------------------------------
REG
      n  missing distinct
  36952        0       12

Value       TR1  TR2  TR3  TR4  TR5  TR6  TR7  TR8  TR9  TRA  TRB
Frequency  4261 2337 4802 2948 3286 3685 2359 2754 1526 2527 2984
Proportion 0.115 0.063 0.130 0.080 0.089 0.100 0.064 0.075 0.041 0.068 0.081

Value       TRC
Frequency  3483
Proportion 0.094
---------------------------------------------------------------------------
SET
      n  missing distinct
  36952        0        2

Value         1     2
Frequency  12997 23955
Proportion 0.352 0.648
---------------------------------------------------------------------------
GEN
      n  missing distinct
  36952        0        2
```

```
Value         1     2
Frequency  19215 17737
Proportion  0.52  0.48
---------------------------------------------------------------------------
AGE
      n  missing distinct
  36952        0        6

Value         1     2     3     4     5     6
Frequency   5556  8580  7563  6536  4375  4342
Proportion 0.150 0.232 0.205 0.177 0.118 0.118
---------------------------------------------------------------------------
HS
      n  missing distinct
  36952        0        5

Value         1     2     3     4     5
Frequency   1113  5989  7703  8550 13597
Proportion 0.030 0.162 0.208 0.231 0.368
---------------------------------------------------------------------------
hh_child
      n  missing distinct
  36952        0        6

Value         1     2     3     4     5     6
Frequency  14386  8396  7458  3308  1598  1806
Proportion 0.389 0.227 0.202 0.090 0.043 0.049
---------------------------------------------------------------------------
HT
      n  missing distinct
  36952        0       10

Value        10    11    12    13     5     6     7     8     9    99
Frequency   4273  4901  3696 11190  1113  3570  2216  5537   449     7
Proportion 0.116 0.133 0.100 0.303 0.030 0.097 0.060 0.150 0.012 0.000
---------------------------------------------------------------------------
GH
      n  missing distinct
  36952        0        2

Value         1     2
Frequency   4448 32504
Proportion  0.12  0.88
---------------------------------------------------------------------------
AH
      n  missing distinct
  36952        0        5

Value         1     2     3     4     5
Frequency     28   824  5493 24930  5677
Proportion 0.001 0.022 0.149 0.675 0.154
---------------------------------------------------------------------------
ELD
      n  missing distinct
  36952        0        2

Value         0     1
Frequency  32610  4342
Proportion 0.882 0.118
---------------------------------------------------------------------------
```

RO_D
n missing distinct
36952 0 2

Value 0 1
Frequency 26038 10914
Proportion 0.705 0.295

HEATS
n missing distinct
36952 0 4

Value 1 2 3 4
Frequency 25143 3848 6862 1099
Proportion 0.680 0.104 0.186 0.030

ENG_H
n missing distinct
36952 0 7

Value 1 2 3 4 5 6 7
Frequency 6736 17747 8444 123 20 1774 2108
Proportion 0.182 0.480 0.229 0.003 0.001 0.048 0.057

SAN1
n missing distinct
36952 0 3

Value 1 2 3
Frequency 33687 1767 1498
Proportion 0.912 0.048 0.041

SAN2
n missing distinct
36952 0 3

Value 1 2 3
Frequency 31145 1854 3953
Proportion 0.843 0.050 0.107

KIT
n missing distinct
36952 0 2

Value 1 2
Frequency 35462 1490
Proportion 0.96 0.04

PWS
n missing distinct
36952 0 2

Value 1 2
Frequency 35955 997
Proportion 0.973 0.027

HWS
n missing distinct
36952 0 2

```
Value         1     2
Frequency  28698  8254
Proportion 0.777 0.223
```

```
ROWW
       n  missing distinct
   36952        0        2

Value         1     2
Frequency  15493 21459
Proportion 0.419 0.581
```

```
INS
       n  missing distinct
   36952        0        2

Value         1     2
Frequency  15077 21875
Proportion 0.408 0.592
```

```
LIGHT
       n  missing distinct
   36952        0        2

Value         1     2
Frequency   9605 27347
Proportion  0.26  0.74
```

```
NOI
       n  missing distinct
   36952        0        2

Value         1     2
Frequency   5679 31273
Proportion 0.154 0.846
```

```
POL
       n  missing distinct
   36952        0        2

Value         1     2
Frequency   9579 27373
Proportion 0.259 0.741
```

```
CRI
       n  missing distinct
   36952        0        2

Value         1     2
Frequency   3517 33435
Proportion 0.095 0.905
```

```
TEL
       n  missing distinct
   36952        0        3

Value         1     2     3
Frequency  20213  7092  9647
Proportion 0.547 0.192 0.261
```

MPHO
n missing distinct
36952 0 3

Value 1 2 3
Frequency 35147 750 1055
Proportion 0.951 0.020 0.029

TV
n missing distinct
36952 0 3

Value 1 2 3
Frequency 36362 357 233
Proportion 0.984 0.010 0.006

COMPINT
n missing distinct
36952 0 5

Value 1 2 3 4 5
Frequency 11812 1569 2014 12222 9335
Proportion 0.320 0.042 0.055 0.331 0.253

WM
n missing distinct
36952 0 3

Value 1 2 3
Frequency 34492 2145 315
Proportion 0.933 0.058 0.009

REF
n missing distinct
36952 0 3

Value 1 2 3
Frequency 36240 665 47
Proportion 0.981 0.018 0.001

DISHW
n missing distinct
36952 0 3

Value 1 2 3
Frequency 15875 15805 5272
Proportion 0.430 0.428 0.143

AIRC
n missing distinct
36952 0 3

Value 1 2 3
Frequency 5165 13219 18568
Proportion 0.140 0.358 0.502

CAR
n missing distinct
36952 0 3

```
Value         1     2     3
Frequency  12109 20609  4234
Proportion 0.328 0.558 0.115
---------------------------------------------------------------------------
EDU
       n  missing distinct
   36952        0        7

Value         0     1     2     3     4     5     6
Frequency  5213  3151 13836  4593  3674  2744  3741
Proportion 0.141 0.085 0.374 0.124 0.099 0.074 0.101
---------------------------------------------------------------------------
MED
       n  missing distinct
   36952        0        8

Value         1     2     3     4     5     6     7     8
Frequency  30225  4394   188   757   367   221   726    74
Proportion 0.818 0.119 0.005 0.020 0.010 0.006 0.020 0.002
---------------------------------------------------------------------------
DEN
       n  missing distinct
   36952        0        8

Value         1     2     3     4     5     6     7     8
Frequency  32456  3124   106   456   157   401   220    32
Proportion 0.878 0.085 0.003 0.012 0.004 0.011 0.006 0.001
---------------------------------------------------------------------------
FS1
       n  missing distinct
   36952        0        6

Value         1     2     3     4     5     6
Frequency  7332 12997  9532  4387  2491   213
Proportion 0.198 0.352 0.258 0.119 0.067 0.006
---------------------------------------------------------------------------
DWE
       n  missing distinct
   36952        0        5

Value         1     2     3     4    99
Frequency  17905  1488  9377  8162    20
Proportion 0.485 0.040 0.254 0.221 0.001
---------------------------------------------------------------------------
HEALTH
       n  missing distinct
   36952        0        5

Value         1     2     3     4     5
Frequency  3993 19147  8156  4818   838
Proportion 0.108 0.518 0.221 0.130 0.023
---------------------------------------------------------------------------
CIL
       n  missing distinct
   36952        0        2

Value         1     2
Frequency  12597 24355
Proportion 0.341 0.659
---------------------------------------------------------------------------
```

FS2
```
      n  missing distinct
  36952        0        4

Value          1     2     3     4
Frequency    518  2808  5976 27650
Proportion 0.014 0.076 0.162 0.748
```

FS3
```
      n  missing distinct
  36952        0        4

Value          1     2     3     4
Frequency   2281 13240 21201   230
Proportion 0.062 0.358 0.574 0.006
```

FS4
```
      n  missing distinct
  36952        0        4

Value          1     2     3     4
Frequency   1668  9474 16266  9544
Proportion 0.045 0.256 0.440 0.258
```

FS5
```
      n  missing distinct
  36952        0        2

Value          1     2
Frequency  12272 24680
Proportion 0.332 0.668
```

FS6
```
      n  missing distinct
  36952        0        2

Value          1     2
Frequency  24480 12472
Proportion 0.662 0.338
```

FS7
```
      n  missing distinct
  36952        0        2

Value          1     2
Frequency  15150 21802
Proportion  0.41  0.59
```

FS8
```
      n  missing distinct
  36952        0        2

Value          1     2
Frequency  24070 12882
Proportion 0.651 0.349
```

FS9
```
      n  missing distinct
  36952        0        2
```

```
Value          1     2
Frequency   7564 29388
Proportion 0.205 0.795
---------------------------------------------------------------------------
FS10
      n  missing distinct
  36952        0        2

Value          1     2
Frequency   5222 31730
Proportion 0.141 0.859
---------------------------------------------------------------------------
FS11
      n  missing distinct
  36952        0        3

Value          1     2     3
Frequency   8468 22275  6209
Proportion 0.229 0.603 0.168
---------------------------------------------------------------------------
FS12
      n  missing distinct
  36952        0        4

Value          1     2     3     4
Frequency   9525 11207  1548 14672
Proportion 0.258 0.303 0.042 0.397
---------------------------------------------------------------------------
ES1
      n  missing distinct
  36952        0        6

Value          1     2     3     4     5     6
Frequency  17753  1405  1405    32    30 16327
Proportion 0.480 0.038 0.038 0.001 0.001 0.442
---------------------------------------------------------------------------
ES2
      n  missing distinct
  36952        0       10

Value          1    10     2     3     4     5     6     7     8     9
Frequency  10517   381   597  5252  1354  1454  1095  3211  2548 10543
Proportion 0.285 0.010 0.016 0.142 0.037 0.039 0.030 0.087 0.069 0.285
---------------------------------------------------------------------------
ES3
      n  missing distinct
  36952        0        7

Value          1     2     3     4     5     6     7
Frequency   8885  1564   708  3708  2829 19199    59
Proportion 0.240 0.042 0.019 0.100 0.077 0.520 0.002
---------------------------------------------------------------------------
OG
      n  missing distinct
  36952        0       11

Value          1    10    11     2     3     4     5     6     7     8     9
Frequency   1372 19199    59  1379  1003   950  2072  4491  2264  1637  2526
Proportion 0.037 0.520 0.002 0.037 0.027 0.026 0.056 0.122 0.061 0.044 0.068
---------------------------------------------------------------------------
```

IND
```
      n  missing distinct
  36952        0       18

Value         1    10    11    12    13    14    17    18    19     2    20
Frequency  5125   185   105   258   584  2452   100   532 19199   135    59
Proportion 0.139 0.005 0.003 0.007 0.016 0.066 0.003 0.014 0.520 0.004 0.002

Value         3     4     5     6     7     8     9
Frequency  2698   143  1297  2327   811   823   119
Proportion 0.073 0.004 0.035 0.063 0.022 0.022 0.003
```

WH
```
      n  missing distinct
  36952        0        5

Value         1     2     3     4     5
Frequency  1499  4778 11417 19199    59
Proportion 0.041 0.129 0.309 0.520 0.002
```

SC
```
      n  missing distinct
  36952        0        4

Value         1     2     3     4
Frequency  9515  8179 19199    59
Proportion 0.257 0.221 0.520 0.002
```

CONTRACT
```
      n  missing distinct
  29707     7245        6

Value         1     2     3     4     5     6
Frequency  8425   981  1032    11 19199    59
Proportion 0.284 0.033 0.035 0.000 0.646 0.002
```

EDUH
```
      n  missing distinct
  36952        0        8

Value         0     1     2     3     4     5     6    99
Frequency  3415  2950 17285  3805  2759  2796  3931    11
Proportion 0.092 0.080 0.468 0.103 0.075 0.076 0.106 0.000
```

ES1H
```
      n  missing distinct
  36952        0        7

Value         1     2     3     4     5     6    99
Frequency 24045  1267   897    24    20 10688    11
Proportion 0.651 0.034 0.024 0.001 0.001 0.289 0.000
```

ES2H
```
      n  missing distinct
  36952        0       11

Value         1    10     2     3     4     5     6     7     8     9    99
Frequency 14503   184   480  8155   873  1322   173  6609  2588  2054    11
Proportion 0.392 0.005 0.013 0.221 0.024 0.036 0.005 0.179 0.070 0.056 0.000
```

```
ES3H
      n  missing distinct
  36952        0        7

Value        1     2     3     4     5     6     7
Frequency  11907  2176  1462  8159   291 12907    50
Proportion 0.322 0.059 0.040 0.221 0.008 0.349 0.001
-------------------------------------------------------------------------------
OGH
      n  missing distinct
  36952        0       11

Value        1    10    11     2     3     4     5     6     7     8     9
Frequency   2719 12907    50  1584  1243   993  2472  5613  3436  2823  3112
Proportion 0.074 0.349 0.001 0.043 0.034 0.027 0.067 0.152 0.093 0.076 0.084
-------------------------------------------------------------------------------
WHH
      n  missing distinct
  36952        0        5

Value        1     2     3     4     5
Frequency   1032  6246 16717 12907    50
Proportion 0.028 0.169 0.452 0.349 0.001
-------------------------------------------------------------------------------
SCH
      n  missing distinct
  36952        0        4

Value        1     2     3     4
Frequency  13890 10105 12907    50
Proportion 0.376 0.273 0.349 0.001
-------------------------------------------------------------------------------
HEALTHH
      n  missing distinct
  36952        0        6

Value        1     2     3     4     5    99
Frequency   2989 18132  9812  5214   794    11
Proportion 0.081 0.491 0.266 0.141 0.021 0.000
```

We note that in the following analysis concerning classification trees by applying the package rpart in R, we will give only the essential statistical outputs.

First step (step1 analysis): detecting multidimensional poverty patterns

In this investigation, the decision trees-based capabilities analysis will be based on the overall deprivation index we will built. First of all, we define a deprivation line for each of the 28 variables. Deprivation line is the criteria of distinguishing who is deprived or not regarding the variable. As we have already mentioned in GHS 2011 case, the utility of each variable representing a different functioning is determined according to universal human rights chart. According to the deprivation threshold, if the person experiences deprivation regarding the variable the he or she takes the score 1, and 0 otherwise. Taking in consideration all variables simultaneously, we count the number of deprivations the person suffers to obtain her/his overall individual deprivation score. Then, as a function of the distribution of overall deprivation index scores in the data, we create several deprivation groups. At last stage, we will proceed with classification tress method to discover the most influential capabilities patterns emerging in the data distinguishing deprivation groups among them. The table below gives information on

deprivation line for each variable used in the index. The table contains 4 columns: first column gives the code name of the variable given that each variable represents a functioning, second column mentions deprivation line regarding the variable, third column indicates the utility of the functioning for being capable of living a decent life according to universal human rights chart, and last column contains remarks about the variable.

Table 33. Deprivation lines and composition of variables for model 1: overall deprivation index and multidimensional poverty patterns, data: Turkey SILC 2011

Variable (each variable represents a functioning)	**Deprivation line[180]**	**The utility of functionings for having the capability of living a decent life according to Universal Human Rights Chart[181]**	**Remark**
RO_D_D	Deprived (RO_D=1) if the number of rooms available per person in the dwelling is inferior than 1	Article 25: functioning of living with an adequate housing.	Excluding kitchen, bathroom, garden, and entranceway, in normal conditions every single person should have at least 1 room for her/his privacy.
HEATS_D	Deprived (HEATS_D=1) if HEATS is 1	Article 25: functioning of living with an adequate housing.	See note below[182]

[180] For the codes and categories, see the table provided in section 4.2.

[181] Please refer to the link below for further information (accessed 12 February 2020) :
https://www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/eng.pdf

[182] We consider stove as a not appropriate heating tool compared to radiator, air conditioner, and flat/combi boiler/radiator in the sense that common or central heating via radiator/combi radiator is the best solution spreading heat waves throughout several radiators implemented in the dwelling, a combi boiler as a radiator depends on the choice of the person if he or she wills to open or not radiator (but still has the same impact as a central heating), and air conditioner seems to be the best tool for protecting environment as it is based on electricity. Stove is in general used in Turkey in rural areas, or poor households. A stove can be placed in a room and it does not provide heating in all rooms of the dwelling; so that, heated air is felt only in nearabout of where stove is placed. Thus, a stove cannot be considered as an optimal way of protecting from cold especially during harsh winter season in some parts of the country (e.g., north-east, east). Coal, gas, natural gas, or electricity can be used for running a stove. Electric stove is a good solution for environment however it can be very costly in the context of the country given that poor households during severe winter (which the majority of country suffers during the autumn/winter season) are more likely to not afford high electricity bills due to financial reasons and an electric stove will require certainly a high amount of bill in areas where winter is strong and the poorest parts of the country experience it. Electric stove can be probably rather preferred in Mediterranean climate areas such as south, and south west of the country.

ENG_H_D	Deprived (ENG_H_D=1) if ENG_H is 1, 2, 4, 5, 7	Article 25: functioning of living with an adequate housing.	We consider wood, coal, fuel-oil, lpg, and cowpat as not appropriate tools as energy source for heating in the dwelling, compared to natural gas and electricity. In the context of the country, poor households are more likely to use wood, cowpat, coal especially in rural areas, and fuel-oil and lpg, and coal in urban areas.
SAN1_D	Deprived (SAN1_D=1) if there is no bathroom or shower in the dwelling, or if bathroom or shower is shared with other dwellings	Article 25: functioning of living with an adequate housing.	
SAN2_D	Deprived (SAN2_D=1) if there is no toilet in the dwelling, or if toilet is shared with other dwellings	Article 25: functioning of living with an adequate housing.	
KIT_D	Deprived (KIT_D=1) if there is no kitchen in the dwelling	Article 25: functioning of living with an adequate housing.	
PWS_D	Deprived (PWS_D=1) if there is no piped water system in the dwelling	Article 25: functioning of living with an adequate housing.	
HWS_D	Deprived (HWS_D=1) if there is no hot water system in the dwelling	Article 25: functioning of living with an adequate housing.	
ROWW_D	Deprived (ROWW_D=1) if there are any problems regarding roof, wall, and window in the dwelling	Article 25: functioning of living with an adequate housing.	
INS_D	Deprived (INS_D=1) if there is a heating problem due to insulation in the dwelling	Article 25: functioning of living with an adequate housing.	
LIGHT_D	Deprived (LIGHT_D=1) if there is not enough light in the dwelling	Article 25: functioning of	

		living with an adequate housing.	
NOI_D	Deprived (NOI_D=1) if there is noise pollution in the dwelling	Article 25: functioning of living with an adequate environment.	
POL_D	Deprived (POL_D=1) if there is air pollution, environmental pollution in the neighborhood/community	Article 25: functioning of living with an adequate environment.	
CRI_D	Deprived (CRI_D=1) if there is an intensive occurrence of crime, violence in the neighborhood/community	Article 3: functioning of having the right to the security.	
TEL_D	Deprived (TEL_D=1) if TEL is 2, 3 (i.e., there is no landline/telephone in the dwelling)	Article 25: functioning of living with an adequate housing in terms of material and social service regarding availability of landline network in the neighborhood where the person lives. It is not necessary that telephone, internet networks are provided by the State, but at least State should make sure that every single citizen has	Due to an increasing number (tendency) of usage of mobile phone, the use of a landline phone at home may decrease or people may find unnecessary the use of a landline phone at home. However, we keep this variable as a one of the (main) tools of communication because still some parts of the country such as rural areas may lack access to mobile phone service, and we don't have information whether household choose mobile phone rather than landline phone and if yes, why[183].

[183] We also note that for material deprivation (tel, mobile phone, tv, and so on), there are 3 answer types determined in the questionnaire: yes, household have it; no, household does not have it due to financial reason/financial shortage; no, household does not have it due to other reason. Other reason option covers many possible reasons such as personal willing to not have it, or a family member does not allow the person to have it, or household head decides for all members of the household, or there is no accessible point/service to buy it, and so on. Because of that, we consider the last 2 answer options as an indicator of deprivation regarding the item. If we had full information whether the person has in fact full liberty to have it if he or she wants but simply he or she does not prefer to have the item; then, we could not consider such possibility as an indicator of deprivation.

		access to such services.	
MPHO_D	Deprived (MPHO_D=1) if MPHO is 2, 3 (i.e., there is no mobile phone in the dwelling)	The same argument mentioned for TEL_D.	
TV_D	Deprived (TV_D=1) if TV is 2,3 (i.e., there is no color TV in the dwelling)	Article 25: functioning of living with an adequate housing in terms of material.	
COMPINT_D	Deprived (COMPINT_D=1) if COMPINT is 4, 5 (i.e., there is no computer in the dwelling)	Article 25: functioning of living with an adequate housing in terms of material.	This variable is composed by two variables: computer and internet. In the questionnaire, it is asked whether the household having a computer have also an internet connection. The internet question was not asked to people having not a computer in their household. According to that, the variable COMPINT is created and it is composed of 5 modalities: 1=computer and internet, 2=computer but no internet for financial reason, 3=computer but no internet for other reason, 4=no computer for financial reason, 5=no computer for other reason
WM_D	Deprived (WM_D=1) if WM is 2, 3 (i.e., there is no washing machine in the dwelling)	Article 25: functioning of living with an adequate housing in terms of material.	
REF_D	Deprived (REF_D=1) if REF is 2, 3 (i.e., there is no refrigerator in the dwelling)	Article 25: functioning of living with an adequate housing	

		in terms of material.	
DISHW_D	Deprived (DISHW_D=1) if DISHW is 2, 3 (i.e., there is no dishwasher in the dwelling)	Article 25: functioning of living with an adequate housing in terms of material.	
AIR_D	Deprived (AIR_D=1) if AIR is 2, 3 (i.e., there is no air conditioner in the dwelling)	Article 25: functioning of living with an adequate housing in terms of material.	
CAR_D	Deprived (CAR_D=1) if CAR is 2, 3 (i.e., there is no car in the dwelling)	Article 25: functioning of living with an adequate housing in terms of material.	
EDU_D	Deprived (EDU_D=1) if EDU is 0, 1, 2, 3 (less than high school level diploma)	Article 26: functioning of having an educational attainment[184].	
MED_D	Deprived (MED_D=1) if MED is 2, 3, 4, 5, 6, 7, 8	Article 25: functioning of the right to medical care for a healthy life.	
DEN_D	Deprived (DEN_D=1) if DEN is 2, 3, 4, 5, 6, 7, 8	Article 25: functioning of the right to medical care for a healthy life.	
FS1_D	Deprived (FS1_D=1) if FS1 is 1, 2, 3, 4	Financial aspect is not mentioned in the chart. In that vein, monetary aspect becomes an	

[184] Article 26 mentions that "Everyone has the right to education. Education shall be free, at least in the elementary and fundamental stages. Elementary education shall be compulsory. Technical and professional education shall be made generally available and higher education shall be equally accessible to all on the basis of merit". However, we argue that at least a general high school diploma or its equivalent in terms of vocational technical high school is a fundamental necessary for people to obtain opportunities for higher professional education and specific job skills.

		indirect tool for detecting capabilities. So that, we consider this indicator as a complementary tool reflecting the financial state of the household. Having an adequate financial resource will indirectly help the household to achieve some functionings and may bring some advantages for achieving some functionings.	

Regarding the overall deprivation index 'DEP' we built based on 28 variables above, 0.08 per cent of the target population (18+) has no deprivation whilst 0.01 percent of them has deprivation in 23 of 28 items. The most frequent groups are those having deprivation in 8 or 9 items, which represent 9.65, and 10 per cent of the population, respectively. We drop 28 persons having 0 deprivation (the most prosperous group) in the data. After that, there are in total 36,924 persons in the data. The summary statistics of DEP is as follows:

```
                    Total Deprivation
-------------------------------------------------------------
      Percentiles      Smallest
 1%             2              1
 5%             3              1
10%             4              1       Obs                36924
25%             6              1       Sum of Wgt.        36924

50%             9                      Mean           9.083198
                        Largest        Std. Dev.      3.764283
75%            12             23
90%            14             23       Variance       14.16982
95%            15             23       Skewness       .2588782
99%            18             23       Kurtosis       2.681643
```

As we did in South Africa GHS 2011 analysis, here also we apply the same methodology and we built a deprivation groups variable namely DEP_GRS regrouping the population by 20th, 40th, 60th, 80th, and 100th percentiles. 1st group (DEP_GRS1) includes people having deprivation in 1 to 5 items; 2nd group (DEP_GRS2) includes people having deprivation in 6 to 7 items: 3rd group (DEP_GRS3) includes people having deprivation in 8 to 9 items; 4th group (DEP_GRS4) includes people having deprivation in 10 to 11 items; and 5th group (DEP_GRS5) includes people having deprivation in 12 to 23 items. When we take into account simultaneously 28 deprivation variables (e.g., HEATS_D, SAN2_D-see the table of deprivation lines above) to find out the capabilities profiles of 36,924 individuals in the data (the variable named PATH28),

the results show that there are 12,474 different profiles which is relatively high amount. For instance, with the most frequent 5 profiles (with at least 100 observations each of them), we can only capture 2.7 per cent of the population. But also, in terms of profiles with at least 100 observations, only 15 of 12,474 profiles surface yielding information for 6.8 per cent of the population as follows:

PATH28	Freq.	Percent	%<=
011000000000001001001111001	261	0.71	0.71
011000000000000001001111001	208	0.56	1.27
000000000000000000000100001	188	0.51	1.78
011000010000001001001111001	187	0.51	2.29
011000001100001001001111001	180	0.49	2.77
000000000000000000000111001	166	0.45	3.22
100000000000000000000100001	151	0.41	3.63
011000011100001001001111001	134	0.36	3.99
000000000000000000000101001	133	0.36	4.35
111000000000000001001111001	125	0.34	4.69
011000000000000000001111001	122	0.33	5.02
011000000000001001001101001	113	0.31	5.33
000000000000000000000110001	112	0.30	5.63
011000000000000000000111001	110	0.30	5.93
011000001110001001001111001	109	0.30	6.23
011000000000000000000101001	103	0.28	6.51
111000000000001001001111001	102	0.28	6.78

The table below shows the number of persons in the group, unique capabilities profiles based on the 28 variables and the ratio of unique profiles by the number of persons in the group.

Table 34. Ratio of unique profiles to the number of persons in the deprivation group (based on the variable DEP_GRS and 28 variables), data: Turkey SILC 2011

Group	Unique capabilities profiles	Number of persons in the group	Ratio (Unique profiles/Number of persons in the group)
DEP_GRS1 (1-5 deprivations)	1,405	7,004	0.200
DEP_GRS2 (6-7 deprivations)	2,212	6,198	0.356
DEP_GRS3 (8-9 deprivations)	2,595	7,263	0.357
DEP_GRS4 (10-11 deprivations)	2,416	6,736	0.358
DEP_GRS5 (12-23 deprivations)	3,846	9,723	0.395
Total (whole sample)	**12,474**	**36,924**	**0.337**

As we can see from the data there is clearly a large variety of living conditions people face in their life. Within first group, the most frequent 15 profiles represent 20.7 per cent of the group; within second group, the most frequent 5 profiles capture 7.4 per cent of the group; within third group, the most frequent 10 profiles provide information for 14.2 per cent of the group; within fourth group, the most frequent 8 profiles we obtain information for 11.5 per cent of the group; and within fifth group, the most frequent 16 profiles lead to capture 8.7 per cent of the group. The analysis of capabilities, therefore, becomes quite complicated due to this plurality. In next paragraphs, we will show how to deal with such complexity by employing classification trees method. The goal is to find out the most informative capabilities patterns distinguishing the 5 deprivation groups. More specifically, we will do in total 6 classification tree analysis predicting how the deprivation groups distinguish among them by 28 deprivation variables of actual capabilities: {DEP_GRS1 vs. DEP_GRS3, DEP_GRS1 vs. DEP_GRS4, DEP_GRS1 vs. DEP_GRS5, DEP_GRS2 vs. DEP_GRS4, DEP_GRS2 vs. DEP_GRS5, DEP_GRS3 vs. DEP_GRS5}. As we have experienced in South Africa GHS 2011 analysis, the analysis between closely neighbor groups does not provide good prediction; so that, we focus only on 6 cases rather than 10.

Table 35. Per centage of groups in the sample of classification target groups, data: Turkey SILC 2011

Target group in classification	Per centage of groups in the sample
DEP_GRS1 vs. DEP_GRS3	{49, 51}
DEP_GRS1 vs. DEP_GRS4	{51, 49}
DEP_GRS1 vs. DEP_GRS5	{42, 58}
DEP_GRS2 vs. DEP_GRS4	{48, 52}
DEP_GRS2 vs. DEP_GRS5	{39, 61}
DEP_GRS3 vs. DEP_GRS5	{43, 57}

As the table above shows we have a relatively balanced data but still we don't have a 100 per cent equal distribution between two groups (e.g., DEP_GRS1: 50 per cent vs. DEP_GRS2: 50 per cent).

CAP112: GR1 vs. GR3 analysis

In this analysis, we focus on DEP_GRS1 vs. DEP_GRS3. We use the data containing only people from the deprivation group 1 and the deprivation group 3. The outcome/target response variable is dichotomous, namely CAP112, taking the value 1 for the group 1 and the value 3 for the group 3 (so that people identified in the group 1 take the value 1 and people identified in the group 3 take the value 3 according to CAP112).

Table 36. Training vs. test samples in CAP112 analysis, data: Turkey SILC 2011

	GR1	GR3
Training sample	4,669	4,842
Test sample	2,335	2,421

As the table above shows, there are 9,511 individuals in the training data (corresponding to 67 per cent of the initial data) and 4,756 individuals in the test sample. Training and test data are obtained by randomly partitioning the initial data/total sample into two subsets by 70 per cent vs. 30 per cent respectively. The data is very balanced with 49 per cent for GR1 and 51 per cent for GR3 in both training and test samples (in the initial data, the proportion is also 49 vs. 51 per cent). Concerning the rpart application we have set cp at 0.025 and we used gini index as split criterion, and we kept other parameters at their default values. Classification rules are discovered on the training data and they will be tested on the test data for evaluating predictive accuracy. The classification tree is called CAP112, and the model is named dtreeCAP112. The results in the box below show that the initial saturated tree given the parameters we used contains 28 splits (i.e., 29 terminal nodes) with a cross-validation error rate around 0.15. This tree is quite big and therefore difficult for interpretation, so we will prune the tree to around 10-11 splits to obtain an optimal interpretation and the pruning process yields also a relatively low level of cross-validation error rate (between 0.22 and 0.21).

```
> printcp(dtreeCAP112)

Classification tree:
rpart(formula = CAP112 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
 [1] AIRC_D    CAR_D     COMPINT_D DISHW_D   ENG_H_D   FS1_D     HEATS_D
 [8] INS_D     LIGHT_D   MED_D     NOI_D     POL_D     ROWW_D    SAN1_D

Root node error: 4669/9511 = 0.49091

n= 9511

          CP nsplit rel error  xerror      xstd
1  0.5939173      0   1.00000 1.00000 0.0104421
2  0.0526879      1   0.40608 0.40608 0.0083448
3  0.0222032      2   0.35339 0.35339 0.0079094
4  0.0162776      5   0.28679 0.29043 0.0073031
5  0.0130649      7   0.25423 0.25787 0.0069454
6  0.0121011      8   0.24117 0.24780 0.0068277
7  0.0104948     10   0.21696 0.22296 0.0065212
8  0.0096380     11   0.20647 0.21589 0.0064295
9  0.0068537     12   0.19683 0.20326 0.0062601
10 0.0059970     13   0.18998 0.20047 0.0062218
11 0.0044978     15   0.17798 0.18398 0.0059871
12 0.0040694     16   0.17348 0.17734 0.0058886
13 0.0038552     17   0.16942 0.17648 0.0058757
14 0.0036410     19   0.16170 0.17263 0.0058172
15 0.0033198     20   0.15806 0.17091 0.0057909
16 0.0029985     22   0.15142 0.16556 0.0057077
17 0.0027843     23   0.14843 0.16449 0.0056908
18 0.0025701     24   0.14564 0.15549 0.0055463
19 0.0025000     28   0.13536 0.15271 0.0055005
```

The CAP112 pruned tree, is composed of 11 terminal nodes providing 4 patterns of GR1, and 7 patterns of GR3. As we can see from the pruned tree below, the discovered patterns from the left to the right side of the tree are as follows: (1) no deprivation regarding heating system of the dwelling (HEATS), dishwasher (DISHW), and insulation state of the dwelling (INS); (2) no deprivation regarding HEATS, DISHW, and car (CAR), but deprivation regarding INS; (3) no deprivation regarding HEATS, and DISHW, but deprivation regarding CAR, and INS; (4) no deprivation regarding HEATS, computer with internet (COMPINT), and pollution (POL), but deprivation regarding DISHW; (5) no deprivation regarding HEATS and COMPINT, but deprivation regarding DISHW, and POL; (6) no deprivation regarding HEATS, but deprivation regarding COMPINT, and DISHW; (7) no deprivation regarding DISHW, COMPINT, ROWW, CAR, but deprivation regarding HEATS; (8) no deprivation regarding DISHW, COMPINT, ROWW, but deprivation regarding CAR, and HEATS; (9) no deprivation regarding DISHW, and COMPINT, but deprivation regarding HEATS, and roof wall, and window conditions (ROWW); (10) no deprivation regarding DISHW, but deprivation regarding HEATS, and COMPINT; (11) deprivation regarding HEATS, and DISHW.

The patterns of GR1 are 1st, 2nd, 4th, and 7th with 71.4, 5.8, 4.6, and 9.55 per centage of coverage respectively (see the table 38 below). For instance, with the 1st pattern, 71.4 per cent of people from GR1 are detected. On the other hand, the patterns of GR3 are 3rd, 5th, 6th, 8th, 9th, 10th, and 11th with 4.8, 2.4, 5.8, 4.13, 5.9, 15, and 49 per centage of coverage respectively (see the table 39 below). For instance, with the 3rd pattern, 4.8 per cent of people from GR3 are captured.

Table 37. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP112 pruned tree, data: Turkey SILC 2011

Pattern no	Coverage of pattern (in %) among the total population of GR1	Cumulative coverage (in %)
1	71.4	71.4
2	5.8	77.2
4	4.6	81.8
7	9.55	91.35~**91**

Table 38. Coverage of GR3's patterns among the total population of GR3 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP112 pruned tree, data: Turkey SILC 2011

Pattern no	Coverage of pattern (in %) among the total population of GR3	Cumulative coverage (in %)
3	4.8	4.8
5	2.4	7.2
6	5.8	13
8	4.13	17.13
9	5.9	23.03

10	15	38.03
11	49	87.03~**87**

In the training data, by using the CAP112 pruned tree classification patterns, we lose 9 per cent of information for GR1 and 13 per cent of information for GR3.

We can also say that through the 4 patterns of GR1, deprivation regarding HEATS, INS, and DISHW concern 9.55, 5.8, and 4.6 per cent of the entire population of GR1 (among 4,669 persons) respectively. On the other hand, concerning GR3, the 7 patterns show that the variables HEATS, DISHW, COMPINT, CAR, ROWW, INS, and POL affect 74, 57, 20.8, 8.9, 5.9, 4.8, and 2.4 per cent of the entire population of GR3 (among 4,842 persons). The 3 deprivation variables, INS, DISHW, and HEATS are the common in terms of deficits for both GR1 and GR3; however, the magnitude of DISHW and HEATS is much higher for GR3 compared to GR1.

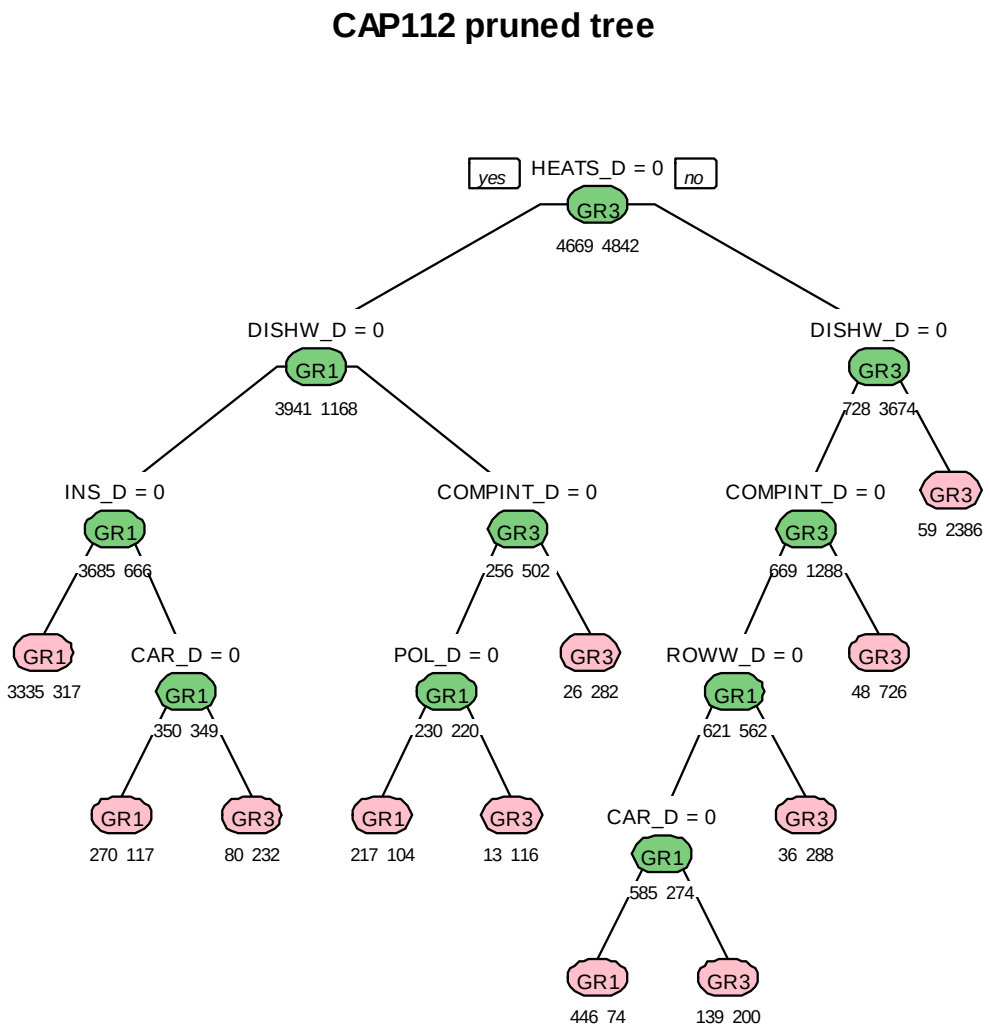


Figure 19. CAP112 pruned tree, data: Turkey SILC 2011

Regarding the confusion matrix statistics mentioned in the box below, we can say that the overall accuracy rate is 88 per cent. With the classification rules found on the training data, 10 per cent of GR1 and 13 per cent of GR3 could not be captured on the test data.

```
Confusion Matrix and Statistics

          Reference
Prediction  GR1  GR3
       GR1 2093  315
       GR3  242 2106

               Accuracy : 0.8829
                 95% CI : (0.8734, 0.8919)
    No Information Rate : 0.509
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7658

 Mcnemar's Test P-Value : 0.002283

            Sensitivity : 0.8964
            Specificity : 0.8699
         Pos Pred Value : 0.8692
         Neg Pred Value : 0.8969
             Prevalence : 0.4910
         Detection Rate : 0.4401
   Detection Prevalence : 0.5063
      Balanced Accuracy : 0.8831

       'Positive' Class : GR1
```

CAP113: GR1 vs. GR4 analysis

In this analysis, we focus on DEP_GRS1 vs. DEP_GRS4. We use the data containing only people from the deprivation group 1 and the deprivation group 4. The outcome/target response variable is dichotomous, namely CAP113, taking the value 1 for the group 1 and the value 4 for the group 4 (so that people identified in the group 1 take the value 1 and people identified in the group 4 take the value 4 according to CAP113).

Table 39. Training vs. test samples in CAP113 analysis, data: Turkey SILC 2011

	GR1	GR4
Training sample	4,669	4,490
Test sample	2,335	2,246

As the table above shows, there are 9,159 individuals in the training data (corresponding to 67 per cent of the initial data) and 4,581 individuals in the test sample. Training and test data are obtained by randomly partitioning the initial data/total sample into two subsets by 70 per cent vs. 30 per cent respectively. The data is very balanced with 51 per cent for GR1 and 49 per cent for GR4 in both training and test samples (in the initial data, the proportion is also 51 vs. 49 per cent).

In first attempt of fitting the model based on the cp equal to 0.025, we obtained a very good tree with 14 splits generating almost zero cross-validated error rate. As we can see from the

information provided in the box below, for a tree with 9 splits, the cross-validated error rate is still almost zero. For that reason, we proceeded with pruning phase to obtain a tree with 10 terminal nodes.

```
printcp(dtreeCAP113)

Classification tree:
rpart(formula = CAP113 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
 [1] AIRC_D    CAR_D     COMPINT_D DEN_D     DISHW_D   HEATS_D   INS_D
 [8] MED_D     POL_D     ROWW_D

Root node error: 4490/9159 = 0.49023

n= 9159

         CP nsplit rel error   xerror      xstd
1 0.7358575      0  1.000000 1.000000 0.0106553
2 0.0413140      1  0.264143 0.264143 0.0071562
3 0.0342984      3  0.181514 0.181960 0.0060754
4 0.0188196      4  0.147216 0.148775 0.0055424
5 0.0097996      6  0.109577 0.111136 0.0048377
6 0.0045657      7  0.099777 0.103563 0.0046791
7 0.0033408      9  0.090646 0.098664 0.0045729
8 0.0025612     12  0.080624 0.089978 0.0043767
9 0.0025000     14  0.075501 0.089978 0.0043767
```

The CAP113 pruned tree is composed of 10 terminal nodes providing 4 patterns of GR1, and 6 patterns of GR4. As we can see from the pruned tree below, the discovered patterns from the left to the right side of the tree are as follows: (1) no deprivation regarding heating system of the dwelling (HEATS), dishwasher (DISHW), and access to dentist for medical need (DEN); (2) no deprivation regarding HEATS, DISHW, and pollution (POL), but deprivation regarding DEN; (3) no deprivation regarding HEATS, and DISHW, but deprivation regarding DEN, and POL; (4) no deprivation regarding HEATS, computer with internet (COMPINT), and roof, wall, and window conditions (ROWW), but deprivation regarding DISHW; (5) no deprivation regarding HEATS, and COMPINT, but deprivation regarding DISHW, and ROWW; (6) no deprivation regarding HEATS, but deprivation regarding DISHW, and COMPINT; (7) no deprivation regarding DISHW, COMPINT, and insulation state of the dwelling (INS), but deprivation regarding HEATS; (8) no deprivation regarding DISHW, and COMPINT, but deprivation regarding HEATS, and INS; (9) no deprivation regarding DISHW, but deprivation regarding HEATS, and COMPINT; (10) deprivation regarding HEATS, and DISHW.

The patterns of GR1 are 1st, 2nd, 4th, and 7th with 76.8, 1.88, 4.5, and 12.8 per centage of coverage respectively (see the table 41 below). For instance, with the 1st pattern, 76.8 per cent of people from GR1 are detected. On the other hand, the patterns of GR4 are 3rd, 5th, 6th, 8th, 9th, and 10th, with 1.2, 1.4, 4.3, 4.43, 12.5, and 71.7 per centage of coverage respectively (see the table 42 below). For instance, with the 3rd pattern, 1.2 per cent of people from GR4 are captured.

Table 40. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP113 pruned tree, data: Turkey SILC 2011

Pattern no	Coverage of pattern (in %) among the total population of GR1	Cumulative coverage (in %)
1	76.8	76.8
2	1.88	78.68
4	4.5	83.18
7	12.8	95.98~**96**

Table 41. Coverage of GR4's patterns among the total population of GR4 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP113 pruned tree, data: Turkey SILC 2011

Pattern no	Coverage of pattern (in %) among the total population of GR4	Cumulative coverage (in %)
3	1.2	1.2
5	1.4	2.6
6	4.3	6.9
8	4.43	11.33
9	12.5	23.83
10	71.7	95.53~**95**

In the training data, by using the CAP113 pruned tree classification patterns, we lose 4 per cent of information for GR1 and 5 per cent of information for GR4.

We can also say that through the 4 patterns of GR1, deprivation regarding DEN, DISHW, and HEATS concern 1.88, 4.5, and 12.8 per cent of the entire population of GR1 (among 4,669 persons) respectively. On the other hand, concerning GR4, the 6 patterns show that the variables HEATS, DISHW, COMPINT, INS, ROWW, and POL affect 88.6, 77.4, 16.8, 4.43, 1.4, and 1.2 per cent of the entire population of GR4 (among 4,490 persons). The 2 deprivation variables, DISHW, and HEATS are the common in terms of deficits for both GR1 and GR4; however, their magnitude is much higher for GR4 compared to GR1.

Regarding the overall accuracy of the mode, as we can see from the confusion matrix statistics, prediction on the test data is very good generating around 95 per cent of accuracy, and sensitivity and specificity rates vary between around 94 and 95 per cent. By using the classification ruled extracted on the training data, 5 per cent of GR1 and 4.5 per cent of GR4 are misclassified on the test data.

CAP113 pruned tree

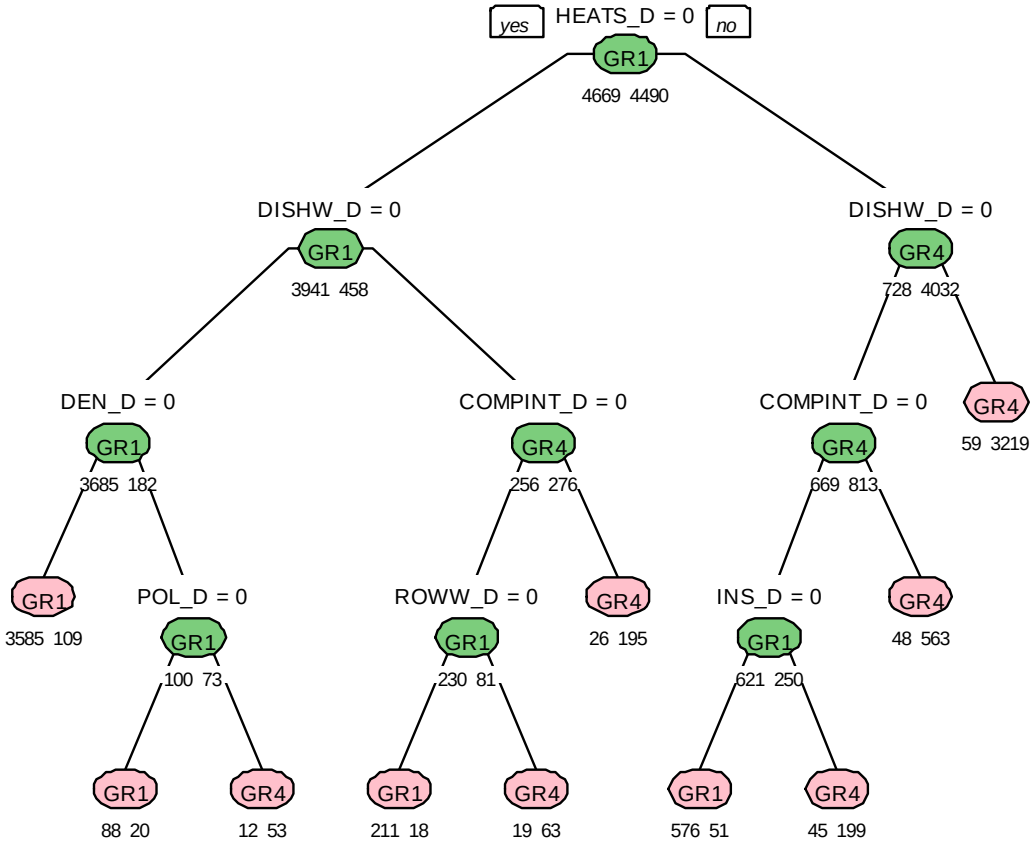


Figure 20. CAP113 pruned tree, data: Turkey SILC 2011

```
Confusion Matrix and Statistics

          Reference
Prediction  GR1  GR4
       GR1 2211  101
       GR4  124 2145

               Accuracy : 0.9509
                 95% CI : (0.9442, 0.957)
    No Information Rate : 0.5097
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.9018

 Mcnemar's Test P-Value : 0.1425

            Sensitivity : 0.9469
            Specificity : 0.9550
         Pos Pred Value : 0.9563
         Neg Pred Value : 0.9454
             Prevalence : 0.5097
         Detection Rate : 0.4826
```

```
   Detection Prevalence : 0.5047
      Balanced Accuracy : 0.9510

       'Positive' Class : GR1
```

CAP114: GR1 vs. GR5 analysis

In this analysis, we focus on DEP_GRS1 vs. DEP_GRS5. We use the data containing only people from the deprivation group 1 and the deprivation group 5. The outcome/target response variable is dichotomous, namely CAP114, taking the value 1 for the group 1 and the value 5 for the group 5 (so that people identified in the group 1 take the value 1 and people identified in the group 5 take the value 5 according to CAP114).

Table 42. Training vs. test samples in CAP114 analysis, data: Turkey SILC 2011

	GR1	GR5
Training sample	4,669	6,482
Test sample	2,335	3,241

As the table above shows, there are 11,151 individuals in the training data (corresponding to 67 per cent of the initial data) and 5,576 individuals in the test sample. Training and test data are obtained by randomly partitioning the initial data/total sample into two subsets by 70 per cent vs. 30 per cent respectively. The data is relatively balanced with 42 per cent for GR1 and 58 per cent for GR5 in both training and test samples (in the initial data, the proportion is also 42 vs. 58 per cent).

Fitting this model with the initial default parameters, and a cp equal to 0.025 has directly produced a not saturated tree with only 11 splits that corresponds to what we are looking for a synthetic information, and with a cross-validated error rate almost zero. Thus, we are also expecting to obtain a very good overall prediction on the test data, and small misclassification rates on the training data. As the initial tree is perfectly match with our requirements, we don't continue with pruning process. The box below contains detailed information on the results of the model.

```
> printcp(dtreeCAP114)

Classification tree:
rpart(formula = CAP114 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
[1] CAR_D     COMPINT_D DISHW_D   HEATS_D   INS_D     MED_D     ROWW_D

Root node error: 4669/11151 = 0.41871

n= 11151

         CP nsplit rel error   xerror      xstd
1 0.8415078      0  1.000000 1.000000 0.0111580
2 0.0368387      1  0.158492 0.158492 0.0056296
3 0.0311630      2  0.121653 0.121653 0.0049728
4 0.0119940      4  0.059327 0.059970 0.0035386
5 0.0033198      5  0.047333 0.047976 0.0031732
6 0.0032127      7  0.040694 0.046263 0.0031171
```

```
7 0.0027843      10   0.031056 0.037481 0.0028110
8 0.0025000      11   0.028272 0.034269 0.0026897
```

CAP114 tree

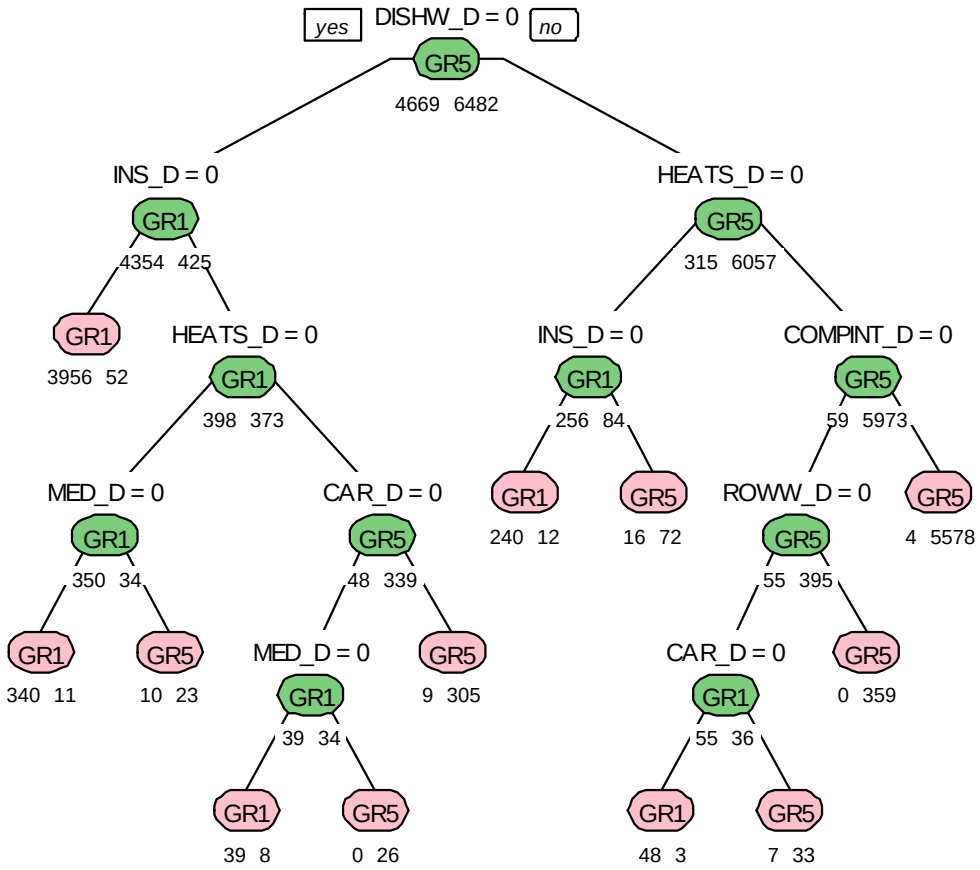


Figure 21. CAP114 tree, data: Turkey SILC 2011

The CAP114 pruned tree is composed of 12 terminal nodes providing 5 patterns of GR1, and 7 patterns of GR5. As we can see from the tree above, the discovered patterns from the left to the right side of the tree are as follows: (1) no deprivation regarding dishwasher (DISHW), and insulation state of the dwelling (INS); (2) no deprivation regarding DISHW, heating system of the dwelling (HEATS), and access to medical services (MED), but deprivation regarding INS; (3) no deprivation regarding DISHW, and HEATS, but deprivation regarding INS, and MED; (4) no deprivation regarding DISHW, car (CAR), and MED, but deprivation regarding INS and HEATS; (5) no deprivation regarding DISHW, and CAR, but deprivation regarding INS, HEATS, and MED; (6) no deprivation regarding DISHW, but deprivation regarding INS, HEATS, and CAR; (7) no deprivation regarding HEATS, and INS, but deprivation regarding DISHW; (8) no deprivation regarding HEATS, but deprivation regarding DISHW, and INS; (9) no deprivation regarding computer with internet (COMPINT), roof, wall, and window conditions (ROWW), and CAR, but deprivation regarding DISHW, and HEATS; (10) no deprivation regarding COMPINT, and ROWW, but deprivation regarding HEATS, DISHW,

and CAR; (11) no deprivation regarding COMPINT, but deprivation regarding HEATS, DISHW, and ROWW; (12) deprivation regarding HEATS, DISHW, and COMPINT.

The patterns of GR1 1st, 2nd, 4th, 7th, and 9th with 84.7, 7.3, 0.8, 5, and 1 per centage of coverage respectively (see the table below 44). For instance, with the 1st pattern, 84.7 per cent of people from GR1 are detected. On the other hand, the patterns of GR5 are 3rd, 5th, 6th, 8th, 10th, 11th, and 12th with 0.35, 0.4, 4.7, 1.1, 0.5, 5.5, and 86 per centage of coverage respectively (see the table 45 below). For instance, with the 3rd pattern,0.35 per cent of people from GR5 are captured.

Table 43. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP114 tree, data: Turkey SILC 2011

Pattern no	**Coverage of pattern (in %) among the total population of GR1**	**Cumulative coverage (in %)**
1	84.7	84.7
2	7.3	92
4	0.8	92.8
7	5	97.8
9	1	98.8~**99**

Table 44. Coverage of GR5's patterns among the total population of GR5 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP114 tree, data: Turkey SILC 2011

Pattern no	**Coverage of pattern (in %) among the total population of GR5**	**Cumulative coverage (in %)**
3	0.35	0.35
5	0.4	0.75
6	4.7	5.45
8	1.1	6.55
10	0.5	7.05
11	5.5	12.55
12	86	98.55~**99**

In the training data, by using the CAP114 tree classification patterns, we lose only 1 per cent of information for both GR1 and GR5.

We can also say that through the 5 patterns of GR1, deprivation regarding INS, HEATS, and DISHW concern 8.1, 1.8, and 6 per cent of the entire population of GR1 (among 4,669 persons) respectively.

On the other hand, concerning GR5, the 7 patterns show that the variables HEATS, DISHW, COMPINT, INS, ROWW, CAR, and MED affect 97, 93, 86, 6.55, 5.5, 5.2, and 0.4 per cent of the entire population of GR5 (among 6,482 persons). The 3 deprivation variables, INS, HEATS, and DISHW are the common in terms of deficits for both GR1 and GR5; however, the magnitude of HEATS, and DISHW is much higher for GR5 compared to GR1.

Regarding the overall accuracy of the mode, as we can see from the confusion matrix statistics, prediction on the test data is very good generating almost 99 per cent of accuracy, and sensitivity and specificity rates are around 99 per cent, and misclassification rates are very small.

```
  Confusion Matrix and Statistics

          Reference
Prediction  GR1  GR5
       GR1 2315   46
       GR5   20 3195

               Accuracy : 0.9882
                 95% CI : (0.985, 0.9908)
    No Information Rate : 0.5812
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9757

 Mcnemar's Test P-Value : 0.002089

            Sensitivity : 0.9914
            Specificity : 0.9858
         Pos Pred Value : 0.9805
         Neg Pred Value : 0.9938
             Prevalence : 0.4188
         Detection Rate : 0.4152
   Detection Prevalence : 0.4234
      Balanced Accuracy : 0.9886

       'Positive' Class : GR1
```

CAP116: GR2 vs. GR4 analysis

In this analysis, we focus on DEP_GRS2 vs. DEP_GRS4. We use the data containing only people from the deprivation group 2 and the deprivation group 4. The outcome/target response variable is dichotomous, namely CAP116, taking the value 2 for the group 2 and the value 4 for the group 4 (so that people identified in the group 2 take the value 2 and people identified in the group 4 take the value 4 according to CAP116).

Table 45. Training vs. test samples in CAP116 analysis, data: Turkey SILC 2011

	GR2	GR4
Training sample	4,132	4,490
Test sample	2,066	2,246

As the table above shows, there are 8,622 individuals in the training data (corresponding to 67 per cent of the initial data) and 4,312 individuals in the test sample. Training and test data are

obtained by randomly partitioning the initial data/total sample into two subsets by 70 per cent vs. 30 per cent respectively. The data is very balanced with 48 per cent for GR2 and 52 per cent for GR4 in both training and test samples (in the initial data, the proportion is also 48 vs. 52 per cent).

By employing the default parameters and a cp equal to 0.025, we obtain a first saturated tree composed of 38 splits reaching to a cross-validation error rate of 0.212 at last split. We prune the tree to around 10-12 splits. The information on the saturated tree is given in the box below.

```
> printcp(dtreeCAP116)

Classification tree:
rpart(formula = CAP116 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
 [1] CAR_D     COMPINT_D DEN_D     DISHW_D   HEATS_D   HWS_D     INS_D
 [8] LIGHT_D   MED_D     POL_D     ROWW_D    SAN1_D    SAN2_D

Root node error: 4132/8622 = 0.47924

n= 8622

          CP nsplit rel error  xerror      xstd
1  0.4603098      0   1.00000 1.00000 0.0112264
2  0.0615924      1   0.53969 0.53969 0.0098403
3  0.0360600      3   0.41651 0.42473 0.0090481
4  0.0124637      5   0.34439 0.34245 0.0083232
5  0.0113746      7   0.31946 0.31849 0.0080817
6  0.0110116      8   0.30808 0.31050 0.0079977
7  0.0072604     10   0.28606 0.29622 0.0078430
8  0.0041142     11   0.27880 0.28049 0.0076654
9  0.0038722     12   0.27469 0.27130 0.0075578
10 0.0034487     14   0.26694 0.26186 0.0074445
11 0.0033882     20   0.24080 0.25194 0.0073219
12 0.0032672     27   0.20353 0.24927 0.0072884
13 0.0031462     30   0.19240 0.24250 0.0072019
14 0.0029042     31   0.18925 0.23475 0.0071008
15 0.0026621     34   0.18054 0.22483 0.0069677
16 0.0025411     35   0.17788 0.21539 0.0068372
17 0.0025000     38   0.16989 0.21297 0.0068030
```

CAP116 pruned tree

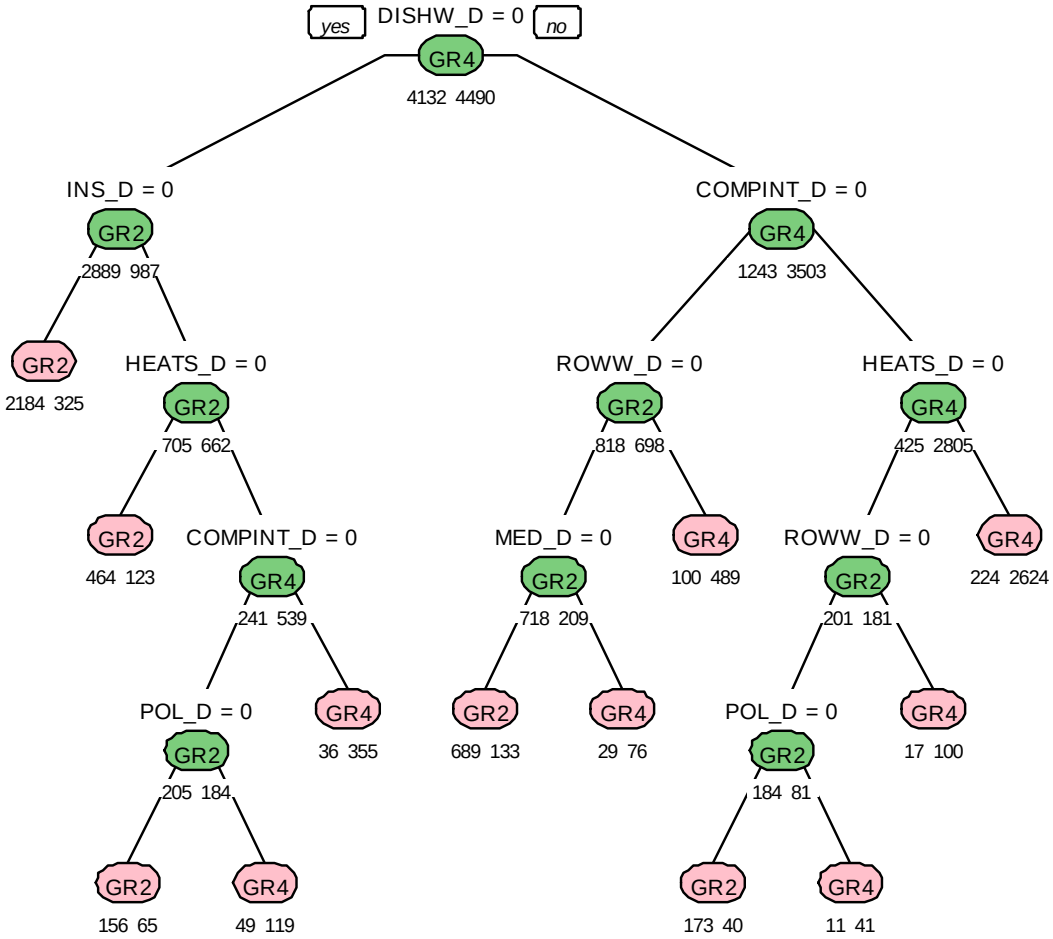


Figure 22. CAP116 pruned tree, data: Turkey SILC 2011

The CAP116 pruned tree is composed of 12 terminal nodes providing 5 patterns of GR2, and 7 patterns of GR4. As we can see from the pruned tree above, the discovered patterns from the left to the right side of the tree are as follows: (1) no deprivation regarding dishwasher (DISHW), and insulation state of the dwelling (INS); (2) no deprivation regarding DISHW, and heating system of the dwelling (HEATS), and deprivation regarding INS; (3) no deprivation regarding DISHW, computer with internet (COMPINT), and pollution (POL), and deprivation regarding INS, and HEATS; (4) no deprivation regarding DISHW, COMPINT, and deprivation regarding INS, HEATS, and POL; (5) no deprivation regarding DISHW, but deprivation regarding INS, HEATS, and COMPINT; (6) no deprivation regarding COMPINT, row, wall, and window conditions (ROWW), access to medical service (MED), but deprivation regarding DISHW; (7) no deprivation regarding COMPINT, and ROWW, but deprivation regarding DISHW, and MED; (8) no deprivation regarding COMPINT, but deprivation regarding DISHW, and ROWW; (9) no deprivation regarding HEATS, ROWW, and POL, but deprivation regarding DISHW, and COMPINT; (10) no deprivation regarding HEATS, and ROWW, but

deprivation regarding DISHW, COMPINT, and POL; (11) no deprivation regarding HEATS, deprivation regarding DISHW, COMPINT, and ROWW; (12) deprivation regarding DISHW, COMPINT, and HEATS.

The patterns of GR2 are 1st, 2nd, 3rd, 6th, 9th with 52.4, 11.2, 3.77, 16.7, 4.2 per centage of coverage respectively (see the table below). For instance, with the 1st pattern, 52.4 per cent of people from GR2 are detected. On the other hand, the patterns of GR4 are 4th, 5th, 7th, 8th, 10th, 11th, and 12th with 2.65, 7.9, 1.7 10.9, 0.9, 2.22, and 58.44 per centage of coverage respectively (see the table below). For instance, with the 4th pattern, 2.65 per cent of people from GR4 are captured.

Table 46. Coverage of GR2's patterns among the total population of GR2 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP116 pruned tree, data: Turkey SILC 2011

Pattern no	Coverage of pattern (in %) among the total population of GR2	Cumulative coverage (in %)
1	52.4	52.4
2	11.2	63.6
3	3.77	67.37
6	16.7	84.07
9	4.2	88.27~**88**

Table 47. Coverage of GR4's patterns among the total population of GR4 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP116 pruned tree, data: Turkey SILC 2011

Pattern no	Coverage of pattern (in %) among the total population of GR4	Cumulative coverage (in %)
4	2.65	2.65
5	7.9	10.55
7	1.7	12.25
8	10.9	23.15
10	0.9	24.05
11	2.22	26.27
12	58.44	84.71~**85**

In the training data, by using the CAP116 pruned tree classification patterns, we lose 12 per cent of information for GR2 and 15 per cent of information for GR4.

We can also say that through the 5 patterns of GR2, deprivation regarding INS, HEATS, DISHW, and COMPINT concern 15, 4, 21, and 4.2 per cent of the entire population of GR2 (among 4,132 persons) respectively.

On the other hand, concerning GR4, the 7 patterns show that the variables DISHW, COMPINT, HEATS, ROWW, INS, POL, and MED affect 74, 69.5, 69, 13, 10.55, 3.55, and 1.7 per cent of the entire population of GR4 (among 4490 persons). Except INS, among the common deprivation variables in terms of deficits for both GR2 and GR4, HEATS, DISHW, and COMPINT appear with a higher magnitude for GR4 compared to GR1.

Regarding the overall accuracy of the model, as we can see from the confusion matrix statistics below, prediction on the test data is relatively good producing an overall 86 per cent of accuracy, and sensitivity and specificity rates are around 87, and 85 per cent respectively, and misclassification rates are not very small, but relatively acceptable given the size of the saturated tree and its complexity.

```
 Confusion Matrix and Statistics

          Reference
Prediction  GR2  GR4
       GR2 1795  342
       GR4  271 1904

               Accuracy : 0.8578
                 95% CI : (0.8471, 0.8681)
    No Information Rate : 0.5209
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7156

 Mcnemar's Test P-Value : 0.004695

            Sensitivity : 0.8688
            Specificity : 0.8477
         Pos Pred Value : 0.8400
         Neg Pred Value : 0.8754
             Prevalence : 0.4791
         Detection Rate : 0.4163
   Detection Prevalence : 0.4956
      Balanced Accuracy : 0.8583

       'Positive' Class : GR2
```

CAP117: GR2 vs. GR5 analysis

In this analysis, we focus on DEP_GRS2 vs. DEP_GRS5. We use the data containing only people from the deprivation group 2 and the deprivation group 5. The outcome/target response variable is dichotomous, namely CAP117, taking the value 2 for the group 2 and the value 5 for the group 5 (so that people identified in the group 2 take the value 2 and people identified in the group 5 take the value 5 according to CAP117).

Table 48. Training vs. test samples in CAP117 analysis, data: Turkey SILC 2011

	GR2	GR5
Training sample	4,132	6,482
Test sample	2,066	3,241

As the table above shows, there are 10,614 individuals in the training data (corresponding to 67 per cent of the initial data) and 5,307 individuals in the test sample. Training and test data are obtained by randomly partitioning the initial data/total sample into two subsets by 70 per cent vs. 30 per cent respectively. The data is relatively balanced with 39 per cent for GR2 and 61 per cent for GR5 in both training and test samples (in the initial data, the proportion is also 39 vs. 61 per cent). With the default parameters and a cp equal to 0.025, we obtain initially a saturated tree with 22 splits as indicated in the box below. Cross-validation error rates are very small. We prune the tree in order to obtain a synthetic information with still a low cross-validation error rate for good prediction on the test data.

```
> printcp(dtreeCAP117)

Classification tree:
rpart(formula = CAP117 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
 [1] CAR_D     COMPINT_D DISHW_D   EDU_D     ENG_H_D   HEATS_D   INS_D
 [8] LIGHT_D   MED_D     NOI_D     POL_D     ROWW_D    SAN1_D    SAN2_D
[15] TEL_D

Root node error: 4132/10614 = 0.3893

n= 10614

          CP nsplit rel error   xerror      xstd
1  0.5989835      0  1.000000 1.000000 0.0121572
2  0.1006776      1  0.401016 0.401016 0.0090499
3  0.0643756      2  0.300339 0.300339 0.0080117
4  0.0285576      3  0.235963 0.235963 0.0072014
5  0.0179090      4  0.207406 0.207406 0.0067928
6  0.0113746      7  0.151985 0.154163 0.0059220
7  0.0110116      8  0.140610 0.145692 0.0057671
8  0.0079864     10  0.118587 0.124395 0.0053523
9  0.0070184     11  0.110600 0.120765 0.0052776
10 0.0038722     12  0.103582 0.111810 0.0050874
11 0.0037512     13  0.099710 0.111810 0.0050874
12 0.0031462     15  0.092207 0.103824 0.0049103
13 0.0027832     16  0.089061 0.100194 0.0048273
14 0.0026621     18  0.083495 0.098984 0.0047992
15 0.0025815     19  0.080833 0.096563 0.0047425
16 0.0025000     22  0.073088 0.094143 0.0046850
```

The CAP117 pruned tree is composed of 11 terminal nodes providing 6 patterns of GR2, and 5 patterns of GR5. As we can see from the pruned tree below, the discovered patterns from the left to the right side of the tree are as follows: (1) no deprivation regarding dishwasher (DISHW), and insulation state of the building (INS); (2) no deprivation regarding DISHW, and computer with internet (COMPINT), but deprivation regarding INS; (3) no deprivation regarding DISHW, and heating system of the dwelling (HEATS), but deprivation regarding INS, and COMPINT; (4) no deprivation regarding DISHW, but deprivation regarding INS, COMPINT, and HEATS; (5) no deprivation regarding COMPINT, and roof, wall, and window conditions (ROWW), but deprivation regarding DISHW; (6) no deprivation regarding COMPINT, but deprivation regarding DISHW, and ROWW; (7) no deprivation regarding HEATS, and ROWW, but deprivation regarding DISHW, and COMPINT; (8) no deprivation

regarding HEATS, but deprivation regarding DISHW, COMPINT, and ROWW; (9) no deprivation regarding car (CAR), and ROWW, but deprivation regarding DISHW, COMPINT, and HEATS; (10) no deprivation regarding CAR, but deprivation regarding DISHW, COMPINT, HEATS, and ROWW; (11) deprivation regarding DISHW, COMPINT, HEATS, and CAR.

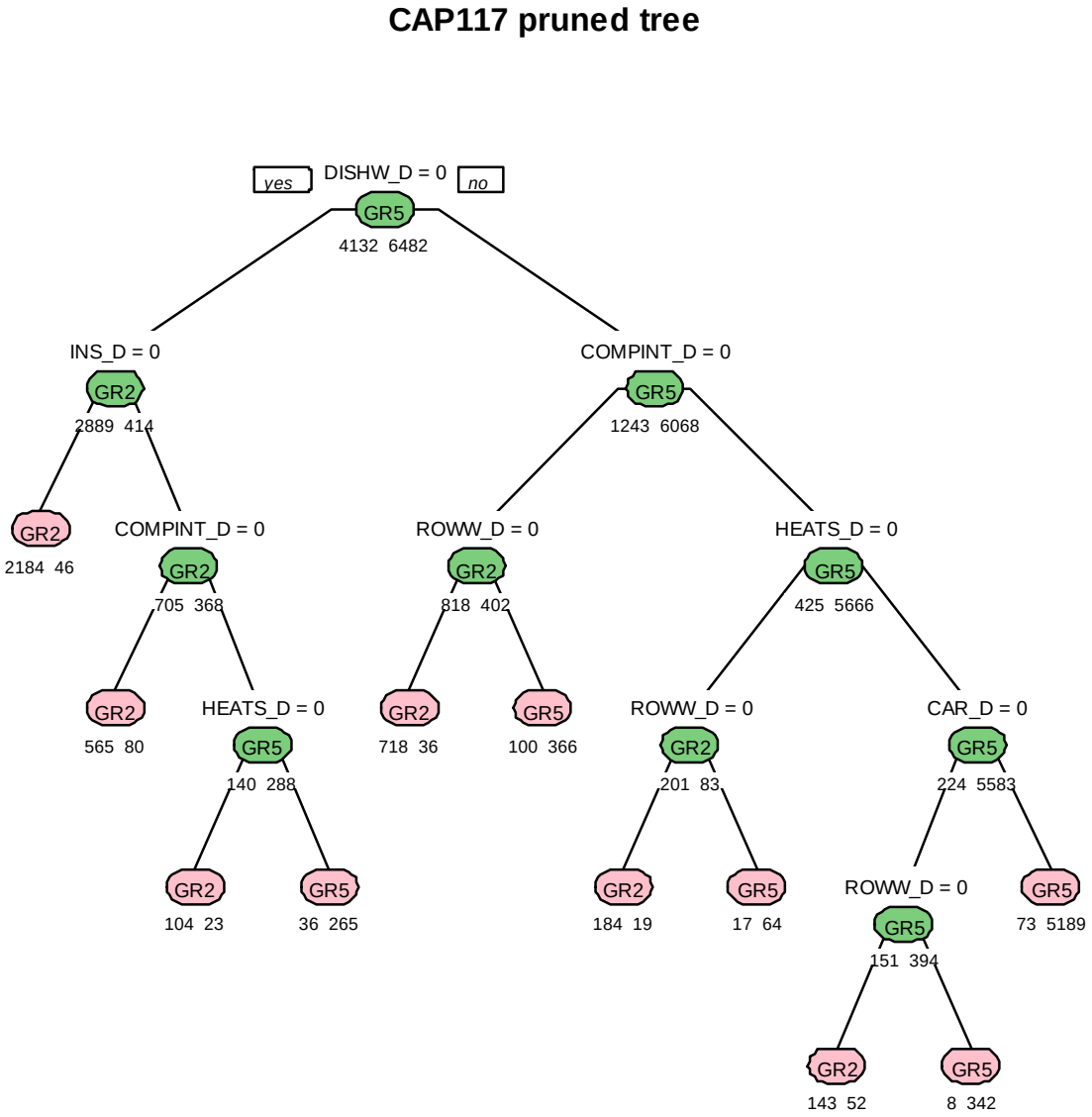


Figure 23. CAP117 pruned tree, data: Turkey SILC 2011

The patterns of GR2 are 1st, 2nd, 3rd, 5th, 7th, and 9th with 52.85, 13.7, 2.5, 17.4, 4.45, and 3.5 per centage of coverage respectively (see the table 50 below). For instance, with the 1st pattern, 52.85 per cent of people from GR2 are detected. On the other hand, the patterns of GR5 are 4th, 6th, 8th, 10th, and 11th with 4, 5.6, 0.98, 5.3, and 80 per centage of coverage respectively (see the table 51 below). For instance, with the 4th pattern, 4 per cent of people from GR5 are captured.

Table 49. Coverage of GR2's patterns among the total population of GR2 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP117 pruned tree, data: Turkey SILC 2011

Pattern no	Coverage of pattern (in %) among the total population of GR2	Cumulative coverage (in %)
1	52.85	52.85
2	13.7	66.55
3	2.5	69.05
5	17.4	86.45
7	4.45	90.9
9	3.5	94.4~**94**

Table 50. Coverage of GR5's patterns among the total population of GR5 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP117 pruned tree, data: Turkey SILC 2011

Pattern no	Coverage of pattern (in %) among the total population of GR5	Cumulative coverage (in %)
4	4	4
6	5.6	9.6
8	0.98	10.58
10	5.3	15.88
11	80	95.88~**96**

In the training data, by using the CAP117 pruned tree classification patterns, we lose 6 per cent of information for GR2 and 4 per cent of information for GR5.

We can also say that through the 6 patterns of GR2, deprivation regarding DINS, COMPINT, DISHW, and HEATS concern 16.2, 10.45, 25.35, and 3.5 per cent of the entire population of GR2 (among 4,132 persons) respectively.

On the other hand, concerning GR5, the 5 patterns show that the variables DISHW, COMPINT, HEATS, CAR, ROWW, and INS affect 92, 90, 89, 80, 12, and 4 per cent of the entire population of GR5 (among 6,482 persons). The 4 deprivation variables, INS, COMPINT, DISHW, and HEATS are the common in terms of deficits for both GR2 and GR5; however, the magnitude of COMPINT, DISHW, and HEATS is much higher for GR5 compared to GR2.

Regarding the overall accuracy of the model, as we can see from the confusion matrix statistics below, prediction on the test data is very good with 95 per cent of accuracy, and sensitivity and specificity rates are quite equal around 94-96 per cent, and misclassification rates are very small.

```
  Confusion Matrix and Statistics

          Reference
Prediction  GR2  GR5
       GR2 1942  138
       GR5  124 3103

               Accuracy : 0.9506
                 95% CI : (0.9445, 0.9563)
    No Information Rate : 0.6107
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.8963

 Mcnemar's Test P-Value : 0.4219

            Sensitivity : 0.9400
            Specificity : 0.9574
         Pos Pred Value : 0.9337
         Neg Pred Value : 0.9616
             Prevalence : 0.3893
         Detection Rate : 0.3659
   Detection Prevalence : 0.3919
      Balanced Accuracy : 0.9487
```

CAP119: GR3 vs. GR5 analysis

In this analysis, we focus on DEP_GRS3 vs. DEP_GRS5. We use the data containing only people from the deprivation group 3 and the deprivation group 5. The outcome/target response variable is dichotomous, namely CAP119, taking the value 3 for the group 3 and the value 5 for the group 5 (so that people identified in the group 3 take the value 3 and people identified in the group 5 take the value 5 according to CAP119).

Table 51. Training vs. test samples in CAP119 analysis, data: Turkey SILC 2011

	GR3	GR5
Training sample	4,842	6,482
Test sample	2,421	3,241

As the table above shows, there are 11,324 individuals in the training data (corresponding to 67 per cent of the initial data) and 5,662 individuals in the test sample. Training and test data are obtained by randomly partitioning the initial data/total sample into two subsets by 70 per cent vs. 30 per cent respectively. The data is quite balanced with 43 per cent for GR3 and 57 per cent for GR5 in both training and test samples (in the initial data, the proportion is also 43 vs. 57 per cent).

By employing default values of parameters, and a cp equal to 0.025, we obtain a saturated tree with 31 splits reaching to a cross-validation error rate of 0.212 at last split. For obtaining a synthetic information, we prune the tree and we obtain a tree with 10 splits. The box below contains information on the saturated tree.

```
> printcp(dtreeCAP119)

Classification tree:
rpart(formula = CAP119 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
 [1] AIRC_D    CAR_D     COMPINT_D DISHW_D   HEATS_D   HWS_D     INS_D
 [8] LIGHT_D   MED_D     MPHO_D    PWS_D     ROWW_D    TEL_D     WM_D

Root node error: 4842/11324 = 0.42759

n= 11324

          CP nsplit rel error  xerror      xstd
1  0.4380421      0   1.00000 1.00000 0.0108728
2  0.0896324      1   0.56196 0.56196 0.0093900
3  0.0739364      2   0.47233 0.47708 0.0088560
4  0.0276745      3   0.39839 0.40314 0.0083010
5  0.0194135      4   0.37071 0.37546 0.0080681
6  0.0179678      5   0.35130 0.36824 0.0080048
7  0.0134242      6   0.33333 0.33457 0.0076950
8  0.0115655      8   0.30648 0.31991 0.0075520
9  0.0105328      9   0.29492 0.30525 0.0074036
10 0.0065056     10   0.28439 0.29409 0.0072870
11 0.0060925     12   0.27138 0.27984 0.0071330
12 0.0055762     15   0.24886 0.26766 0.0069966
13 0.0043371     16   0.24329 0.25506 0.0068507
14 0.0039240     19   0.23028 0.24494 0.0067297
15 0.0036486     20   0.22635 0.23585 0.0066180
16 0.0028914     23   0.21541 0.22945 0.0065375
17 0.0026848     26   0.20673 0.21892 0.0064016
18 0.0026160     27   0.20405 0.21375 0.0063333
19 0.0025000     31   0.19290 0.21252 0.0063168
```

CAP119 pruned tree

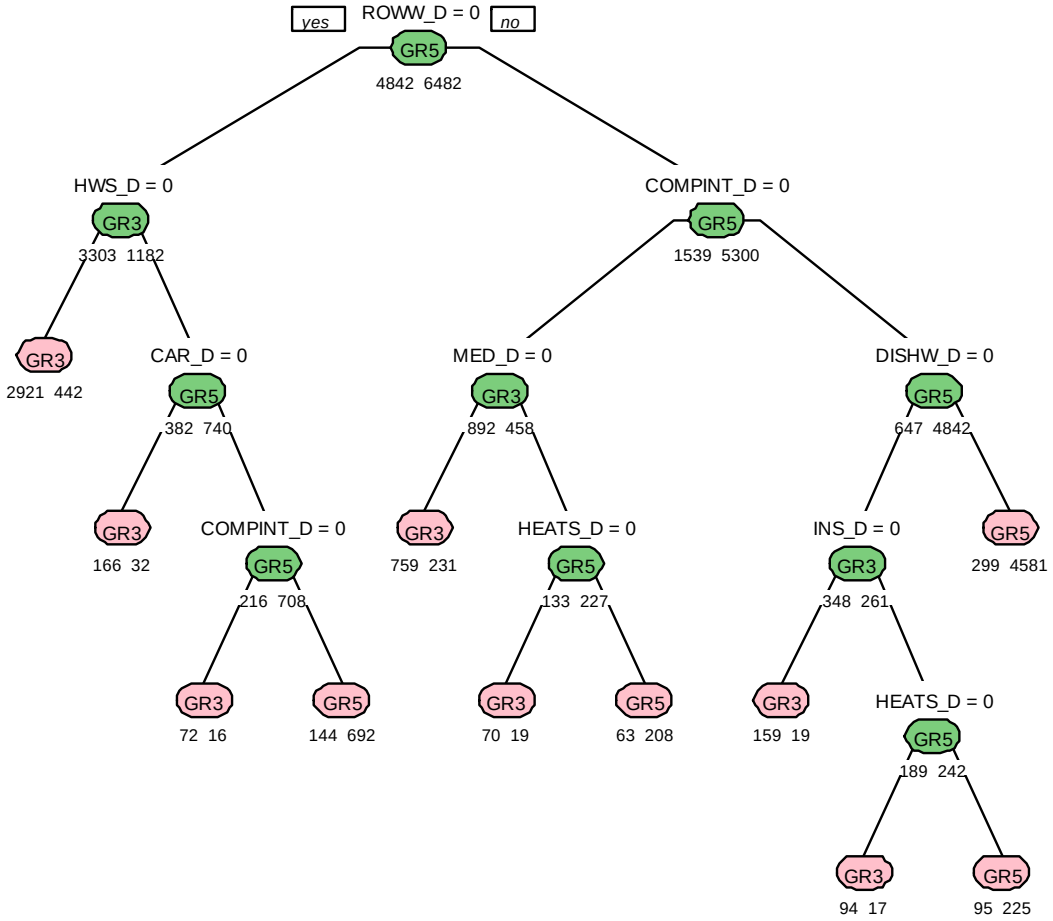


Figure 24. CAP119 pruned tree, data: Turkey SILC 2011

The CAP119 pruned tree is composed of 11 terminal nodes providing 7 patterns of GR3, and 4 patterns of GR5. As we can see from the pruned tree above, the discovered patterns from the left to the right side of the tree are as follows: (1) no deprivation regarding roof, wall, and window conditions (ROWW), and hot water system in the dwelling (HWS); (2) no deprivation regarding ROWW, and car (CAR), but deprivation regarding HWS; (3) no deprivation regarding ROWW and computer with internet (COMPINT), but deprivation regarding HWS, and CAR; (4) no deprivation regarding ROWW, but deprivation regarding HWS, CAR, and COMPINT; (5) no deprivation regarding COMPINT, and access to medical service (MED), but deprivation regarding ROWW; (6) no deprivation regarding COMPINT, and heating system of the dwelling (HEATS), but deprivation regarding ROWW and MED; (7) no deprivation regarding COMPINT, but deprivation regarding ROWW, MED, and HEATS; (8) no deprivation regarding dishwasher (DISHW), and insulation state of the dwelling (INS), but deprivation regarding ROWW, and COMPINT; (9) no deprivation regarding DISHW, and HEATS, but deprivation regarding ROWW, COMPINT, and INS; (10) no deprivation

regarding DISHW, but deprivation regarding ROWW, COMPINT, INS, and HEATS; (11) deprivation regarding ROWW, COMPINT, and DISHW.

The patterns of GR3 are 1st, 2nd, 3rd, 5th, 6th, 8th, and 9th with 60.3, per centage of coverage respectively (see the table 53 below). For instance, with the 1st pattern, 60.3, 3.4, 1.5, 15.7, 1.44, 3.3, and 1.9 of per cent of people from GR3 are detected. On the other hand, the patterns of GR5 are 4th, 7th, 10th, and 11th with 10.7, 3.2, 3.5, and 70.7 per centage of coverage respectively (see the table 54 below). For instance, with the 4th pattern, 70.7 per cent of people from GR5 are captured.

Table 52. Coverage of GR3's patterns among the total population of GR3 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP119 pruned tree, data: Turkey SILC 2011

Pattern no	Coverage of pattern (in %) among the total population of GR3	Cumulative coverage (in %)
1	60.3	60.3
2	3.4	63.7
3	1.5	65.2
5	15.7	80.9
6	1.44	82.34
8	3.3	85.64
9	1.9	87.54~**87.5**

Table 53. Coverage of GR5's patterns among the total population of GR5 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP117 pruned tree, data: Turkey SILC 2011

Pattern no	Coverage of pattern (in %) among the total population of GR5	Cumulative coverage (in %)
4	10.7	10.7
7	3.2	13.9
10	3.5	17.4
11	70.7	88.1~**88**

In the training data, by using the CAP119 pruned tree classification patterns, we lose 12.5 per cent of information for GR3, 12 per cent of information for GR5.

We can also say that through the 7 patterns of GR3, deprivation regarding HWS, CAR, ROWW, MED, COMPINT, and INS concern 5, 1.5, 22, 1.4, 5, and 2 per cent of the entire population of GR3 (among 4,842 persons) respectively.

On the other hand, concerning GR5, the 4 patterns show that the variables COMPINT, ROWW, DISHW, HWS, CAR, HEATS, INS, and MED affect 85, 77, 70.7, 10.7, 10.7, 5.5, 3.5, and 3.2 per cent of the entire population of GR5 (among 6,482 persons). The 5 deprivation variables,

HWS, ROWW, MED, COMPINT, and INS are the common in terms of deficits for both GR3 and GR5; however, the magnitude of ROWW, and COMPINT is greatly higher for GR5 compared to GR3.

Regarding the overall accuracy of the model, as we can see from the confusion matrix statistics below, prediction on the test data is relatively good with 88 per cent of accuracy, and sensitivity and specificity rates are between 87 and 88 per cent, and misclassification rates are relatively small.

```
Confusion Matrix and Statistics

          Reference
Prediction  GR3  GR5
       GR3 2107  376
       GR5  314 2865

               Accuracy : 0.8781
                 95% CI : (0.8693, 0.8865)
    No Information Rate : 0.5724
    P-Value [Acc > NIR] : < 2e-16

                  Kappa : 0.7519

 Mcnemar's Test P-Value : 0.02022

            Sensitivity : 0.8703
            Specificity : 0.8840
         Pos Pred Value : 0.8486
         Neg Pred Value : 0.9012
             Prevalence : 0.4276
         Detection Rate : 0.3721
   Detection Prevalence : 0.4385
      Balanced Accuracy : 0.8771

       'Positive' Class : GR3
```

Comparative summary

After presented in detail the results of classification tress through 5 different deprivation groups, we will provide in this sub-section a summary of the discovered patterns regarding content and predictive accuracy.

CAP112

Based on the training data of GR1 and GR3 in this model, we can see that people living in a dwelling with an appropriate heating system, a good insulation state, and possessing a dishwasher are most likely to be part of GR1. Subsequently, people living in a dwelling with a not appropriate heating system but having a dishwasher, a computer with an internet, a car, and good roof, wall, and window conditions are at risk of being in GR1. In addition, people living in a dwelling with insulation problems, but having an appropriate heating system, a car and a dishwasher are likely to be in GR1. On the other hand, for detecting GR3 compared to GR1, more patterns are extracted. If we take into account the most relevant four patterns with a per centage higher than 5 in terms of coverage of GR3, we can observe that people living in a household with no dishwasher, and a not appropriate heating system are at most risk of being in GR3. Subsequently, people not lacking a dishwasher, but not having a computer with an internet, and an appropriate heating system in their dwelling are also likely being in GR3.

Another combination, being deprived regarding heating system, roof, wall, and window conditions, but possessing both a computer with an internet, and a dishwasher leads likely to being in GR3. Or, not having a computer with an internet, a dishwasher but living in a dwelling with good heating system makes likely the person to be in GR3 compared to GR1. For both GR1 and GR3, problems due to heating system of the dwelling appear as the most important deprivation item affecting most the capabilities of people. In terms of material, computer with internet, and dishwasher emerge as the most influential ones for GR3. Notably, among GR1 being deprived regarding dishwasher, and/or heating system is less likely in opposition to GR3 in which 57 per cent of them experience lack of a dishwasher, and 74 per cent experience not appropriate heating system in their dwelling. Thus, GR3 distinguishes primarily from GR1 in terms of deprivation in heating system, dishwasher, and computer with an internet.

CAP113

In this model, compared to CAP112, we expect to see more severe deprivation among GR4 as the number of deprivations is higher for GR4 compared to GR3, and the inequality normally accentuates between GR1 and GR4 compared to what is happening between GR1 and GR3. Experiencing a life with a not appropriate heating system and no dishwasher in the dwelling put people at most risk of being in GR4. This result is also valid for detecting GR3 compared to GR1, as CAP112 has shown. But the impact of these two deprivations is higher among GR4 compared to GR3, notably for not possessing a dishwasher in the dwelling. On the other hand, for who have a dishwasher but living in a dwelling with heating problems and lacking a computer with an internet is likely associated with GR4. For GR1, not suffering regarding heating system, dishwasher, and access to dentist is the most relevant capability set. However, 12.8 per cent of GR1 suffer of not living in an appropriate heating system donated dwelling even if they possess a dishwasher, a computer with internet, and even if they do live in a dwelling with no insulation problems. In line with the findings of CAP112, here also a relatively small part of GR1 emerge as suffering living in a dwelling with a not appropriate heating system, and/or not having a dishwasher their dwelling. However, if we compare how GR3 and GR4 distinguish from GR1, we can say that lacking an appropriate heating system, and/or a dishwasher, and/or a computer with an internet is very relevant.

CAP114

The poorest group, GR5, is most likely experiencing deprivation regarding heating system, dishwasher, and computer with an internet. In terms of heating system, and dishwasher, similar result was also found for GR3 and GR4 in CAP112 and CAP113 models, which means that there is less heterogeneity among GR3, GR4, and GR5 when it comes to be distinguished from the less poor group, GR1. However, the magnitude of these two variables is much higher among GR5. It is important to note that 86 per cent of GR5 is detected through the pattern of being deprived simultaneously regarding dishwasher, heating system, and computer with internet. In addition, 5.5 per cent of GR5 appears living in a dwelling with bad roof, wall, window conditions with a not appropriate heating system, and not having a dishwasher even if they have a computer with internet access. On the other hand, 84.7 per cent of GR1 live in a household with a dishwasher and without insulation problems in the dwelling. But it is important to remark that, 7.3 per cent of them suffer regarding insulation state of their dwelling but at the same time they do not lack a dishwasher, bad heating system, neither access to medical services. And 5 per cent of GR1 do not possess a dishwasher but live in a dwelling with an appropriate heating system and good insulation state.

CAP116

Compared to GR2, 58.44 per cent of GR4 emerges as being deprived in terms of dishwasher, computer with internet, and heating system of the dwelling. In addition to that, 10.9 per cent of GR4 lack a dishwasher and live in a dwelling with not adequate roof, wall, and window conditions but knowing that they have a computer with internet, and 7.9 per cent of GR4 suffer heating problems in their dwelling both in terms of heating system and insulation state, do not have a computer with an internet but they do have a dishwasher. In CAP113 analysis, 71.7 per cent of GR4 was emerged as being deprived in terms of heating system and dishwasher; but also 12.5 per cent of them have a problem with heating system and do not have a computer with an internet. When we compare these two analyses from the perspective of GR4, we can observe that deprivation regarding heating system and dishwasher are still dominant for GR4 no matter with GR1 or GR2 they are in comparison. On the other hand, half of GR2 is most likely experiencing no deprivation in both dishwasher and heating problems due to insulation in their dwelling. However, almost one of 5 persons among GR2 lack a dishwasher, and 15 per cent of GR2 suffer heating problems due to insulation state of their dwelling. The magnitude of not having a dishwasher is almost 3.5 times higher, of not having a computer and internet is 16.5 times higher and having a not adequate heating system is 17.25 times higher among GR4 compared to GR2 in terms of per centage points.

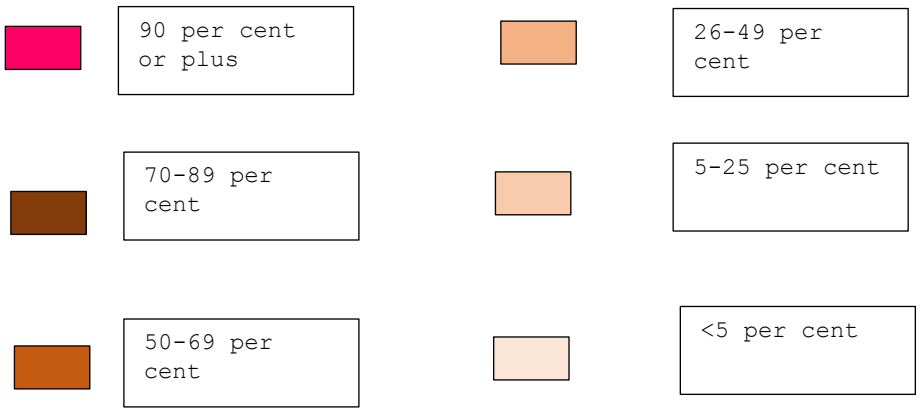
CAP117

In this analysis, it is found that 80 per cent of GR5 experience deprivation simultaneously in terms of dishwasher, computer with an internet, heating system, and car. In addition to that, 5.6 per cent of GR5 don't have a dishwasher in their dwelling and the roof, wall, and window conditions of their dwelling are not good even if they do have a computer with an internet. And, 5.3 per cent of GR5 is deprived regarding dishwasher, computer with internet, heating system, roof, wall and window conditions even if they do have a car. On the other hand, 52.85 per cent of GR2 don't experience deprivation both in having a dishwasher, and heating problems due to insulation in their dwelling. Nevertheless, 17.4 per cent of them don't have a dishwasher even if they live in a dwelling with good roof, wall, window conditions and do not experience deprivation regarding computer with internet. And 13.7 per cent of GR2 suffer also regarding heating problems due to insulation state of their dwelling even if they do have a dishwasher and a computer with internet.

CAP119

When GR5 is compared with GR3, which can be considered as in the middle of 5 deprivation groups, we can see that 70.7 per cent of GR5 suffer deprivation in terms of computer with internet, dishwasher, and roof, wall, and window conditions of their dwelling. Compared to the results of GR2 vs. GR5, here the heating system variable (HEATS) does not emerge as very relevant affecting only 5.5 per cent of GR5. This result is probably due to the fact that GR3 suffer more in terms of heating system similarly as GR5 compared to GR2. Moreover, 10.7 per cent of GR5 lack hot water system, car, computer and internet in their dwelling but they live in a dwelling with adequate roof, wall, and window conditions. And, in particular, the difference in terms of bad roof, wall, and window conditions of the dwelling becomes more relevant between GR3 and GR5 compared to GR2 vs. GR5. On the other hand, 60 per cent of GR3 don't have deprivation in terms of hot water system, and roof, wall, and window conditions of their dwelling. But still, 15.7 per cent of them live in not adequate roof, wall, window conditions dwelling even if they do not suffer regarding computer with internet, and access to medical services.

As we have already explained in detail the utility of Figure 25 below in GHS analysis. Just for recall, this figure helps to catch the overall importance of variables in terms of deprivation people experience regarding them on the training data across all employed models. There are 29 columns labelled as model, group, and variables-from RO_D to FS1, respectively. The rows in the figure include the name of 6 models (e.g., CAP111) and the groups existing in the model/target groups (e.g., GR1 vs. GR2). The 6 different criteria, corresponding each of them to a different interval of per centage of population in the training data experiencing deprivation regarding the variable, are arbitrarily set as follows: 90 per cent +, 70-89 per cent, 50-69 per cent, 26-49 per cent, 5-25 per cent, and < 5 per cent. For instance, in CAP112 model, per centage of GR3 population in the training data suffering in terms of deprivation regarding the variable HEATS is in the interval of 70-89 per cent. Each criterion is presented in a different color as follows:



From the information below, we can see that the variables RO_D, ENG_H, SAN1, SAN2, KIT, PWS, LIGHT, NOI, CRI, TEL, MPHO, TV, WM, REF, AIRC, EDU, and FS1 are not used in any pruned tree to distinguish the deprivation groups. It is important to mention that the variables appearing in the pruned trees are the most influential ones in terms of information gain to distinguish the two groups of the target response variable. However, it does not mean that the groups in the model are not deprived regarding other variables that do not appear in the tree.

Based on the results displayed in Figure 25, we can observe that across the three models, CAP112, CAP113, CAP114, not having an appropriate heating system and a dishwasher in the dwelling emerges with a considerably high frequency among the population of GR3, GR4, and GR5, which shows clearly an important item of discrepancy between these 3 groups and the most well-off (GR1) population. More specifically, the deprivation frequency regarding these two items becomes more intense among GR5 in CAP114 model compared to GR1. Among the list of materials included in the analysis, the results show that after dishwasher, the variable of having a computer and internet in the dwelling surfaces as the second most important item. Notably, for reducing inequality between GR1 and GR5, the priority should be given to the items of heating system and dishwasher. In other words, heating system of the poorest dwellings should be revised and a specific programme for diminishing the number of poor households

lacking a dishwasher should be lanced, which in turn will help to decrease inequality between GR1 and GR5 regarding material deprivation and dwelling conditions. During the phase of monitoring heating systems existing in poor households, roof, window, and wall conditions, and insulation state of these dwellings can be checked as well. Another dwelling conditions variable the hot water system emerges also in the patterns of GR3, and GR5 in CAP119 model. Hot water system does not show up in CAP114 model, which is probably due to the fact that heating system variable emerges significantly in the pattern of GR5 compared to GR1 in CAP114. In technical terms, heating system and hot water system in the dwelling are quite correlated, so that these two elements should be considered simultaneously for increasing living standards. Heating problems due to insulation state of the dwelling appear affecting all groups with relatively low per centages compared to notably dishwasher, and heating system. Deprivation in terms of computer with an internet occur in GR2, GR3, GR4, and GR5 with higher frequency among GR4 and GR5. As a summary, we can say that deprivation in terms of heating system (HEATS), dishwasher (DISHW), and insulation state of the dwelling (INS) concern all groups with different magnitudes. Notably, HEATS, DISHW, and COMPINT become problematic for GR3, GR4, and GR5; and ROWW, hot water system (HWS), car (CAR), access to medical services (MED) for GR5 should be prominently considered; and in addition to HEATS, the insulation state of the dwellings can be simultaneously examined when heating system available in dwellings is considered.

In CAP112 model, GR3 is greatly distinguished by GR1 in terms of high deficit in access to adequate heating system, dishwasher in the dwelling as well as in terms of moderate deficit in computer with an internet and a car. In CAP113 model, the situation of GR4 is found greatly deprived in terms of heating system, and dishwasher in the dwelling compared to GR1. If we compare the results of CAP112 and CAP113 models, we can see that discrepancy between GR1 and GR4 becomes more striking in terms of lacking an appropriate heating system and dishwasher in the dwelling compared to the discrepancy between GR1 and GR3. In addition, in terms of deprivation regarding computer with an internet, roof, wall, window conditions, and pollution, discrepancy between GR1 and GR4 appears less than those between GR1 and GR3, and deprivation in terms of car does not emerge in case of GR1 vs. GR4, and deprivation in terms of insulation state of the dwelling remains at similar level for both GR3 and GR4 in CAP112 and CAP113 models. In other words, deprivation per centages regarding computer with internet, roof, wall, and window conditions, pollution diminishes for GR4 in CAP113 compared to GR3 in CAP112 because deprivation in heating system and dishwasher becomes more important for GR4 including already COMPINT, ROWW, and POL deprivation profiles. In CAP114 model, the situation of GR5 emerges significantly worse in terms of deprivation in heating system, dishwasher, and computer with an internet. The discrepancy between GR1 and GR5 in these 3 items but also regarding insulation state of the dwelling, roof, wall, and window conditions becomes worse compared to the discrepancy between GR1 and GR3. This result means that inequality in terms of these items increases from GR1 (the most prosperous) to GR5 (the poorest). According to CAP116 model, the discrepancy between GR2 and GR4 is found higher in terms of deprivation in computer with an internet, roof, wall, window conditions, and insulation state of the dwelling whilst it is found lower when it comes to deprivation in dishwasher, and heating system of the dwelling compared to the discrepancy between GR1 and GR4 in CAP113 model. This is due to the fact that in terms of COMPINT, ROWW, and INS, the inequality between GR2 and GR4 manifests still whereas inequality diminishes in terms of HEATS and DISHW. In CAP117 model, it is found that GR5 suffers at a very high-level the

deprivation in terms of dishwasher, computer with internet, heating system of the dwelling, and car. And if we compare these results with GR1 vs. GR5 results we can observe that inequality between GR2 and GR5 in terms of car and roof, wall, and window conditions becomes higher in detriment of GR5. And, based on the results of CAP119 model, we can say that inequality between GR3 and GR5 becomes important in terms of roof, wall, window conditions, access to medical services, hot water system in the dwelling, and car in detriment of GR5.

In Figure 26, patterns of groups are shown per model. The green color means no deprivation and the red color means deprivation. Rows are the models, and columns include the title of model (Mod), groups (Gr), pattern (Pat), frequency of pattern (Fre) or in other words the coverage of pattern among the training population in per centage, and variables from HEATS to DEN, respectively (including only variables emerging in the patterns). For instance, the first pattern (P1) of GR1 in CAP112 concerns 71.4 per cent of GR1 people in the training data. This pattern includes no deprivation regarding HEATS, DISHW, and INS at the same time. In CAP112, deprivation regarding HEATS and DISHWASHER at the same time surfaces as the biggest issue for GR3. Then, for those having a dishwasher, deprivation regarding HEATS and COMPINT at the same time comes out as the second most important deficit. Besides, for those having an accurate heating system, deprivation regarding COMPINT and DISHW comes out in the model; and for those having a dishwasher, and a computer with internet, deprivation regarding HEATS and ROWW at the same time emerges. In CAP119, findings show that 60.3 per cent of GR3 does not suffer deprivation in terms of bad roof, window, wall conditions (ROWW) and hot water system (HTW), given that such pattern is the most important one distinguishing GR3 as the more prosperous than the poorest, GR5, population in the country. In other words, GR3 appears as at relatively low risk of suffering in terms of ROWW, and HTW compared to the poorest households but at the same time half of them lack an appropriate heating system and a dishwasher, and 15 per cent of them is poor regarding HEATS and COMPINT compared to the least poor GR1 population.

From the results of CAP113, we can see that the combination of deprivation regarding HEATS and DISHW constitutes 71.7 per cent of GR4 population, given that this pattern is the most informative one in terms of inequality between GR1 and GR4. Then, 12.5 per cent of GR4 not suffering deprivation regarding DISHW, suffer nevertheless regarding heating system and computer with internet. After that, 4.43 per cent of GR4 not suffering regarding DISHW and COMPINT, suffer however regarding HEATS, and INS. We can see that the biggest discrepancy between GR1 and GR4 rests on HEATS, and DISHW as a material deprivation. Not having a computer with an internet, and car comes as the two other material deprivation issues among GR4 compared to GR1. Concerning the poorest, GR5, population from CAP114 model, we can observe that 86 per cent of GR5 is separated from GR1 in terms of capabilities deficit regarding simultaneously HEATS, COMPINT, and DISHW. After that, 5.5 per cent of GR5 does not have a proper heating system, adequate roof, walls, windows, and a dishwasher, despite they do have a computer with internet. Throughout these findings, we can say that HEATS, and DISHW are the two main problems that should be solved for GR3, GR4, and GR5. After that, computer and internet issue should be checked for improving access to information system for all households across the country. Moreover, in CAP117 model when GR2 is compared with the poorest, GR5, the results show that a slightly higher of half of the GR2 population do not suffer in terms of insulation state of their dwelling, and dishwasher. However, 17.4 per cent of them lack a dishwasher even if they are not deprived regarding COMPINT, and

ROWW. On the other hand, 13.7 per cent of them do have a computer with internet but they have insulation problem in their dwelling. Throughout these results, capabilities deficit concern mostly dwelling standard issues in terms of heating system, insulation, roof, wall, window conditions, hot water system, and material possession issues in terms of dishwasher, computer, internet, and car. In this line, we recommend primarily taking in consideration the monitoring program specifically focusing on the state of dwellings across the country. Checking water access system will be an advantage, and notably the use of dishwasher in the dwellings, notably in large size households, will be an important point for a sustainable development and preserve of water stock. So that, government may work on the promotion of spreading the use of dishwasher. Moreover, as the access to internet has become an important functioning; for instance, the information on how to get an access to state services such as national social security system, having appointments in public hospitals, are available on various web sites of government/state, government may work on the promotion of the use and possession of computer and internet for all households across the country. In second model, we will more specifically focus on capabilities of individuals in terms of access to medical services, educational attainment, and financial situation for bringing additional information to the summary view we get from the overall pattern analysis.

Table 55 below shows the main confusion matrix statistics clarifying the predictive accuracy of our 6 models. In the table, columns are the models and rows correspond to accuracy (acc), sensitivity (sens), specificity (spec), positive predictive value (pos pred value), negative predictive value (neg pred value), balanced accuracy (balanced acc) computed on the test data, and misclassification rate (misc rate) computed in the training data, respectively. From the results, we can see that the models containing the groups that are close in terms of total number of deprivations have less accuracy. The models CAP113 (GR1 vs. GR4), CAP114 (GR1 vs. GR5), CAP117 (GR2 vs. GR5) have the best predictive accuracy. The 3 other models CAP112 (GR1 vs. GR3), CAP116 (GR2 vs. GR4), and CAP119 (GR3 vs. GR5) can be considered having a relatively good predictive accuracy above 85 per cent. In this line, it will be adequate to focus on **CAP112** for *GR3*, **CAP113** for *GR4*, **CAP114** for both *GR1* and *GR5*, and **CAP117** for *GR2* for further analysis. It is important to emphasize the GHS analysis have also shown the same tendencies regarding that point.

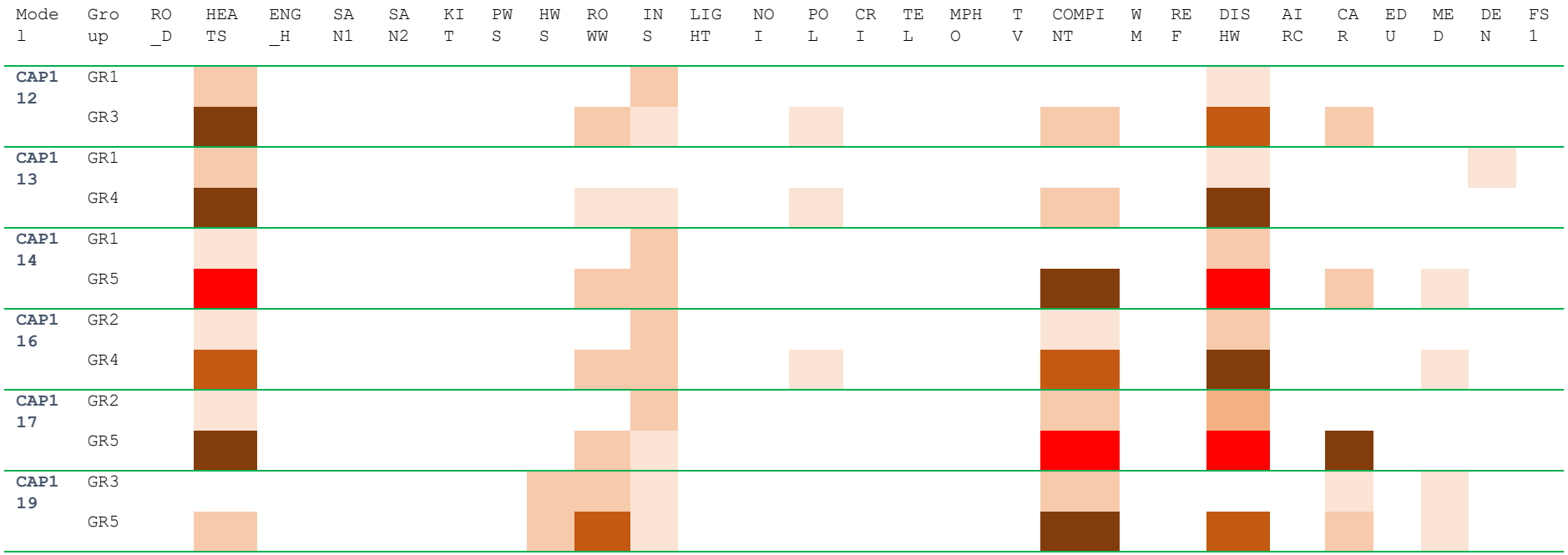


Figure 25. ***Deprivation per centages represented by different colors in terms of variables emerging in the patterns by group within each model, data: Turkey SILC 2011***

Mod	Gr	Pat	Freq	HEATS	HWS	ROWW	INS	POL	COMPINT	DISHW	CAR	MED	DEN
CAP112	1	P1	71.4										
		P7	9.55										
		P2	5.8										
		P4	4.6										
	3	P11	49										
		P10	15										
		P9	5.9										
		P6	5.8										
		P3	4.8										
		P8	4.13										
		P5	2.4										
CAP113	1	P1	76.8										
		P7	12.8										
		P4	4.5										
		P2	1.88										
	4	P10	71.7										
		P9	12.5										
		P8	4.43										
		P6	4.3										
		P5	1.4										
		P3	1.2										

Mod	Gr	Pat	Freq	HEATS	HWS	ROWW	INS	POL	COMPINT	DISHW	CAR	MED	DEN
CAP114	1	P1	84.7										
		P2	7.3										
		P7	5										
		P9	1										
	5	P12	86										
		P11	5.5										
		P6	4.7										
		P8	1.1										
CAP116	2	P1	52.4										
		P6	16.7										
		P2	11.2										
		P9	4.2										
		P3	3.77										
	4	P12	58.44										
		P8	10.9										
		P5	7.9										
		P4	2.65										
		P11	2.22										
		P7	1.7										
CAP117	2	P1	52.85										
		P5	17.4										

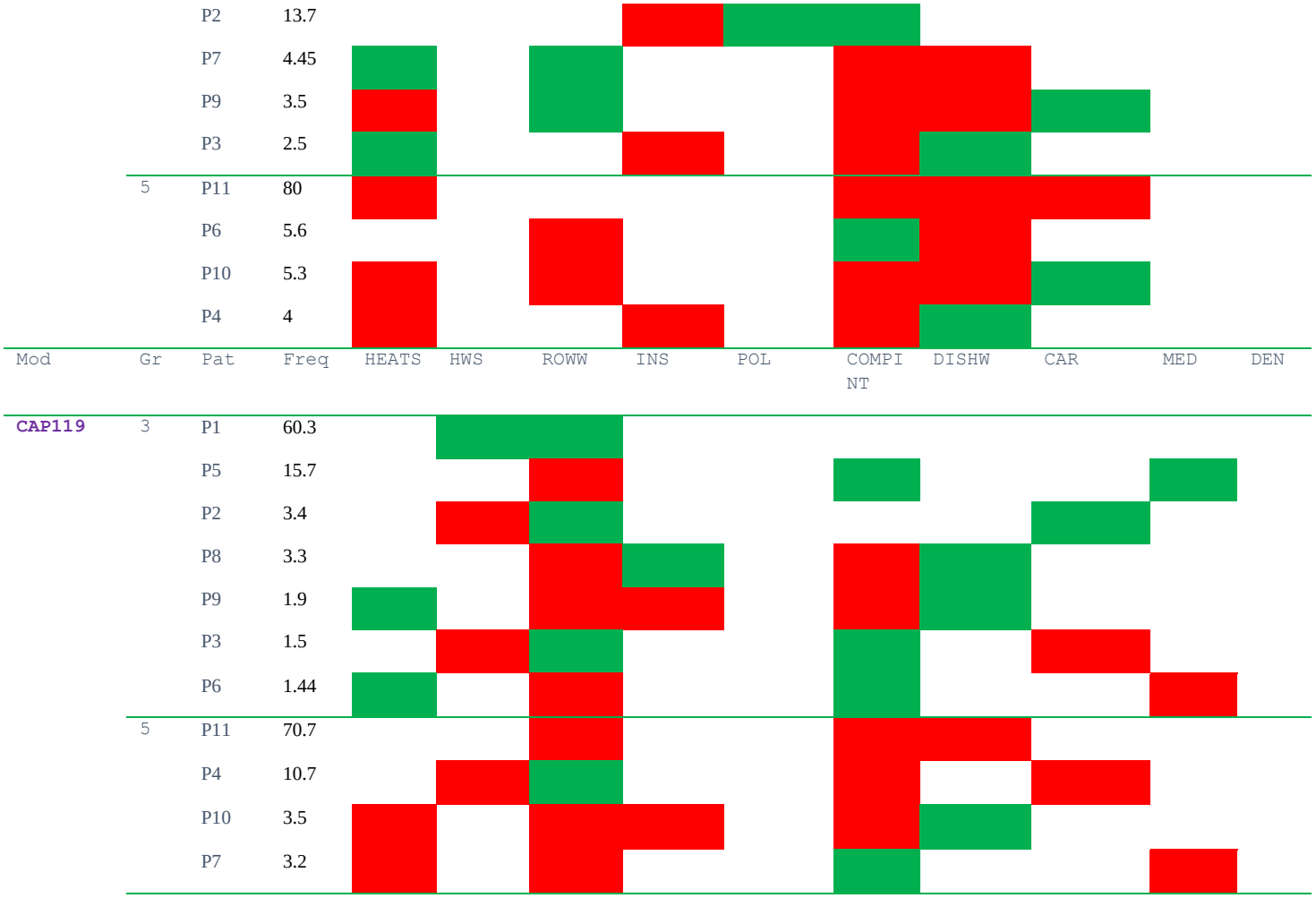


Figure 26. Frequency and composition of classification patterns per model, data: Turkey SILC 2011.

Table 54. Main confusion matrix statistics per model, data: Turkey SILC 2011

	CAP112 (GR1 vs. GR3)	CAP113 (GR1 vs. GR4)	CAP114 (GR1 vs. GR5)	CAP116 (GR2 vs. GR4)	CAP117 (GR2 vs. GR5)	CAP119 (GR3 vs. GR5)
Acc	0.882	0.950	0.988	0.857	0.950	0.878
Sens	0.896	0.946	0.991	0.868	0.940	0.870
Spec	0.869	0.955	0.985	0.847	0.957	0.884
Pos pred value	0.869	0.956	0.980	0.840	0.933	0.848
Neg pred value	0.896	0.945	0.993	0.875	0.961	0.901
Balanced acc	0.883	0.951	0.988	0.858	0.948	0.877
No-Misc rate (test data)	89.64 87	94.7 95.5	99.15 98.6	86.8 84.7	94 95.7	87 88.4

Second step (step2 analysis): detecting dimensional poverty profiles of poverty groups

In this sub-section, we will in particular focus on GR3, GR4, and GR5 in terms of their dimensional profiles. By taking in consideration the performance of models (i.e., predictive accuracy), we can see that for GR3 the model CAP112, for GR4 the model CAP113, and for GR5 the model CAP114 bring the most efficient information concerning their most relevant capabilities patterns. Inside each of these three models, by selecting the most influential patterns, we will attempt to find out the variety of profiles of the three groups regarding dwelling, environment, material, education, health, and financial situation aspects respectively. The aim of doing this is to detect more information on living conditions of these three most deprived groups.

The dimension of *dwelling conditions* is separated into two sub-sets: (i) DWE1: RO_D, KIT, ROWW, SAN1, and SAN2; (ii) DWE2: HEATS, ENG_H, INS, PWS, HWS, and LIGHT. The first covers deprivation regarding number of room per person in the household, whether a kitchen exists or not in the welling, whether roof, wall, window conditions are bad in the dwelling, whether a bathroom/shower exists or not in the dwelling, and whether a toilet exists or not in the dwelling, respectively. The second includes deprivation heating system, energy source for heating, whether heating problems exist due to the insulation state of the dwelling, whether a piped water system exists or not in the dwelling, whether a hot water system exists or not in the dwelling, and whether there is a problem of lighting or not in the dwelling respectively. The main reason behind this separation is profiles become easily complex to interpret when many variables are included, as we have seen in the beginning of the analysis. The dimension of *environment* (ENV) includes deprivation regarding whether noise pollution (NOI), air pollution (POL), and crime, and violence (CRI) exist or not in the community, neighborhood where the dwelling is located. The dimension of *material* (MAT) is divided into two sub-variables: (i) MAT1: TEL, MPHO, TV, COMPINT, CAR; (ii) MAT2: WM, REF, DISHW, AIRC. The first group comprises deprivation regarding available landline telephone, mobile phone, television, computer and internet, and car in the dwelling. The second concerns deprivation regarding available washing machine, refrigerator, dishwasher, air-condition in the dwelling. The dimension of *education* (EDUC) includes only deprivation regarding educational attainment. The dimension of *health* (HEALTH) is composed of two variables, namely, access to medical services when it was needed (MED) and access to dental services when it was needed (DEN). And the dimension of *financial situation* (FINANCE) includes only deprivation regarding monthly financial situation of the household.

We start off by detecting the population of GR3, GR4, and GR5 conforming to the most informative patterns (see below) in the full data (training plus test data), and then we find out their dimensional profiles according to the variables indicated above. This exercise is performed in statistical software STATA, for each group separately.

GR3 dimensional profiles based on CAP112 model

With this model, we saw that the following 6 patterns capture 85 per cent of population of GR3 in the training data: (1) deprivation regarding HEATS, and DISHW; (2) no deprivation regarding DISHW, but deprivation regarding HEATS, and COMPINT; (3) no deprivation regarding DISHW, and COMPINT, but deprivation regarding HEATS, and roof wall, and window conditions (ROWW); (4) no deprivation regarding HEATS, but deprivation regarding COMPINT, and DISHW; (5) no deprivation regarding HEATS, and DISHW, but deprivation regarding CAR, and INS; (6) no deprivation regarding DISHW, COMPINT, ROWW, but

deprivation regarding CAR, and HEATS. When we applied these 6 classification rules in the full data, we lose 15 per cent of information reaching to 85 per cent (as in the case of training data), which shows that the model predicts well GR3.

The table below illustrates the dimensional profiles of GR3 based on the extracted 6 patterns. In GHS analysis, we already note that only profiles concerning at least 5 per cent of the pattern population were shown; however, in case of GR3 here, we employ in addition two patterns contributing around 4 per cent of information in order to gather higher information at dimensional level for GR3. The table has 6 columns representing pattern number, and the most substantial dwelling conditions (DWE1, and DWE2), environment (ENV), and material (MAT1, and MAT2) profiles per each pattern. Due to space limitation in the table, the profiles of the dimensions of education (EDUC), health (HEALTH), and financial situation (FINANCE) per each pattern will be given directly in the text. In first column, next to the pattern number, the frequency of the pattern is mentioned in parenthesis in terms of per centage. For instance, 1st pattern concern 49 per cent of population of GR3 in the entire data. Profiles are composed of 0 and 1 where 0 means no deprivation and 1 means deprivation regarding the variable. The profiles are composed in the following way:

DWE1= RO_D_D+KIT_D+ROWW_D+SAN1_D+SAN2_D where the 5 variables are string character according to STATA language, which allows the combination of values and not necessary summing the values to obtain a numeric total value. For instance, the profile 10001 means having deprivation regarding RO_D, and SAN2 but no deprivation regarding KIT, ROWW, and SAN1. Please see the table of deprivation variables presented in the beginning of the case study.

DWE2= HEATS_D+ENG_H_D+INS_D+PWS_D+HWS_D+LIGHT_D.

ENV=NOI_D+POL_D+CRI_D.

MAT1=TEL_D+MPHO_D+TV_D+COMPINT_D+CAR_D.

MAT2=WM_D+REF_D+DISHW_D+AIRC_D.

EDU=EDU_D

HEALTH=MED_D+DEN_D.

FINANCE=FS1_D.

In the table, next to the profiles, their frequency (i.e., Freq) within the pattern population is mentioned in parenthesis in terms of per centage. For instance, the profile 00000 in DWE1 concerns 53.15 per cent of population of pattern 1. The sum of frequencies of dimensional profiles in each column for each pattern equals maximum 100 per cent. In case of DWE1 profile for the 1st pattern, there are 4 profiles selected as the most substantial ones with at least 100 observations. The last row in the table indicates the total information captured in each dimension for the ensemble of GR3. For instance, for the DWE1 profile, we do computation in following way:

(49*0.934) + (14.7*0.883) + (6.1*0.78) + (6*0.741) + (4.8*0.65) +(4*0.625) = **73.5**

where 49, 14.7, 6.1, 6, 4.8 and 4 are the frequencies of each single 6 patterns of GR3 respectively. And 0.934 means 93.4 per cent corresponding to the sum of the frequencies (in

per centage) of the DWE1 profiles for the 1st pattern (53+19.2+17.4+3.8). Or, 0.883 means 88.3 per cent corresponding to the sum of frequencies (in per centage) of the DWE1 profiles for the 2nd pattern (33.6+32.5+22.2), and so on. At the end, through the selected DWE1 profiles based on the 6 patterns of GR3 in CAP112, we reach to gather information for 73.5 per cent of the population of GR3 in the full data. We don't necessarily obtain 100 per cent of information. As we know due to variety of patterns, it is quite complex to capture full information. In this regard, the selected DWE1, MAT1, MAT2 profiles bring a quite good amount of information, at least 70 per cent. The DWE2, and ENV profiles reach at least 65 per cent of information.

Concerning DWE1 profile throughout the 6 patterns, we can say that no deprivation regarding simultaneously the 6 components is the most relevant combination covering 35.6 per cent of GR3. After that, being deprived only regarding roof, window, wall conditions (ROWW) comes by affecting 19 per cent of GR3 and being deprived only regarding the number of rooms available per person in the dwelling (RO_D) emerges by covering 15.5 per cent of GR3. And 1.8 per cent of GR3 lacks a toilet in their dwelling. Moreover, 73.5 per cent of GR3 have a kitchen and a bathroom/shower in their dwelling; these are the 2 functionings with the highest proportions of non-deprivation compared to other 3 items of DWE1.

On the other hand, concerning DWE2 profile according to the collected information based on the 6 patterns for GR3, we can say that the most relevant combination is being deprived only regarding heating system (HEATS) and energy source for heating (ENG_H) in the dwelling, affecting 36.5 per cent of GR3. Besides, being deprived regarding HEATS, ENG_H, and insulation state (INS) of the dwelling covers 12.4 per cent of GR3; being deprived regarding HEATS, ENG_H, and lighting state of the dwelling (LIGHT) includes 4.7 per cent of GR3; being deprived regarding HEATS, ENG_H, and hot water system (HWS) concerns 5.4 per cent of GR3. Moreover, lacking an appropriate heating system touches 60.7 per cent of GR3, lacking an appropriate energy source for heating affects 59 per cent of GR3, living in a not adequate insulation based dwelling leading to heating problems concerns 16.55 per cent of GR3, living in a dwelling with lighting problem concerns 6.35 per cent of GR3, but also 5.4 per cent of GR3 lives in a dwelling with being deprived regarding hot water system. In addition, everyone detected via the 6 patterns in GR3 have no deprivation regarding piped water system (PWS).

Regarding ENV profile, 59.2 per cent of GR3 detected through the 6 patterns don't have any deprivation in 3 items within that dimension. In addition, 7 per cent of GR3 suffer air pollution, and 1.4 per cent of GR3 suffer noise pollution. On the other hand, all persons of GR3 throughout the 6 patterns don't suffer regarding crime or violence in their community, neighborhood where they live.

When it comes to materials, in terms of MAT1 dimension, 35 per cent of GR3 don't have a computer and internet, and a car, and 15 per cent of GR3 don't have a telephone, a computer and internet, and a car. Throughout all combinations, we can observe that in total 50.2 per cent of GR3 don't have a computer and internet, 54 per cent of GR3 don't have a car, and 30 per cent of GR3 don't have a landline telephone in their dwelling. On the other hand, everyone detected via the 6 patterns among GR3 do not experience deprivation regarding mobile phone and television. Concerning the MAT2 dimension, all people of GR3 captured by the 6 patterns don't suffer regarding washing machine and refrigerator. But 74 per cent of GR3 don't have an air-condition, 54 per cent of GR3 don't have a dishwasher, and 49 per cent of GR3 don't have both an air-condition and a dishwasher.

Based on these results, we can say again that heating related problems including heating system, insulation state, main energy source available for heating, and available hot water system in the dwelling is the main problem to be solved for GR3 in order to improve their living conditions. In addition to that, roof, wall, window conditions including lighting problem and available room number by person in the dwelling should be checked simultaneously for making better GR3 people's living standards in their dwelling. Lacking a landline telephone may not be considered at first stage as the use of mobile phone is very high. Nevertheless, government may implement programmes for increasing the possession of dishwasher, computer and internet, which in turn may contribute to sustainable development for avoiding excess use of water, and to increase communication capabilities given that internet has become a fundamental element of life for reaching to information both at national, and international level. Depending on geographical context, specific programs for increasing the possession of car might be considered however it should take in consideration possible air pollution that can seriously affect health of people and in particular agricultural works; so that, improving public transports seems a better option. Deprivation regarding air-condition appears also as one of the most important material deprivations; however, it is quite a delicate issue because air-condition requires a substantial amount of use of electricity as well as the addition of air-condition at external side of buildings may cause technical problems reducing solidity of building given that many parts of the country risk violent earthquakes.

Regarding educational attainment, there is in total 4,836 persons found deprived in that aspect of life based on the 6 patterns of GR3, which corresponds to 66.6 per cent of total population of GR3. Among these 4,836 persons, 59.8 per cent comes from the first pattern, 17.3 per cent are from the second pattern, 6.22 per cent are from the third pattern, 7.6 per cent are from the fourth pattern, 4.5 per cent are from the fifth pattern, and another 4.5 per cent are from the sixth pattern. In addition, there are 701 persons from GR3 who are not detected by the 6 patterns being deprived regarding educational attainment-which corresponds to 9.65 per cent of total population of GR3. The sum of 4,836 with 701 leads to the fact that in total 76.2 per cent of GR3 population is deprived regarding educational attainment.

In health dimension, there are two indicators: access to medical services (MED), and access to dental services (DEN). When we look at their combinations, we find that 91.7 per cent of GR3 detected via the first pattern do not have deprivation regarding the two aspects, and 4.8 per cent of them experience deprivation regarding MED. Throughout the 6 patterns, we can observe that in total 75 per cent of GR3 population do not experience deprivation regarding both MED and DEN.

And, concerning the financial situation of the household (FS1), 93 per cent of GR3 from the first pattern, 94 per cent of GR3 from the second pattern, 96.6 per cent of GR3 from the third pattern, 96 per cent of GR3 from the fourth pattern, 96.25 per cent of GR3 from the fifth pattern, and 99.6 per cent of GR3 from the sixth pattern have difficulties to afford financially their monthly needs.

These findings show that financial difficulties emerge as the biggest issue amongst all items included in dimensional analysis. In GR4, and GR5 dimensional analysis, we will see how the magnitude of FS1 varies. Besides, educational attainment has also not brilliant results; based on the 6 patterns we can say that around 7 of 10 persons from GR3 don't have a sufficiently enough or obligatory education. Concerning health dimension, the results are positive in the

sense that capabilities deficit in these two items emerge relatively less among GR3; however, it is important to get information on whether the interviewee is conscient of importance of visiting dentist and doctor when he or she needs or also for a general routine control in year.

Table 55. Dimensional profiles of GR3 based on the CAP112 model, data: Turkey SILC 2011

Pattern number of GR3 (Freq in % in full data)	DWE1 profile (Freq in % in full data)	DWE2 profile (Freq in % in full data)	ENV profile (Freq in % in full data)	MAT1 profile (Freq in % in full data)	MAT2 profile (Freq in % in full data)
1 (49)	00000 (53), 00100 (19.2), 10000 (17.4), 00001 (3.8)	110000 (58), 111000 (13.4), 110010 (11), 110001 (6.6), 100000 (3.3),	000 (82), 010 (10), 100 (2.9)	00011 (25), 10011 (16.8), 00001 (14.15), 10010 (11), 00010 (11), 00000 (8.5), 10001 (7.7), 10000 (3.6)	0011 (88.2), 0010 (10.6)
2 (14.7)	00000 (33.6), 10000 (32.5), 00100 (22.2)	110000 (45.2), 111000 (25), 110001 (10.2)	000 (70.25), 010 (14)	00011 (38.7), 10011 (29), 10010 (16), 00010 (14)	0001 (84.1), 0000 (15.7)
3 (6.1)	00100 (78),	111000 (36)	000 (54),	00001 (47.2), 00000 (23.7)	0001 (84.65),
4 (6)	10000 (37.6), 00000 (36.5)	000000 (30.5)	000 (64)	10011 (46.4), 00011 (30)	0011 (94)
5 (4.8)	00100 (36), 10100 (29)	001000 (52), 001001 (33.7)	000 (33.4)	00001 (39)	0001 (86.2)
6 (4)	00000 (62.5)	110000 (37.2)	-	00001 (55), 10001 (41.2)	0001 (91.2)
Total info/GR4 population coverage in %	(49*0.53) + (49*0.192) + (49*0.174) + (49*0.038) + (14.7*0.336) + (14.7*0.325) + (14.7*0.222) + (6.1*0.78) + (6*0.376) + (6*0.365) + (4.8*0.36) + (4.8*0.29) + (4*0.625) = **73.5**	(49*0.558) + (49*0.134) + (49*0.11) + (49*0.066) + (49*0.033) + (14.7*0.452) + (14.7*0.25) + (14.7*0.102) + (6.1*0.36) + (6*0.305) + (4.8*0.52) + (4.8*0.337) + (4*0.372) = **65.6**	(49*0.82) + (49*0.1) + (49*0.029) + (14.7*0.7025) + (14.7*0.14) + (6.1*0.54) + (6*0.64) + (4.8*0.334) = **67.7**	(49*0.25) + (49*0.168) + (49*0.1415) + (49*0.11) + (49*0.11) + (49*0.085) + (49*0.077) + (49*0.036) + (14.7*0.387) + (14.7*0.29) + (14.7*0.16) + (14.7*0.14) + (6.1*0.472) + (6.1*0.464) + (6*0.30) + (4.8*0.39) + (4*0.55) + (4*0.412) = **75.5**	(49*0.882) + (49*0.106) + (14.7*0.841) + (14.7*0.157) + (6.1*0.846) + (6*0.94) + (4.8*0.862) + (4*0.912) = **81.6**

GR4 dimensional profiles based on CAP113 model

With this model, we saw that the following 4 patterns capture 93 per cent of population of GR4 in the training data: (1) deprivation regarding HEATS, and DISHW; (2) no deprivation regarding DISHW, but deprivation regarding HEATS, and COMPINT; (3) no deprivation regarding DISHW, and COMPINT, but deprivation regarding HEATS, and INS; (4) no deprivation regarding HEATS, but deprivation regarding DISHW, and COMPINT. When we applied these 4 classification rules in the full data, we lose 7 per cent of information reaching to 93 per cent (as in the case of training data), which shows that the model predicts very well GR4. We note that, in GHS analysis, we took consideration only profiles concerning at least 5 per cent of the pattern population; nevertheless, in case of GR4 here (as we did also in GR3 dimensional analysis), in addition to the first two patterns we use two other patterns contributing 4.43, and 4 per cent of information respectively in order to gather higher information at dimensional level for GR4.

Regarding DWE1 dimension, throughout the 4 extracted patterns of GR4, the most relevant four combinations of DWE1 functionings are as follows: (i) deprivation regarding roof, window, wall conditions in the dwelling and non-deprivation in other items (31 per cent of GR4); (ii) no deprivation regarding all items of GR4 at the same time (22 per cent of GR4); (iii) deprivation regarding number of room available per person in the dwelling and non-deprivation in other items (12 per cent of GR4), and (iv) deprivation regarding roof, window, wall conditions in the dwelling, as well as less than one room per person in the dwelling (7.9 per cent of GR4). Moreover, 6.8 per cent of GR4 are deprived regarding toilet facilities in the dwelling, and in total 39.5 per cent of GR3 do not live in an appropriate dwelling in terms of roof, wall, and window conditions, and 20 per cent of GR3 do live in a dwelling with less than one room available per person. If we think that in the first and second patterns, people lack also an appropriate heating system in their dwelling, so that we can say that 39.5 per cent of GR4 lack both an appropriate heating system and do live in a dwelling with bad roof, wall, and window conditions. But also, 7.9 per cent of GR4 live in a dwelling lacking an appropriate heating system, good roof, wall, and window conditions, and in a dwelling with less than one room available per person. Similar to GR3 situation, here also importance should be given to roof, wall, window conditions, and available room per person in the dwelling. Nevertheless, we can see that compared to GR3, access to adequate sanitation (toilet) facilities becomes more problematic for GR4 touching around 7 per cent of them.

Regarding DWE2 dimension, the most frequent combination of DWE2 functionings is living in a dwelling with not appropriate heating system, not a proper source of energy for heating, and not a proper insulation state (and no deprivation in other items) concerning 21.45 per cent of GR4. After that, it comes living in a dwelling with not appropriate heating system and not a proper source of energy for heating (and no deprivation in other items), which affects 20.76 per cent of GR4. Other combinations, living in a dwelling lacking an appropriate heating system and energy source for heating and hot water system in the dwelling (and no deprivation in other items) concerns 12.7 per cent of GR4. Living in a dwelling lacking an appropriate system of heating, energy source for heating, and insulation state with lacking an appropriate lighting system in the dwelling (and no deprivation in other items) touches 11.6 per cent of GR4. Across all combinations discovered through the 4 patterns mentioned in the table below, we can observe that major problem for GR4 regarding DWE2 dimension is experiencing at the same time not adequate heating system and source of energy for heating in their dwelling (79 per cent of GR4). Subsequently, living in a dwelling with heating problems due the insulation state of

the dwelling comes with 36 per cent. But also, 19.6 per cent of GR4 don't have a hot water system in their dwelling, and 19 per cent of GR4 live in a dwelling with lighting problems. Based on these results, we can say that heating problems due to especially heating system and energy source for heating available in the dwelling is the major concern for that group regarding DWE2 dimension, which is worse compared to GR3. Moreover, insulation state of the dwelling, hot water system in the dwelling, and lighting capacity of the dwelling are other deficits affecting capabilities of people in GR4. It is important to mention that deprivation regarding hot water system, and insulation state of the dwelling emerge substantially here compared to GR3, but also people living in a dwelling with lighting problems is almost multiplied by 3 among GR4 compared to GR3.

When it comes to environment of where the dwelling is located, 59.86 per cent of GR4 live in a community, neighborhood with no problems in terms of noise, air pollution, crime and violence. Nevertheless, 11 per cent of GR4 live in an environment witnessing air pollution (and no deprivation in other items). When all combinations of ENV revealed in the table below based on the 4 patterns are considered, we can say that in total 13 per cent of GR4 are deprived regarding air pollution, and 5.5 per cent of GR4 experience noise pollution.

In terms of material deprivation, we start off by the dimension MAT1. Regarding that aspect, we can see that the most encountered combination is not having a telephone, computer and internet, and car in the household (and no deprivation in other items), which corresponds to 33.86 per cent of GR4. Subsequently, lacking both a car and a computer and internet in the dwelling (and no deprivation in other items) comes with 19 per cent among GR4; lacking a car (and no deprivation in other items) concerns 8 per cent of GR4; lacking a landline telephone and a computer with internet (and no deprivation in other items) concerns 8.6 per cent of GR4; and lacking a telephone and a car (and no deprivation in other items) includes 4 per cent of GR4. When we take in consideration all combinations mentioned in the table below, we arrive at the following results: 69.75 per cent of GR4 don't have a computer and internet in their dwelling; 68.8 per cent of GR4 don't have a car in their dwelling; and 46.5 per cent of GR4 don't have a telephone in their dwelling. When it comes to mobile phone and television, deprivation flourishes at very low level, or almost no deprivation. On the other hand, for the dimension MAT2, 68.3 per cent of GR4 don't have both a dishwasher and air conditioner (And no deprivation in other items), and 15 per cent lacks only an air conditioner. Throughout all combinations, we can say that in total 74 per cent don't have a dishwasher, and 85 per cent don't have an air conditioner. The per centage of GR4 lacking a washing machine and/or a refrigerator is very low, or almost zero. Compared to GR3, the content of material deprivation types is similar, but the magnitude of deprivation is higher among GR4.

Concerning educational attainment, among GR4 detected through the 4 patterns 87.75 per cent do not have a minimum necessary education grade (at least 12 grades as determined by the government across the country). Among them, 78.7 per cent comes from the first pattern, 12.7 per cent are from the second pattern, 4.3 per cent are from the third pattern, and 4.3 per cent are from the fourth pattern. We note that GR4's educational attainment capabilities is worser than those of GR3.

Based on the 4 patterns, the capabilities of the households of GR4 in terms of access to medical and dental services are found as follows: 72 per cent of GR4 don't lack access to both medical and dental services when it was needed; 10 per cent of GR4 lack access to medical services,

and 3.5 per cent of GR4 lack both two services. These results show quite good performance in terms of access to medical and dental services for GR4 as well as GR3 as we have seen previously. However, deprivation for GR4 is slightly higher than GR3.

And, concerning monthly financial situation of the households of GR4, we can observe that 97.6 per cent of them suffer deprivation regarding that issue. The situation for GR4 is slightly higher than GR3, about 2-2.5 per centage points.

Table 56. Dimensional profiles of GR4 based on the CAP113 model, data: Turkey SILC 2011

Pattern number of GR4 (Freq in % in full data)	DWE1 profile (Freq in % in full data)	DWE2 profile (Freq in % in full data)	ENV profile (Freq in % in full data)	MAT1 profile (Freq in % in full data)	MAT2 profile (Freq in % in full data)
1 (72)	00100 (33.5), 00000 (27.3), 10000 (14.4), 10100 (8), 00001 (5.25), 00101 (4.2)	110000 (25.3), 111000 (23.8), 110010 (17.6), 111001 (10.5), 110001 (7.7), 111010 (6.7), 110011 (2.8)	000 (73.8), 010 (12.3), 100 (4.5), 110 (3), 001 (2.8)	10011 (36.4), 00011 (21.8), 10010 (9.8), 00001 (8.7), 00010 (6.6), 10001 (5.6), 01011 (4.8), 00000 (2.5)	0011 (90.3), 0010 (6.13), 1011 (2.8)
2 (12.4)	00100 (35.9), 00000 (19), 10100 (17.2), 10000 (16)	111000 (34.8), 110000 (20.5), 111001 (17.5)	000 (54.3), 010 (17.7)	10011 (46.6), 00011 (27.5), 10010 (12.24)	0001 (88.6)
3 (4.5)	00100 (59)	110001 (42.6)	-	00001 (46.2)	0001 (90.2)
4 (4)	-	-	-	10011 (47)	0011 (81.8)
Total info/GR4 population coverage in %	(72*0.926) + (12.4*0.881) + (4.5*0.59) = **80**	(72*0.944) + (12.4*0.728) + (4.5*0.426) = **79**	(72*0.964) + (12.4*0.72) = **78.3**	(72.9*0.962) + (12.4*0.863) + (4.5*0.462) + (4*0.47) = **84.8**	(72.9*0.992= + (12.4*0.886) + (4.5*0.902) + (4*0.818) = **90.6**

GR5 dimensional profiles based on CAP114 model

With this model, we saw that the following 3 patterns capture 96 per cent of population of GR5 in the training data: (1) deprivation regarding HEATS, DISHW, and COMPINT; (2) no deprivation regarding COMPINT, but deprivation regarding HEATS, DISHW, and ROWW; (3) no deprivation regarding DISHW, but deprivation regarding INS, HEATS, and CAR. When we applied these 3 classification rules in the full data, we lose only 4 per cent of information reaching to 96 per cent (as in the case of training data), which shows that the model predicts very well GR5.

Concerning DWE1 dimension, 31.9 per cent of GR5 has a deprivation regarding roof, wall, and window conditions of their dwelling and no deprivation regarding the other 4 items of DWE1. Subsequently, 10.6 per cent of GR5 has deprivation regarding both roof, wall, window conditions of their dwelling and number of available rooms per person in the dwelling, and no deprivation regarding 3 other items of DWE1. Besides, 10.5 per cent of GR5 has deprivation regarding both roof, wall, window conditions of their dwelling and sanitation (toilet) facilities in the dwelling, and no deprivation regarding the other 3 items if DWE1. When we look at all combinations revealed in the table below based on the 4 patterns of GR5, we can say that 15.7 per cent of GR5 live in a dwelling with less than one room available per person; 71 per cent of GR5 live in a dwelling with not adequate roof, wall, and window conditions; 32 per cent don't have an adequate sanitation (toilet) facilities in their dwelling; 15.4 per cent don't have an adequate bathroom/shower in their dwelling; and 5.6 per cent live in a dwelling with no kitchen. Compared to GR4, the situation of GR5 has become much worse in terms of roof, wall, window conditions of the dwelling, but especially in terms of sanitation including both toilet and bathroom/shower facilities in the dwelling.

Concerning DWE2 dimension, the following 5 combinations of 6 items of DWE2 are found: (i) deprivation regarding heating system of the dwelling, energy source for heating (HEATS), insulation state of the dwelling (INS), and hot water system (HWS), and no deprivation regarding piped water system (PWS), and lighting state of the dwelling (LIGHT) (17 per cent of GR5); (ii) deprivation regarding HEATS, ENG_H, INS, HWS, and LIGHT, and no deprivation regarding PWS (15.25 per cent of GR5); (iii) deprivation regarding HEATS, ENG_H, INS, and LIGHT, and no deprivation regarding PWS, and HWS (17 per cent of GR5); (iv) deprivation regarding HEATS, ENG_H, and INS, but no deprivation regarding PWS, HWS, and LIGHT (12 per cent of GR5); (v) deprivation regarding HEATS, ENG_H, and HWS, and no deprivation regarding INS, PWS, and LIGHT (8.4 per cent of GR5). Moreover, we can also observe that across all combinations shown in the table below: 89 per cent of GR5 live in a dwelling with no adequate heating system and energy source for heating neither; 68.5 per cent live in a dwelling with heating problems due to insulation state of the dwelling; 9.26 per cent lack piped water system in their dwelling; 54 per cent lack hot water system; and 41 per cent experience lighting problems in the dwelling. As it was the case for GR4, here also we can see that for GR5 heating system, energy source for heating, and insulation state of the dwelling are the three main issues. However, we can also see that the capabilities of GR5 become more restrained in terms of hot water system (2.76 times higher compared to GR4), lighting state of the dwelling (2 times higher than GR4), and piped water system in the dwelling respectively.

Regarding the environment features of community, neighborhood where the household lives, 55 per cent of GR5 don't experience deprivation regarding the 3 items of ENV. However, 24 per cent of GR5 suffer air pollution, 13 per cent of GR5 suffer noise pollution, and 8.4 per cent of GR5 encounter violence, crime related events in their neighborhood, community. Compared to GR4, deprivation issues regarding air pollution, and noise pollution are almost doubled in case of GR5. And violence and crime related environment deprivation surfaces with a slightly important per centage in case of GR5 compared to other groups.

When we look at the dimension MAT1, 56.7 per cent of GR5 live with a combination of deprivation regarding telephone, computer and internet, and car, and no deprivation regarding mobile phone, and tv. Across all combinations shown in the table below, we can also say that 70.73 per cent of GR5 don't have a telephone; 86.73 per cent of GR5 don't have a computer

and internet; 86.5 per cent don't have a car; 10.7 per cent of GR5 don't have a mobile phone; and 3.5 per cent don't have a tv in their dwelling. Compared to GR4, the situation for GR5 is worse concerning all items of MAT1, especially rates of not having a telephone, computer and internet, car is very high. And 5.4 per cent of GR5 lack simultaneously a telephone and a mobile phone in their dwelling. On the other hand, concerning the dimension MAT2, 65.7 per cent of GR5 live with a combination of deprivation regarding dishwasher and air conditioner, and no deprivation regarding washing machine, and refrigerator. Besides, 16 per cent of GR5 don't possess a washing machine, dishwasher, and air conditioner, and 5 per cent of GR5 live in a complete deprivation regarding the 4 items of DWE2. If take into account all combinations appearing in the table below, we can say that 91 per cent of GR5 lack an air conditioner, 87 per cent of GR5 don't have a dishwasher, 21.4 per cent of GR5 don't possess a washing machine. Compared to GR4, the deprivation situation of GR5 regarding air conditioner, and dishwasher is getting worse, and even a considerably high per centage of lacking a washing machine appears.

Concerning educational attainment, 94 per cent of GR5 suffers. It is quite high amount. We can say that across GR3, GR4, and GR5, 8 of 10 persons on average don't have an obligatory educational attainment, grade 12, defined by the government.

Moreover, regarding the access to medical and dental services, 54 per cent don't experience any deprivation. However, 19 per cent of GR5 have deprivation regarding both the two items of health dimension; 17 per cent of GR5 have experienced deprivation regarding medical services and no deprivation regarding dental services. Health related statistics are appeared relatively good in case of GR3 and GR4 as we have seen previously; but we can see that At least 3 of 10 persons in GR5 have experienced an accessibility problem to medical services.

And, concerning the monthly financial situation of the household, statistics show that 99 per cent of GR5 revealed in the table below by the 3 patterns have experienced a deficit on that issue. This functioning is extremely important for GR3, GR4, and GR5 in which the rates are all above 95 per cent.

Table 57. Dimensional profiles of GR5 based on the CAP114 model, data: Turkey SILC 2011

Pattern number of GR5 (Freq in % in full data)	DWE1 profile (Freq in % in full data)	DWE2 profile (Freq in % in full data)	ENV profile (Freq in % in full data)	MAT1 profile (Freq in % in full data)	MAT2 profile (Freq in % in full data)
1 (86.2)	00100 (31.4), 00101 (12.2), 10100 (10.9), 00111 (9.2), 00000 (6.7), 01111 (6.5), 10000 (3.77), 01101 (3), 00001 (2.3), 00011 (2.2), 10101 (2.2)	111010 (20), 111011 (17.7), 111001 (15.7), 111000 (12.6), 110010 (9.75), 111110 (5.3), 110011 (4.4), 110000 (4.2), 110001 (3), 110110 (2.85), 111111 (2.6)	000 (60.4), 010 (16.8), 110 (6.2), 100 (5.5), 111 (3.75), 011 (3.16), 001 (2.8)	10011 (63), 00011 (15), 01011 (6.14), 10010 (5.3), 11011 (4.3), 10111 (2.11), 11111 (2)	0011 (70.5), 1011 (18.8), 1111 (6), 0010 (2.5)

2 (5.4)	00100 (54.8)	111001 (30)	000 (31.4), 010 (21.8)	10001 (41.3), 00001 (39.2)	0011 (90.8)
3 (4.6)	00100 (40.6), 10100 (25.7)	111001 (42.9), 111000 (26.6)	000 (26.6)	10011 (51.8)	0001 (88.4)
Total info/GR5 population coverage in %	(86.2*0.904) + (5.4*0.548) + (4.6*0.663) = **84**	(86.2*0.981) + (5.4*0.30) + (4.6*0.695) = **89.4**	(86.2*0.986) + (5.4*0.532) + (4.6*0.266) = **89**	(86.2*0.978) + (5.4*0.805) + (4.6*0.518) = **91**	(86.2*0.978) + (5.4*0.908) + (4.6*0.884) = **93**

Third step (step3 analysis): detecting various profiles of poverty groups

In this sub-section, we will proceed with various sub analysis in order to examine in more detail the extent of capabilities of GR3, GR4, and GR5 in the following aspects respectively: educational attainment, financial situation of the household, economic activity status, and geographical situation.

GR3 sub-group profiles based on CAP112 model

In terms of educational attainment, we already saw in dimensional analysis that around 7 of 10 persons in GR3, GR4, and GR5 don't have an obligatory grade 12 level diploma or equivalent. An adequate educational attainment constitutes a vital functioning for people for leading a pleasant life or at least such functioning is more likely to increase the quality of life of people for several aspects such as earning more, having a better job regarding working conditions and status, having a rich vision for life, etc. Of course, it is also important to note some limits of such variable. The variable we use to measure educational attainment of persons comes from the question of what the highest degree/grade is the person has achieved. With this question, we can only know the grade of the person. Education given at school, at university are very important but there is also an education given by parents to child outside of school notably when child is growing up (childhood and teenage years). Technical learning is not enough to have a complete vision for life, or for being social with community, with people in large sense, or for being aware of the nature to preserve it and for respecting animals for instance. Unfortunately, we don't have a more detailed information about educational attainment except the grade information. We will assume that grade or diploma also plays an important role notably in the labour market for increasing living conditions. In that vein, we recommend that highest educational attainment will bring a higher technical level skills and knowledge, and possibilities of enriching life standards. We also know that in developing countries, or even in developed countries, not every person graduated with a university diploma finds the job the persons desires or the job that is convenient according to her/his technical knowledge and skills. So that, the added value of university or vocational technical school diplomas is another debate, which we won't enter here. This is an issue that governments should consider for creating new jobs or enabling sustainable good living conditions to their citizens. We will just focus on the distribution of type of diploma people have achieved among GR3, GR4, and GR5 in order to have a more detailed view on achieved highest grades.

In the data, there are in total 7,263 individuals from GR3. Among them, 14.7 per cent is found to have graduated from primary school, 6 per cent from secondary, vocational secondary or primary education school, 15 per cent are illiterate, 10 per cent have a high school diploma, 49.2 per cent are literate but without a diploma, 4.8 are graduated from vocational or technical high school, and 4 per cent have a higher education diploma such university or equivalent. In this schema, we can say that the rate of illiterate people is quite important, and notably the rate of higher education graduates is very low within the group.

In dimensional analysis, we saw that a significant number of people in GR3 live in a household financially suffering in the sense that the household's monthly living status regarding the total household income is not easy. If we look at into detail, we can observe that 35.7 per cent of GR3 live in a household, which tries to manage with difficulty the monthly consumption or in other words these people find difficult to live with their total household income to afford their monthly needs. In addition, 30.6 per cent of GR3 find a bit hard, 16 per cent perceive very difficult, and 12 per cent think it is a bit easy to afford their monthly consumption with their total monthly household income. On the other hand, only 5 per cent find it is easy, and 0.5 per cent says it is very easy their monthly living status in their household. We can see that in total 82 per cent of households find very difficult, or difficult, or a bit hard their monthly living status. Besides a relatively high rate of unfulfillment of minimum necessary educational attainment, monthly financial situation of individuals should be considered as one of the top priority elements to reduce poverty among GR3 across the country. In addition to that global monthly aspect of financial situation of the households, there are other 9 indicators, namely, FS2, FS3, FS4, FS5, FS6, FS7, FS8, FS9, and FS10. The first three ones concern various arrears as follows:

FS2: Arrears on mortgage, loan repayments or rent payments in the last 12 months?
FS3: Arrears on utility bills (house electricity, water and gas invoices) in the last 12 months?
FS4: Arrears on hire purchase instalments, credit cards or other loan payments in the last 12 months?

When we take in consideration the 5 relevant patterns of GR3 employed in dimensional analysis, we arrive at the following results in terms of combinations of the 3 arrears related variables: 44 per cent of GR3 did not have deprivation regarding the 3 types of arrears at the same time, 13 per cent of GR3 had deprivation regarding arrears on utility, and hire purchase instalments, credit cards or other loan payments and no deprivation regarding FS2, and 9.7 per cent of GR3 had deprivation regarding utility bills and no deprivation regarding FS2 and FS3. Across all relevant combinations (with at least 100 observations each of them) of GR3 based on the 6 patterns, we can say that 24.7 per cent of GR3 had experienced arrears on utility bills, and 18.45 per cent of GR3 had experienced arrears on hire purchase instalments, credit cards, or other loan payments. On the other hand, experiencing arrears on mortgage, loan repayments or rent payments flourished relatively low, under 2 per cent, among GR3 based on the 6 patterns.

Moreover, 17.5 per cent of GR3 lack an appropriate heating system in their dwelling and they have faced both FS3 and FS4 types of arrears; 13 per cent of GR3 lack an appropriate heating system in their dwelling and they have faced FS3 type arrear; 13.4 per cent of GR3 don't have a dishwasher in their dwelling and they have faced both FS3 and FS4 types of arrears; and 9.8

per cent of GR3 don't have a dishwasher in their dwelling and they have faced FS3 type of arrear. As we have shown previously the variables of heating system, dishwasher, and computer and internet are the 3 most influential deprivation elements among GR3; for those are deprived regarding these items, ability of paying utility bills as well as the rate of other types of bills are important to be considered.

Regarding financial aspect, the other 6 indicators, namely, FS5, FS6, FS7, FS8, FS9, and FS10, are as follows:

Financial state V (FS5)	Is the household able to afford unexpected financial expenses (approximately 400TL)?
Financial state VI (FS6)	Is the household able to afford adequately the heating needs of the dwelling? (ability to keep home adequately warm)
Financial state VII (FS7)	At least 3 days a week, when required or needed, is the household able to afford a meal with meat, chicken, fish? (equivalent food for vegetarians)
Financial state VIII (FS8)	Is the household able to afford new clothes, except secondhand? (replace worn out clothes by some new (not second hand) ones)
Financial state XI (FS9)	Is the household able to renew/replace old or worn out furniture?
Financial state X (FS10)	Is the household able to afford, when required or needed, a one-week annual holiday away from home with all household members? (e.g., hotels, motels, lodgings, summer houses)

Based on these 6 variables, we looked at its combinations (i.e., profiles) per each of 6 relevant patterns of GR3. It is important to note that a large variety of profiles is surfaced, which does not allow to collect a substantial amount of information on the ensemble of GR3. In other words, we reach only capturing 44 per cent of GR3, given that each profile contains at least 100 observations. In this condition, we find that the most frequent profile is the combination of deprivation regarding FS5, FS7, FS9, and FS10, and no deprivation regarding FS6, and FS8 (12.4 per cent of GR3). Subsequently, the combination of deprivation regarding all 6 variables comes with 6 per cent. Moreover, when we take in consideration all combinations, we reach the following statistics: 31.7 per cent of GR3 suffer regarding FS5; 10.4 per cent of GR3 suffer regarding FS6; 25.75 per cent of GR3 suffer regarding FS7; 10.3 per cent of GR3 suffer regarding FS8; 38.76 per cent of GR3 suffer regarding FS9; and 41.63 per cent of GR3 suffer regarding FS10. We can result that deprivation regarding FS6 and FS8 relatively touch less persons compared to the other 4 items; more specifically deprivation in terms of not being able to go annual holiday at least for one week a year with all household members, not being able to buy new furniture, not being able to afford unexpected financial expenses, and not being able to afford a meal with meat, chicken, fish at least 3 days a week when required or needed should be rather taken in consideration as priorities compared to deprivation in terms of not being able to finance heating costs, and buying new clothes. It is important to note that as we have seen deprivation in terms of not being able to afford heating costs did not flourish with high frequency despite, we reach only almost half of GR3 population. Through classification tree

analysis, we saw that deprivation regarding heating system available in the dwelling is one of the most important type of capability failure for GR3, GR4, and GR5. So that we can say that it is possible there are many households in the country giving priority to heating costs especially during the winter from their household budget, and use stove as heating system, and wood, coal, cowpat, or fuel-oil, lpg as source of energy for heating. This is an important issue for the country but also in terms of air pollution, and preservation of nature. To check the situation, we do a cross-tabulation between two variables HEATS and FS6.

72 per cent of GR3 have an ability to afford heating costs vs. 28 per cent of GR3 cannot afford such costs. And as we argued 76.65 per cent of GR3 affording heating costs use a stove as the heating system in their dwelling; but also 73 per cent of GR3 not being able to afford heating costs use a stove. When we cross tabulate ENG_H and FS6, we find that 59 per cent of GR3 households affording heating system use coal as energy source for heating, and 20.36 per cent use wood. But also, 48.6 per cent of GR3 households cannot affording heating costs use coal, 26 per cent use natural gas, which can be considered a fairly good tool, and 17.74 per cent use wood. Among the GR3 households affording heating costs and using a stove as heating system in their dwelling, 64 per cent prefer using coal, and 26.6 per cent prefer using wood. As we can see, it becomes extremely important for government to find solutions for switching from the use of stove to radiator, flat combi, or air condition for heating purpose, but also for switching from the use of wood, coal to natural gas or electricity, given that therefore households will be needing subvention for being able to finance heating costs based on natural gas or electricity, which are expensive compared to other heating energy sources.

As we have seen the great majority of GR3 households have difficulties to afford monthly living expenses. At this point, it becomes crucial to check economic activity related situation of individuals. We will create an employment profile as a combination of the following 4 variables respectively: ES1 (economic activity status), SC (social security), WH (Working hours), and ES3 (employment status). When we look at employment profile of GR3 based on their most relevant 6 patterns, we find that amongst the profiles with at least 50 observations (that we call 'pool'), the most relevant profile is being not economically active with 40 per cent. After that, the profile of being employed, registered to social security, working more than 43 hours per week, and being paid, salaried (regular wageworker/employee) comes with 11.6 per cent. When we consider only the economically active population found in this pool, we can see that 77.5 per cent of them work long hours (more than 43 hours per week), 55.4 per cent of them work without being registered to national social security system. Among whom are not registered to social security, 47.4 per cent are unpaid family worker, 29 per cent are own account worker, 12.8 per cent are casual employee, and 10.77 per cent are paid, salaried. On the other side, 61.84 per cent of whom are not registered to social security are working long hours, 26.6 per cent work between 28-42 hours per week, and 11.56 per cent work less than 27 hours per week. In that vein, we can observe that working long hours and not being registered to social security concerns one quart of economically active population of GR3 detected through the 6 relevant patterns. But in general sense, long hours of working, no registration to social security, and in particular unpaid family workers situation should be considered by government for making decent working conditions.

As we have already explained in GHS analysis, for discovering geographic situation of people from deprivation groups the best way is to cross tabulate the two main variables: REG (region) and SET (setting) for each deprivation group. The combination of these 2 variables produces a

very large variety of profiles, which makes not enable to cover a substantial share of people in the group. However, according to the cross table between REG and SET we can say the following results. Among GR3, 65 per cent live in urban areas whilst the rest 35 per cent live in rural areas. In terms of region, 14.69 per cent of GR3 live in the region of Aegean, 10.5 per cent live in Istanbul, and 10.7 per cent live in Mediterranean. These 3 regions are the most frequented regarding GR3. On the other hand, 4.9 per cent live in East Black Sea, 5.18 per cent live in Northeast Anatolia, 6.10 per cent live in Middle East Anatolia, and 6.68 per cent live in Central Anatolia. In the other 5 regions, per centage varies between 7.5 and 8.9. When we look at the combination of REG and SET, we can observe that the most frequented is urban areas in Istanbul (10.3 per cent), then urban areas in Aegean (8.64 per cent), followed by urban areas in West Anatolia (7.3 per cent), urban areas in Mediterranean (7.3 per cent), urban areas in East Marmara (6.5 per cent), rural areas in Aegean (5.9 per cent), and urban areas in Southeast Anatolia (5.6 per cent). On the other hand, the least frequented locations are rural areas in Istanbul (0.19 per cent), and rural areas in West Anatolia (1.3 per cent). Among who live in rural areas, the most frequented region is Aegean (17 per cent), then West Black Sea (13 per cent), and followed by Mediterranean with 10.6 per cent. The least frequented rural areas among who live in rural areas are Istanbul (0.57 per cent), followed by West Anatolia (3.77 per cent). Among who live in urban areas, the most frequented region is Istanbul with 16 per cent, then Aegean with 13.3 per cent, then West Anatolia with 11.2 per cent, and followed by Mediterranean with 10 per cent. In opposite, the least frequented urban areas among who live in urban areas are East Black Sea with 2.8 per cent, then Northeast Anatolia with 3.86 per cent. These results show that the majority of GR3 live in urban areas, and notably in Istanbul, Aegean, West Anatolia, Mediterranean, and East Marmara. In addition, rural areas in Aegean are also preferred, which is possibly due to rich, fruitful land of this mostly coast region notably in terms of very good weather conditions compared to the East Black Sea, and Northeast Anatolia regions witnessing the hardest winter conditions in the entire country.

GR4 sub-group profiles based on CAP113 model

In terms of educational attainment measured by the highest grade obtained by the person, among 6,736 persons in the data, it is found that 45.6 per cent of them are graduated from primary school, 16.4 per cent are illiterate, 14.6 per cent from secondary, vocational secondary or primary education school, 10 per cent are literate but without a diploma, 7.84 per cent have a high school diploma, 7.75 per cent are literate but without a diploma, 4.2 are graduated from vocational or technical high school, and only 2.3 per cent have a higher education diploma such university or equivalent. Regarding these results, we can say that the rate of illiterate people manifests quite important, and notably the rate of higher education graduates is very low within the group. Compared to GR3, the situation of GR4 is worse in that aspect because notably the rate of higher education graduates decreases sharply and the rate of high school graduates is lower, and the rate of illiteracy is higher.

In dimensional analysis, we saw that a significant number of people in GR4 live in a household they have some difficulties to make ends meet at the end of the month. When we check in detail the distribution of the variable FS1 among GR4, we can observe that 41 per cent of them find difficult to live with the household's total monthly income. But also, 23.6 per cent find very difficult, and 23.9 per cent find a bit hard to live given their household's total monthly income. On the other hand, 9 per cent find a bit easy, 2.24 per cent find easy, and 0.09 per cent almost no one find very easy their monthly living status in their household. Thus, in total 88.6 per cent of households evaluate very difficult, or difficult, or a bit hard their monthly living status. As

we mentioned in GR3 analysis, here also we can say that educational attainment, and household's monthly financial status should be determined as one of the top priorities in poverty reduction programs of government.

To complement the financial aspect, we will see now the situation regarding other 9 indicators, namely, FS2, FS3, FS4, FS5, FS6, FS7, FS8, FS9, and FS10.

When we take in consideration the 4 relevant patterns of GR4 employed in dimensional analysis, we arrive at the following results in terms of combinations of the 3 arrears related variables (FS2, FS3, and FS4): 38 per cent of GR4 don't have deprivation regarding the 3 types of arrears at the same time, 20.5 per cent of GR4 have deprivation regarding arrears on utility, and hire purchase instalments, credit cards or other loan payments, and no deprivation regarding FS2, and 16 per cent of GR4 have deprivation regarding utility bills and no deprivation regarding FS2 and FS3. Moreover, across all relevant combinations (with at least 100 observations each of them) of GR4 based on the 4 patterns, we can say that 42 per cent of GR4 had experienced arrears on utility bills, and 28.3 per cent of GR4 had experienced arrears on hire purchase instalments, credit cards, or other loan payments. On the other hand, experiencing arrears on mortgage, loan repayments or rent payments flourished relatively low, under 4 per cent, among GR4 based on the 4 patterns. We can see that deprivation regarding arrears on mortgage, loan repayments or rent payments is very few compared to the 2 other types of arrears, we remind that this was also the case for GR3; this is probably due to the fact that GR3, GR4 population are not very able to apply and/or use mortgage credits for buying a dwelling and/or people pay attention to pay rent first for keeping their shelter. Furthermore, 24 per cent of GR4 lack an appropriate heating system in their dwelling and they have faced both FS3 and FS4 types of arrears; 16 per cent of GR4 lack an appropriate heating system in their dwelling and they have faced (only) FS3 type arrear; 20.4 per cent of GR4 don't have a dishwasher in their dwelling and they have faced both FS3 and FS4 types of arrears; and 14 per cent of GR4 don't have a dishwasher in their dwelling and they have faced (only) FS3 type of arrear. Compared to GR3, both arrears on utility bills and hire purchase instalments, credit cards or other loan payments are higher among GR4. We note that households suffering notably arrears on utility bills will be probably suffering as well regarding material deprivation such as possessing a dishwasher, or a computer, or even they won't be really able to afford natural gas or electricity as main source of energy or radiator, or combi, or air conditioner as a heating system in their dwelling and therefore they will be probably preferring to use a stove with coal, or wood as these two are cheaper versions of heating inside the dwelling.

Regarding the other 6 indicators of financial aspect, namely, FS5, FS6, FS7, FS8, FS9, and FS10, when we look at its combinations (i.e., profiles) per each of 4 relevant patterns of GR4, we reach only capturing 69 per cent of GR4, given that each profile contains at least 100 observations. The data contains a large variety of profiles, which does not really allow to collect a substantial amount of information on the ensemble of GR4. According to that, the most frequent profile is the combination of deprivation regarding all 6 items (19.4 per cent of GR4). Subsequently, the combination of deprivation regarding FS5, FS7, FS9, and FS10, and no deprivation regarding FS6, and FS8 comes with 15 per cent. Moreover, when we take in consideration all combinations, we reach the following statistics: 58 per cent of GR4 suffer regarding FS5; 27 per cent of GR4 suffer regarding FS6; 50.6 per cent of GR4 suffer regarding FS7; 31 per cent of GR4 suffer regarding FS8; 65.7 per cent of GR4 suffer regarding FS9; and 67.5 per cent of GR4 suffer regarding FS10. Based on these results, we can say that deprivation

regarding FS6 and FS8 relatively touch less persons compared to the other 4 items (the same finding for GR3 as well); more specifically deprivation in terms of not being able to go annual holiday at least for one week a year with all household members, not being able to buy new furniture, not being able to afford unexpected financial expenses, and not being able to afford a meal with meat, chicken, fish at least 3 days a week when required or needed should be rather taken in consideration as priorities compared to deprivation in terms of not being able to finance heating costs, and buying new clothes. As we mentioned in GR3 analysis, we also remark here as well that deprivation in terms of not being able to afford heating costs did not emerge with high frequency given the fact that through classification tree analysis, we saw that deprivation regarding heating system available in the dwelling was found as one of the most important type of capability failure for GR3, GR4, and GR5. When we do a cross-tabulation between two variables HEATS and FS6, we find that 91.5 per cent of GR4 affording heating costs use a stove as the heating system in their dwelling; but also 88.6 per cent of GR4 not being able to afford heating costs use a stove. When we cross tabulate ENG_H and FS6, we find that 59.3 per cent of GR4 households affording heating system use coal as energy source for heating, and 27 per cent use wood. But also, 56 per cent of GR4 households cannot affording heating costs use coal, 23.2 per cent use wood, and 11.8 per cent use natural gas. Among the GR4 households affording heating costs and using a stove as heating system in their dwelling, 60.7 per cent prefer using coal, and 29.4 per cent prefer using wood. And, among the GR4 households cannot affording heating costs and using a stove as a heating system in their dwelling, 61.5 per cent prefer using coal, and 26 per cent prefer using wood.

When it comes to economic activity related situation of individuals, which is an extremely important driver of moving out of extreme poverty and of improving quality of life, we will focus the employment profile composed of the following 4 variables respectively: ES1 (economic activity status), SC (social security), WH (working hours), and ES3 (employment status). We note that it is important to have a job certainly, but it is also very important to work under decent condition such adequate working hours, being registered to social security in order to benefit from the national health insurance system. In line with that, based on the most relevant 4 patterns of GR3, we find that amongst the profiles with at least 50 observations, the most relevant profile is being not economically active with 41 per cent. After that, the profile of being employed, registered to social security, working more than 43 hours per week, and being paid, salaried (regular wageworker/employee) comes with 10 per cent. There is actually found 34.8 per cent of GR4 being economically active persons in this pool (corresponding to 2,373 people), and there is a variety of profiles within. There is also 12.7 per cent of GR4 economically active population not detected through the 4 relevant patterns. When we consider only the economically active population found in this pool, we can see that 73 per cent of them work long hours (more than 43 hours per week), 82.5 per cent of them work without being registered to national social security system. Among whom are not registered to social security, 43 per cent are unpaid family worker, 33 per cent are own account worker, 11 per cent are casual employee, and 13 per cent are paid, salaried. On the other side, 63 per cent of who are not registered to social security are working long hours, 22.7 per cent work between 28-42 hours per week, and 14 per cent work less than 27 hours per week. In these conditions, we can observe that working long hours and not being registered to social security concerns dramatically GR4, worse than GR3 situation.

Now, we will focus on geographical situation composed of 2 variables: REG and SET. Among GR4, 55.6 per cent live in urban areas whilst the rest 44.4 per cent live in rural areas. We can

see that compared to GR3, the population living in rural areas is relatively higher in case of GR4. In terms of region, 12.7 per cent of GR4 live in the region of Aegean, 11.45 per cent live in Mediterranean, and 10.38 per cent live in Southeast Anatolia. These 3 regions are the most frequented regarding GR4. On the other hand, 5.55 per cent live in East Black Sea, 6.22 per cent live in West Marmara, 6.64 per cent live in Northeast Anatolia, and 6.70 per cent live in West Anatolia. In the other 5 regions, per centage varies between 7 and 9.14. In case of GR3, the region of Istanbul was the second most frequented one; however, here we can see that notable Southeast Anatolia manifests as the third most frequented region. When we look at the combination of REG and SET, we can observe that the most frequented location is urban areas in Istanbul (7.25 per cent), then urban areas in Mediterranean (6.8 per cent), followed by urban areas in Southeast Anatolia (6.7 per cent), urban areas in Aegean (6.6 per cent), rural areas in Aegean (6 per cent). On the other hand, the least frequented locations are rural areas in Istanbul (0.11 per cent), urban areas in East Black Sea (1.6 per cent), rural areas in West Anatolia (1.9 per cent), urban areas in Northeast Anatolia (2.4 per cent), and rural areas in East Marmara (2.4 per cent). Among who live in rural areas, the most frequented region is Aegean (13.7 per cent), then followed by Middle East Anatolia (12.10 per cent), West Black Sea (11.16 per cent), and Mediterranean with 10.6 per cent. The least frequented rural areas (among who live in rural areas) are Istanbul (0.26 per cent), followed by West Anatolia (4.23 per cent), and East Marmara (5.43 per cent). Among the persons living in urban areas, the most frequented region is Istanbul with 13 per cent, then Mediterranean with 12.34 per cent, then Southeast Anatolia with 12.1 per cent, and followed by Aegean with 11.9 per cent.

In opposite, the least frequented urban areas (among who live in urban areas) are East Black Sea with 2.8 per cent, then Northeast Anatolia with 4.31 per cent, and West Marmara with 5.12 per cent. These results show that there is not a significant difference between the per centage of GR4 living in urban and rural areas. But we can say that Istanbul is the most preferred location for living an urban life, and Aegean is for living rural life. Compared to GR3, the population in Mediterranean is higher here, and those living Aegean is less among GR4. And, most importantly, the region of Southeast Anatolia emerges among GR4 as one of the top frequented regions for an urban setting.

The cross table between the variables EDU and SET for GR1 shows that 91 per cent of GR1 live in urban areas. For those are living in urban areas, 32 per cent have a higher education diploma (university or equivalent), 22 per cent are primary school graduate, 15 per cent are vocational or technical high school graduate, 14 per cent are high school graduate. The rates of illiterate and literate without a diploma are very few. However, we can remark that the percentage of primary school graduate is quite important, the gap between this rate and the rate of higher education graduate is only about 10 per centage points. On the other hand, concerning GR4, 55.6 per cent of GR4 live in urban areas vs. 44.4 per cent of GR4 live in rural areas. So that, the gap between urban and rural areas in terms of population living there is only about 11.2 per centage points, which is very low compared to GR1. Among GR4 living in urban areas, 48 per cent are primary school graduate, 16 per cent are secondary, vocational secondary or primary education school graduate, 12 per cent are illiterate, 9 per cent are literate without a diploma, 9.6 per cent are high school graduate, and only 8 per cent are vocational or vocational high school (5 per cent), or university or equivalent (3 per cent) graduate. The rate of higher education graduate (university or equivalent) is significantly low compared to GR1 where the higher education takes the top position as a most achieved educational degree. But also, the per centage of illiterate, or the total of who don't have an obligatory educational attainment is about

82.4 per cent, which can be considered quite something (in opposition to 38 per cent of GR1 living in urban areas don't have an obligatory educational attainment). Concerning GR4 people living in rural areas, almost 90 per cent don't have an obligatory educational attainment, and the rate of university or equivalent graduate persons is only 1.7 per cent.

In terms of region, 23.4 per cent of GR1 live in Istanbul, 15.65 per cent live in Aegean, 15 per cent live in West Anatolia, 9 per cent live in East Marmara, and 9 per cent live in Mediterranean. Especially, in East Black Sea, Northeast Anatolia, Southeast Anatolia, West Black Sea, and Middle East Anatolia, the per centage of the least poor GR1 people is very few. So that, the majority of GR1 live in coast regions or around Istanbul in the region of Marmara, or at the West Anatolia part of the country. The region where university or equivalent degree graduates live most is Istanbul, Aegean, and West Anatolia. 26 per cent of those living in Istanbul from GR1, 35 per cent of Aegean, and 35 per cent of West Anatolia are university or equivalent degree graduate. In the regions of East Black Sea, Southeast Anatolia, and Northeast Anatolia, the rate of university or equivalent degree graduate is very low. On the other hand, 12.7 per cent of GR4 live in Aegean, 11.45 per cent live in Mediterranean, 10 per cent live in Southeast Anatolia, 7.3 per cent live in Istanbul, and the lowest score is 5.5 per cent for East Black Sea. The most illiterate region regarding GR4 is Southeast Anatolia, then Middle East Anatolia, Mediterranean, and Northeast Anatolia. Clearly, among GR4, the rate of people having achieved obligatory educational attainment is significantly low compared to GR1. Herewith, we can say that the level of higher education is clearly very low among the poorer group GR4 compared to GR1; nevertheless, deprivation regarding educational attainment remains a national problem across the country no matter deprivation group degrees.

GR5 sub-group profiles based on CAP113 model

In GR3, and GR4 analysis, we saw that deprivation regarding educational attainment emerges as an important item of capabilities failure. Now, we will see its distribution among the poorest group GR5. The most frequented educational achievement corresponds to primary school concerning 39.4 per cent of GR5. Besides, 28 per cent of GR5 are illiterate, 15 per cent are literate but without a diploma, and 11.65 per cent are secondary, vocational secondary or primary education school graduate. On the other hand, only 0.77 per cent are university or equivalent graduate, 1.87 per cent have achieved vocational or technical high school degree, and 3 per cent are high school graduate. Compared to GR3, and GR5, we can clearly observe that the situation regarding that aspect of life is worse, notably the per centage of illiterate people is dramatically alarming. As already shown in the literature of economics, or in socio-economics in a broad sense, educational attainment constitutes one of the key factors of earning more money, or living better than parents' living standards, and most importantly having a real liberty (i.e., real liberty allows the person to find a job he or she values in healthy and non-dangerous conditions; to have access to local, or global level information through different sources of media; to express and defend his or her rights). But of course, national and global context over periods affect the labour market situation; for instance, not every university graduate arrives to find the job he or she values in the country or in the world in a global sense. Despite challenges and varying word conditions, achieving a vocational or technical high school, or a university or equivalent is fundamental for the person to increase her/his real liberty.

In terms of household financial situation, we will first focus on the distribution of FS1 indicator among GR5. The results show that only 0.78 per cent of GR5 find easy or very easy to make

ends meet with their household's total monthly income. Besides, 4.8 per cent find a bit easy, 16.5 per cent find a bit hard, 35 per cent find very difficult, and 43 per cent find difficult to live with their household's total monthly income. These numbers are quite frightening, and clearly around 95 per cent of GR5 have serious problems about this financial aspect.

If we look at arrears related financial profile of GR5 including at least 100 observations based on the 4 most relevant patterns of the group discovered via the CAP113 model in first model analysis (generating 90 per cent of coverage of GR5), we can see that 30 per cent of GR5 have deprivation regarding FS3 and FS4 and no deprivation regarding FS2. In addition, 28 per cent of GR5 have no deprivation regarding FS2, FS3, and FS4 simultaneously; and 20.4 per cent have only deprivation regarding FS3. Across all extracted the profiles, we reach to the fact that 58 per cent of GR5 have deprivation regarding FS3, and 39 per cent have deprivation regarding FS4. Similar to the cases of GR3 and GR4, here also arrears on utility bills and arrears on hire purchase instalments, credit cards or other loan payments flourish as important elements of deficit in financial capabilities.

Moreover, when we consider the other 6 variables of household's financial situation, namely FS5, FS6, FS7, FS8, FS9, and FS10 respectively, we find the following results of GR5 profiles including at least 100 observations based on the 3 most relevant patterns of the group discovered via the CAP113 model: 39 per cent of GR5 have deprivation regarding all these 6 items; 13 per cent have only no deprivation regarding ability to keep home adequately warm; 9.8 per cent have only no deprivation regarding ability to keep home adequately warm and to afford new clothes; and 8.7 per cent have only no deprivation regarding ability to afford new clothes. Moreover, when all profiles are considered, we find that 79 per cent of GR5 have deprivation regarding FS5 (afford unexpected financial expenses); 58.6 per cent have deprivation regarding FS6 (ability to keep home adequately warm); 73.5 per cent have deprivation regarding FS7 (able to afford a meal with meat, chicken, fish, at least 3 days a week, when required or needed); 55.7 per cent have deprivation regarding FS8 (affording new clothes); 83.4 per cent have deprivation regarding both FS9 and FS10 (renew/replace old or worn out furniture, and afford, when required or needed, a one-week annual holiday away from home with all household members). It is important to remark that almost 4 of 10 people from the poorest GR5 group have deprivation in the 6 items; but also, deprivation regarding each item becomes more severe compared to GR3 and GR4 notably deprivation regarding FS5, FS7, FS9, and FS10. As at least 90 per cent of GR5 has problems in affording their needs with the household's monthly income, affording meat, holiday, furniture, unexpected expenses, heating expenses, and new clothes must be intrinsically difficult as well. In GR3, and GR4 analysis, we saw that despite many households are found deprived regarding heating system, energy source used for heating, or heating problems due to the insulation state of the household, ability of keeping adequately harm the dwelling in financial aspect is found relatively higher than other financial abilities.

When we plot the variables HEATS and FS6 we can see that 56 per cent of GR5 is not able to finance heating expenses. Among these people, 98 per cent use a stove in their dwelling. On the other hand, among who can finance heating expenses, 97.5 per cent use also a stove. Concerning main source of energy for heating, 27.8 per cent of who are deprived regarding FS6 use wood, 53 per cent use coal, and 15.8 per cent use cowpat; besides, 50.4 per cent of who are not deprived regarding FS6 use coal, 30 per cent use wood, 16.8 per cent use cowpat. We can see that the use of cowpat is emerged among GR5 compared to GR3 and GR4. And, even among who can afford heating expenses, the use of coal or wood is high compared to the use of

electricity or natural gas. Thus, the picture is no matter who can finance or not heating expenses, the use of stove as a heating system in the dwelling concern more than 95 per cent of the entire GR5, and the use of coal, wood, and cowpat is used most as energy source for heating in the dwelling respectively.

Regarding economic activity status, 47.5 per cent of GR5 is economically active vs. 43 per cent is not, 5 per cent is unemployed, and 3.8 per cent did not seek a job in last 4 weeks but available for work within 2 weeks in case of a job opportunity. To see in more details the combination of the variables ES1, SC, WH, and ES3 respectively, we computed economic activity profiles composed of these 4 variables based on the most relevant 3 patterns of GR5 and including at least 50 observations. As expected, 41.4 per cent of GR5 is found not economically active across the extracted profiles; otherwise, 38.6 per cent of GR5 is found economically active. Among these 38.6 per cent of GR5 people, the most frequent profile is being economically active, registered to national security system, working long hours more than 43 hours per week, and being wage worker (corresponding to 15 per cent of economically active GR5, in other words 6.8 per cent of GR5). Through the extracted profiles, we also find that 56.7 per cent of economically active GR5 population is not registered to social security; 35.7 per cent are not registered to social security and work long hours; 14 per cent are not registered to social security and work between 28 and 42 hours per week; 12.6 per cent are not registered to social security and work long hours and are own account workers; 9 per cent are not registered to social security, work long hours, and are casual employees; 7.6 per cent are not registered to social security, work long hours, and are unpaid family workers; and 6 per cent are not registered to social security, work long hours, and are wage workers. Hereby, we can say that being not enrolled to social security system, and/or work long hours are the two main issues among economically active GR5 population. Being not registered to social security system makes them unable to receive old age pension from state but also not adequately benefit public and state subsidized private health services across the country. If we think also that these people work o average long hours, and most of them are not able to finance annual holiday, not able to eat meat, chicken, or fish; so that these conditions may lead them to have health problems in terms of both physical and mental. And if they become unable or disadvantaged to visit health services due to non-enrollment to social security system, then their health situation can get worse.

Information on geographical situation may facilitate the implementation of poverty reduction programs at national level. For instance, necessary financial contribution from national budget might be directly sent to the regions suffering severe poverty. At least, such information might help to build a more tailored poverty elimination programs at local level, with a more specific eye. Once, the poorest regions are found, then more detailed poverty statistics may be later computed to amplify the content of programs for different target groups living in the region. When we look at the cross tabulation of the variables REG and SET for GR5, we can see that 56 per cent of GR5 live in rural areas vs. 44 live in urban areas. In terms of region, 17 per cent of GR5 live in Southeast Anatolia, 14 per cent live in Middle East Anatolia, 12 per cent live in Northeast Anatolia, 10 per cent live in Mediterranean, 9 per cent live in Aegean, 7 per cent live in West Black Sea. In Istanbul, only 4 per cent live, which means that Istanbul is the region harboring less the poorest people of the country. Compared to GR3 and GR4, the proportion of people living in rural areas is much higher among GR5. In terms of combination of REG and SET, we can say that the most frequented are: rural areas in Middle East Anatolia (9 per cent), urban areas in Southeast Anatolia (8.9 per cent), rural areas in Northeast Anatolia (8.8 per cent),

rural areas in Southeast Anatolia (8 per cent). Notably, the region of Southeast Anatolia emerges as one of the places harboring the poorest of the country both at its rural and urban areas.

Fourth step (step4 analysis): detecting multidimensional poverty profiles of specific groups

In this sub-section, the focus will be on elderly, unemployed, the most prosperous (no deprivation), deprivation group 1 (DEP_GRS1), deprivation group 2 (DEP_GRS2) populations as well as some descriptive statistics regarding some variables that were not used in constructing overall deprivation index. It is important to note that (i) the non-weighted numbers (i.e., no estimation at national level by using weighting function) are presented in the following descriptive statistics; and (ii) computations are based on the same data used for building overall deprivation index and poverty analysis above; in other words, the data is without missing values.

Elderly

Initially in the data there are 4.434 elderly people (65+). However, after removing missing values and the most prosperous group (zero deprivation), 4,341 elderly people remain in the data that corresponds to 11.75 per cent of total population. Among them, 36 per cent of them are in GR5, 20 per cent are in GR3, 20.7 are in GR4, 14 are in GR2, and 9.5 per cent are in GR1. We can see that elderly people's living conditions are quite poor as the majority are living in poorer groups such as GR4, and GR5, or at the middle such as GR3. When we look at whether elderly live in households suffering financially in terms of the household's total monthly income, by employing the indicator FS1 we can observe that 33.45 per cent live in a household with a difficult monthly living status, 26 per cent live in a household with a bit hard monthly living status, 18.3 per cent live in a household with a very difficult monthly living status, 13 per cent live in a household with a bit easy monthly living status, 8.4 per cent live in a household with an easy monthly living status, and only 1 per cent live in a household with a very easy monthly living status. If we sum the first three cases, we obtain a total of 77.75 as a per centage of elderly live in a household suffering clearly affording their needs at the end of the month. In terms of the health status in general, 37.5 per cent of them answered that they are bad, 35.2 per cent are not bad, 17 per cent are good, 9.5 per cent are very bad, and only 0.6 per cent are in very good situation. That means that the majority suffer some health problems. Moreover, 78 per cent are suffering from any chronic illness/long term illness or condition such as diabetes, hypertension, asthma, kidney failure, rheumatic diseases, etc. Notably, as the ratio of chronic illness is quite high, it becomes extremely important to have an easy access to medical services. Positively, 80.6 per cent were not deprived regarding access to medical examination or treatment during the last 12 months, and 90.5 per cent were not deprived regarding access to dentist examination or treatment during the last 12 months. Among chronically ill elderly people, 79 per cent are found not deprived regarding access to medical services; on the other hand, 14 per cent could not afford it, and 3 per cent live in somewhere too far to travel or there are no means of transportation to access to medical services. Among elderly having a bad or a very bad health conditions, 26 per cent are found deprived regarding access to medical service, and more specifically 18.8 per cent could not afford it and 4 per cent live in somewhere too far to travel or there are no means of transportation to access to medical services. In total, 12.8 per cent of elderly could not have access to medical services because they could not afford it, 2.8 per cent live in somewhere too far to travel or there are no means of transportation to access to medical services, and 1.87 per cent wanted to wait and see if problem got better on its own.

Focusing on elderly people could not have access to medical services during the last 12 months, 65 per cent of them are in GR5, 19.7 per cent are in GR4, and 10.7 per cent are in GR3, and the main reason is due to financial problems, and secondary it is due to distance to health services and availability of transports to reach the centers. In terms of location, 13 per cent of them live in West Black Sea, 10.8 per cent live in Aegean, 10.45 per cent live in Southeast Anatolia, 9.4 per cent live in Mediterranean, and 9 per cent live in Middle East Anatolia; besides, 63.4 per cent live in rural areas vs. 36.5 per cent live in urban areas. 16 per cent of these elderly people live in rural areas and could not get to health institutions due to distance or lack of transportation (i.e., too far to travel/no means of transportation). And, if we look at some features of the households of this population, in terms of economic activity status of the head of the household, we can say that 64 per cent of them come from a household headed by an economically not active person, 34 per cent come from a household headed by an economically active person, 24 per cent come from a household headed by an economically active person but he or she is not registered to national social security system. This last point might be important because in some specific cases parents may benefit of the social security of their children.

Unemployment

In the data, 48 per cent of 18+ individuals are economically active vs. 44 per cent are not economically active, 3.80 per cent are unemployed, 3.80 per cent did not seek a job in last 4 weeks but available for work within 2 weeks in case of a job opportunity. We will focus on the last 2 categories: (i) 35 per cent are from GR5, 17 per cent are from GR4, 17 are from GR3, 13 per cent are from GR2, and 17 per cent are from GR1; (ii) 50.1 per cent are female vs. 49.9 per cent are male. Among female jobseekers, 32.6 per cent are between 25-34 yrs. old, 27.7 per cent are between 18 and 24 yrs. old, 26 per cent are between 35 and 44 yrs. old. So that, 60 per cent of female job seekers are youth, between 18 and 34 yrs. old. On the other hand, among male jobseekers, 29 per cent are between 18 and 24 yrs. old, 23.8 per cent are 25-34 yrs. old, 16.5 per cent are 35-44 yrs. old, 17 per cent are between 45 and 54 yrs. old. In that case, 53 per cent of male job seekers are youth, between 18 and 34 yrs. old. It is interesting to note that 12.6 per cent of job seekers are university or equivalent degree graduates, 32 per cent are primary school graduates, 17.6 per cent are secondary, vocational secondary or primary education school graduates, 17.3 per cent are high school graduates. At this point, it will be more adequate to seek for information on the length of unemployment and the reason of seeking a job for more than 4 months among university of equivalent degree graduates, to understand whether the higher education graduates have difficulty or not to search for job. In some cases, employment rate might be higher among lower education graduates but their working conditions in terms of salary, dangerous working conditions, registration to social security, working hours may be worse. And university of equivalent degree graduates may be more selective to accept an offer in the market. Among university or equivalent degree graduates, 74 per cent are economically active and 9.5 per cent are job seekers; among illiterate, only 23 per cent are economically active, 74 per cent are economically not active and only 2.6 per cent are job seekers; among literate but with a no diploma, 60 per cent are not economically active, 33.6 per cent are economically active, and 6 per cent are job seekers; among primary school graduates, 50 per cent are economically active, 43.6 are not economically active, 6.5 per cent are job seekers; among secondary, vocational secondary, or primary education school graduates, 53.45 per cent are economically active, 35.5 per cent are not economically active, and 10.7 per cent are job seekers; among high school graduates, 47 per cent are economically active, 39.6 per cent are not economically active, 13 per cent are job seekers; among vocational or technical high school

graduates, 60 per cent are economically active, 30 per cent are not economically active, and 9 per cent are job seekers. In this picture, we can say that clearly vocational or technical high school, and university or equivalent are most economically active population vs. illiterate, literate but with a no diploma people are the most disadvantaged and they are mostly not active in the labour market. The tendency is that people with higher achievement in educational attainment (with a higher diploma) are more advantaged and are most active in the labour market. To reach a health institution when it is needed risks to become harder when the person is unemployed due to notably financial issues; the data shows that 17 per cent of job seekers could not access to medical services because they could not afford it, and 2 per cent could not because of lack of necessary transportation or distance from home to the institution.

In terms of location, 13.6 per cent live in Aegean, 12.64 per cent of job seekers live in Mediterranean, 11.15 per cent live in Middle East Anatolia, 11 per cent live in Southeast Anatolia, and 10.3 per cent live in Istanbul. Among people living in rural areas, 55 per cent of them are economically active, 38.45 are not economically active, and 6 per cent are job seekers. Combined with setting information, we can also say that 10.3 per cent of job seekers live in urban areas of Istanbul, 10.1 per cent live in urban areas of Aegean, 9.5 per cent live in urban areas of Mediterranean, 7.4 per cent live in urban areas of West Anatolia, and 6.8 per cent live in urban areas of Southeast Anatolia. Compared to urban areas, the per centage of job seekers is lower in rural areas. It is important to note the country has witnessed a substantial migration from rural to urban areas, but also from the east to the west part of country such as Istanbul as the largest metropolitan city or to Marmara region where Istanbul is located given that the region of Marmara has known a relatively higher industrial investment. In some regions, the rate of job seekers may be low than others, but it does not necessarily mean that the regions with low rate have good performance regarding offering jobs in the market because it is possible that in some regions if there is not much available jobs then women or man may prefer be housewife/men or become discouraged and do not seek a job. For that reason, we will look at more specifically people who are not economically active neither job seekers.

In terms of deprivation groups, 24.7 per cent of economically inactive 18+ population come from the poorest GR5, 21 per cent from GR3, 19 per cent from GR4, 18 per cent from GR2, and 17 per cent from GR1. Regarding that, there is not a dramatical difference between groups given that the per centage is slightly higher among GR5 compared to others with a maximum 7 per centage point gap. In fact, 58.5 per cent of these economically inactive population are reported effectuating household chores, taking care of children, elderly, sick, etc., 18 per cent are retired, early retired or quit; 15 per cent are elderly, disabled or unable to work, 5.7 per cent are apprentice/trainee, or simply they are student. Concerning 58.5 per centage, 99. 97 per cent are female; besides, 43 per cent of these women are primary school graduate, 12.5 per cent are literate but with no diploma, 22 per cent are illiterate, 10 per cent are secondary vocational secondary or primary school education graduates. Among 18+ population, the rate of primary school graduates is quite similar between male and female; but, notably in terms of illiteracy, vocational or technical high school diploma, and university or equivalent degree, the rate of male graduates is at least 30 per cent higher than female graduates across the country. It is important to emphasize that 12.8 per cent of economically inactive and taking care of household tasks, or elderly, or children, etc. women are between 18 and 24 yrs. old, 24 per cent are between 25 and 34 yrs. old. It becomes, therefore, important for government to promote programs and create opportunities, and adequate conditions to increase employment rate of women across the country given that only 30 per cent of economically active population is female, but also it is

crucial to increase possibilities for girls to enroll to school and to have at least obligatory national diploma and technical skills for being able to perform a job in their future. Moreover, training programs for illiterate women or specific programs for offering opportunities to finish high school or continue to higher education will increase certainly capabilities of women across the country.

The most prosperous group (GR0)

In the data, 0.08 per cent of 18+ belongs to the most prosperous group. This figure is almost zero. For that reason, it does not make very sense of examining such case.

GR1

We will examine the profiles of these groups in terms of economic activity status, geographical situation, educational attainment, access to medical services, and household monthly living status respectively[185].

From employment perspective, the 3 following profiles represent 77 per cent of GR1: (i) not being economically active (41 per cent); (ii) being economically active, registered to national social security system, working long hours (at least 43 hours per week), and being wage-worker (19.6 per cent); (iii) being economically active, registered to social security system, working between 28 and 42 hours per week, and being wage-worker (16 per cent). In addition, 3.5 per cent did not seek a job in last 4 weeks but available for work within 2 weeks in case of a job opportunity; 3.2 per cent are unemployed. There is in total 45 per cent of GR1 are economically active; among them, 43 per cent are registered to social security, work long hours, and are wage-workers, 35.4 per cent are registered to social security, work between 28 and 42 hours per week, and are wage-workers, 7 per cent are registered to social security, work long hours, and are employer, 4.8 per cent are registered to social security, work long hours, and are self-employed or own account workers, 4.6 per cent are not registered to social security, work long hours and are wage-workers, 3 per cent are not registered to social security, work long hours and are self-employed and own account workers, and 1.64 per cent are not registered to social security, work long hours, and are employer. Through these economically active profiles, we can remark that all of them work long hours, which is clearly the most important problem needed to be solved out. Moreover, among these profiles, 9.4 per cent are not registered to social security. On the other hand, 75.6 per cent of economically not active population is female vs. 24.4 are male. Among female not economically active population, 72.5 per cent are reported doing household chores, or taking care of children, elderly, sick, or they are busy with care of people whilst 68 per cent of male not economically active population are reported being retired, early retired or quit their jobs, and interestingly any of them are reported doing any household chores or are busy of taking care of people. For increasing empowerment of women, which means in other words increasing their capabilities for leading a decent life with a freedom, being

[185] We don't represent the statistics in terms of age and gender because the finding shows that there is not much difference between male and female population across the age groups (AGE) within the GR1 in terms of per centage. However, we can point out that the population of the age group 25-34 is the highest, which is followed by 35-44, 45-54, 18-24, and 65+ respectively. Such result seems not very much interesting because elderly population due to retirement, 18-24 population due to education or formation or as junior workers at the beginning of their career are likely to be disadvantaged in terms of financial aspects. But also, the second half of the group 45-54 might encounter early retirement. In this perspective, the age group 35-44 seems the most advantaged in terms of earnings, and experience in professional life.

economically active is vital. Unfortunately, even among the least poor group (GR1), we see that the rate of economically not active female population is considerably high; such situation puts them in a risk of becoming not only dependent of their partners but also vulnerable to life in a broad sense. For instance, a woman in such position won't receive old age pension and will be at risk of not having an economic freedom for building a new life in case she and her partner split up. And, among male not economically active population having been retired, or early retired, or left their jobs, 64 per cent are between 45 and 64 years old vs. only 35 per cent are aged 65 or over. This last point indicates the fragility of labour market, and unemployment risk for a person over 40 yrs. old when he or she quits or has to quit or lost her/his job, and in such situation these people might be finding themselves not been able to receive an old age pension and to trap into vicious life cycle with significantly reduced capabilities.

In geographical terms, the majority are from urban areas. The following urban locations in terms of region are the most frequented 11 ones comprising 89.6 per cent of GR1, respectively: Istanbul (23 per cent), West Anatolia (14.8 per cent), Aegean (13 per cent), Mediterranean (8 per cent), East Marmara (8 per cent), Central Anatolia (4.5 per cent), Middle East Anatolia (4.12 per cent), West Marmara (3.8 per cent), West Black Sea (3.6 per cent), Southeast Anatolia (3.4 per cent), and Northeast Anatolia (2.7 per cent). If we consider the west side of the country[186] with the regions of Istanbul, West Marmara, Aegean, East Marmara, West Anatolia, Mediterranean, and West Black Sea, we can see that the majority of least poor people live in West part of the country. On the other hand, regarding rural areas, the most advantaged region is Aegean that harbors 2.35 per cent of GR1, then West Marmara that includes 1 per cent of GR1; the per centages of other rural areas are negligible.

When it comes to educational attainment, 30 per cent of GR1 have a higher education diploma (university or equivalent), 21.7 per cent are primary school graduates, 15 per cent are vocational or technical high school graduates, 20 per cent are high school graduates, 9 per cent are secondary, vocational secondary or primary education school graduates, 2.77 per cent are illiterate, and 2 per cent are literate but with no diploma. In total, 45 per cent of GR1 have a diploma oriented to a specific métier (higher education plus vocational or technical high school). For a least poor group, this number is can be considered low. Again, we remark that the level of educational attainment should be revised and subject to new programs across the country for any improvement on that.

In terms of access to medical services, a hundred per cent of capability has emerged, everyone in GR is reported to have had access to medical services when it was needed, except only 1 per cent could not because of their work/family time schedule. Concerning household's financial status, 30 per cent find a bit hard, 22.4 per cent find difficult, 20 per cent find a bit easy, 19 per cent find easy, and only 1.85 per cent find very easy, and most importantly 6 per cent find very difficult to make ends meet. Normally, for such group, it is expected to find out higher proportions for the categories easy and very easy; nevertheless, it is not the case. This indicator

[186] The terminology of West Side is different than Europe vs. Asia parts of the country. European part of the country, in geographical terms, is composed of European side of Istanbul and East Marmara region. We also not that East and West Marmara are the most advanced in industrial investments, and Mediterranean, Aegean are the 2 advantaged regions in terms of weather conditions, agricultural activities, and tourism; in opposite, East Side of the country notably the regions of East Black Sea, Northeast Anatolia, and Middle East Anatolia, and Southeast Anatolia are the most disadvantaged regions in terms of whether conditions, agricultural and industrial activities.

is subjective including a personal evaluation on the ability of household to afford monthly need with their total monthly income. We cannot really know what the objectives or how least poor people are spend their money, with such indicator. A person earning a relatively good level of income might also say that he or she encounters difficulties, which can depend on emotions, personal state, character, social environment of the person but also on global context; for instance, if the person consumes a considerable number of imported goods he or she might be find also very difficult or difficult to afford her/his needs at the month. The financial situation of GR3, and notably GR4 and GR5 should be taken into account at first place; but it is also recommended to investigate consumption patterns of GR1 and GR2 to understand why they find relatively difficult to live with their household's total monthly income.

GR2

We will follow the same steps as we did for GR1 above. So that, we will examine GR2 in terms of economic activity status, geographical situation, educational attainment, access to medical services, and household monthly financial situation[187].

In terms of economic activity, based on the combination of the variables ES1, SEC, WH, ES3, the 14 profiles containing at least 50 observations each of them arrives of capturing 94 per cent of GR2 (corresponding to 6,198 persons). Among these profiles, 12 of them are various derivatives of economically active profile capturing in total 2,578 persons corresponding to 41.6 per cent of GR2. The most frequented profile of the economically active population is being registered to national social security system, working long hours, and being a wage-worker (42 per cent), then the profile of being registered to social security working between 28 and 42 hours per week, and being a wage-worker comes with 20 per cent. From the economically active population profiles, we can highlight that working long hours emerges also as a biggest problem affecting 70 per cent of them. Moreover, 29 per cent of them are found to not be registered to social security: 41.6 per cent are wageworker, 10.9 per cent are unpaid family worker, 9.7 per cent are self-employed or own account worker. Compared to GR1, here unpaid family workers emerge with a relatively higher amount. Through all detected profiles, we can observe that in total 19 per cent of economically active population is both not registered to social security and work long hours. It is important to remark that even in such relatively less poor group, being employee and not registered to social security emerges, which should be seriously taken into account by the government for increasing the rights of workers across the country. On the other hand, 46.8 per cent of 6.046 persons are economically not active, 2.86 per cent are unemployed, and 3.78 per cent did not seek a job in last 4 weeks but available for work within 2 weeks in case of a job opportunity. Concerning economically not active population, 77 per cent are female vs. 23 per cent are male. Among these female population, 77 per cent are reported effectuating household chores, or taking care of children, elderly, sick people, etc., or are busy with care of people, 9.35 per cent are elderly, disabled or unable to work, 7.2 per cent are retired, early retired, or have quit their job, 4.6 per cent are apprentice/trainee, or they continue their education. As it is found in case of GR1, here also no men are economically inactive due to household chores or taking care of people. On the other side, among male economically inactive population, 73 per cent are reported retired, early retired or having quit their job, 14 per cent are apprentice/trainee or they continue their

[187] As the same reasons we mentioned in GR1 analysis, here also we don't represent the statistics in terms of age and gender because the finding shows that there is not much difference between male and female population across the age groups (AGE) within the GR2 in terms of per centage.

education, and 8 per cent are elderly, disabled or unable to work. And 60 per cent of male population being retired, early retired or having quite their job are between 45 and 64 yrs. old.

Regarding geographical situation, we obtain several profiles as a combination of the variables REG and SET respectively, which reach to capture 98.5 per cent of GR2. Based on these profiles, we find that 22 per cent of total population of GR2 are living in rural areas and 78 per cent are living in urban areas. The most frequent urban areas are Istanbul, West Anatolia, Aegean, East Marmara, Mediterranean respectively, composing 51 per cent of GR2 population.

When it comes to educational attainment, 37 per cent of GR2 are primary school graduates, 14 per cent are high school graduates, 13.9 per cent have university or equivalent degree, 11.2 per cent are vocational or technical high school graduates, 6 per cent are illiterate, and 4.7 per cent are literate with no diploma. We can remark that the percentage of illiteracy is quite high for such group who does not witness severe poverty as GR3, GR4, and GR5, and only 24 per cent of GR2 have an education with a special focus on a métier (vocational or technical high school, university or equivalent degree).

In terms of access to medical services, 7.7 per cent are found to be deprived; among them, 33 per cent said could not afford it, 27 per cent said could not find time, 24 per cent wanted to wait and see if problem got better on its own, 7 per cent said fear of doctor/hospitals/examination/treatment, and 5.5 per cent were in waiting list. Compared to GR1, we can see that there is clearly an increase in lacking access to medical services due to financial reasons. Related to that, if we look at the monthly living status of the household, 31.7 per cent find a bit hard, 30 per cent find difficult, 16.65 per cent find a bit easy, 11.5 per cent find very difficult, 9 per cent find easy, and only 0.67 per cent find very easy to afford their needs with their household's total monthly income. As in GR1, here also there is still relatively high share of people finding difficult to make ends meets with the household's total monthly income.

Additional variables

Here, we will see some descriptive statistics regarding some additional variables that we did not employ previously in our main analysis. We start off by 3 economic activity related variables, namely, OG (occupational group), IND (the main activity code of the workplace, organization where the individual works), and CONTRACT (type of contract for employees). From GR1 to GR5, among economically active population, we can observe that the rate of professionals, legislators, senior officials and managers, technicians and associated professionals, clerks diminish, the rate of service workers and shop and market sales workers remain at quite similar rate, and the rate of skilled agriculture and fishery workers, craft and related trade workers, plant and machine operators and assemblers, and elementary occupation increase. Notably, in GR1, the rate of professionals; and in GR4 and GR5, the rate of skilled agriculture and fishery workers are the most relevant. On the other hand, in terms of industry, the rate of economically active population in public administration and defence; and the rate of people working in agriculture, forestry and fishing are the most significant compared to others. From GR1 to GR5, there is a clear increasing tendency of people working in agriculture, forestry and fishing; the rate of working in the industry of manufacturing, transportation and storage are quite similar; and the rate of working in financial and insurance activities, professional, scientific and technical activities, public administration and defense, and information and communication decreases. In terms of type of contract, permanent job/work

contract of unlimited duration is the most common but with a decreasing tendency from GR1 to GR5; on the other hand, there is a increasing tendency of temporary job/work contract of limited duration, and of working occasionally without contract from GR1 to GR5.

We already examined the household's financial situation with several indicators; nevertheless, in the data there are also another two variables focusing on financial burden of the total housing cost (FS11), and financial burden of the repayment of debts from hire purchases or loans excluding housing costs (FS12). When we look at their distribution among the 5 deprivation groups, we find that from GR1 to GR5 there is clear tendency of people living in households reporting a heavy burden the total housing cost, given that in all 5 groups the most frequent answer is 'a slight burden' except in GR5 the difference between the answers 'a heavy burden' and 'a slight burden' is relatively small compared to other groups. On the other hand, concerning FS12, the rate of people with no related with such kind of repayment increases from GR1 to GR5, and there is an increasing tendency of people living in households finding a heavy burden' such type of payment from GR1 to GR5. And households finding 'not burden at all' FS12 type of payments are mostly living in GR1 with an increasing tendency from GR5 to GR1.

Concerning dwelling standards, there is also a variable of the type of main dwelling (DWE) that we did not use in the overall index neither in dimensional analysis because the answer options to this variable in the survey questionnaire does not make clear about the quality of dwellings. In GR1, the majority live in apartments whilst in GR4 and GR5 the most frequent dwelling type is a detached house. From GR1 to GR5, there is an increasing tendency of detached house and twin or row houses, and a decreasing tendency of apartments.

In terms of household structure, we will use the variable computed and made available in the data by the Turkish Institute of Statistics. The variable is called HB050 and mentioned in the household questionnaire. We start off by the least poor GR1 where the most frequent 3 profiles are (see section 4.2 for full information on available modalities of this variable): other households with dependent children (24.5 per cent), other households without dependent children (20.5 per cent), and two adults with two dependent children (19 per cent). For GR2, the profiles of other households with dependent children (24.5 per cent), other households without dependent children (27 per cent); for GR3, the profiles of other households with dependent children (30.8 per cent), other households without dependent children (16 per cent); for GR4, the profile of other households with dependent children (32 per cent); and for the poorest GR5, the profiles of other households with dependent children (35 per cent), and two adults with three or more dependent children (15 per cent) emerge as the most occurred ones. We can see that a larger household in terms of at least 3 children exists in GR5 with a relatively important per centage compared to other groups where the rate of such household type varies between 6.9 and 10.5 in an increasing tendency from GR1 to GR4 respectively. Otherwise, in GR1, two adults with one dependent children emerges with 15 per cent; in GR2, two adults with two dependent children is about 14.8 per cent, and two adults with one dependent children is 14 per cent; in GR3, two adults with two dependent children corresponds to 12 per cent, two adults, no dependent children, both adults below 65 years concerns 11 per cent, and two adults with 1 dependent children constitutes 11 per cent; in GR4, two adults with two dependent children is 12 per cent, two adults with one dependent children is 11 per cent, and two adults with three or more dependent children concerns 10.5 per cent; in GR5, Other households without dependent children concerns 9.6 per cent, two adults with two dependent children

includes 9.6 per cent, two adults with one dependent child covers 8.4 per cent, and two adults, no dependent children, both adults below 65 years is about 8 per cent. Across the 5 groups, notably the per centage of single person and single person with one dependent children households are quite small; nevertheless, their per centages increase from GR1 to GR5 given that in GR5, 2 per cent of households are composed of one person and one dependent children, and 4.8 per cent of households is composed of one person. Previously, we already saw that the 26.4 per cent of elderly population live in GR1 and GR2; so that, the majority of them suffer a higher level of capabilities deficit. In terms of the number of dependent children in the household, the findings show that among the households in GR5 have a higher number of children notably at least 3. In other words, notably, households with two dependent children are the most occurred among the more prosperous groups compared to GR5. As we have seen, in the data, the per centages of other type of households are quite frequent- which does not yield precise information unfortunately on the structure of the households.

To finalize our Turkey analysis, we will lastly point out the economic activity status information on the head of the household concerning GR5. By combining the variables ES1H, SCH, WHH, ES3H respectively, we obtain 17 profiles capturing 98 per cent of GR5. Among this population reveled by the profiles, 29 per cent live in households headed by a person economically not active, 6.5 per cent live in households headed by an unemployed, 1.5 per cent live in households headed by a person who did not seek a job in last 4 weeks but available for work within 2 weeks in case of a job opportunity, and the rest 61 per cent live in households headed by an economically active person giving a total of 5,716 persons. Moreover, 65.5 per cent live in households headed by an economically active person but not registered to national social security system, 71.7 per cent live in households headed by an economically active person but working long hours, 42.7 per cent live in households headed by an economically active person not registered to social security and work long hours. And the per centages of GR5 people living in households headed by a person working as a skilled agriculture and fishery workers (39.5 per cent), performing elementary occupations (19 per cent), working as a craft and related trade workers (15 per cent), and as a plant and machine operators and assemblers (+0.5 per cent) are the most relevant ones. We can again clearly see that working long hours and no registration to social security emerge as the biggest problems for those are economically active.

If we recapitulate our findings, we can summarize that the per centage of population living in urban areas is significantly higher compared to rural areas, but also the west side of the country experience less deprivation than the east side of the country. In terms of dwelling situation, heating system, insulation state of the dwelling, energy source used for heating, roof, wall, and window conditions of the dwelling should be revised notably for GR3, GR4, and GR5. Besides, hot water system, available sanitation facilities in the dwelling should be checked for GR4 and GR5. In terms of material deprivation, lacking dishwasher and computer and internet emerged very relevant in GR3, GR4, and GR5. Concerning educational attainment, the new programs promoting its importance and efforts for increasing quality and quantity of educational institutions should be primarily implemented across the country. Notably, the share of illiterate, literate with no diploma, and primary school graduates are very substantial. It is extremely important for future generations to have at least a diploma of vocational or technical high school in order to specialize towards a specific métier. Among the economically active population, working long hours and not being registered to social security are clearly big problems to be urgently solved out. Working long hours may have impacts on the health status of the person and not being registered to social security prevent receiving old age pension. Interestingly, the

per centage of persons living in rural areas is relatively few but on the other hand, people working as a skilled agriculture or as a fishery worker represent the highest proportion among the economically active population. As a country having relatively advantaged land notably in Mediterranean, Aegean, West North Black Sea costs, new high tech and trainings in agricultural and fishing activities should be promoted by the state and people should be encouraged to stay in rural areas for better agricultural activities. Because a high rate of migration from rural to urban areas risks reducing the agricultural production, which can become a problem for a relatively populated country and the country may become dependent on agricultural products importation if necessary efforts are not made. High primary goods price will make poorer citizens notably for those are in GR3, GR4, and GR5. Working in agricultural activities does not mean that high school or higher education diploma are not necessary; given the emerging climate change problems and raise of new technologies, it is still important to have at least technical high school diploma. Industrial activities may be extended outside of the region of Marmara and Istanbul for alleviating the density of urban population in these locations. Moreover, women should be encouraged for working and contributing to the household's income for both economic and social aspects, and the situation of unpaid family workers should be revised. On the other hand, it is also important to emphasize that the country experience good performance in access to medical services, environmental issues in terms of noise, air pollution, and crime and violence, available piped water system, kitchen in the dwelling, having a television, a mobile phone, a refrigerator, a washing machine in the dwelling across all deprivation groups. Lastly, the monthly living status of the households with their total income requires to be examine deeply in the sense that even in GR1 and GR2, at least 90 per cent of them don't still find easy or very easy to make ends meet. Such result is very surprising for relatively prosperous groups such as GR1 and GR2 compared to GR3, GR4, and GR5.

9.5 Case study of Mexico

Introduction

In case of Mexico, we will use the Socio-Economic Conditions Survey of 2012. The original name of the survey is in Spanish namely Modulo de Condiciones Socioeconómicas (MCS) de la Encuesta Nacional de Ingresos y Gastos de los Hogare (ENIGH)s, carried out by the National Statistical office of Mexico (Instituto Nacional de Estadística y Geografía-INEGI).

As we did in the cases of South Africa, and Turkey, here also in Mexico analysis we will follow the same methodology. In that vein, based on the ENIGH 2012 data, we will carry out the following analysis to measure multidimensional poverty, respectively: detecting multidimensional poverty patterns, dimensional poverty profiles, sub-group profiles, and some additional profiles such as elderly, unemployed. The list of variables is given in section 4.2. Depending on the purpose of the model, we will select indicators. The short names of variables (see the codes in parenthesis in first column of the table of selected indicators) will be displayed in the outputs including graphics. Before describing models and presenting results, we will yield a brief information on ENIGH and then the quality report of the variables.

The MCS 2012 collects information on various aspects of life including income, health, education, social security, housing and living standards (e.g., whiteware and materials such as television, car, computer), and food. The data is collected from 27 August to 2 December of

2012[188]. This survey is especially designed for measuring multidimensional poverty in accordance with the requirements of the Social Development Law (Ley de Desarollo Social) in a joint collaboration of INEGI and National Council for Social Development Policy Evaluation (CONEVAL[189]).

In our analysis, we will employ several variables relating to geographic situation, individual characteristic, household composition, dwelling situation, material situation, educational attainment, health, economic activity (labour), and food. Some variables are constructed based on the information collected through the individual questionnaire where the questions are asked for each of the household member, whilst some other variables are based on the information collected through the household questionnaire where the questions concern the household as a whole or the dwelling where all household members live (i.e., the household level variables' scores are the same for all members of the household in the data)[190].

Quality report of variables

As a first fundamental step, we created unique household (hid) and personal identifiers (id) in order to merge data together. More specifically, 3 data are used: personal ('Pobla12'), dwelling ('Vivienda'), and household ('Hogares'), respectively. The first data includes information on individual characteristic such as age, gender, educational attainment, health status; the second data covers information on the house or apartment or room in which the household lives such as wall, roof conditions of the dwelling; and the third data contains information on some socio-economic features of the household as a whole such as household assets (e.g., car, radio, tv), household income, food situation of the household.

In dwelling data, there are 56,179 unique dwellings; in personal data there are 212,881 individuals including children; in economic activity data, there are 49,984 individuals (questions are asked only to 12+); and in household data there are 56,179 unique households and 57,274 records. Compared to previous South Africa, and Turkey data, in case of Mexico data is collected based on dwellings ('vivienda'). This process is due to the fact that in some dwellings, more than one household live together; nevertheless, the per centage of such case appear quite few in the data (only in 1.6 per cent of 56,179 dwellings, more than one household live). At first stage, the dwelling data is merged with the personal data by the unique dwelling identifier (called *dwid*) provided in the data by INEGI. After that, at second step, we created a unique household (called *hid*) and a unique personal identifier (called *id*) by employing the unique dwelling identifier, unique household identifier, and serial number of each member within their household. The merge of dwelling and personal data produced a data composed of 212,881 individuals across 56,179 dwellings, and 57,274 households (i.e., all individuals were merged correctly according to their dwellings and households). At third step, this merge data is combined with the household data through the unique household identifier, which led to a new data composed of 212,881 individuals across 56,179 dwellings, and 57,274 households (i.e., all

[188] Information presented in this paragraph come from the official document on the description of MCS' data bases provided by the INEGI.

[189] CONEVAL is coordinated by the Mexican Ministry of Social Development (SEDESOL). Please see the links below for more information (accessed 5 December 2019):
https://www.coneval.org.mx/Paginas/principal.aspx
https://web.archive.org/web/20110722125903/http://www.diputados.gob.mx/LeyesBiblio/pdf/264.pdf

[190] For facilitating the lecture and comprehension, we employ the term *household* as the unit of the variables coming from the dwelling questionnaire, instead of the term *dwelling*.

cases are matched correctly). We note that in fact there is also labour data ('trabajos') provided by the INEGI as a part of the MCS; however, this data contains some duplicates in the sense that for some individuals the information is collected for their two jobs described in subsequent two different rows (first row is for the first job, and second row is for the second job of the same person). In case we merge this labour data under such format with the other merged data, duplicates will be kept and therefore, the results will be overestimated. To avoid it, we will simply ignore the detailed variables on labour conditions and effectuate the poverty analysis based on the final merged data composed of dwelling, personal, and household data (given that the final merged data include the fundamental variable of economic activity and another variable on the number of jobs performed by the person-these 2 variables are available in personal data). Besides, we also remark that there is no information on geographical situation in terms of federal states and setting (urban vs. rural) in the data. There is one variable named ubica_geo in the data that contains federal, municipality, locality, AGEB[191], control number codes respectively but unfortunately it does not allow to find out profiles ate federal and setting levels.

Based on the final merged data, all indicators (i.e., derived variables) defined in the methodology (section 4.2) are created and later derived data including both initial and derived variables is produced. Derived data and its subsets form the basis for our analysis. 32 variables are taken to construct the overall deprivation index, which means that all persons not having a full information of these variables were dropped from the derived data. The 32 variables, with their short names, are as follows: WAL, ROF, FLO, KIT, WATD, SAN2, SINK, SAN1, HOTW, ELC, HEATS, ENG_C, DR, RUB, AIRC, TEL, MPHO, INT, CAR, RAD, REF, STONE, WM, IRON, VC, COMP, EDU, MC, AF1, AF2, and AF3 (see the table in first model sub-section below for information on these variables).

Our analysis will cover persons aged 18 or over. In the data, there are 140,624 persons in this category, which constructs 66 per cent of the data. We remove in total 1,038 individuals aged 18 or over because they have missing values regarding at least one of the variables, ENG_C, MC, VC, WM, IRON. We lose only 0.74 per cent of information, which can be considered quite negligible. Thus, the final data on which the analysis will be done includes 139,586 persons aged 18 or over across 56,581 households and 55,504 dwellings in the country. Additional to 32 variables mentioned in first paragraph, we created also several other variables for decomposition and supplementary analysis. There are in total 73 variables[192] (including PATH variable that corresponds to capabilities profile of each individual aged 18 or more in the data) based on the main 32 variables that are used in building the overall deprivation index. The variable PATH is a string variable reflecting the combination of 32 variables, as follows:

PATH=WAL+ROF+FLO+KIT+WATD+SAN2+SINK+SAN1+HOTW+ELC+HEATS+ENG_C+DR+RUB+AIRC+TEL+MPHO+INT+CAR+RAD+TV+REF+STONE+WM+IRON+VC+COMP+EDU+MC+AF1+AF2+AF3.

It is found 89,985 different capabilities profile (PATH) among 139,586 individuals in the data. In other words, 64 % of the 18+ population in the derived data has her own singular capability

[191] A specific variable mentioned in the dictionary of the variables provided by the INEGI.

[192] The variables namely age (numerical variable indicating age of the person), dwid, hid, id, serial number amongst others for decomposition and supplementary analysis are included.

profile/path. Such result is very complex to examine. So, we need to find out the most influential ones, which we will find out by employing classification tree analysis.

We obtained the quality report of 73 variables in statistical software R via the package Hmisc[193]. We note also that descriptive statistics and analysis is based on the non-weighted data[194]. The short name of the variables corresponds to their codes mentioned in section 4.2.

Descriptive statistics outputs:

```
MX2012derived[, 1:73]
 73  Variables      139586  Observations
---------------------------------------------------------------------------
dwid
       n  missing distinct
  139586        0    55504

lowest : 011001 011002 011003 011004 011005, highest: 329696 329697 329698 329699 329700
---------------------------------------------------------------------------
hid
       n  missing distinct
  139586        0    56581

lowest : 0110010 0110020 0110030 0110040 0110050
highest: 3296960 3296970 3296980 3296990 3297000
---------------------------------------------------------------------------
id
       n  missing distinct
  139586        0   139586

lowest : 011001001 011001002 011001003 011002001 011002002
highest: 329699001 329700001 329700002 329700003 329700004
---------------------------------------------------------------------------
serial
       n  missing distinct
  139586        0       21

lowest : 01 02 03 04 05, highest: 17 18 19 20 21
---------------------------------------------------------------------------
HS
       n  missing distinct
  139586        0        5

Value          1     2     3     4     5
Frequency   5882 18540 24932 32343 57889
Proportion 0.042 0.133 0.179 0.232 0.415
---------------------------------------------------------------------------
age
```

[193] See South Africa GHS 2011 analysis for more technical information on statistical software and packages. But also, we remark that missing values are coded 9 and not applicable cases are coded 99 in the data, except the variables having 9 categories/modalities (see section 4.2-table of variables). We already dropped 198 individuals from the data; however, concerning decomposition and supplementary analysis variables, there are still few missing values and not applicable cases such as receiving medical care concern only people having got sick or suffered from any pain, physical discomfort or accident.

[194] To obtain estimated numbers at national level typically for the national reports, the obtained figures can be multiplied by the appropriate weight provided in the data by the Turkey Institute of Statistics. But, in our analysis, we use the national sample, which is also an appropriate representative of the population in the country.

```
      n  missing distinct     Info     Mean      Gmd      .05      .10
 139586        0       80        1     41.5    19.25       19       21
    .25      .50      .75      .90      .95
     27       39       53       66       74

lowest : 18 19 20 21 22, highest: 93 94 95 96 97
---------------------------------------------------------------------------
hh_child
      n  missing distinct
 139586        0        6

Value          1     2     3     4     5     6
Frequency  49530 33806 30345 16063  6056  3786
Proportion 0.355 0.242 0.217 0.115 0.043 0.027
---------------------------------------------------------------------------
WAL
      n  missing distinct
 139586        0        8

Value          1      2      3      4      5      6      7      8
Frequency    132    289    709    554    666   6432  13505 117299
Proportion 0.001  0.002  0.005  0.004  0.005  0.046  0.097  0.840
---------------------------------------------------------------------------
ROF
      n  missing distinct
 139586        0        9

Value          1     2     3     4     5     6     7     8     9
Frequency    139  2263 18792  8847  1306  4097  1967  3358 98817
Proportion 0.001 0.016 0.135 0.063 0.009 0.029 0.014 0.024 0.708
---------------------------------------------------------------------------
FLO
      n  missing distinct
 139586        0        3

Value          1     2     3
Frequency   4526 78196 56864
Proportion 0.032 0.560 0.407
---------------------------------------------------------------------------
KIT
      n  missing distinct
 139586        0        2

Value          1      2
Frequency  128980  10606
Proportion  0.924  0.076
---------------------------------------------------------------------------
WATD
      n  missing distinct
 139586        0        7

Value          1     2     3     4     5     6     7
Frequency  77621 19836 31641   428  1142  1338  7580
Proportion 0.556 0.142 0.227 0.003 0.008 0.010 0.054
---------------------------------------------------------------------------
SAN2
      n  missing distinct
 139586        0        3

Value          1      2      3
```

Frequency 126594 8459 4533
Proportion 0.907 0.061 0.032

SINK

n	missing	distinct
139586	0	2

Value	1	2
Frequency	79934	59652
Proportion	0.573	0.427

SAN1

n	missing	distinct
139586	0	2

Value	1	2
Frequency	93946	45640
Proportion	0.673	0.327

HOTW

n	missing	distinct
139586	0	2

Value	1	2
Frequency	61035	78551
Proportion	0.437	0.563

ELC

n	missing	distinct
139586	0	5

Value	1	2	3	4	5
Frequency	137453	97	394	658	984
Proportion	0.985	0.001	0.003	0.005	0.007

HEATS

n	missing	distinct
139586	0	2

Value	1	2
Frequency	2970	136616
Proportion	0.021	0.979

ENG_C

n	missing	distinct
139584	2	5

Value	1	2	3	4	5
Frequency	24798	504	105125	8382	775
Proportion	0.178	0.004	0.753	0.060	0.006

DR

n	missing	distinct
139586	0	5

Value	1	2	3	4	5
Frequency	97794	29838	1307	576	10071
Proportion	0.701	0.214	0.009	0.004	0.072

RUB

```
      n  missing distinct
 139586        0        8

Value          1      2      3      4      5      6      7      8
Frequency  109226   2069   7022  19425    590    699    508     47
Proportion  0.782  0.015  0.050  0.139  0.004  0.005  0.004  0.000
--------------------------------------------------------------------------------
EDU
      n  missing distinct
 139586        0        9

Value          0     1     2     3     4     5     6     7     8
Frequency  10671  1818    27 39730 35524 22527  1509  6371 21409
Proportion 0.076 0.013 0.000 0.285 0.254 0.161 0.011 0.046 0.153
--------------------------------------------------------------------------------
MC
      n  missing distinct
 139586        0        4

Value          1     2     3     4
Frequency  49435  6826 52502 30823
Proportion 0.354 0.049 0.376 0.221
--------------------------------------------------------------------------------
AIRC
      n  missing distinct
 139586        0        2

Value          1      2
Frequency   20700 118886
Proportion  0.148  0.852
--------------------------------------------------------------------------------
TEL
      n  missing distinct
 139586        0        2

Value          1     2
Frequency  56155 83431
Proportion 0.402 0.598
--------------------------------------------------------------------------------
MPHO
      n  missing distinct
 139586        0        2

Value          1     2
Frequency  96458 43128
Proportion 0.691 0.309
--------------------------------------------------------------------------------
INT
      n  missing distinct
 139586        0        2

Value          1      2
Frequency   35165 104421
Proportion  0.252  0.748
--------------------------------------------------------------------------------
CAR
      n  missing distinct
 139586        0        2

Value          0      1
```

Frequency 101738 37848
Proportion 0.729 0.271

RAD

n	missing	distinct
139586	0	2

Value	0	1
Frequency	114783	24803
Proportion	0.822	0.178

TV

n	missing	distinct
139586	0	2

Value	0	1
Frequency	8415	131171
Proportion	0.06	0.94

REF

n	missing	distinct
139586	0	2

Value	0	1
Frequency	19692	119894
Proportion	0.141	0.859

STOVE

n	missing	distinct
139586	0	2

Value	0	1
Frequency	14758	124828
Proportion	0.106	0.894

WM

n	missing	distinct
139586	0	2

Value	0	1
Frequency	45492	94094
Proportion	0.326	0.674

IRON

n	missing	distinct
139586	0	2

Value	0	1
Frequency	26929	112657
Proportion	0.193	0.807

VC

n	missing	distinct
139586	0	2

Value	0	1
Frequency	130209	9377
Proportion	0.933	0.067

COMP

```
      n  missing distinct
 139586        0        2

Value          0     1
Frequency  96981 42605
Proportion 0.695 0.305
---------------------------------------------------------------------------
AF1
      n  missing distinct
 139586        0        2

Value          1      2
Frequency   20203 119383
Proportion  0.145  0.855
---------------------------------------------------------------------------
AF2
      n  missing distinct
 139586        0        2

Value          1     2
Frequency  67124 72462
Proportion 0.481 0.519
---------------------------------------------------------------------------
AF3
      n  missing distinct
 139586        0        2

Value          1     2
Frequency  51891 87695
Proportion 0.372 0.628
---------------------------------------------------------------------------
GEN
      n  missing distinct
 139586        0        2

Value          1     2
Frequency  73292 66294
Proportion 0.525 0.475
---------------------------------------------------------------------------
AGE
      n  missing distinct
 139586        0        6

Value          1     2     3     4     5     6
Frequency  26861 29607 28550 23003 15531 16034
Proportion 0.192 0.212 0.205 0.165 0.111 0.115
---------------------------------------------------------------------------
ETH
      n  missing distinct
 139586        0        2

Value          1     2
Frequency  40903 98683
Proportion 0.293 0.707
---------------------------------------------------------------------------
LNG
      n  missing distinct
  40903    98683        3

Value          1     2     3
```

Frequency 10178 2910 27815
Proportion 0.249 0.071 0.680

ELD
n missing distinct
139586 0 2

Value 0 1
Frequency 123552 16034
Proportion 0.885 0.115

DWE
n missing distinct
139586 0 6

Value 1 2 3 4 5 9
Frequency 132181 4384 2553 137 136 195
Proportion 0.947 0.031 0.018 0.001 0.001 0.001

HSS
n missing distinct
139586 0 2

Value 1 2
Frequency 38604 100982
Proportion 0.277 0.723

MC2
n missing distinct
139586 0 6

Value 1 2 3 4 5 99
Frequency 33730 2589 1096 1025 164 100982
Proportion 0.242 0.019 0.008 0.007 0.001 0.723

MC2_1
n missing distinct value
180 139406 1 01

Value 1
Frequency 180
Proportion 1

MC2_2
n missing distinct value
638 138948 1 02

Value 2
Frequency 638
Proportion 1

MC2_3
n missing distinct value
1413 138173 1 03

Value 3
Frequency 1413
Proportion 1

MC2_4

```
      n  missing distinct    value
    454   139132        1       04

Value        4
Frequency  454
Proportion   1
--------------------------------------------------------------------------------
MC2_5
      n  missing distinct    value
    480   139106        1       05

Value        5
Frequency  480
Proportion   1
--------------------------------------------------------------------------------
MC2_6
      n  missing distinct    value
    339   139247        1       06

Value        6
Frequency  339
Proportion   1
--------------------------------------------------------------------------------
MC2_7
      n  missing distinct    value
    298   139288        1       07

Value        7
Frequency  298
Proportion   1
--------------------------------------------------------------------------------
MC2_8
      n  missing distinct    value
     21   139565        1       08

Value        8
Frequency  21
Proportion  1
--------------------------------------------------------------------------------
MC2_9
      n  missing distinct    value
    875   138711        1       09

Value        9
Frequency  875
Proportion   1
--------------------------------------------------------------------------------
MC2_10
      n  missing distinct    value
   1057   138529        1       10

Value        10
Frequency  1057
Proportion    1
--------------------------------------------------------------------------------
MC2_11
      n  missing distinct    value
    671   138915        1       11

Value       11
```

Frequency 671
Proportion 1

MC2_12
n missing distinct value
280 139306 1 12

Value 12
Frequency 280
Proportion 1

MC2_13
n missing distinct value
101 139485 1 13

Value 13
Frequency 101
Proportion 1

MC2_14
n missing distinct value
285 139301 1 14

Value 14
Frequency 285
Proportion 1

MC2_15
n missing distinct value
119 139467 1 15

Value 15
Frequency 119
Proportion 1

MC2_16
n missing distinct value
181 139405 1 16

Value 16
Frequency 181
Proportion 1

MC2_17
n missing distinct value
12 139574 1 17

Value 17
Frequency 12
Proportion 1

MC2_18
n missing distinct value
1116 138470 1 18

Value 18
Frequency 1116
Proportion 1

MC2_19

```
      n  missing distinct    value
    218   139368        1       19

Value       19
Frequency  218
Proportion   1
---------------------------------------------------------------------------
ES1
      n  missing distinct
 139586        0        9

Value         1     2     3     4     5     6     7     8     9
Frequency  86052   501  3969   139  4339 32536  6656  2830  2564
Proportion 0.616 0.004 0.028 0.001 0.031 0.233 0.048 0.020 0.018
---------------------------------------------------------------------------
JOB
      n  missing distinct
 139482      104        3

Value         1     2    99
Frequency  81338  5111 53033
Proportion 0.583 0.037 0.380
---------------------------------------------------------------------------
GH
      n  missing distinct
 139586        0        2

Value          1      2
Frequency   30586 109000
Proportion  0.219  0.781
---------------------------------------------------------------------------
AH
      n  missing distinct
 139586        0        5

Value         1     2     3     4     5
Frequency     19  4453 17552 91635 25927
Proportion 0.000 0.032 0.126 0.656 0.186
---------------------------------------------------------------------------
EDUH
      n  missing distinct
 139586        0        9

Value         0     1     2     3     4     5     6     7     8
Frequency  12815  2742    24 51600 31944 15160  1699  5310 18292
Proportion 0.092 0.020 0.000 0.370 0.229 0.109 0.012 0.038 0.131
---------------------------------------------------------------------------
ES1H
      n  missing distinct
 139586        0        9

Value          1      2      3      4      5      6      7      8      9
Frequency  105950    632   3847    285   9259  12257    543   2984   3829
Proportion  0.759  0.005  0.028  0.002  0.066  0.088  0.004  0.021  0.027
---------------------------------------------------------------------------
PATH
      n  missing distinct
 139586        0    89985
```

```
lowest : 1111312221232122221111101000121 1111312221232122221111101003121
1111312221232122221111101004121 1111512221233722120110110003422
2
1111512221233722120110110004422
2
highest: 8932712221212122220000000003122 8932712221212122220000000004122
1
8932712221232422121011111004122 8932712221232422121011111005122
2
8932712225221122220010000000411
1
---------------------------------------------------------------------------
```

We note that in the following analysis concerning classification trees by applying the package rpart in R, we will give only the essential statistical outputs.

First step (step1 analysis): detecting multidimensional poverty patterns

In this investigation, the decision trees-based capabilities analysis will be based on the overall deprivation index we will built. First of all, we define a deprivation line for each of the 32 variables. Deprivation line is the criteria of distinguishing who is deprived or not regarding the variable. As we have already mentioned in GHS 2011 case, the utility of each variable representing a different functioning is determined according to universal human rights chart. According to the deprivation threshold, if a person experiences deprivation regarding the variable the he or she takes the score 1, and 0 otherwise for that variable. Taking in consideration all variables simultaneously, we count the number of deprivations the person suffers to obtain her/his overall individual deprivation score. Then, as a function of the distribution of all person's overall deprivation index scores in the data, we create several deprivation groups. At last stage, we will proceed with classification tress method to discover the most influential capabilities patterns emerging in the data distinguishing deprivation groups among them. The table below gives information on deprivation line for each variable used in the index. The table contains 4 columns: first column gives the code name of the variable given that each variable represents a functioning, second column mentions deprivation line regarding the variable, third column indicates the utility of the functioning for being capable of living a decent life according to universal human rights chart, and last column contains remarks about the variable.

Table 58. Deprivation lines and composition of variables for model 1: overall deprivation index and multidimensional poverty patterns, data: Mexico MCS 2012

Variable (each variable represents a functioning)	**Deprivation line[195]**	**The utility of functionings for having the capability of living a decent life according to Universal Human Rights Chart[196]**	**Remark**
WAL_D	Deprived (WAL_D=1) if WAL is 1,2,3,4,5,6,7	Article 25: functioning of living with an adequate housing.	Waste material, cardboard sheet, reed, bamboo, palm, wattle, daub, wood, adobe, asbestos, metal foil are considered as not appropriate wall materials

[195] For the codes and categories, see the table provided in section 4.2.
[196] Please refer to the link below for further information (accessed 12 February 2020) : https://www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/eng.pdf

			compared to partition wall, brick, block, stone, quarry, cement or concrete.
ROF_D	Deprived (ROF_D=1) if ROF is 1,2,3,4,5,6	Article 25: functioning of living with an adequate housing.	Waste material, cardboard sheet, sheet metal, asbestos, palm, straw, wood are considered as not appropriate roof materials compared to rooftop with beams, floor framing, brick red, specific floor or joists with small vault.
FLO_D	Deprived (FLO_D=1) if FLO is 1	Article 25: functioning of living with an adequate housing.	Land is considered as not appropriate floor material compared to cement, firm, wood, mosaic, or other coating.
KIT_D	Deprived (KIT_D=1) if there is not a kitchen in the dwelling	Article 25: functioning of living with an adequate housing.	
WATD_D	Deprived (WATD_D=1) if piped water system is not inside the house or there is a piped water system inside the dwelling, but the water does not arrive daily	Article 25: functioning of living with an adequate housing.	This variable is composed of the variables WAT1 and WAT2 (see section 4.2).
SAN2_D	Deprived (SAN2_D=1) if there is no toilet in the dwelling	Article 25: functioning of living with an adequate housing.	
SINK	Deprived (SINK_D=1) if there is no sink (fregadero or tarja) in the dwelling	Article 25: functioning of living with an adequate housing.	
SAN1_D	Deprived (SAN1_D=1) if there is no shower (regadera) in the dwelling	Article 25: functioning of living with an adequate housing.	
HOTW_D[197]	Deprived (HOTW_D=1) if there is no boiler, no hot	Article 25: functioning of living	

[197] The variable HOTW_D is the same of the variable HWS_D used in Turkey SILC 2011 analysis. These two variables mean the same thing. Due to a mistake, in Mexico MCS 2012 case, the code HOTW is used instead of HWS.

	water system (calentador de agua) in the dwelling	with an adequate housing.	
ELC_D	Deprived (ELC_D=1) if there is no electricity in the dwelling	Article 25: functioning of living with an adequate housing.	
HEATS_D	Deprived (HEATS_D=1) if there is no heating system (calefaccion) in the dwelling	Article 25: functioning of living with an adequate housing.	
ENG_C_D	Deprived (ENG_C_D=1) if energy source for cooking is firewood, coal, or tank gas (i.e., ENG_C is 1,2,3)	Article 25: functioning of living with an adequate housing.	
DR_D	Deprived (DR_D=1) if the dwelling does not have a drain or drain is not connected to the public network (i.e., DR is 2,3,4,5	Article 25: functioning of living with an adequate housing.	
RUB_D	Deprived (RUB_D=1) if RUB is 2,4,5,6,7,8	Article 25: functioning of living with an adequate housing.	
AIRC_D	Deprived (AIRC_D=1) if the household does not have an air conditioner	Article 25: functioning of living with an adequate housing in terms of material.	
TEL_D	Deprived (TEL_D=1) if the household does not have a landline telephone	Article 25: functioning of living with an adequate housing in terms of material and social service regarding availability of landline network in the neighborhood where the person lives. It is not necessary that telephone, internet networks are provided by the State, but at least State should make sure that every single citizen has	

		access to such services.	
MPHO_D	Deprived (MPHO_D=1) if the household does not have a mobile phone	The same argument mentioned for TEL_D.	
INT_D	Deprived (INT_D=1) if the household does not have an internet	Article 25: functioning of living with an adequate housing in terms of material and social service regarding availability of internet service in the neighborhood where the person lives.	
CAR_D	Deprived (CAR_D=1) if the household does not have a car	Article 25: functioning of living with an adequate housing in terms of material.	
RAD_D	Deprived (RAD_D=1) if the household does not have a radio	Article 25: functioning of living with an adequate housing in terms of material.	
TV_D	Deprived (TV_D=1) if the household does not have a television	Article 25: functioning of living with an adequate housing in terms of material.	
REF_D	Deprived (REF_D=1) if the household does not have a refrigerator	Article 25: functioning of living with an adequate housing in terms of material.	
STOVE_D	Deprived (STOVE_D=1) if the household does not have a gas/electric stove	Article 25: functioning of living with an adequate housing in terms of material.	
WM_D	Deprived (WM_D=1) if the household does not have a washing machine	Article 25: functioning of living with an adequate housing in terms of material.	

IRON_D	Deprived (IRON_D=1) if the household does not have an iron	Article 25: functioning of living with an adequate housing in terms of material.	
VC_D	Deprived (AIR_D=1) if the household does not have a vacuum cleaner	Article 25: functioning of living with an adequate housing in terms of material.	
COMP_D	Deprived (COMP_D=1) if the household does not have a computer	Article 25: functioning of living with an adequate housing in terms of material.	
EDU_D	Deprived (EDU_D=1) if EDU is 0,1,2,3,4 (less than high school level diploma)	Article 26: functioning of having an educational attainment[198].	
MC_D	Deprived (MC_D=1) if the person is not covered by a medical coverage: no affiliation in Popular Insurance neither in other institution	Article 22: functioning of benefiting social security. Article 25: functioning of the right to medical care for a healthy life.	This variable is composed of the 2 variables INS1 and MC1 (see section 4.2)
AF1_D	Deprived (AF1_D=1) if in the last 3 months, due to lack of money or resources, the household has run out of food (i.e., AF1 is 1)	Article 25: functioning of living with an adequate nutrition (food).	
AF2_D	Deprived (AF2_D=1) if in the last 3 months, due to lack of money or resources, the household head worried about the food running out for the entire household (i.e., AF2 is 1)	Article 25: functioning of living with an adequate nutrition (food).	

[198] Article 26 mentions that "Everyone has the right to education. Education shall be free, at least in the elementary and fundamental stages. Elementary education shall be compulsory. Technical and professional education shall be made generally available and higher education shall be equally accessible to all on the basis of merit". However, we argue that at least a general high school diploma or its equivalent in terms of vocational technical high school is a fundamental necessary for people to obtain opportunities for higher professional education and specific job skills.

AF3_D	Deprived (AF3_D=1) if in the last 3 months, the household has run out of money ore resources to get a healthy and varied diet (i.e., AF3 is 1)	Article 25: functioning of living with an adequate nutrition (food).	

Regarding the overall deprivation index 'DEP' we built based on 32 variables above, 0.03 per cent of the target population (18+) has no deprivation whilst 0.01 percent of them has deprivation in 32 items. The most frequent groups are those having deprivation in 10 or 11 items, which represent 6.5, and 6.4 per cent of the population, respectively. We drop 48 persons having 0 deprivation (the most prosperous group) in the data. After that, there are in total 139,586 persons in the data. The summary statistics of DEP is as follows:

```
                    Total Deprivation
-------------------------------------------------------------
      Percentiles      Smallest
 1%             3              1
 5%             5              1
10%             6              1       Obs              139538
25%             9              1       Sum of Wgt.      139538

50%            13                      Mean           13.12541
                        Largest        Std. Dev.        5.7585
75%            17             32
90%            21             32       Variance       33.16032
95%            23             32       Skewness       .3227952
99%            27             32       Kurtosis       2.467471
```

As we did in previous country analysis, here also we apply the same methodology and we built a deprivation groups variable namely DEP_GRS regrouping the population by 20th, 40th, 60th, 80th, and 100th percentiles. 1st group (DEP_GRS1) includes people having deprivation in 1 to 7 items; 2nd group (DEP_GRS2) includes people having deprivation in 8 to 10 items: 3rd group (DEP_GRS3) includes people having deprivation in 11 to 13 items; 4th group (DEP_GRS4) includes people having deprivation in 14 to 17 items; and 5th group (DEP_GRS5) includes people having deprivation in 18 to 32 items. When we take into account simultaneously 32 deprivation variables (e.g., HEATS_D, SAN2_D-see the table of deprivation lines above) to find out the capabilities profiles of 139,528 individuals in the data (the variable named PATH32), the results show that there are 47,421 different profiles which is relatively high amount. For instance, with the most frequent 5 profiles (with at least 100 observations each of them), we can only capture 2.2 per cent of the population. But also, in terms of profiles with at least 100 observations, only 45 of 47,421 profiles surface yielding information for 7 per cent of the population as follows:

```
+-------------------------------------------------------------+
|                           PATH32   Freq.   Percent    %<= |
|-------------------------------------------------------------|
| 00000000001100100001000001000000    1222      0.88   0.88 |
| 00000000001100100001000000000000     546      0.39   1.27 |
| 00000000001100100011000001000000     479      0.34   1.61 |
| 00000000001100000001000001000000     456      0.33   1.94 |
| 00000000001100110111000001110000     357      0.26   2.19 |
```

00000000001100100001000001001000	303	0.22	2.41
00001000001100100001000001000000	294	0.21	2.62
00000000001100100001000001010000	288	0.21	2.83
00000000001000100001000001000000	275	0.20	3.02
00000000001100000001000000000000	260	0.19	3.21
00000000001100101111000001110000	253	0.18	3.39
00000000001100100011000001010000	250	0.18	3.57
00000000001100100000000001000000	246	0.18	3.75
00000000001100110111000001100000	233	0.17	3.91
00000000001100100111000001110000	219	0.16	4.07
00000000101100110111000001110000	207	0.15	4.22
00000000001100100000000000000000	205	0.15	4.37
00000000001100000011000001000000	192	0.14	4.50
00000000001100100001000000001000	187	0.13	4.64
00000000001100110101000001000000	169	0.12	4.76
00001000001100100011000001000000	167	0.12	4.88
00000000001000100001000000000000	158	0.11	4.99
00000000001100100101000001100000	154	0.11	5.10
00000000001100100101000001110000	153	0.11	5.21
00000000001110000001000001000000	153	0.11	5.32
00000000001100110111000001110011	149	0.11	5.43
00000000001100100011000001001000	137	0.10	5.53
00000000001100110101000001100000	136	0.10	5.62
00000000001100100111000001100000	135	0.10	5.72
00000000001100000000000001000000	134	0.10	5.82
00001000001100100001000000000000	133	0.10	5.91
00000000001100000001000001001000	132	0.09	6.01
00000000001100110101000001110000	129	0.09	6.10
00000000001100110111000001000000	127	0.09	6.19
00000000101100000001000001000000	119	0.09	6.28
00000000000100000001000000000000	117	0.08	6.36
00000000001100100101000001000000	117	0.08	6.44
00000000001100000001000001010000	116	0.08	6.53
00000000001100110001000001000000	116	0.08	6.61
00000000101100110111000001110011	116	0.08	6.69
00000000001100100111000001000000	114	0.08	6.77
00000000001000000001000001000000	111	0.08	6.85
00000000101100100001000001000000	111	0.08	6.93
00000000000000000001000001000000	110	0.08	7.01
00000000001100000000000000000000	108	0.08	7.09

The table below shows the number of persons in the group, unique capabilities profiles based on the 32 variables and the ratio of unique profiles by the number of persons in the group.

Table 59. Ratio of unique profiles to the number of persons in the deprivation group (based on the variable DEP_GRS and 32 variables), data: Mexico MCS 2012

Group	Unique capabilities profiles	Number of persons in the group	Ratio (Unique profiles/Number of persons in the group)
DEP_GRS1 (1-7 deprivations)	3,727	25,865	0.144
DEP_GRS2 (8-10 deprivations)	7,009	24,798	0.282
DEP_GRS3 (11-13 deprivations)	9,363	26,240	0.356
DEP_GRS4 (14-17 deprivations)	12,611	29,746	0.423
DEP_GRS5 (18-32 deprivations)	14,711	32,889	0.447
Total (whole sample)	**47,421**	**139,538**	**0.339**

As we can see from the data there is clearly a large variety of living conditions people face in their life. Within first group, the most frequent 10 profiles represent 17 per cent of the group; within second group, the most frequent 10 profiles capture 7.4 per cent of the group; within third group, the most frequent 10 profiles provide information for 4 per cent of the group; within fourth group, the most frequent 7 profiles we obtain information for 1.6 per cent of the group; and within fifth group, the most frequent 10 profiles lead to capture 1.4 per cent of the group. The analysis of capabilities, therefore, becomes quite complicated due to this plurality. In next paragraphs, we will show how to deal with such complexity by employing classification trees method. The goal is to find out the most informative capabilities patterns distinguishing the 5 deprivation groups. More specifically, we will do in total 4 classification tree analysis predicting how the deprivation groups distinguish among them by 32 deprivation variables of actual capabilities: {DEP_GRS1 vs. DEP_GRS3, DEP_GRS1 vs. DEP_GRS4, DEP_GRS1 vs. DEP_GRS5, DEP_GRS2 vs. DEP_GRS5}. As we have experienced in previous case studies, the analysis between closely neighbor groups does not provide good prediction; so that, we focus only on 4 cases rather than 10[199].

[199] In Turkey SILC 2011 analysis, we also analyzed GR2 vs. GR4, and GR3 vs. GR5; however, we saw that the predictions are less accurate compared to the 4 mentioned above. For GR2, the best results come from GR2 vs. GR5, and for GR3, the best results come from GR1 vs. GR3 analysis. Such result was also found in South Africa GHS 2011 examination as well.

Table 60. Per centage of groups in the sample of classification target groups, data: Mexico MCS 2012

Target group in classification	Per centage of groups in the sample
DEP_GRS1 vs. DEP_GRS3	{49.6, 50.4}
DEP_GRS1 vs. DEP_GRS4	{46.6, 53.4}
DEP_GRS1 vs. DEP_GRS5	{44, 56}
DEP_GRS2 vs. DEP_GRS5	{43, 57}

As the table above shows we have a relatively balanced data but still we don't have a 100 per cent equal distribution between two groups.

CAP112: GR1 vs. GR3 analysis

In this analysis, we focus on DEP_GRS1 vs. DEP_GRS3. We use the data containing only people from the deprivation group 1 and the deprivation group 3. The outcome/target response variable is dichotomous, namely CAP112, taking the value 1 for the group 1 and the value 3 for the group 3 (so that people identified in the group 1 take the value 1 and people identified in the group 3 take the value 3 according to CAP112).

Table 61. Training vs. test samples in CAP112 analysis, data: Mexico MCS 2012

	GR1	GR3
Training sample	17,243	17,493
Test sample	8,622	8,747

As the table above shows, there are 34,736 individuals in the training data (corresponding to 67 per cent of the initial data) and 17,369 individuals in the test sample. Training and test data are obtained by randomly partitioning the initial data/total sample into two subsets by 70 per cent vs. 30 per cent respectively. The data is quite balanced with 49.6 per cent for GR1 and 50.4 per cent for GR3 in both training and test samples (in the initial data, the proportion is also the same).

Concerning the rpart application we have set cp at 0.025 and we used gini index as split criterion, and we kept other parameters at their default values. Classification rules are discovered on the training data and they will be tested on the test data for evaluating predictive accuracy. The classification tree is called CAP112, and the model is named dtreeCAP112. The results in the box below show that the initial saturated tree given the parameters we used contains 16 splits (i.e., 17 terminal nodes) with a cross-validation error rate around 0.105. This tree is quite big and therefore difficult for interpretation, so we will prune the tree to around 10-11 splits to obtain an optimal interpretation -which yields also a relatively low level of cross-validation error rate (around 0.13).

```
Classification tree:
rpart(formula = CAP112 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
 [1] AF2_D  AF3_D  CAR_D  COMP_D HOTW_D INT_D  SAN1_D SINK_D TEL_D  WATD_D

Root node error: 17243/34736 = 0.4964

n= 34736

          CP nsplit rel error  xerror      xstd
1  0.7289915      0   1.00000 1.00000 0.0054043
2  0.0368845      1   0.27101 0.27101 0.0036882
3  0.0241837      2   0.23412 0.23412 0.0034641
4  0.0192542      4   0.18576 0.18576 0.0031272
5  0.0114539      5   0.16650 0.16650 0.0029763
6  0.0043496      7   0.14359 0.14359 0.0027810
7  0.0041756      8   0.13924 0.13988 0.0027476
8  0.0040306     13   0.11332 0.13258 0.0026800
9  0.0039436     15   0.10526 0.11947 0.0025530
10 0.0025000     16   0.10132 0.10578 0.0024109
```

The CAP112 pruned tree, is composed of 14 terminal nodes providing 4 patterns of GR1, and 10 patterns of GR3. As we can see from the pruned tree below, the discovered patterns from the left to the right side of the tree are as follows: (1) no deprivation regarding having an internet (INT), a sink in the dwelling (SINK), and run out of money or resources to get a healthy and varied diet in the household (AF3); (2) no deprivation regarding INT, SINK, a car (CAR) in the dwelling, and deprivation regarding AF3; (3) no deprivation regarding INT, SINK, and deprivation regarding CAR, and AF3; (4) no deprivation regarding INT, and deprivation regarding SINK; (5) no deprivation regarding COMP, and deprivation regarding INT, and having a hot water system (HOTW) in the dwelling; (6) no deprivation regarding COMP, HOTW, and deprivation regarding INT, and worrying about the food running out due to lack of money or resources in the household (AF2); (7) no deprivation regarding COMP, HOTW, SINK, and AF2, and deprivation regarding INT; (8) no deprivation regarding COMP, HOTW, and AF2, and deprivation regarding INT, and SINK; (9) deprivation regarding INT, COMP, and CAR; (10) no deprivation regarding CAR, and deprivation regarding INT, COMP, and a landline telephone (TEL) in the dwelling; (11) no deprivation regarding CAR, TEL, but deprivation regarding IT, COMP, and AF2; (12) no deprivation regarding CAR, TEL, AF2, and deprivation regarding INT, COMP, and HOTW; (13) no deprivation regarding CAR, TEL, HOTW, AF2, and having a regular piped drinking water system (WATD), and deprivation regarding INT, and COMP; (14) no deprivation regarding CAR, TEL, HOTW, AF2, and deprivation regarding COMP, INT, WATD.

The patterns of GR1 are 1st, 7th, 13th, and 2nd with 81, 8, 3, and 2.8 per centage of coverage respectively (see the table below). For instance, with the 1st pattern, 81 per cent of people from GR1 are detected. On the other hand, the patterns of GR3 are 9th, 10th, 5th, 4th, 3rd, 6th, 11th, 12th, 14th, and 8th with 63, 11, 6, 5, 3, 2.5, 1.5, 1, 0.6, and 0.5 per centage of coverage respectively (see the table below). For instance, with the 9th pattern, 63 per cent of people from GR3 are

captured. In the training data, by using the CAP112 pruned tree classification patterns, we lose 5 per cent of information for both GR1 and GR3.

Table 62. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP112 pruned tree, data: Mexico MCS 2012

Pattern no	Coverage of pattern (in %) among the total population of GR1	Cumulative coverage (in %)
1	81	81
7	8	89
13	3	92
2	2.8	94.8~**95**

Table 63. Coverage of GR3's patterns among the total population of GR3 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP112 pruned tree, data: Mexico MCS 2012

Pattern no	Coverage of pattern (in %) among the total population of GR3	Cumulative coverage (in %)
9	63	63
10	11	74
5	6	80
4	5	85
3	3	88
6	2.5	90.5
11	2.5	93
12	1	94
14	0.6	94.6
8	0.5	95.1~**95**

We can also say that through the 4 patterns of GR1, deprivation regarding AF3, INT, COMP concern 11, 3, 2.8 per cent of the entire population of GR1 (among 17,243 persons) respectively. On the other hand, concerning GR3, the 10 patterns show that the variables INT, COMP, CAR, TEL, HOTW, SINK, AF2, AF3, and WATD affect 87, 78, 66, 11, 7, 5.5, 5, 3, and 0.6 per cent of the entire population of GR3 (among 17,493 persons). The 3 deprivation variables, INT, COMP, and AF3 are the common in terms of deficits for both GR1 and GR3; however, the magnitude of the two former is much higher for GR3, compared to GR1.

```
Confusion Matrix and Statistics

          Reference
Prediction  GR1  GR3
       GR1 8027  402
       GR3  595 8345

               Accuracy : 0.9426
                 95% CI : (0.939, 0.946)
    No Information Rate : 0.5036
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.8852

 Mcnemar's Test P-Value : 1.197e-09

            Sensitivity : 0.9310
            Specificity : 0.9540
         Pos Pred Value : 0.9523
         Neg Pred Value : 0.9334
             Prevalence : 0.4964
         Detection Rate : 0.4621
   Detection Prevalence : 0.4853
      Balanced Accuracy : 0.9425

       'Positive' Class : GR1
```

Regarding confusion matrix statistics, this model has a good overall accuracy of around 94 per cent, and the rate of sensitivity is 93 per cent, and the rate of specificity is 95 per cent.

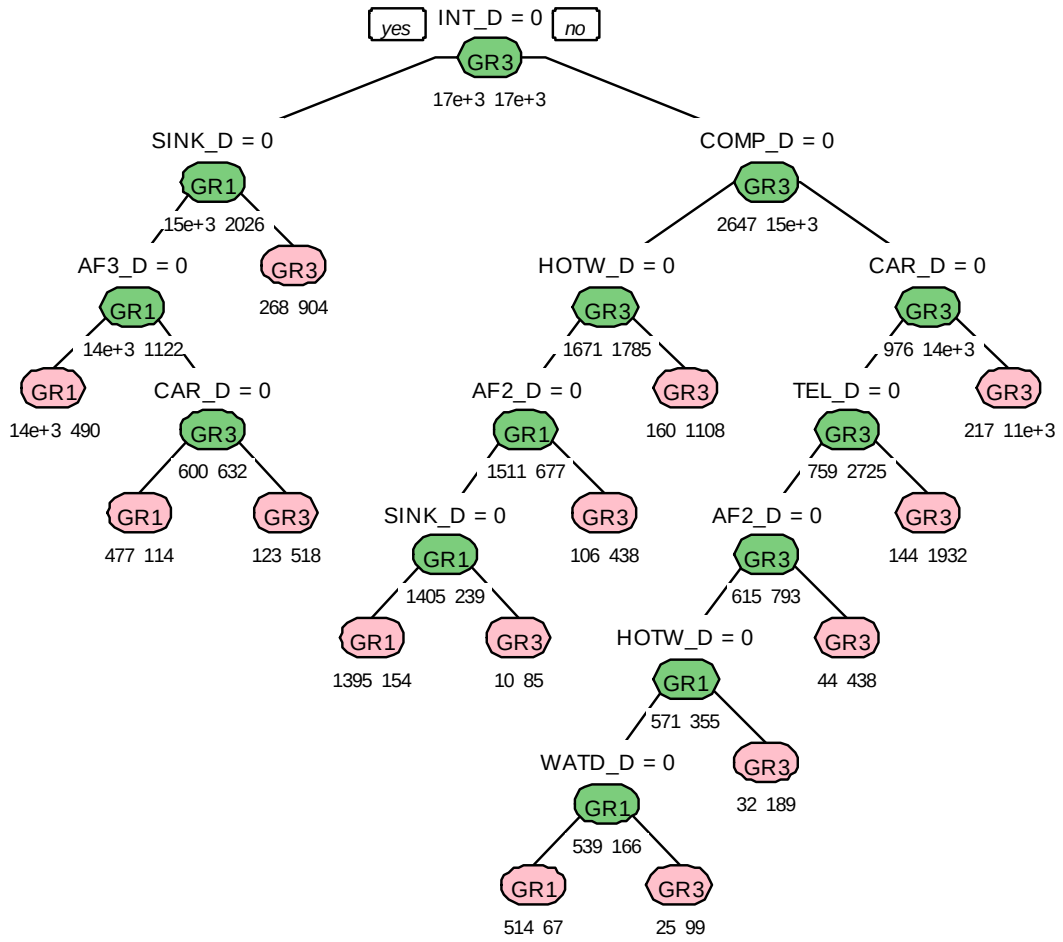


Figure 27. CAP112 pruned tree, data: Mexico MCS 2012

CAP113: GR1 vs. GR4 analysis

In this analysis, we focus on DEP_GRS1 vs. DEP_GRS4. We use the data containing only people from the deprivation group 1 and the deprivation group 4. The outcome/target response variable is dichotomous, namely CAP113, taking the value 1 for the group 1 and the value 4 for the group 4 (so that people identified in the group 1 take the value 1 and people identified in the group 4 take the value 4 according to CAP113).

Table 64. Training vs. test samples in CAP113 analysis, data: Mexico MCS 2012

	GR1	GR4
Training sample	17,243	19,830
Test sample	8,622	9,916

As the table above shows, there are 37,073 individuals in the training data (corresponding to 67 per cent of the initial data) and 18,538 individuals in the test sample. Training and test data are obtained by randomly partitioning the initial data/total sample into two subsets by 70 per cent

vs. 30 per cent respectively. The data is very balanced with 46.6 per cent for GR1 and 53.4 per cent for GR4 in both training and test samples (in the initial data, the proportion is the same as well).

In first attempt of fitting the model based on the cp equal to 0.025, we obtained a very good tree with 12 splits generating almost zero cross-validated error rate. As we can see from the information provided in the box below, for a tree with 10 splits, the cross-validated error rate is still approximately zero. For that reason, we proceeded with pruning phase to obtain a tree with 11 terminal nodes.

```
> printcp(dtreeCAP113)

Classification tree:
rpart(formula = CAP113 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
[1] AF3_D  CAR_D  COMP_D HOTW_D INT_D  SAN1_D SINK_D

Root node error: 17243/37073 = 0.46511

n= 37073

          CP nsplit rel error   xerror      xstd
1  0.8227107      0  1.000000 1.000000 0.0055696
2  0.0420460      1  0.177289 0.177289 0.0030715
3  0.0296352      2  0.135243 0.136461 0.0027224
4  0.0118599      3  0.105608 0.106826 0.0024264
5  0.0085832      6  0.069999 0.071217 0.0019984
6  0.0068434      7  0.061416 0.063794 0.0018947
7  0.0063794      8  0.054573 0.056661 0.0017887
8  0.0044656      9  0.048193 0.048077 0.0016510
9  0.0037117     10  0.043728 0.044772 0.0015945
10 0.0028997     11  0.040016 0.041060 0.0015283
11 0.0025000     12  0.037117 0.037580 0.0014633
```

The CAP113 pruned tree is composed of 11 terminal nodes providing 4 patterns of GR1, and 7 patterns of GR4. As we can see from the pruned tree below, the discovered patterns from the left to the right side of the tree are as follows: (1) no deprivation regarding having a computer (COMP), a sink (SINK), and a shower (SAN1) in the dwelling; (2) no deprivation regarding COMP, and SINK, and deprivation regarding SAN1; (3) no deprivation regarding COMP, having a hot water system (HOTW), and an internet (INT) in the dwelling, and deprivation regarding SINK; (3) no deprivation regarding COMP, HOTW, and INT, but deprivation regarding SINK; (4) no deprivation regarding COMP, and HOTW, but deprivation regarding SINK, and INT; (5) no deprivation regarding COMP, but deprivation regarding SINK, and HOTW; (6) no deprivation regarding INT, and SINK, but deprivation regarding COMP; (7) no deprivation regarding INT, but deprivation regarding COMP, and SINK; (8) deprivation regarding COMP, INT, and having a car in the household (CAR); (9) no deprivation regarding CAR, but deprivation regarding CAR, but deprivation regarding COMP, INT, and HOTW; (10) no deprivation regarding CAR, HOTW, and running out of money or resources to get a healthy and varied diet in the dwelling (AF3); (11) no deprivation regarding CAR, and HOTW, but deprivation regarding COMP, INT, and AF3.

The patterns of GR1 are 1st, 6th, 10th, and 3rd with 87, 4.3, 4, and 1.25 per centage of coverage respectively (see the table below). For instance, with the 1st pattern, 87 per cent of people from GR1 are detected. On the other hand, the patterns of GR4 are 8th, 9th, 5th, 11th, 7th, 2nd, and 4th with 86, 6.8, 4.5, 1, 0.8, 0.7, and 0.2 per centage of coverage respectively (see the table below). For instance, with the 8th pattern, 86 per cent of people from GR4 are captured.

Table 65. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP113 pruned tree, data: Mexico MCS 2012

Pattern no	Coverage of pattern (in %) among the total population of GR1	Cumulative coverage (in %)
1	87	87
6	4.3	91.3
10	4	94.3
3	1.25	95.55~**95**

Table 66. Coverage of GR4's patterns among the total population of GR4 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP113 pruned tree, data: Mexico MCS 2012

Pattern no	Coverage of pattern (in %) among the total population of GR4	Cumulative coverage (in %)
8	86	86
9	6.8	92.8
5	4.5	97.3
11	1	98.3
7	0.8	99.1
2	0.7	99.8
4	0.2	**100**

In the training data, by using the CAP113 pruned tree classification patterns, we lose only 4 per cent of information for GR1 and 0 per cent of information for GR4.

We can also say that through the 4 patterns of GR1, deprivation regarding SINK, COMP, and INT concern 8.3, 4, and 1.25 per cent of the entire population of GR1 (among 17, 243 persons) respectively. On the other hand, concerning GR4, the 7 patterns show that the variables COMP, INT, CAR, HOTW, SINK, AF3, and SAN1 affect 95, 94, 86, 11, 5.7, 1, 0.7 per cent of the entire population of GR4 (among 19,830 persons). The 3 deprivation variables, COMP, INT, and SINK are the common in terms of deficits for both GR1 and GR4; however, the magnitude of the two formers are very higher for GR4 compared to GR1.

Regarding the overall accuracy of the mode, as we can see from the confusion matrix statistics, prediction on the test data is very good generating around 98 per cent of accuracy, and

sensitivity and specificity rates vary between around 97.5 and 98 per cent, and misclassification rates are very small.

```
Confusion Matrix and Statistics

          Reference
Prediction  GR1  GR4
       GR1 8412  188
       GR4  210 9728

               Accuracy : 0.9785
                 95% CI : (0.9763, 0.9806)
    No Information Rate : 0.5349
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.9568

 Mcnemar's Test P-Value : 0.2925

            Sensitivity : 0.9756
            Specificity : 0.9810
         Pos Pred Value : 0.9781
         Neg Pred Value : 0.9789
             Prevalence : 0.4651
         Detection Rate : 0.4538
   Detection Prevalence : 0.4639
      Balanced Accuracy : 0.9783

       'Positive' Class : GR1
```

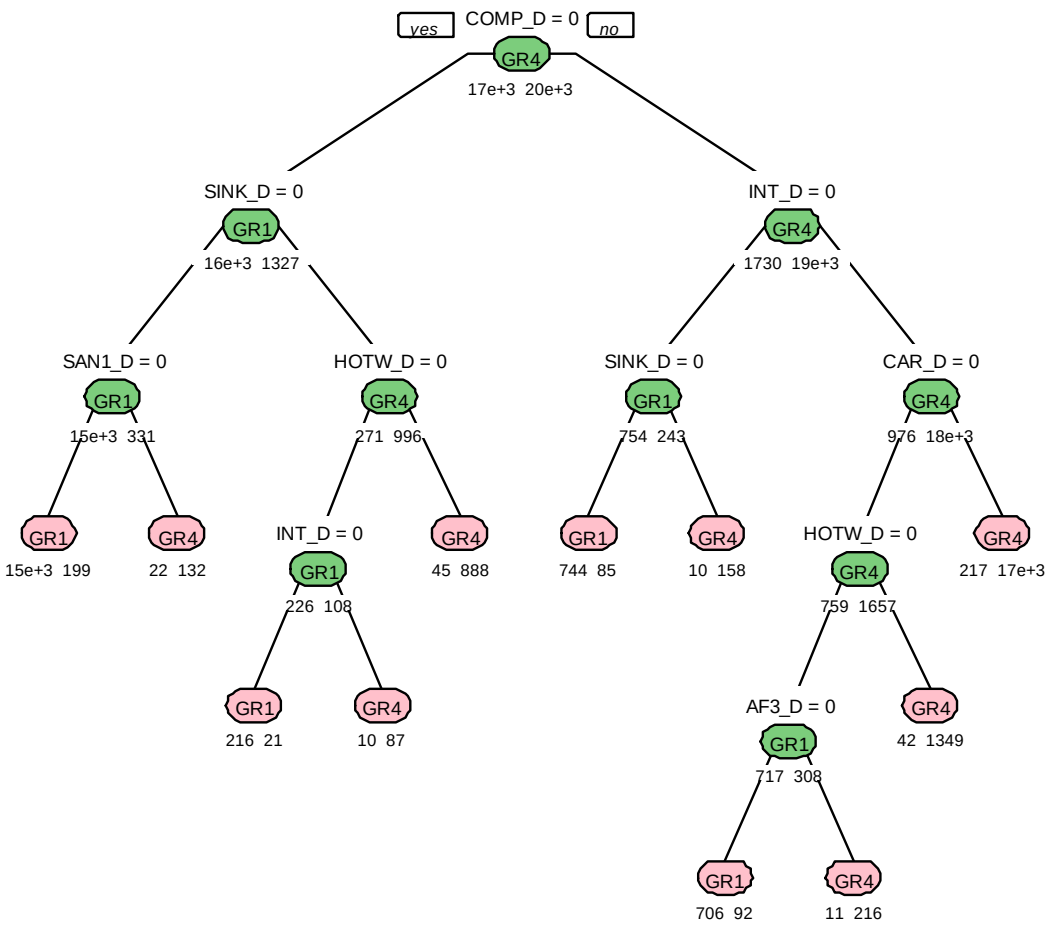


Figure 28. CAP113 pruned tree, data: Mexico MCS 2012

CAP114: GR1 vs. GR5 analysis

In this analysis, we focus on DEP_GRS1 vs. DEP_GRS5. We use the data containing only people from the deprivation group 1 and the deprivation group 5. The outcome/target response variable is dichotomous, namely CAP114, taking the value 1 for the group 1 and the value 5 for the group 5 (so that people identified in the group 1 take the value 1 and people identified in the group 5 take the value 5 according to CAP114).

Table 67. Training vs. test samples in CAP114 analysis, data: Mexico MCS 2012

	GR1	GR5
Training sample	17,243	21,926
Test sample	8,622	10,963

As the table above shows, there are 39,169 individuals in the training data (corresponding to 67 per cent of the initial data) and 19,585 individuals in the test sample. Training and test data are obtained by randomly partitioning the initial data/total sample into two subsets by 70 per cent

vs. 30 per cent respectively. The data is relatively balanced with 44 per cent for GR1 and 54 per cent for GR5 in both training and test samples (in the initial data, the proportion is also the same).

Fitting this model with the initial default parameters, and a cp equal to 0.025 has directly produced a not saturated tree with only 8 splits that corresponds to what we are looking for a synthetic information, and with a cross-validated error rate almost zero. Thus, we are also expecting to obtain a very good overall prediction on the test data, and small misclassification rates on the training data. As the initial tree is perfectly match with our requirements, we don't continue with pruning process. The box below contains detailed information on the results of the model.

```
> printcp(dtreeCAP114)

Classification tree:
rpart(formula = CAP114 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
[1] AF1_D  COMP_D INT_D  RUB_D  SAN1_D SINK_D

Root node error: 17243/39169 = 0.44022

n= 39169

         CP nsplit rel error   xerror       xstd
1 0.8915502      0 1.0000000 1.000000 0.00569773
2 0.0637940      1 0.1084498 0.108450 0.00244729
3 0.0112509      2 0.0446558 0.044656 0.00159339
4 0.0111929      3 0.0334049 0.036943 0.00145177
5 0.0042916      4 0.0222119 0.022212 0.00112941
6 0.0037117      5 0.0179203 0.017804 0.00101216
7 0.0027257      6 0.0142087 0.014209 0.00090491
8 0.0025000      8 0.0087572 0.010381 0.00077414
```

The CAP114 pruned tree is composed of 9 terminal nodes providing 4 patterns of GR1, and 4 patterns of GR5. As we can see from the tree below, the discovered patterns from the left to the right side of the tree are as follows: (1) no deprivation regarding having a sink (SINK), a shower (SAN1) in the dwelling, running out of food due to lack of money or resources (AF1), and removal of rubbish/garbage of the house (RUB); (2) no deprivation regarding SINK, SAN1, AF1, and having a computer (COMP) in the dwelling, but deprivation regarding RUB; (3) no deprivation regarding SINK, SAN1, AF1, but deprivation regarding RUB, and COMP; (4) no deprivation regarding SINK, SAN1, COMP, but deprivation regarding AF1; (5) no deprivation regarding SINK, SAN1, but deprivation regarding AF1, and COMP; (6) no deprivation regarding SINK, but deprivation regarding SAN1; (7) no deprivation regarding INT, and SAN1, but deprivation regarding SINK; (8) no deprivation regarding INT, but deprivation regarding SINK, and SAN1; (9) deprivation regarding SINK, and INT.

The patterns of GR1 are 1st, 7th, 2nd, and 4th with 98.6, 1.5, 0.4, and 0.4 per centage of coverage respectively (see the table below). For instance, with the 1st pattern, 98.6 per cent of people from GR1 are detected. On the other hand, the patterns of GR5 are 9th, 6th, 5th, 3rd, and 8th with 91, 5, 1.2, 0.45, and 0.3 per centage of coverage respectively (see the table below). For instance, with the 9th pattern, 91 per cent of people from GR5 are captured.

Table 68. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP114 tree, data: Mexico MCS 2012

Pattern no	Coverage of pattern (in %) among the total population of GR1	Cumulative coverage (in %)
1	98.6	98.6
7	1.5	~**100**
2	~0	100
4	~0	100

Table 69. Coverage of GR5's patterns among the total population of GR5 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP114 tree, data: Mexico MCS 2012

Pattern no	Coverage of pattern (in %) among the total population of GR5	Cumulative coverage (in %)
9	91	91
6	5	96
5	1.2	97.2
3	0.45	97.65
8	0.3	97.95~**98**

In the training data, by using the CAP114 tree classification patterns, we lose only 2 per cent of information for GR5 and 0 per cent of information for GR1.

We can also say that through the 4 patterns of GR1, deprivation regarding SINK concern 1.5 per cent of the entire population of GR1 (among 17,243 persons). On the other hand, concerning GR5, the 5 patterns show that the variables SINK, INT, SAN1, COMP, AF1, and RUB affect 91.3, 91, 5.3, 1.65, 1.2, and 0.45 per cent of the entire population of GR5 (among 21,926 persons). The deprivation variables, SINK, is the common in terms of deficits for both GR1 and GR5; however, its magnitude is significantly higher for GR5 compared to GR1.

CAP114 tree

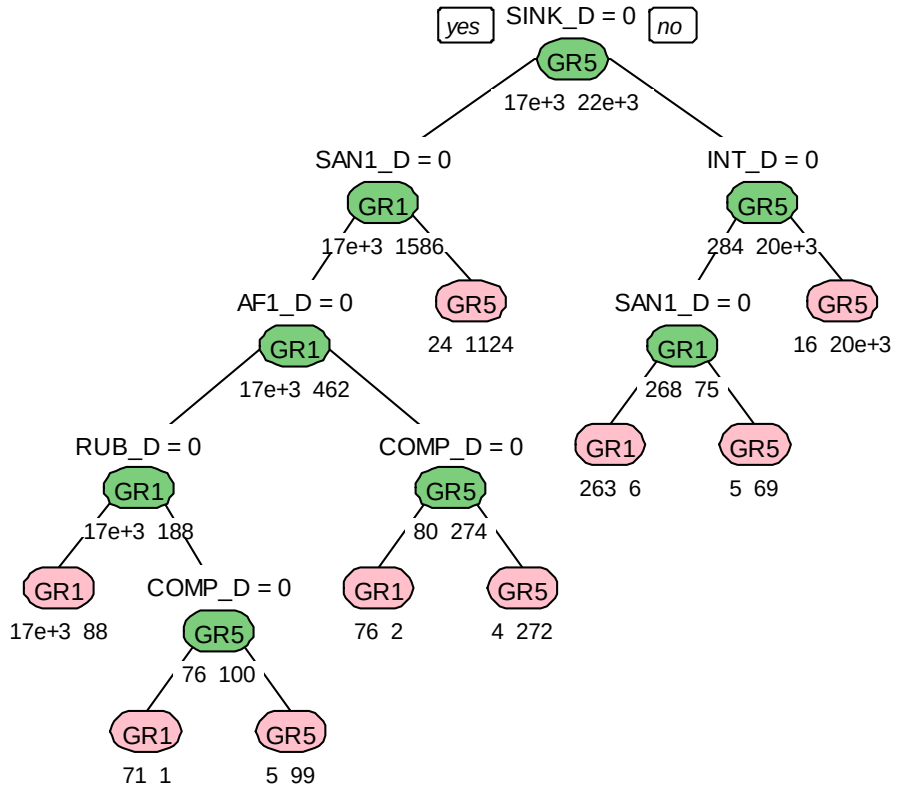


Figure 29. CAP114 tree, data: Mexico MCS 2012

Regarding the overall accuracy of the mode, as we can see from the confusion matrix statistics, prediction on the test data is perfectly good generating almost 99.6 per cent of accuracy, and sensitivity and specificity rates are around 99 per cent, and misclassification rates are very small.

```
   Confusion Matrix and Statistics

          Reference
Prediction   GR1   GR5
       GR1  8593    43
       GR5    29 10920

               Accuracy : 0.9963
                 95% CI : (0.9954, 0.9971)
    No Information Rate : 0.5598
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.9925

 Mcnemar's Test P-Value : 0.1255

            Sensitivity : 0.9966
            Specificity : 0.9961
         Pos Pred Value : 0.9950
         Neg Pred Value : 0.9974
```

```
            Prevalence : 0.4402
        Detection Rate : 0.4388
  Detection Prevalence : 0.4409
     Balanced Accuracy : 0.9964

      'Positive' Class : GR1
```

CAP117: GR2 vs. GR5 analysis

In this analysis, we focus on DEP_GRS2 vs. DEP_GRS5. We use the data containing only people from the deprivation group 2 and the deprivation group 5. The outcome/target response variable is dichotomous, namely CAP117, taking the value 2 for the group 2 and the value 5 for the group 5 (so that people identified in the group 2 take the value 2 and people identified in the group 5 take the value 5 according to CAP117).

Table 70. Training vs. test samples in CAP117 analysis, data: Mexico MCS 2012

	GR2	GR5
Training sample	16,532	21,926
Test sample	8,266	10,963

As the table above shows, there are 38,458 individuals in the training data (corresponding to 67 per cent of the initial data) and 19,229 individuals in the test sample. Training and test data are obtained by randomly partitioning the initial data/total sample into two subsets by 70 per cent vs. 30 per cent respectively. The data is relatively balanced with 43 per cent for GR2 and 57 per cent for GR5 in both training and test samples (in the initial data, the proportion is also the same).

With the default parameters and a cp equal to 0.025, we obtain initially a saturated tree with 9 splits, and a cross-validated error rate almost zero as indicated in the box below. Thus, we are also expecting to obtain a very good overall prediction on the test data, and small misclassification rates on the training data. As the initial tree is perfectly match with our requirements, we don't continue with pruning process. The box below contains detailed information on the results of the model.

```
Classification tree:
rpart(formula = CAP117 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
[1] AF2_D  AF3_D  COMP_D DR_D   INT_D  REF_D  SAN1_D SINK_D WM_D

Root node error: 16532/38458 = 0.42987

n= 38458

         CP nsplit rel error   xerror      xstd
1 0.8233124      0  1.000000 1.000000 0.0058725
2 0.0485120      1  0.176688 0.176688 0.0031426
3 0.0113719      3  0.079664 0.079664 0.0021573
4 0.0105855      5  0.056920 0.061336 0.0019006
5 0.0047484      6  0.046334 0.050750 0.0017329
6 0.0035083      8  0.036838 0.036293 0.0014701
7 0.0025000      9  0.033329 0.035023 0.0014445
```

The CAP117 pruned tree is composed of 10 terminal nodes providing 5 patterns of GR2, and 5 patterns of GR5. As we can see from the pruned tree below, the discovered patterns from the left to the right side of the tree are as follows: (1) no deprivation regarding having a shower (SAN1), a sink (SINK), and a refrigerator (REF) in the dwelling; (2) no deprivation regarding SAN1, SINK, and running out of money or resources to get a healthy and varied diet (AF3), and deprivation regarding (REF); (3) no deprivation regarding SAN1, SINK, but deprivation regarding REF, and AF3; (4) no deprivation regarding SAN1, and having a computer (COMP) in the dwelling, but deprivation regarding SINK; (5) no deprivation regarding SAN1, but deprivation regarding SINK, COMP, and worrying about the food running due to lack of money or resources (AF2); (6) no deprivation regarding SAN1, AF2, but deprivation regarding SINK, COMP, and having a washing machine (WM) in the dwelling; (7) no deprivation regarding SAN1, AF2, WM, and having an appropriate drainage system (DR) in the dwelling, but deprivation regarding SINK, and COMP; (8) no deprivation regarding SAN1, AF2, WM, but deprivation regarding SINK, COMP, and DR; (9) no deprivation regarding having an internet in the dwelling (INT), but deprivation regarding SAN1; (10) deprivation regarding SAN1, and INT.

The patterns of GR2 are 1st, 4th, 7th, 9th, and 2nd. The 1st pattern brings 90.7, the 4th captures 5.7 per cent of GR2 reaching in total 96 per cent of the group (the contribution of the other 3 patterns are not significant). On the other hand, the patterns of GR5 are 10th, 5th, 6th, 3rd, and 8th. The 10th pattern detects 86.65, and the 5th pattern captures 8 per cent of GR5 reaching in total 95 per cent of the group (the contribution of the other 3 patterns are not significant).

CAP117 tree

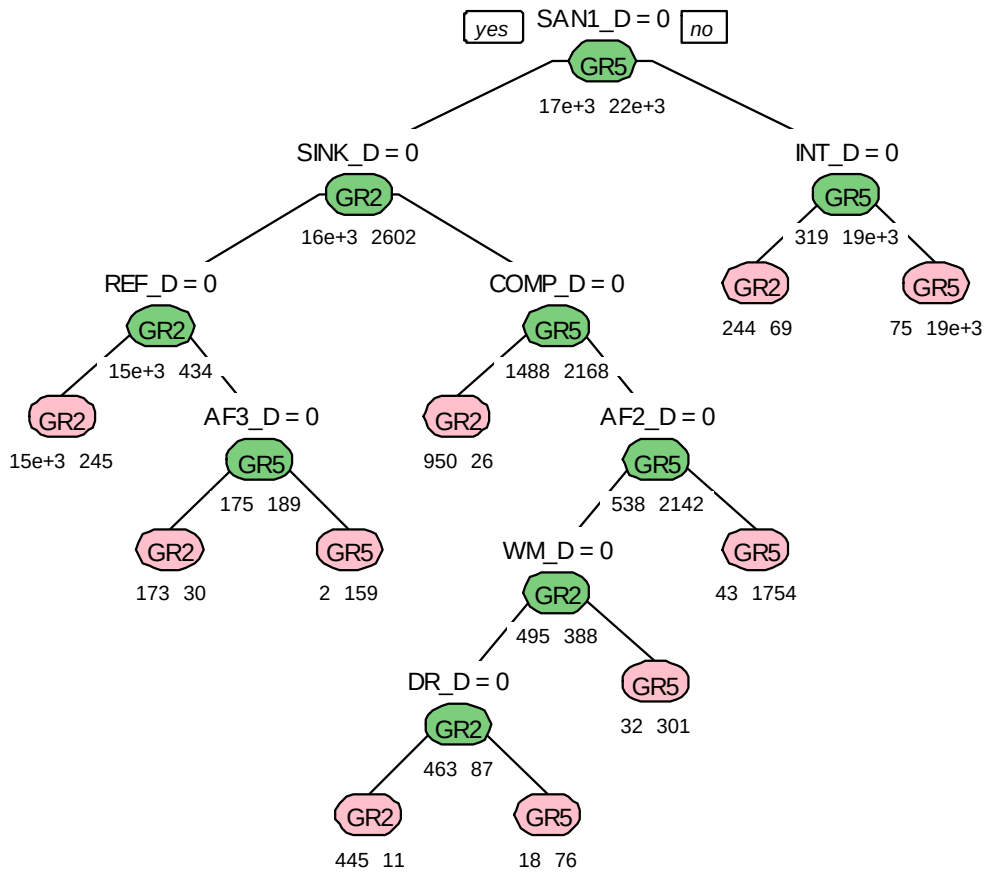


Figure 30. CAP117 pruned tree, data: Mexico MCS 2012

Table 71. Coverage of GR2's patterns among the total population of GR2 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP117 pruned tree, data: Mexico MCS 2012

Pattern no	Coverage of pattern (in %) among the total population of GR2	Cumulative coverage (in %)
1	90.7	90.7
4	5.7	**96**

Table 72. Coverage of GR5's patterns among the total population of GR5 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP117 pruned tree, data: Mexico MCS 2012

Pattern no	Coverage of pattern (in %) among the total population of GR5	Cumulative coverage (in %)
10	86.65	86.65
5	8	**95**

In the training data, by using the CAP117 tree classification patterns, we lose only 4 per cent of information for GR2 and 5 per cent of information for GR5.

We can also say that through the 2 patterns of GR2, deprivation regarding SINK concern 5.7 per cent of the entire population of GR2 (among 16,532 persons). On the other hand, concerning GR5, the 2 patterns show that the variables SAN1, INT, SINK, COMP, and AF2 affect 86.65, 86.65, 8,8,8 per cent of the entire population of GR5 (among 21,926 persons).

Regarding the overall accuracy of the model, as we can see from the confusion matrix statistics below, prediction on the test data is very good with 98 per cent of accuracy, and sensitivity and specificity rates are quite equal around 98-99 per cent, and misclassification rates are very small.

```
Confusion Matrix and Statistics

          Reference
Prediction   GR2   GR5
       GR2  8194   243
       GR5    72 10720

               Accuracy : 0.9836
                 95% CI : (0.9817, 0.9854)
    No Information Rate : 0.5701
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9667

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.9913
            Specificity : 0.9778
         Pos Pred Value : 0.9712
         Neg Pred Value : 0.9933
             Prevalence : 0.4299
         Detection Rate : 0.4261
   Detection Prevalence : 0.4388
      Balanced Accuracy : 0.9846

       'Positive' Class : GR2
```

Comparative summary

After presented in detail the results of classification tress through 5 different deprivation groups, we will provide in this sub-section a summary of the discovered patterns regarding content and predictive accuracy.

CAP112

Based on the training data of GR1 and GR3 in this model, we can see that people living in a dwelling having an internet and a sink, and did not experience run out of money or resources to get a healthy and varied diet in the last 3 months are most likely to be part of GR1. However, it is also found that people living in a dwelling lacking an internet but having a computer, a hot water system, a sink, and living without being ever worried about the food running out due to lack of money or resources in the last 3 months are likely to be in GR1. On the other hand, people living in a dwelling deprived of an internet, a computer, and a car are at most risk of being in GR3. Subsequently, people living in a dwelling deprived of an internet, a computer, a landline telephone but having a car in their household are also at risk of being in GR3. Moreover, people living in a dwelling deprived of an internet and a hot water system but having a computer, and people living in a dwelling deprived of a sink but having an internet are also slightly at risk of being in GR3 compared to GR1. We can say that a simultaneous deprivation in terms of internet, computer, and car represents the most important pattern of poverty experienced by GR3, but also it corresponds to a major inequality element to the detriment of GR3 compared to GR1.

CAP113

In this model, compared to CAP112, we expect to see more severe deprivation among GR4 as both the number of deprivations is higher for GR4 compared to GR3, and the inequality normally accentuates between GR1 and GR4 compared to what is happening between GR1 and GR3. It is found that people experiencing no deprivation regarding having a computer, a sink, and a shower in their dwelling are most likely to be in GR1. Besides, people experiencing a deprivation regarding having a computer in the dwelling but being not deprived regarding having an internet and a sink in the dwelling; and people living in a dwelling possessing a hot water system but not a computer neither an internet, and in a household not lacking a car and have not ever run out of money or resources to get a healthy and varied diet in the last 3 months are likely being in GR1. On the other hand, people living with a no computer, a no internet in their dwelling, a no car in their household are at most risk of being in GR4. Moreover, people lacking a computer, an internet, and a hot water system in their dwelling but having a car in their household; and people living in a dwelling lacking a sink and a hot water system but possessing a computer are at risk of being in GR4 compared to GR1. Similar to the main CAP112 result for GR3, here we can also observe that a simultaneous deprivation in terms of internet, computer, and car represents the most important pattern of poverty experienced by GR4, but also it corresponds to a major inequality element to the detriment of GR4 compared to GR1. But, the per centage of people experiencing such configuration of poverty is expected to be higher among GR4 compared to GR3.

CAP114

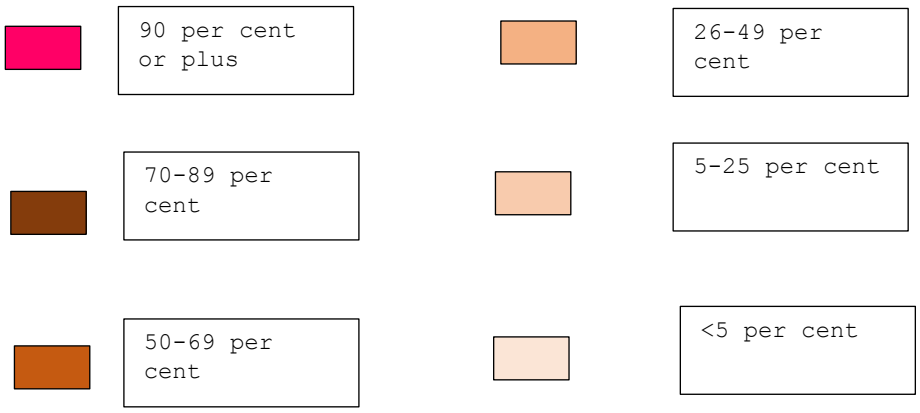
The findings for this model show that people living in a dwelling having a sink, a shower, and an accurate rubbish removal system, and in a household haven’t ever run out of food due to lack of money or resources in the last 3 months are great likely to be part of GR1. On the other hand, people living in a dwelling lacking a sink and an internet are at great risk of being part of GR5. Moreover, people living in a dwelling lacking a shower but not a sink are likely to be in GR5. Compared to GR3 and GR4, the combination of deprivation regarding a sink and an internet in the dwelling flourishes as the most important pattern of poverty distinguishing in the most remarkable way the poorest group (GR5) from the least poor group (GR1) in the population.

CAP117

It is found that people living in a dwelling having a shower, a sink, and a refrigerator are most likely to be in GR2. In addition to that, people living in a dwelling lacking a sink but not a shower and a computer are likely to be in GR2. On the other hand, people living in a dwelling lacking a shower and an internet are at most risk of being part of GR5. Moreover, people living in a dwelling having a shower but not a sink neither a computer, and in a household worried about the food running out due to lack of money or resources in the last 3 months are likely to be in GR5. Similar to the main CAP114 result for GR5, the poorest group greatly distinguishes from relatively less poor group (GR2) by living in a dwelling lacking a shower and an internet.

We remark that compared to previous South Africa and Turkey studies, in case of Mexico, we obtained better predictions with relatively higher overall accuracy as well as sensitivity and specificity rates on the test data. And, as we can see the discovered patterns reflect combinations of some material such as car, internet, computer, refrigerator, landline telephone and some essential living items such as sink, shower, hot water system, accurate rubbish removal system and access to food.

Figure 31 below is a kind of snapshot view on the overall importance of variables in terms of deprivation people experience regarding them on the training data across all employed models. There are 34 columns labelled as model, group, and variables-from WAL to AF3, respectively. The rows in the figure include the name of 4 models (e.g., CAP112) and the groups existing in the model/target groups (e.g., GR1 vs. GR3). The 6 different criteria, corresponding each of them to a different interval of per centage of population in the training data experiencing deprivation regarding the variable, are arbitrarily set as follows: 90 per cent +, 70-89 per cent, 50-69 per cent, 26-49 per cent, 5-25 per cent, and < 5 per cent. For instance, in CAP112 model, per centage of GR3 population in the training data suffering in terms of deprivation regarding the variable INT is in the interval of 70-89 per cent. Each criterion is presented in a different color as follows:



From the information above, we can see that the variables WAL, ROF, FLO, KIT, SAN2, ELC, HEATS, ENG_C, EDU, MC, AIRC, MPHO, RAD, TV, STOVE, IRON, and VC are not used in any selected tree to distinguish the deprivation groups. It is important to mention that the variables appearing in the selected trees are the most influential ones in terms of information gain to distinguish the two groups of the target response variable. However, it does not mean

that the groups in the model are not deprived regarding other variables that do not appear in the tree.

Based on the results displayed in Figure 31, we can observe that the poorest group (GR5) distinguishes greatly from the least poor groups (GR1 and GR2) in terms of the 2 essential sanitation tools must have in the dwelling that are a sink and a shower; but also in terms of access to internet in the dwelling. Otherwise, we can also say that some households of GR5 suffer regarding access to food, drainage system, rubbish removal, refrigerator, and washing machine compared to GR1 and GR2. When we look at the variables influentially emerging in GR1 vs. GR4 analysis, we can see that GR4 experiences greater deprivation in terms of having a car, a computer, and an access to internet services in their dwelling/household compared to GR1. Given that GR5, which is poorer than GR4, experience notably deprivation regarding a sink and an internet available in the dwelling, GR5 is more appearing as being deprived not very much regarding essential sanitation tools must have in the dwelling but more regarding some electronic materials such as car and computer. It is found that some households of GR4 lack also a hot water system, a sink, and a shower as well as experience food deprivation but the magnitude of poverty in these 3 essential elements of life are less violent among GR4 compared to GR5. But still, for reducing the discrepancy between the well-off groups such GR1 and GR2 and poorer groups including GR4 and GR5, we can recommend that public policies aiming at increasing adequate sanitation facilities in the dwellings of people from GR4 and GR5 and at increasing access to food for GR4 and GR5 households should take priority. Concerning the middle-deprived group (GR3), we can see that they also experience deprivation in terms of computer, car, and internet like GR4 compared to GR1; nevertheless, the magnitude of such deprivations appears slightly less among GR3 despite their magnitude are quite considerable. In addition to these items, some households of GR3 are poor regarding drinking water, a landline telephone, and worrying about food running due to financial reasons in the household; but, the fact that these 3 types of deprivation did not appear among GR4 in CAP113 analysis is due to the higher deprivation regarding car, computer, and internet among GR4. We can summarize that computer, car, and internet emerge definitely as the most important deprivation elements among GR3, GR4, and GR5 compared to the well-off groups, and even if GR5 whiteness greater deprivation regarding sanitation facilities and access to food, GR3 and GR4 experience capabilities deficits in these 2 main aspects of life. Lastly, some households from the most well-off group (GR1) lack a computer and an internet in their dwellings/households, and interestingly some of them suffer also regarding sink, and food.

In Figure 32, patterns of groups are shown per model. The green color ▭ means no deprivation and the red color ▭ means deprivation. Rows are the models, and columns include the title of model (Mod), groups (Gr), pattern (Pat), frequency of pattern (Fre) or in other words the coverage of pattern among the training population in per centage, and variables from WATD to AF3, respectively (including only variables emerging in the patterns). For instance, the first pattern (P1) of GR1 in CAP112 concerns 81 per cent of GR1 people in the training data. This pattern includes no deprivation regarding SINK, INT, and AF3 at the same time. For GR1, in terms of capabilities deficit, the 2 material items computer and internet appear as the most influential ones; however, 8 per cent of GR1 has only deprivation regarding access to internet in their dwelling and they do have a computer. Other indicators appearing in CAP112 tree affect GR3 and/or their impact on GR1 population are very negligible concerning less than 3 per cent of GR1. Concerning GR3, the number of patterns emerging in the tree is higher compared to GR1, and even compared to GR4 and GR5 in other models. This means that the structure of

data for GR3 contains a more variety of combinations and therefore more complexity corresponding to a more heterogenous structure. 63 per cent of GR3 lack a computer, an internet, and a car. Subsequently, 11 per cent of this group have a car but do not have a computer, an internet, and a landline telephone in their dwelling. Moreover, 6 per cent of the group have a computer but lack an internet and more interestingly their dwelling lack a hot water system. And, 5 per cent live in a dwelling having an internet but lacking a sink. With this model, we see that GR3 is especially in disadvantaged position compared to GR1 in terms of computer, internet, car, landline telephone. But still, it is very important to highlight that despite GR3 is a kind of middle poverty group, some people (a slightly low per centage) do live in dwellings lacking a hot water system, or a sink, or even suffering access to a healthy and varied diet due to financial reasons. In CAP113 model comparing GR1 vs. GR4, 4 per cent of GR1 flourishes as lacking both a computer and an internet. On the other hand, 86 per cent of GR4 lack simultaneously a computer, an internet, and a car. This pattern was also emerged for GR3 compared to GR1 in CAP112 model, but we can see that its concern is much higher among GR4. 6.8 per cent of GR4 do have a car but lack an internet and a computer and most importantly a hot water system, which is one of the fundamental items the dwelling must have. Moreover, 4.5 per cent of GR4 have a computer but lack a sink and a hot water system in their dwelling. Compared to GR3, we can clearly see here that GR4 suffer more in terms of a certain essential items such as hot water system and sink lacking in their dwelling. When it comes to the poorest group (GR5), 91 per cent of them lack an internet and a sink in their dwelling, and 5 per cent lack a shower compared to GR1. Furthermore, compared to GR2, CAP117 model shows that 86.65 per cent have an internet in their dwelling but lack a shower, which is extremely important. And 8 per cent of GR5 have a shower but lack a sink and a computer and have worried about the food running out due to financial reasons compared to GR2. It is certainly important to have a computer and an internet in the dwelling with an equal access to every member of the household notably for being able to connect to the news, and if we consider that most of information on government and its services (e.g., social welfare) are available on the internet it becomes substantial these 2 materials. However, for being able to live a decent and healthy life, having adequate sanitation and living standards such as hot water system, shower or bathroom in the dwelling where the household lives and having regularly a healthy and varied food for all household members are vital. In the light of these findings, the government should definitely take note on the last point we mentioned for not only increasing the living conditions of the middle and low classes in terms of socio-economic capabilities but also for decreasing inequality between the well-off and poorer groups across the country.

Table 74 below shows the main confusion matrix statistics clarifying the predictive accuracy of our 6 models. In the table, columns are the models and rows correspond to accuracy (acc), sensitivity (sens), specificity (spec), positive predictive value (pos pred value), negative predictive value (neg pred value), balanced accuracy (balanced acc) computed on the test data, and misclassification rate (misc rate) computed in the training data, respectively. From the results, we can see that the 4 models have a very good performance for predicting the groups. In this line, we will use the most relevant patterns of **CAP112** for *GR3*, **CAP113** for *GR4*, **CAP114** for both *GR1* and *GR5*, and **CAP117** for *GR2* for further analysis.

Table 73. Main confusion matrix statistics per model, data: Mexico MCS 2012

	CAP112 (GR1 vs. GR3)	CAP113 (GR1 vs. GR4)	CAP114 (GR1 vs. GR5)	CAP117 (GR2 vs. GR5)
Acc	0.942	0.978	0.996	0.983
Sens	0.931	0.975	0.996	0.991
Spec	0.954	0.981	0.996	0.977
Pos pred value	0.952	0.978	0.995	0.971
Neg pred value	0.933	0.978	0.997	0.993
Balanced acc	0.942	0.978	0.996	0.984
No-Misc rate (test data)	93 95	97.5 98	99.6 99.6	99 97.8

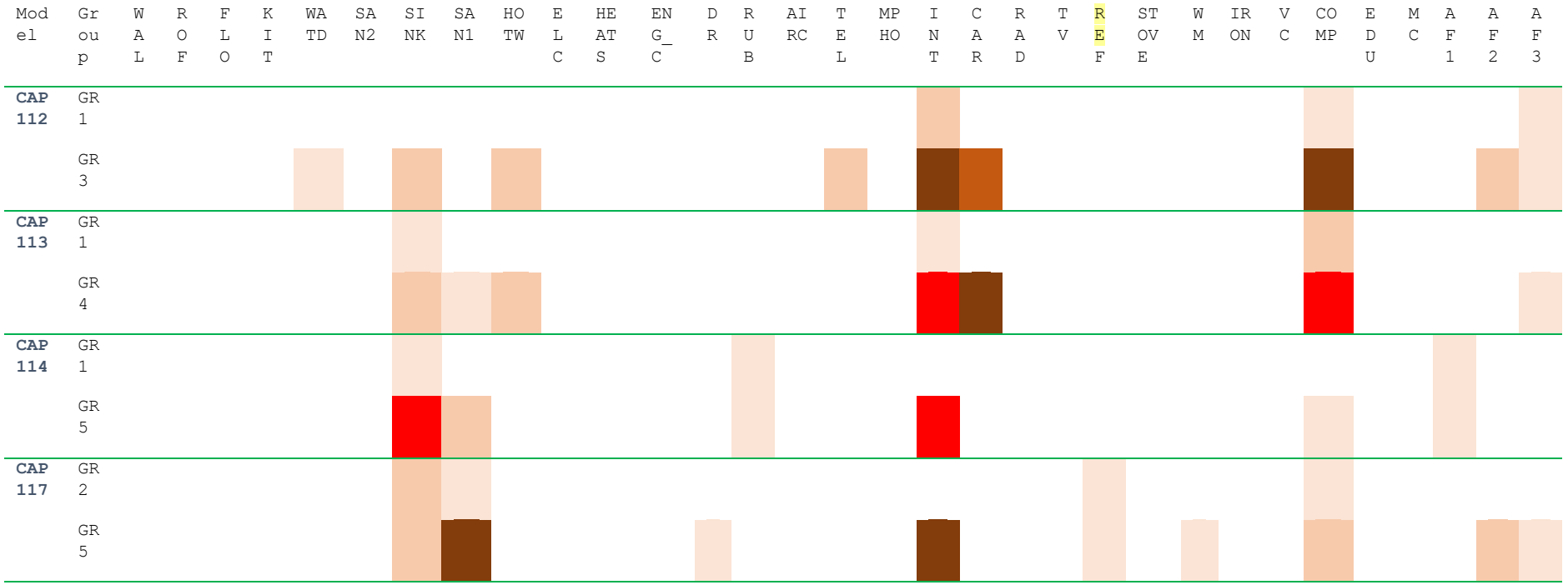


Figure 31. Deprivation per centages represented by different colors in terms of variables emerging in the patterns by group within each model, MCS Mexico 2012

Figure 32. Frequency and composition of classification patterns per model, Mexico MCS 2012

Second step (step2 analysis): detecting dimensional poverty profiles of poverty groups

In this sub-section, we will in particular focus on GR3, GR4, and GR5 in terms of their dimensional profiles. By taking in consideration the performance of models (i.e., predictive accuracy), we can see that for GR3 the model CAP112, for GR4 the model CAP113, and for GR5 the model CAP114 bring the most efficient information concerning their most relevant capabilities patterns. Inside each of these three models, by selecting the most influential patterns, we will attempt to find out the variety of profiles of the three groups regarding dwelling, material, education, health, and food aspects respectively. The aim of doing this is to detect more information on living conditions of these three most deprived groups.

The dimension of *dwelling conditions* is separated into two sub-sets: (i) DWE1: WAL, ROF, FLO, KIT, ENG_C, ELC, and HEATS; (ii) DWE2: WATD, SAN2, SINK, SAN1, HOTW, DR, and RUB. The first covers deprivation regarding the main wall material, the main roof material, the main floor material of the dwelling, whether a kitchen exists or not in the dwelling, energy source for cooking in the dwelling, whether the dwelling has an electricity, and whether there is a heating system in the dwelling respectively. The second includes deprivation in terms of whether the dwelling has a daily access to drinking water via a piped water system inside the house, whether the dwelling has a toilet, a sink, a shower, a hot water system, an accurate drainage system connected to the public network, and whether the garage of the house is picked up by a truck or garbage cart or thrown in a container or a deposit, respectively. The main reason behind this separation is profiles become easily complex to interpret when many variables are included, as we have seen in the beginning of the analysis. The dimension of *material* (MAT) is also divided into two subsets: (i) MAT1: REF, STOVE, WM, IRON, VC, and AIRC; (ii) MAT2: RAD, TV, TEL, MPHO, COMP, INT, CAR. The first group comprises deprivation regarding whether there is a refrigerator, a stove, a washing machine, an iron, a vacuum cleaner, an air condition in the dwelling. The second concerns deprivation regarding whether there is a radio, a television, a landline telephone, a mobile phone, a computer, an internet, and a car in the dwelling. The dimension of *education* (EDUC) includes only deprivation regarding educational attainment. The dimension of *health* (HEALTH) is also composed of one variable reflecting medical coverage. And, the dimension of *nutrition* (NUTRITION) includes deprivation regarding access to food via the 3 variables namely AF1, AF2, AF3 reflecting whether the household ever run out of food due to lack of money or resources, did ever worry about the food running out because of lack of money or resources, have ever run out of money or resource to get a healthy and varied diet, respectively.

We start off by detecting the population of GR3, GR4, and GR5 conforming to the most informative patterns (see below) in the full data (training plus test data), and then we find out their dimensional profiles according to the variables indicated above. This exercise is performed in statistical software STATA, for each group separately.

GR3 dimensional profiles based on CAP112 model

With this model, we saw that the following 5 patterns capture 88 per cent of population of GR3 in the training data: (i) deprivation regarding INT, COMP, and CAR; (ii) no deprivation regarding CAR, but deprivation regarding INT, COMP and TEL; (iii) no deprivation regarding COMP, but deprivation regarding INT, and HOTW; (iv) no deprivation regarding INT, but deprivation regarding SINK; and (v) no deprivation regarding INT, and SINK, but deprivation regarding CAR, and AF3. When we applied these 5 classification rules in the full data, we lose

12 per cent of information reaching to 88 per cent (as in the case of training data), which shows that the model predicts well GR3.

The table below illustrates the dimensional profiles of GR3 based on the extracted 5 patterns touching each of them at least 3 per cent of GR3. The table has 6 columns representing pattern number, and the most substantial dwelling conditions (DWE1, and DWE2), material (MAT1, and MAT2), and nutritional profiles per each pattern. Due to space limitation in the table, the profiles of the dimensions of education (EDUC), health (HEALTH) per each pattern will be given directly in the text. In first column, next to the pattern number, the frequency of the pattern is mentioned in parenthesis in terms of per centage. For instance, 1st pattern concern 63 per cent of population of GR3 in the entire data. Profiles are composed of 0 and 1 where 0 means no deprivation and 1 means deprivation regarding the variable. The profiles are composed in the following way:

DWE1= WAL_D+ROF_D+FLO_D+KIT_D+ENG_C_D+ELC_D+HEATS_D where the 7 variables are string character according to STATA language, which allows the combination of values and not necessary summing the values to obtain a numeric total value. For instance, the profile 1000111 means having deprivation regarding WAL, ENG_C, ELC and HEATS but no deprivation regarding ROF, FLO, and KIT. Please see the table of deprivation variables presented in the beginning of the case study.

DWE2= WATD_D+SAN2_D+SINK_D+SAN1_D+HOTW_D+DR_D+RUB_D.

MAT1=REF_D+STOVE_D+WM_D+IRON_D+VC_D+AIRC_D.

MAT2=TEL_D+MPHO_D+COMP_D+INT_D+CAR_D+RAD_D+TV_D.

EDU=EDU_D

HEALTH=MC_D.

NUTRITION=AF1_D+AF2_D+AF3_D.

In table, next to the profiles, their frequency (i.e., Freq) within the pattern population is mentioned in parenthesis in terms of per centage. For instance, the profile 0000101 in DWE1 concerns 75.3 per cent of population of pattern 1. The sum of frequencies of dimensional profiles in each column for each pattern equals maximum 100 per cent. In case of DWE1 profile for the 1st pattern, there are 6 profiles are selected as the most substantial ones with at least 100 observations. The last row in the table indicates the total information captured in each dimension for the ensemble of GR3.

Concerning DWE1 profile throughout the 5 patterns, we reach to capture 83 per cent of GR3 (given that the 5 patterns explain 88 per cent of GR3). There are in total 15 DWE1 profiles. Among them, the profile of deprivation regarding simultaneously the main source of energy used for cooking and heating system in the dwelling, combined with no deprivation regarding all other 5 components of DWE1 is the most relevant one covering 65 per cent of GR3. After that, being deprived regarding the roof material of the dwelling as well as the main source of energy used for cooking and heating system in the dwelling combined with no deprivation regarding all other 4 components of DWE1 surfaces affecting 8 per cent of GR3. Besides, the profile of deprivation regarding the wall material of the dwelling as well as the main source of energy used for cooking and heating system in the dwelling combined with no deprivation

regarding all other 4 components of DWE1 emerges concerning 4.44 per cent of GR3. Throughout the 15 DWE1 profiles, we can say that 81 per cent of GR3 encounters deprivation regarding the main source of energy used for cooking in the dwelling, and 83 per cent of GR3 experience deprivation regarding heating system in the dwelling. Moreover, 6.6 per cent of GR3 live in a dwelling with a not adequate wall in terms of material, 10 per cent live in a dwelling with a not adequate roof in terms of material. In this context, it becomes very important for government to elaborate new programs focusing on increasing the use of better energy source for cooking and heating in the dwellings. On the other hand, concerning the floor material, existence of kitchen, electricity in the dwelling, there is no deprivation or a few/negligible share of GR3 population experience deprivation regarding these 3 items.

Concerning DWE2 profile based on the 5 patterns, there is more complexity compared to DWE1, we could reach to capture 66 per cent of GR3 via 29 profiles shown in the table below. The most relevant profiles are: (i) no deprivation regarding all 7 components of DWE2 (17.7 per cent of GR3); (ii) deprivation regarding only the hot water system in the dwelling and no deprivation regarding the other 6 variables of DWE2 (19 per cent of GR3); (iii) deprivation regarding only the drinking water in the dwelling and no deprivation regarding the other 6 variables of DWE2 (7 per cent of GR3). Given the number of various combinations flourishing in case of GR3, it will be convenient to focus on the deprivation per centage regarding each single variable of DWE2 rather than regarding each profile of DWE2. In that vein, we can see that the most important deprivation is in terms of hot water system in the dwelling concerning 33.5 per cent of GR3. After that, deprivation regarding drinking water, and deprivation regarding having a sink in the dwelling come affecting each of them 15 per cent of GR3. Thus, the access to the 3 sanitation items namely, hot water system, drinking water, and sink in the dwelling should be examined in order to increase living standards of GR3. On the other hand, the per centage of GR3 affected by deprivation regarding shower, drainage system, and rubbish collection of the dwelling is zero or very low.

When it comes to materials, in terms of MAT1 dimension, via the extracted 15 profiles based on the 5 patterns shown in the table below we detect 77 per cent of GR3. The most relevant profile is deprivation regarding vacuum cleaner and air conditioner combined with no deprivation regarding the other 4 variables of MAT1, affecting 53 per cent of GR3. Subsequently, the profile of deprivation regarding washing machine, vacuum cleaner, air conditioner combined with no deprivation regarding the other 3 variables of MAT1 comes covering 17 per cent of GR3. When we consider the 15 profiles, we find out that 77 per cent of GR3 lack a vacuum cleaner, 71 per cent lack an air conditioner, and 14 per cent of GR3 lack a washing machine in their dwelling.

On the other hand, regarding the MAT2 dimension, there is more complexity compared to MAT1. In total, 23 profiles based on the 5 patterns give information on 83 per cent of GR3. Among these 23 profiles, the most relevant is the combination of deprivation regarding a landline telephone, a computer, an internet, a car, and a radio, and non-deprivation regarding a mobile phone and a television, which concerns 27.6 per cent of GR3. Besides, 9 per cent of GR3 have a landline telephone and a television and don't have the other 5 items of MAT2; and another 9 per cent of GR3 have a landline telephone, a mobile phone, and a television and don't have the other 4 items of MAT2. In CAP112 analysis, we already showed that 63 per cent of GR3 lack a car, a computer, and an internet simultaneously. Thus, these 3 items of MAT2 appear as the most important ones in terms of capabilities deficit. Given the numerous profiles

emerging across the 5 patterns, it will be more efficient to compute directly the per centage of deprivation regarding each single item of MAT2. In that vein, we find out that 72.75 per cent of GR3 don't have a computer, 78 per cent don't have an internet, 71 per cent don't have a car. Besides, 55.6 per cent of GR3 don't have a landline telephone vs. 21 per cent don't have a mobile phone in their dwelling. And, 73.7 per cent don't have a radio vs. a considerably low per cent of GR3 lack a television in their dwelling. A landline telephone might be preferred by a mobile phone and a radio might be preferred by a television nowadays; nevertheless, landline telephone and radio remain still some of the important communication tools for being staying in touch with the society.

Regarding the food access profiles emerging based on the 5 patterns of GR3, we can say that 43.65 per cent of GR3 haven't run out of food due to lack of money or resources, did not worry about the food running out due to lack of money or resources, and haven't run out of money or resources to get a healthy and varied diet. However, the majority of GR3 have encountered deprivation regarding at least one of these 3 items. Notably, 25 per cent of GR3 worried about the food running out and even sometimes have run out of money or resources to get a healthy and varied diet; 6.5 per cent have encountered deprivation regarding AF1, AF2, and AF3. Between the 3 items of this dimension, the most encountered one is the second (AF2)- worrying about the food running out due to lack of money or resources, which concerns 40 per cent of GR3. Subsequently, 27 per cent of GR3 have run out of money or resources to get a healthy and varied diet, and the worst situation-run out of food due to lack of money resources has affected 6.5 per cent of GR3.

The educational attainment of GR3 appears quite low given that 32.5 per cent of them have at least a high school diploma or equivalent. In other words, 31.6 per cent have a secondary school diploma, 30 per cent have a primary school diploma, 18 per cent have a high school diploma, 9.4 per cent have a higher education diploma including 'carrera technical or commercial', and university degree, 4 per cent have normal superior education, 4 per cent are illiterate, and 1 per cent are literate but with no diploma.

Moreover, 22 per cent of GR3 don't have a medical coverage, 39 per cent are not enrolled to popular insurance but affiliated with other institution for medical coverage, 32 per cent are enrolled to popular insurance, and 6 per cent are enrolled to both popular insurance and another institution for medical coverage.

As a result of these discovered profiles and the patterns of GR3 from the CAP112 analysis, we conclude that the following items should be considered as priority in poverty reduction programs in Mexico notably for reducing the discrepancy between GR3 and GR1: (i) dwelling standards: energy source for cooking, heating system, hot water system, availability and source of drinking water, and availability of a sink in the dwelling; (ii) materials: computer, internet, car, vacuum cleaner, air conditioner, landline telephone, washing machine; (iii) food security; (iv) educational attainment, having at least a high school diploma; (v) medical coverage for those are not yet covered by popular insurance system or other institution.

Table 74. Dimensional profiles of GR3 based on the CAP112 model, data: Mexico MCS 2012

Pattern number of GR3 (Freq in % in full data)	DWE1 profile (Freq in % in full data)	DWE2 profile (Freq in % in full data)	MAT1 profile (Freq in % in full data)	MAT2 profile (Freq in % in full data)	NUTRITION profile (Freq in % in full data)
1 (62.5)	0000101 (75.3), 0100101 (7.85), 1000101 (6.45), 0000001 (3.85), 1100101 (2.8), 0001101 (2.11)	0000000 (25), 0000100 (21), 1000000 (9), 0010000 (6.2), 0010100 (4.8), 0000110 (4.4), 1010000 (4.1), 1000100 (3.7), 0001100 (2.1)	000011 (61), 001011 (17), 000010 (7), 000111 (4.7)	1011110 (44.2). 0111110 (14.8), 0011110 (14.4), 1011100 (8.9), 1111110 (6.9), 0111100 (4.3), 1111100 (2.6), 0011100 (2.6)	000 (53.5), 011 (18.9), 010 (17.2), 111 (5.4), 001 (3.7)
2 (11)	0000101 (73), 0100101 (11.5), 1000101 (3.8), 1100101 (3.6)	0000100 (17.5), 0000000 (14.3), 1000000 (8.5), 0000110 (6.6), 0010100 (6.6), 1000100 (5), 0010000 (4.7), 1010000 (4.2)	000011 (54.3), 001011 (15.6), 000010 (9.4), 000111 (6.8)	1011010 (74.4), 1011000 (11.6), 1111010 (11)	000 (47), 011 (21.2), 010 (19), 111 (9.3)
3 (6.5)	0000101 (68.9), 0100101 (17)	0000100 (23), 0000110 (11), 1000100 (10), 0010100 (9)	000011 (61.3), 001011 (12), 000010 (11)	1001110 (45.3), 1001010 (17), 0001110 (12.6), 1001100 (7)	000 (48.2), 011 (20.9), 010 (18.3), 111 (8)
4 (5)	0000101 (66), 0100101 (14.5)	0010100 (17.5), 1011100 (14.2), 0011100 (10.8), 1010000 (10.7), 1010100 (9.9)	000011 (61.5), 001011 (16.3)	0000110 (27.3), 1000110 (13.6), 0100110 (12.7), 0000010 (8.4)	000 (38.35), 011 (26), 010 (20.6), 111 (10)
5 (3)	0000101 (72)	0000100 (32), 0000000 (17), 1000000 (13.8)	000011 (61.3), 001011 (14.6)	0000110 (27.5), 1000110 (20.8), 0100110 (17.8), 0010110 (13.3)	011 (57.3), 111 (37)
Total info/GR4 population coverage in %	(62.5*0.983) + (11*0.919) + (6.5*0.859) + (5*0.805) + (3*0.72) = **83**	(62.5*0.803) + (11*0.674) + (6.5*0.53) + (5*0.631) + (3*0.628) = **66**	(62.5*0.897) + (11*0.861) + (6.5*0.843) + (5*0.778) + (3*0.759) = **77**	(62.5*0.987) + (11*0.97) + (6.5*0.819) + (5*0.62) +	(62.5*0.987) + (11*0.965) + (6.5*0.954) + (5*0.949)

				(3*0.794) = **83**	+ (3*0.943) = **86**

GR4 dimensional profiles based on CAP113 model

With this model, we saw that the following 3 patterns capture 97 per cent of population of GR4 in the training data: (i) deprivation regarding COMP, INT, and CAR; (ii) no deprivation regarding CAR, but deprivation regarding COMP, INT, and HOTW; (iii) no deprivation regarding COMP, but deprivation regarding SINK, and HOTW. When we applied these 3 classification rules in the full data, we lose only 5 per cent of information reaching to 95 per cent (very close to what we obtained on the training data), which shows that the model predicts very well GR4.

Based on the 3 extracted patterns of GR4 in CAP113, we find out that with the 11 DWE1 profiles shown in the table below, 88 per cent of GR4 can be captured. The most relevant DWE1 profiles are: (i) deprivation regarding the main source of energy available in the dwelling for cooking, and the heating system of the dwelling combined with no deprivation regarding the 5 other items of DWE1, which concerns 53 per cent of GR4; (ii) deprivation regarding the roof material of the dwelling, the main source of energy available in the dwelling for cooking, and the heating system of the dwelling combined with no deprivation regarding the 4 other items of DWE1, which affects 18.4 per cent of GR4; (iii) deprivation regarding the roof, wall materials of the dwelling, the main source of energy available in the dwelling for cooking, and the heating system of the dwelling combined with no deprivation regarding the 3 other items of DWE, which covers 6 per cent of GR4. By taking in consideration the 11 DWE1 profiles, we result that 88 per cent of GR4 lack both a heating system, and a proper source of energy for cooking in their dwelling. Moreover, 24 per cent of GR4 are deprived in terms of their dwelling's roof quality, and 12.6 per cent of GR4 are deprived in terms of their dwelling's wall quality. On the other hand, deprivation regarding dwelling's floor material quality, kitchen in the dwelling, and an electricity in the dwelling are zero or quite negligible among GR4. These findings show that the major capabilities deficit for GR4 regarding the dimension DWE1 is lacking both a proper source of energy for cooking and a heating system in the dwelling, which was also the case for GR3 but the per centage of GR4 population affected by such deprivation is higher (by 23 per centage points) compared to GR3. Deprivation regarding wall, and roof quality of the dwelling were also found in the case of GR3; but the per centages of population of GR4 affected by these deprivations are doubled up compared to GR3.

Regarding DWE2 dimension, we can see from the table below that there is more complexity as we can only capture 65 per cent of GR4 via the 24 profiles. The most relevant profiles are: (i) deprivation regarding drinking water, sink, shower, and hot water system in the dwelling combined with no deprivation regarding the other 3 items of DWE2, which concerns 11 per cent of GR4; (ii) deprivation regarding drinking water, sink, shower, hot water system, and drainage system in the dwelling combined with no deprivation regarding the other 2 items of DWE2, which covers 6 per cent of GR4. When we take into account the 24 profiles, we find that 34 per cent of GR4 live in a dwelling lacking a regular piped system-based drinking water, 40.5 per cent of GR4 live in a dwelling lacking a sink, 30 per cent of GR4 live in a dwelling lacking a shower, 50.5 per cent live in a dwelling lacking a hot water system, and 5.7 per cent

live in a dwelling lacking an adequate drainage system. On the other hand, deprivation regarding toilet, and rubbish collection corresponds to zero or exist at negligible level among GR4. Compared to the results of GR3, we can say that GR4 also suffer regarding WATD, SINK, HOTW but at much higher levels. Besides, GR4 emerges as affected by deprivation regarding SAN2, DR compared to GR3.

In terms of material deprivation, by employing 12 profiles of MAT1, we capture 84 per cent of GR4. The most frequent 4 profiles are: (i) deprivation regarding vacuum cleaner, and air conditioner combined with no deprivation regarding the 4 other items of DWE1, which concerns 40 per cent of GR4; (ii) deprivation regarding washing machine, vacuum cleaner, and air conditioner combined with no deprivation regarding the 3 other items of DWE1, which affects 24 per cent of GR4; (iii) deprivation regarding iron, vacuum cleaner, and air conditioner combined with no deprivation regarding the 3 other items of DWE1, which covers 7.6 per cent of GR4; and (iv) deprivation regarding washing machine, iron, vacuum cleaner, and air conditioner combined with no deprivation regarding the 2 other items of DWE1, which touches 5 per cent of GR4. By taking in consideration 12 profiles of DWE1, we can conclude that 83.6 per cent of GR4 live in a dwelling lacking both a vacuum cleaner and an air conditioner, 33.8 per cent of GR4 live in a dwelling lacking a washing machine, and 13 per cent of GR4 live in a dwelling lacking an iron. The per centage of GR4 population living in a dwelling lacking a refrigerator and/or a stove corresponds to zero or is at negligible level. In light of these results, we can observe that the per centages of GR4 population experiencing deprivation regarding VC, AIRC, WM are higher than GR3.

On the other hand, concerning MAT2 dimension, with 15 profiles shown in the table below, 89 per cent of GR4 is captured. There is a rich variety of profiles emerging in this dimension. For that reason, we will point out the per centages of variables instead of profiles. In that vein, we can say that 73.5 per cent of GR4 lack a landline telephone, 31.6 per cent lack a mobile phone, 86.8 per cent lack a computer, 88 per cent lack an internet, 82.6 per cent lack a car, 76 per cent lack a radio, and no one lacks a television in their dwelling. Except television, in all 6 items of MAT2, deprivation among GR4 is much higher than GR3.

When it comes to nutritional aspect, we find that 31.5 per cent of GR4 live in a household not having suffered in food shortage in terms of both quantity and quality. However, 16 per cent of GR4 live in household having experienced food shortage regarding AF1, AF2, and AF3 at the same time. Besides, 15.8 per cent of GR4 live in households having had run out of food due to lack of money or resources, 57 per cent live in households having had worried about the food running out due to lack of money or resources, and 46.5 per cent live in households having had run out of money or resources to get a healthy and varied diet. Thus, we can say that compared to GR3, the per centage of GR4 living in households having had experienced food deprivation regarding at least one of the 3 items is around 14 per centage points higher.

Concerning educational attainment, 79.5 per cent of GR4 is deprived; such result is quite striking and about 10 per centage points higher than GR3. More specifically, 36.65 per cent of GR4 have a primary school diploma, 33 per cent have a secondary school diploma, 13.8 per cent have a high school diploma, 7.7 per cent are illiterate, and 4.4 per cent have a higher education diploma.

Moreover, 23 per cent of GR4 don't have any medical coverage. 46 per cent of GR4 are enrolled to popular insurance, 24.6 per cent are affiliated with another institution for medical coverage, and 6 per cent have both a popular insurance and an insurance at another institution.

According to the findings explained above, we can summarize that deprivation regarding nutrition, education, energy source for cooking, and heating system in the dwelling emerge as the most important ones. Deprivation in terms of medical coverage remains at same level compared to GR3. The deprivation gap between GR3 and GR4 in terms of roof, wall material of the dwelling, access to drinking water in the dwelling, having a sink, a hot water system in the dwelling is between 15 and 25 per centage point in detriment of GR4. Besides, compared to GR3, lacking a shower and/or an iron in the dwelling is emerged for GR4. And, in terms of material deprivation, car, internet, and computer remain as the most important items (and the deprivation is higher compared to GR3); however, regarding radio and landline telephone as well the deprivation per centages are quite high among GR4. In CAP114 model, we saw that for reducing inequality between GR1 and GR4, the focus should be given to deprivation regarding computer, internet, car, and hot water system. As the profiles shown above are based on the patterns discovered via CAP114 model, educational attainment (given that , educational deprivation is significantly higher among GR4 compared to GR1), food shortage due to household's financial problems, dwelling structure in terms of roof, wall, main source of energy used for cooking, heating system, access to drinking water, sanitation facilities in the dwelling should be considered as well. Moreover, even if medical coverage has a relatively good performance, the efforts for increasing enrollment to popular insurance should be continued for the poorer population. The use of air conditioner can be considered problematic for its possible effect on higher use of electric, which can elevate easily the electricity bill for the household; but the efforts can be done for increasing the use of washing machine, and vacuum cleaner among GR4.

Table 75. Dimensional profiles of GR4 based on the CAP113 model, data: Mexico MCS 2012

Pattern number of GR4 (Freq in % in full data)	DWE1 profile (Freq in % in full data)	DWE2 profile (Freq in % in full data)	MAT1 profile (Freq in % in full data)	MAT2 profile (Freq in % in full data)	NUTRITION profile (Freq in % in full data)
1 (84)	0000101 (57), 0100101 (18.85), 1000101 (8), 1100101 (6.4), 0001101 (4.3)	1011100 (12), 0000100 (7.6), 0011100 (6.6), 0010100 (6), 1011110 (5.5), 1010000 (4.2), 1010100 (4), 0000000 (3.6), 0001100 (2.75), 0010000 (2.73), 1000100 (2.63), 1001100 (2.5), 0000110 (2.5), 1000000 (2.5), 1011111 (2)	000011 (41.9), 001011 (25.5), 000111 (8.3), 001111 (5.8), 101011 (5.2), 100011 (2.8)	1011110 (49), 1111110 (17), 0111110 (10), 1011100 (8), 0011110 (5), 1111100 (4.4), 0111100 (2.2)	000 (33.6), 011 (28), 111 (16.7), 010 (15.6), 001 (4.3)

2 (7)	0000101 (52.4), 0100101 (22.5), 1100101 (8), 0001101 (6)	1011100 (15), 1011110 (10.2), 1011111 (7.3), 1010100 (5.8), 0010100 (5.3), 0000100 (5)	000011 (39), 001011 (23.8), 000111 (9), 001111 (6)	1011010 (61.7), 1111010 (11), 1011000 (9.4), 0111010 (6.5), 0011010 (5.4)	011 (31.6), 000 (29.9), 010 (17.4), 111 (17)
3 (4)	0000101 (43.6), 0100101 (26)	1011100 (24.7), 1011110 (18)	000011 (48.8), 001011 (26)	1001110 (37), 0000110 (11.6), 1000110 (9.6)	011 (34.3), 000 (30.3), 111 (15), 010 (14.5)
Total info/GR4 population coverage in %	(84*0.945) + (7*0.889) + (4*0.696) =**88**	(84*0.718) + (7*0.486) + (4*0.427) =**65**	(84*0.895) + (7*0.778) + (4*0.748) =**84**	(84*0.956) + (7*0.94) + (4*0.582) =**89**	(84*0.982) + (7*0.959) + (4*0.941) =**93**

GR5 dimensional profiles based on CAP114 model

With this model, we saw that the following 2 patterns capture 96 per cent of population of GR5 in the training data: (i) deprivation regarding SINK and INT; (ii) no deprivation regarding SINK, but deprivation regarding SAN1. When we applied these 2 classification rules in the full data, we lose only 2.5 per cent of information reaching to 97.5 per cent, which shows that the model predicts very well GR5.

Concerning DWE1 dimension, with the 13 profiles shown in the table below, we reach to detect 86 per cent of GR5. There are 4 most relevant profiles that are: (i) deprivation regarding the roof material of the dwelling, energy source used for cooking in the dwelling, and heating system of the dwelling combined with no deprivation regarding the 4 other items of DWE1, which affects 21 per cent of GR5; (ii) deprivation regarding source of energy used for cooking in the dwelling, and heating system of the dwelling, and no deprivation regarding the other 5 items of DWE1, which concerns 20 per cent of GR5; (iii) deprivation regarding the roof, wall materials of the dwelling, energy source used for cooking in the dwelling, and heating system of the dwelling combined with no deprivation regarding the 3 other items of DWE1, which touches 17 per cent of GR5; and (iv) deprivation regarding the wall material of the dwelling, energy source used for cooking in the dwelling, and heating system of the dwelling, combined with no deprivation regarding the 4 other items of DWE1, which covers 6 per cent of GR5. When the 13 profiles are taken in consideration, we find out that 31 per cent of GR5 are deprived regarding the wall material of the dwelling, 53 per cent of GR5 are deprived regarding the roof material of the dwelling, 86 per cent of GR5 are deprived regarding both energy source used for cooking in the dwelling, and the heating system of the dwelling. In addition, 5.8 per cent of GR5 are deprived regarding the floor material of the dwelling, and 15.5 per cent are deprived regarding kitchen in the dwelling. On the other hand, no one is found deprived regarding electricity supply in the dwelling.

Concerning DWE2 dimension, with the 14 profiles shown in the table below 81 per cent of GR5 are captured. From the table below, we can see that in fact there is a more variety of profiles compared to DWE1. However, the following 4 profiles are the most relevant: (i) no deprivation regarding toilet in the dwelling, but deprivation regarding the 6 other items of

DWE2, which concerns 25 per cent of GR5; (ii) no deprivation regarding toilet in the dwelling, and dwelling's rubbish removal system, but deprivation regarding the 5 other items of DWE2, which touches 12.6 per cent of GR5; (iii) no deprivation regarding toilet in the dwelling, drainage system of the dwelling, and dwelling's rubbish removal system, but deprivation regarding the 4 other items of DWE2, which includes 11.8 per cent of GR5; and (iv) deprivation regarding the all 7 items of DWE2, which covers 11 per cent of GR5. By taking in consideration all profiles of DWE2, it is found that 78 per cent of GR5 are deprived regarding access to regular piped drinking water in the dwelling, 77.4 per cent of GR5 are deprived regarding sink in the dwelling, 79 per cent of GR5 are deprived regarding shower in the dwelling, 81 per cent of GR5 are deprived regarding hot water system in the dwelling, 58.8 per cent of GR5 are deprived regarding drainage system of the dwelling, 43.6 per cent of GR5 are deprived regarding dwelling rubbish removal system, and 20 per cent of GR5 are deprived regarding toilet in the dwelling.

When we look at the dimension MAT1, we find that 19 profiles as shown in the table below, 89 per cent of GR5 are detected. The following profiles are the most relevant: (i) deprivation regarding washing machine, vacuum cleaner, and air condition, combined with no deprivation regarding the 3 other variables of MAT1, which concerns 13.4 per cent of GR5; (ii) deprivation regarding vacuum cleaner, and air condition, combined with no deprivation regarding the 4 other variables of MAT1, which affects 10.4 per cent of GR5; (iii) deprivation regarding the 5 items of MAT1 except stove, which concerns 8 per cent of GR5; (iv) deprivation regarding all items of GR5, which covers 16 per cent of GR5; and (v) deprivation regarding refrigerator, washing machine, vacuum cleaner, and air conditioner, combined with no deprivation regarding stove and iron, which affects 7 per cent of GR5. Throughout the 19 profiles, we can summarize that 88.6 per cent of GR5 are deprived regarding both vacuum cleaner and air conditioner, 68.4 per cent are deprived in terms of washing machine, 43.6 per cent are deprived of an iron in their dwelling, 38.7 per cent are deprived of a refrigerator in their dwelling, and 33 per cent are deprived of a stove in their dwelling.

On the other hand, regarding MAT2 dimension, by using 12 profiles as shown in the table below, information on 89 per cent of GR5 can be obtained. The most relevant profiles are: (i) deprivation regarding the 6 items of MAT2 except television, which concerns 31 per cent of GR5; (ii) deprivation regarding the 5 items of MAT2 except mobile phone, and television, which concerns 27 per cent of GR5; (iii) deprivation regarding all items of MAT2, which covers 15.8 per cent of GR5; and (iv) deprivation regarding the 5 items of DWE2 except radio, and television, which includes 6 per cent of GR5. Throughout the 12 profiles of DWE2, we find out that 89 per cent of GR5 live in a dwelling lacking a computer, and internet, and a car. Moreover, 86 per cent of GR5 live in a dwelling lacking a landline telephone, 55 per cent live in a dwelling lacking a mobile phone, 76 per cent live in a dwelling lacking a radio, and 18 per cent live in a dwelling lacking a television.

When it comes to nutrition, with the 9 profiles shown in the table below, 94 per cent of GR5 are detected. We find out that 33 per cent of GR5 worried about the food running out and have sometimes run out of money or resources to get a healthy and varied diet, 30 per cent are deprived regarding AF1, AF2, and AF3 in the last 3 months vs. only 15 per cent have never experienced AF1, AF2, and AF3 in the last 3 months.

In terms of educational attainment, 91.4 per cent of GR5 are deprived. More specifically, 43 per cent have a primary school diploma, 26 per cent have a secondary school diploma, 20 per cent are illiterate, 6.5 per cent have a high school diploma, and only 1.44 per cent have a higher education diploma.

And, for medical coverage, 27.6 per cent of GR5 lack a medical coverage. On the other part, 60.4 per cent of GR5 are enrolled to popular insurance, 8.5 per cent are affiliated with both popular insurance and another coverage from another institution, and 8.5 per cent are affiliated with another institution than popular insurance for medical coverage.

Given the findings we mentioned above, we can conclude that living conditions are deteriorated for GR5 compared to GR3 and GR4 in all aspects. Notably, given the statistics above, priority should be given to education, nutrition, source of energy for cooking, heating system, access to drinking water, sanitation facilities and hot water system in the dwelling, and drainage system of the dwelling.

Table 76. Dimensional profiles of GR5 based on the CAP114 model, data: Mexico MCS 2012

Pattern number of GR5 (Freq in % in full data)	DWE1 profile (Freq in % in full data)	DWE2 profile (Freq in % in full data)	MAT1 profile (Freq in % in full data)	MAT2 profile (Freq in % in full data)	NUTRITION profile (Freq in % in full data)
1 (91)	0100101 (21.7), 0000101 (20.8), 1100101 (17.6), 0001101 (7.25), 1000101 (6.6), 0101101 (5.75), 1110101 (4.2), 1101101 (4), 0110101 (2.2)	1011111 (27.5), 1011110 (13.9), 1011100 (13.1), 1111111 (12.3), 1111110 (5.5), 1111100 (4), 1011101 (3.6), 0011100 (2.8), 1010111 (2.35)	111111 (17.35), 001011 (13.8), 000011 (10.7), 001111 (9.2), 101111 (8.4), 101011 (7.4), 111011 (5.55). 011111 (5.45), 000111 (5.35), 011011 (4.9), 010011 (2.6), 100011 (2.35)	1111110 (32.6), 1011110 (28.45), 1111111 (12), 1111100 (6.5), 1011100 (3.8), 1011111 (3.7), 0111110 (3.3), 1111101 (3.2)	011 (34), 111 (31.3), 000 (16.4), 010 (12), 001 (4)
2 (5)	0100101 (27.5), 0000101 (22), 1100101 (20.5), 1000101 (7.3)	1001111 (31), 1001110 (16.7), 1001100 (12.4), 1101111 (8.2), 0001100 (6.6)	001011 (16.7), 000011 (14), 001111 (13.8), 101111 (10.6), 111111 (8.5), 101011 (8.4), 000111 (7.8)	1111110 (31), 1011110 (29), 1111111 (9.8), 1111100 (6.5)	011 (39), 111 (32.5), 010 (11.3), 000 (11)
Total info/GR5 population coverage in %	(91*0.901) + (5*0.773) = **86**	(91*0.850) + (5*0.749) = **81**	(91*0.93) + (5*0.798) = **89**	(91*0.935) + (5*0.763) = **89**	(91*0.977) + (5*0.938) = **94**

Third step (step3 analysis): detecting various profiles of poverty groups

In this sub-section, we will proceed with sub-analysis in order to examine in more detail the extent of capabilities of GR3, GR4, and GR5 in terms of economic activity status. We note that there is no information on geographical situation, and we don't have also more detailed

economic activity related information due to data structure as we explained in the beginning of the case study. Distribution of educational attainment across GR3, GR4, and GR5 are already shown in second model section. And due to complexity of the data, we cannot proceed with individual characteristics, and household composition profiles. The classification model composed of the variables of these two former aspects do not yield as well efficient information in the sense that cross-validation error rate appears around 0.5, which is quite high for obtaining a good prediction. Thus, we will look at only economic activity related profiles here.

GR3 sub-group profiles based on CAP112 model

In terms of economic activity profile, we show in the table below the relevant profiles of GR3 based on its 5 relevant patterns found in CAP112 analysis, by combining the 2 variables, ES1 and JOB, respectively. 23 economic activity profiles are shown in the table below, which capture 85 per cent of GR3.

Herewith, we can observe that 51 per cent of GR3 are employed performing only one job, 21 per cent are not economically active due to household tasks, 2.3 per cent are retired, 2.3 per cent are unemployed, 3.3 per cent are not economically active because of their studying, 2.3 per cent are employed performing more than one job, and 1.5 per cent are not economically active due to disability. From these profiles, notably the per centage of people performing household tasks are quite considerable. We know that these profiles capture 85 per cent of GR3 as they are based on the 5 relevant patterns of GR3 from CAP112 analysis. Nevertheless, when we look at the distribution of ES1 variable (economic activity status) among the entire GR3 population, we also obtain the same tendencies such as majority around 60.9 per cent are economically active performing only one job, 24.3 per cent are not economically active due to household tasks. So that, there is no significant difference in terms of tendencies between the investigation of the entire population of GR3 and the sample of GR3 based on its 5 relevant patterns.

We note that in the table below, for those are not economically active, the last 2 digits in their profile are 99 because simply they do not have any job.

Table 77. Economic activity profiles of GR3 based on the CAP112 model, data: Mexico MCS 2012

Pattern number of GR3 (Freq in % in full data)	Economic activity profile (Freq in % in full data)
1 (62.5)	11 (57), 699 (25.6), 599 (3.7), 399 (3.1), 799 (3), 12 (2.6), 899 (2.5)
2 (11)	11 (63), 699 (23.7), 399 (3.5), 799 (2.9), 12 (2.3)
3 (6.5)	11 (63), 699 (18), 799 (7.7), 12 (4)

4 (5)	11 (60), 699 (19.4), 799 (8.5), 12 (4)
5 (3)	11 (58), 699 (22.4), 799 (8)
Total info/GR3 population coverage in %	(62.5*0.975) +(11*0.954) +(6.5*0.927) + (5*0.919) +(3*0.884) = **85**

GR4 sub-group profiles based on CAP113 model

In terms of economic activity profile, we show in the table below the relevant profiles of GR4 based on its 3 relevant patterns found in CAP113 analysis, by combining the 2 variables, ES1 and JOB, respectively. 16 economic activity profiles are shown in the table below, which capture around 93 per cent of GR4. In light of this information, we can see that the majority of GR4, 55 per cent, are economically active performing only one job. Besides, 24.5 per cent are not economically active due to household tasks, 3.4 per cent are economically active performing more than one job, 3 per cent are unemployed, 2 per cent are not economically active due to disability, and 2.6 per cent are not economically active due to their studies. When we cross-check the distribution of the variable ES1 among the entire population of GR4, we also obtain the same tendencies including that 61 per cent, i.e., the majority of the group, are economically active performing one job, 25.7 per cent are not economically active due to household tasks. Here also, the per centage of people performing household tasks and not working actually because of that becomes striking. Compared to GR3, the distribution of ES1 is similar among GR4 as well.

We note that in the table below, for those are not economically active, the last 2 digits in their profile are 99 because simply they do not have any job.

Table 78. Economic activity profiles of GR4 based on the CAP113 model, data: Mexico MCS 2012

Pattern number of GR4 (Freq in % in full data)	Economic activity profile (Freq in % in full data)
1 (84)	11 (57.5), 699 (26), 12 (3.5), 399 (3.4), 899 (2.5), 799 (2.5), 999 (2.2)
2 (7)	11 (59), 699 (26.5), 12 (3.8), 399 (3), 799 (2.6)

3 (4)	11 (59.6), 699 (20), 799 (8), 12 (5.6)
Total info/GR4 population coverage in %	(84*0.979) +(7*0.949) +(4*0.932) = **92.6**

GR5 sub-group profiles based on CAP113 model

In terms of economic activity profile, we show in the table below the relevant profiles of GR5 based on its 2 relevant patterns found in CAP114 analysis, by combining the 2 variables, ES1 and JOB, respectively. 11 economic activity profiles are shown in the table below, which capture around 93.5 per cent of GR5. In that vein, we can deduce that 51.6 per cent of GR5 are economically active performing one job, 26 per cent are not economically active due to household tasks, 5.4 per cent are economically active performing more than one job, 2.6 per cent are not economically active due to disability, and 2.5 per cent are not economically active because of studies. When we look at the distribution of the variable ES1 for the entire population of GR5, we can see that the results show the same tendencies such as the majority of the group, 62.3 per cent, are economically active performing one job, and 27.3 per cent are not economically active due to household tasks.

When we compare GR3, GR4, and GR5 regarding 2 categories of ES1, namely, being economically active and performing one job, and not economically active due to household tasks, we remark that the per centages are very similar between the groups. Moreover, in terms of performing more than one job, the per centage of population in such situation increases from GR3 to GR5 such as the difference between GR3 and GR5 is about 3 per centage points, and the difference between GR3 and GR4 is about 1 per centage point.

We note that in the table below, for those are not economically active, the last 2 digits in their profile are 99 because simply they do not have any job.

Table 79. Economic activity profiles of GR5 based on the CAP114 model, data: Mexico MCS 2012

Pattern number of GR5 (Freq in % in full data)	Economic activity profile (Freq in % in full data)
1 (91)	11 (56.7), 699 (27.35), 12 (5.7), 899 (2.7), 999 (2.6), 399 (2.6)
2 (5)	11 (55.7), 699 (27.4), 12 (4.8), 399 (3.4), 899 (3.3)

Total info/GR5 population coverage in %	(91*0.976) +(5*0.946) = **93.5**

Fourth step (step4 analysis): detecting multidimensional poverty profiles of specific groups

To finalize the analysis, we will briefly focus on the case of elderly, unemployed, GR1, GR2 population respectively, and after that we will produce some cross-tabulations regarding some additional variables.

Elderly

In terms of the five deprivation groups, 28.5 per cent of elderly population live in the poorest group, 21 per cent live in GR4, 19.5 per cent live in GR3, 17.6 per cent live in GR2, and 13.2 per cent live in the least poor group. Besides, 14 per cent live alone, 32 per cent live in household composed of 2 persons, 17.9 per cent live in a household composed of 3 persons, and others live in a household composed at least by 4 persons. Besides, 15.6 per cent are deprived of medical coverage. For more tailored programs, questions in the survey should be asked accordingly taking in consideration the specific needs. Otherwise, we can see that the per centage of elderly people living in poorer groups such as GR3, GR4, and GR5 is higher. The implementation of policies in line with the findings on the deprivation groups' discovered capabilities patterns will certainly increase the quality of life of elderly population despite it is more convenient to address their specific needs in a particular module in the national level survey.

Unemployment

When we consider unemployed population based on the ES1 variable in the data, we can see that only 2.84 per cent of 18+ population is in such situation. Among this population, 25.55 per cent live in GR4, 22.4 per cent live in GR5, 21 per cent live in GR3, 17 per cent live in GR2, and 14 per cent live in GR1. So that, unemployed are likely to be living in poorer households. Most importantly, 51 per cent of unemployed are deprived of medical coverage, and 59 per cent do not have at least a high school diploma. Besides, only 4 per cent live alone, 10 per cent live in a household composed of 2 persons, the rest are in household composed of at least 3 persons. Moreover, 73.5 per cent are aged between 18 and 44 yrs. old, and 82 per cent are male.

On the other hand, economically not active population spending most of their time to performing household tasks (this is why they are reported to not being economically active in the labour market) are mostly composed of female population (97.5 per cent are female), but there is no significant difference in terms of age group. In terms of educational attainment, 11 per cent are illiterate, which is quite considerable rate, 36.65 per cent are primary school graduate, 28 per cent are secondary school graduate, and 12 per cent are high school graduate, and interestingly 10 per cent are vocational school or higher education graduates. Besides, 19 per cent of them are deprived of medical coverage. According to the 5 deprivation groups, 27.6 per cent are in GR5, 23.5 per cent are in GR4, 19.6 per cent are in GR3, and the rest are in GR2 or GR1.

GR1

We will briefly see the profiles of GR1 in terms of educational attainment, economic activity, medical coverage, and nutrition respectively. Regarding the first aspect, 18 per cent are

deprived. More specifically, 47 per cent have higher education diploma (university degree or equivalent), 22 per cent have high school diploma, 10.3 per cent have secondary school diploma, 10.2 per cent have vocational school degree, 6.8 per cent have primary school degree, and the per centage of illiterate in this group is almost zero. Compared to GR3, GR4, and GR5, there is a clear divergence in the sense that the per centage of higher education graduates is significantly higher among GR1. Concerning economic activity status, 59 per cent are economically active performing one job, 16.5 per cent are not economically active due to household tasks, 10.3 per cent are not economically active due to studies, 6 per cent are retired, 3 per cent are economically active performing more than one job, and 2 per cent are unemployed. Herewith, we can see that the share of economically active people performing one job among GR1 is around 7 per centage point higher than GR5, and the proportion of economically not active people due to household tasks among GR1 is about 10 per centage point lower compared to GR5. 97 per cent of that group is female similar to GR3, GR4, and GR5 cases; however, around 21 per cent of them have higher education degree. Moreover, 17.6 per cent of GR1 are deprived regarding medical coverage, which is 10 per centage point lower than GR5. And, regarding food availability in the household, 88.5 per cent of GR1 live in a household having not experienced any food shortage in terms of AF1, AF2, and AF3 indicators during the last 3 months. But it is important to highlight that 10 per cent live in a household having worried about the food running out due to lack of money or resources in the last 3 months.

GR2

Regarding educational attainment, 47.3 per cent of GR2 are found deprived. This rate is quite high for a relatively less poor group, and there is 29 per centage point difference in detriment of GR2 compared to GR1. Otherwise, 23.4 per cent are high school graduate, 24 per cent are secondary school graduate, 20 per cent are primary school, 19.8 per cent are higher education graduate, and 7.8 per cent are vocational school graduate. The rate of illiterate people is about 2 per cent in GR1, which is significantly lower than GR3, GR4, and GR5; but we can see that the rate of people achieving higher education diploma is halved among GR2 compared to GR1. In terms of economic activity, 58.7 per cent are economically active performing one job, 21 per cent are not economically active due to household tasks, 6.4 per cent are not economically active due to studies, 4.8 per cent are retired, 2.7 per cent are unemployed, and 2.5 per cent are economically active performing more than one job. 97 per cent of people economically not active due to household tasks are female, and 6.5 per cent of them have higher education diploma, and 66 per cent of them do not have at least a high school diploma, and 17 per cent of them do not have a medical coverage. Besides, 19 per cent of GR2 are deprived in terms of medical coverage. And, concerning food availability in the household, 68.7 percent of GR2 live in a household having not experienced any food shortage in terms of AF1, AF2, and AF3 indicators during the last 3 months. However, 2.4 per cent live in a household having experienced AF1, AF2, and AF3, and 29 per cent live in a household having worried about the food running out due to lack of money or resources, and 15.9 per cent live in household having had run out of money or resources to get a healthy and varied diet. Compared to GR1, the food situation in the household is worser, notably deprivation regarding AF2 is doubled.

Additional variables

We will finalize the Mexico MCS 2012 analysis by doing some cross-tables regarding the variables DWE, ETH, LNG, HSS, HS, and hh_child.

Regarding DWE, it is found that 94.70 per cent of population live in an independent house, 3.14 per cent live in an apartment in building, and 1.83 per cent live in a housing in neighborhood 'vivienda en vecindad'. Across the 5 deprivation groups, we can observe that independent house is the most used dwelling type; however, it is important to highlight that the per centage of living in vecindad increases from GR1 to GR5, and on the other the per centage of living in an apartment in building decreases from GR5 to GR1. Among the poorest group, 2.5 per cent of them live in vecindad vs. 6.3 per cent of the least poor group live in an apartment in building.

Concerning ethnicity (ETH), 70.7 per cent of population don't consider him or herself indigenous. Among the population considering him or herself indigenous, 68 per cent don't understand or don't speak indigenous language or a dialect. The rate of indigenous population, based on the answers to ETH, gradually increases from the least poor to the poorest group. For instance, 12.7 per cent of GR1 is indigenous whilst 49.7 per cent of GR5 does (among GR4 and GR3, the rate is around 33).

In terms of household size, households with at least 5 members construct 45.4 per cent of GR5 vs. 35 per cent among GR1. The rate of large size household increases gradually from the least poor to the poorest group given that the per centage of such type of household is quite considerable among the less poor groups. On the other hand, 35.5 per cent of population live in households with no child, 24 per cent live in households with one child, 22 per cent live in households with two children, and 7 per cent live in households with at least 4 children. If we look at deprivation groups level, we can observe that the per centage of people living in households with at least 4 children is the highest among GR5 corresponding to 13.2 per cent vs. 1.7 per cent among GR1, and 3.87 per cent among GR2; and the per centage of people living in households with no child is the highest among GR1 corresponding to 44.5 per cent. Besides, the per centage of people living in households with two children is quite similar across all the 5 deprivation groups.

Finally, we will check the situation across deprivation groups in terms of access to medical care/services when it was needed during the last 12 months (for those answered yes to the question 'Have you got sick or suffered from any pain, physical discomfort or accident that prevented/blocked you from doing your daily activities?'). Related to that, people almost never, or never, or a few times received medical care despite the need are considered deprived. Only 2.1 per cent of medically needy people from GR1 are found in deprivation whilst the rate increases to 9.8 among GR5, and 6.7 among GR4. If we look at specifically on GR5, we find that 42.75 per cent of deprived people did not have enough money to benefit the service, 18.4 per cent found the service expensive, 16.8 per cent had to wait a longtime to be treated, 10 per cent could not because the medical unit, clinic, or hospital is far from the dwelling, and 10 per cent reported that service was not provided even if the person attended the medical unit.

9.6 Case study of Brazil

Introduction

In case of Brazil, we will present in the following paragraphs the results of poverty analysis based on National Household Sample Survey (Pesquisa Nacional por Amostra de Domicilios-PNAD onwards) of 2011. We start off by describing shortly the aim and content of PNAD. The

information given below is provided by the national institute of statistics and geography of Brazil (Instituto Brasileiro de Geografia e Estatistica-IBGE onwards)[200].

As we did in the cases of South Africa, Turkey, and Mexico, here also in Brazil analysis we will follow the same methodology. In that vein, based on the PNAD 2011 data, we will carry out the following analysis to measure multidimensional poverty, respectively: detecting multidimensional poverty patterns, dimensional poverty profiles, sub-group profiles, and some additional profiles such as elderly, unemployed. The list of variables is given in section 4.2. Depending on the purpose of the model, we will select indicators. The short names of variables (see the codes in parenthesis in first column of the table of selected indicators) will be displayed in the outputs including graphics. Before describing models and presenting results, we will yield a brief information on PNAD and then the quality report of the variables.

The PNAD was first started in 1967, and last carried out by for the year 2015, which is now followed by Continuous National Sample Survey (Continuous PNAD)[201]. This survey is developed as a crucial element for designating and implementing living conditions improving programs/policies across the country, aiming at collecting various socio-economic, demographic, geographical information such as education, economic activity, unemployment, income, dwelling conditions, specific modules according to the context of the country (e.g., victimization and justice), for the country as a whole as well as major regions, federation units, 9 metropolitan regions (Belem, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, Sao Paulo, Curitiba, and Porto Alegre).

In our analysis, we will employ several variables relating to geographic situation, individual characteristic, household composition, dwelling situation, material situation, educational attainment, and economic activity (labour). Some variables are constructed based on the information collected through the individual questionnaire where the questions are asked for each of the household member (e.g., education, economic activity), whilst some other variables (e.g., dwelling conditions, material situation of the household) are based on the information collected through the household questionnaire where the questions concern the household as a whole or the dwelling where the all household members live (i.e., the household level variables' scores are the same for all members of the household in the data).

Quality report of variables

As a first fundamental step, we created unique household (hid-composed of 4 variables including, area, setting, control, and series numbers) and personal identifiers (id-composed of hid, serial number, and other 4 variables including gender, anniversary year, month, and age of the person) in order to merge data together. More specifically, 2 data are used: individual ('PES2011'), and household ('DOM2011'), respectively. The first data includes information on individual characteristic such as age, gender, educational attainment, economic activity; the second data covers information on the house or apartment or room in which the household lives such as wall, roof materials of the dwelling, whether a toilet, a bathroom exist or not in the

[200] Please refer to the link below for more information (accessed 25 December 2019):
https://www.ibge.gov.br/en/statistics/social/labor/20620-summary-of-indicators-pnad2.html?=&t=o-que-e

[201] Please refer to the link below for more information on the new PNDA survey from 2016 (accessed 25 December 2019):
https://www.ibge.gov.br/en/statistics/social/population/16833-monthly-dissemination-pnadc1.html?=&t=o-que-e

dwelling, and some socio-economic features of the household as a whole such as household assets (e.g., car, radio, tv), household income.

In household data, there are 111,986 households (after removing 34,221 records due to uncompleted survey results as indicated in the data), and in individual data there are 358,919 individuals including children. However, 306 duplicates are found in the household data; in other words, some households are found repeated in the data- nevertheless, one household should *only* be included only once in the household data. Thus, after dropping these duplicates, the household data contain 111,680 unique households. On the other hand, in individual data, there are 358,919 unique individuals across 111,833 households. In order to merge correctly household and individual data, i.e., for avoiding overlapping, 306 duplicates are therefore dropped from the household data[202]. When we merge household and individual data via the unique household identifier (hid), 358.017 individuals are found correctly matched whilst 902 individuals do not have information in household data. Thus, we also drop these 902 individuals in the merge data, which leads to a merged data composed of 358,017 unique individuals across 111,680 households. We also keep only permanent private domiciles as the dwelling related questions were asked only to this type of domiciles, which means that 1,0111 records are deleted. And, finally in the data there are 357,006 unique individuals across 111,305 unique households. According to statistics given by the World Bank (data source is United Nations World Population Prospects), the population of Brazil in 2011 was 197,514,534. Based on our final data, we obtain an estimation of 194,266,779 (by employing the individual weight provided in the data by IBGE), which corresponds to 98.3 per cent of the estimation published by the World Bank[203].

Based on the final merged data, all indicators (i.e., derived variables) defined in the methodology (section 4.2) are created and later derived data including both initial and derived variables is produced. Derived data and its subsets form the basis for our analysis. 17 variables are taken to construct the overall deprivation index, which means that all persons not having a full information of these variables were dropped from the derived data. The 17 variables, with their short names, are as follows: WAL, ROF, RO1, WATD, SAN, RUB, ENG_L, ENG_C, TEL, MPHO, RAD, TV, REF, WM, CAR, COMPINT, and EDUs (see the table in first model sub-section below for information on these variables).

Our analysis will cover persons aged 18 or over. In the data, there are 252,498 persons in this category, which constructs 70.72 per cent of the data. We remove in total 2,353 individuals aged 18 or over because they have missing values regarding at least one of the variables, WAL, ROF, RUB, ENG_L, and ENG_C. We lose only 0.93 per cent of the data, which can be considered quite negligible. Thus, the final data on which the analysis will be done includes 250,145 persons aged 18 or over across 110,166 households in the country. Additional to 32 variables mentioned in first paragraph, we created also several other variables for

[202] It is important to note that when we create unique household identifier based on 10 variables including strata, municipality, sector, domicile, psu numbers next to the control, setting, area, series numbers, there is no duplicates in the household data. But problem is that we are obligated to use only the *common* variables to create a *common* household identifier in both household and individual data because these 2 data can only be merged via the common household identifier. This issue of duplicates is due to the structure of the data, which can only be solved by IBGE.

[203] Please refer to the link below for further information (accessed 27 December 2019):
https://data.worldbank.org/country/brazil

decomposition and supplementary analysis. There are in total 53 variables[204] including PATH variable that corresponds to capabilities profile of each individual aged 18 or more in the data based on the main 17 variables that are used in building the overall deprivation index. The variable PATH is a string variable reflecting the combination of 17 variables, as follows:

PATH=WAL+ROF+RO1[205]+WATD+SAN+RUB+ENG_L+ENG_C+EDU+TEL+MPHO+RAD+TV+REF+WM+CAR+COMPINT.

It is found 15,175 different capabilities profile (PATH) among 250,145 individuals in the data. In other words, 6 % of the 18+ population in the derived data has her own singular capability profile/path. Such result is very complex to examine. So, we need to find out the most influential ones which we will discover by employing classification tree analysis. We obtained the quality report of 53 variables in statistical software R via the package Hmisc[206]. We note also that descriptive statistics and analysis is based on the non-weighted data[207]. The short name of the variables corresponds to their codes mentioned in section 4.2.

Descriptive statistics outputs:

```
BR2011derived[, 1:53]

 53  Variables      250145  Observations
---------------------------------------------------------------------------
hid
      n  missing distinct
 250145        0   110166

lowest : 111500017601 111500017602 111500017603 111500017604 111500017605
highest: 385200397209 385200397210 385200397212 385200397214 385200397215
---------------------------------------------------------------------------
id
      n  missing distinct
 250145        0   250145

lowest :  11150001760101207060 11150001760102415030 11150001760103223010
11150001760104223010 11150001760201214050
highest: 38520039721201204060 38520039721202414090 38520039721401229050
38520039721402429050 38520039721501420090
---------------------------------------------------------------------------
case_string
      n  missing distinct
 250145        0       20
```

[204] The variables, namely age (numerical variable indicating age of the person), dwid, hid, id, serial number amongst others for decomposition and supplementary analysis are included.

[205] The variable RO is numeric representing the fraction of number of rooms in the household by the household size. If RO is inferior than 1, we consider deprived the household regarding this aspect (this deprivation variable is called RO_D). However, for building the PATH variable, we transformed RO_D to a string variable, and we named it RO1. In the quality report, the information for the variable RO1 will be given.

[206] See South Africa GHS 2011 analysis for more technical information on statistical software and packages. But also, we remark that missing values are coded 99 in the data.

[207] To obtain estimated numbers at national level typically for the national reports, the obtained figures can be multiplied by the appropriate weight provided in the data by the Turkey Institute of Statistics. But, in our analysis, we use the national sample, which is also an appropriate representative of the population in the country.

1 (110014, 0.440), 2 (88516, 0.354), 3 (30090, 0.120), 4 (12959, 0.052), 5 (5096, 0.020), 6 (1989, 0.008), 7 (817, 0.003), 8 (353, 0.001), 9 (172, 0.001), 10 (65, 0.000), 11 (27, 0.000), 12 (23, 0.000), 13 (8, 0.000), 14 (8, 0.000), 15 (3, 0.000), 19 (1, 0.000), 20 (1, 0.000), 25 (1, 0.000), 26 (1, 0.000), 28 (1, 0.000)

HS

n	missing	distinct
250145	0	5

Value	1	2	3	4	5
Frequency	13883	48322	63999	61080	62861
Proportion	0.055	0.193	0.256	0.244	0.251

GEN

n	missing	distinct
250145	0	2

Value	1	2
Frequency	131481	118664
Proportion	0.526	0.474

---

hh_child

n	missing	distinct
250145	0	6

Value	1	2	3	4	5	6
Frequency	114455	69214	41658	15580	5538	3700
Proportion	0.458	0.277	0.167	0.062	0.022	0.015

WAL

n	missing	distinct
250145	0	5

Value	1	2	3	4	5
Frequency	226659	20146	2095	1177	68
Proportion	0.906	0.081	0.008	0.005	0.000

ROF

n	missing	distinct
250145	0	6

Value	1	2	3	4	5	6
Frequency	189458	54991	448	4023	79	1146
Proportion	0.757	0.220	0.002	0.016	0.000	0.005

RO1

n	missing	distinct
250145	0	2

Value	0	1
Frequency	29820	220325
Proportion	0.119	0.881

WATD

n	missing	distinct
250145	0	4

Value	1	2	3	4

```
Frequency  205895  14309  29069    872
Proportion  0.823  0.057  0.116  0.003
---------------------------------------------------------------------------
SAN
      n  missing distinct
 250145        0        3

Value          1      2      3
Frequency  241550   1729   6866
Proportion  0.966  0.007  0.027
---------------------------------------------------------------------------
RUB
      n  missing distinct
 250145        0        5

Value          1      2      3      4      5
Frequency  208832  14388  23931   2751    243
Proportion  0.835  0.058  0.096  0.011  0.001
---------------------------------------------------------------------------
ENG_L
      n  missing distinct
 250145        0        2

Value          1      3
Frequency  248752   1393
Proportion  0.994  0.006
---------------------------------------------------------------------------
ENG_C
      n  missing distinct
 250145        0        4

Value          1      2      3      4
Frequency    2614   5643      3 241885
Proportion  0.010  0.023  0.000  0.967
---------------------------------------------------------------------------
EDU
      n  missing distinct
 250145        0       12

Value        0     1    10    11     2     3     4     5     6     7     8
Frequency  22538 25073  3635 28903   734     6    39 28936 12127 57145 68604
Proportion 0.090 0.100 0.015 0.116 0.003 0.000 0.000 0.116 0.048 0.228 0.274

Value          9
Frequency   2405
Proportion 0.010
---------------------------------------------------------------------------
TEL
      n  missing distinct
 250145        0        2

Value          1      2
Frequency  102536 147609
Proportion   0.41   0.59
---------------------------------------------------------------------------
MPHO
      n  missing distinct
 250145        0        2

Value          1      2
```

Frequency 223699 26446
Proportion 0.894 0.106

RAD
      n missing distinct
  250145   0   2

Value   1   2
Frequency 243279   6866
Proportion 0.973 0.027

TV
      n missing distinct
  250145   0   2

Value   1   2
Frequency 243334   6811
Proportion 0.973 0.027

REF
      n missing distinct
  250145   0   2

Value   1   2
Frequency 241001   9144
Proportion 0.963 0.037

WM
      n missing distinct
  250145   0   2

Value   1   2
Frequency 128715 121430
Proportion 0.515 0.485

CAR
      n missing distinct
  250145   0   2

Value   1   2
Frequency 104658 145487
Proportion 0.418 0.582

COMPINT
      n missing distinct
  250145   0   3

Value   1   2   3
Frequency   96940 16905 136300
Proportion 0.388 0.068 0.545

AGE
      n missing distinct
  250145   0   6

Value   1   2   3   4   5   6
Frequency 41313 58853 50098 42759 29440 27682
Proportion 0.165 0.235 0.200 0.171 0.118 0.111

FED

```
      n  missing distinct
 250145        0       27

lowest : 11 12 13 14 15, highest: 43 50 51 52 53
---------------------------------------------------------------------------
SET
      n  missing distinct
 250145        0        8

Value          1      2      3      4      5      6      7      8
Frequency  212518   3252   1441    887   5074    191    299  26483
Proportion  0.850  0.013  0.006  0.004  0.020  0.001  0.001  0.106
---------------------------------------------------------------------------
SET1
      n  missing distinct
 250145        0        2

Value          1      2
Frequency   32934 217211
Proportion  0.132  0.868
---------------------------------------------------------------------------
REG
      n  missing distinct
 250145        0        5

Value          1     2     3     4     5
Frequency  36162 71995 75229 40604 26155
Proportion 0.145 0.288 0.301 0.162 0.105
---------------------------------------------------------------------------
POP
      n  missing distinct
 250145        0        5

Value          1      2      3      4      5
Frequency  113991  23326   1365 110303   1160
Proportion  0.456  0.093  0.005  0.441  0.005
---------------------------------------------------------------------------
DWE
      n  missing distinct
 250145        0        3

Value          1      2      3
Frequency  224050  25302    793
Proportion  0.896  0.101  0.003
---------------------------------------------------------------------------
WFIL
      n  missing distinct
 250145        0        2

Value          1      2
Frequency  129421 120724
Proportion  0.517  0.483
---------------------------------------------------------------------------
ES1
      n  missing distinct
 250145        0        2

Value          1      2
Frequency  161463  88682
Proportion  0.645  0.355
```

```
---------------------------------------------------------------------------
JOB
      n  missing distinct
 250145        0        4

Value          0      1      2      3
Frequency   88682 155079   5924    460
Proportion  0.355  0.620  0.024  0.002
---------------------------------------------------------------------------
OG
      n  missing distinct
 250145        0       11

Value          0     1     2     3     4     5     6     7     8     9    99
Frequency  88682  7424 14193 11405 14351 32159 18467 22695 39230  1447    92
Proportion 0.355 0.030 0.057 0.046 0.057 0.129 0.074 0.091 0.157 0.006 0.000
---------------------------------------------------------------------------
IND
      n  missing distinct
 250145        0       13

Value          0     1    10    11     2     3     4     5     6     7     8
Frequency  88682 22974 11557  6254  1344 19226 14018 29054  7930  9089  9808
Proportion 0.355 0.092 0.046 0.025 0.005 0.077 0.056 0.116 0.032 0.036 0.039

Value          9    99
Frequency  15646 14563
Proportion 0.063 0.058
---------------------------------------------------------------------------
ES3
      n  missing distinct
 250145        0       12

Value          0     1    10    11     2     3     4     5     6     7     8
Frequency  88682 62602   160  4056   464 12373 23437  3588  7969 35622  5590
Proportion 0.355 0.250 0.001 0.016 0.002 0.049 0.094 0.014 0.032 0.142 0.022

Value          9
Frequency   5602
Proportion 0.022
---------------------------------------------------------------------------
WH
      n  missing distinct
 250145        0        6

Value          0     1     2     3     4     5
Frequency  88682  8121 31405 71763 23936 26238
Proportion 0.355 0.032 0.126 0.287 0.096 0.105
---------------------------------------------------------------------------
SC
      n  missing distinct
 250145        0        3

Value          0     1     2
Frequency  88682 95895 65568
Proportion 0.355 0.383 0.262
---------------------------------------------------------------------------
BEN_1
      n  missing distinct
 250145        0        4
```

```
Value          0     1     2     3
Frequency   88682  4650 105783 51030
Proportion  0.355 0.019 0.423 0.204
---------------------------------------------------------------------------
BEN_2
       n  missing distinct
  250145        0        4

Value          0     1     2     3
Frequency  88682 45950 64483 51030
Proportion 0.355 0.184 0.258 0.204
---------------------------------------------------------------------------
BEN_3
       n  missing distinct
  250145        0        4

Value          0     1     2     3
Frequency  88682 45103 65330 51030
Proportion 0.355 0.180 0.261 0.204
---------------------------------------------------------------------------
BEN_4
       n  missing distinct
  250145        0        4

Value          0     1     2     3
Frequency   88682  2522 107911 51030
Proportion  0.355 0.010 0.431 0.204
---------------------------------------------------------------------------
BEN_5
       n  missing distinct
  250145        0        4

Value          0     1     2     3
Frequency  88682 19631 90802 51030
Proportion 0.355 0.078 0.363 0.204
---------------------------------------------------------------------------
UPW
       n  missing distinct
  250145        0        4

Value          0     1     2     3
Frequency   88682  5002 36210 120251
Proportion  0.355 0.020 0.145 0.481
---------------------------------------------------------------------------
CNPJ
       n  missing distinct
  250145        0        5

Value          0     1     2     3    99
Frequency  88682 71642 40538 42553  6730
Proportion 0.355 0.286 0.162 0.170 0.027
---------------------------------------------------------------------------
INV
       n  missing distinct
  250145        0        5

Value          0     1     2     3    99
Frequency  88682 54672 11767 83765 11259
Proportion 0.355 0.219 0.047 0.335 0.045
```

PAYC
n missing distinct
250145 0 5

Value 0 1 2 3 99
Frequency 88682 54143 17485 83765 6070
Proportion 0.355 0.216 0.070 0.335 0.024

DR
n missing distinct
250145 0 8

Value 0 1 2 3 4 5 6 99
Frequency 6866 122964 24144 43619 44846 3494 3946 266
Proportion 0.027 0.492 0.097 0.174 0.179 0.014 0.016 0.001

GH
n missing distinct
250145 0 2

Value 1 2
Frequency 86978 163167
Proportion 0.348 0.652

AH
n missing distinct
250145 0 5

Value 1 2 3 4 5
Frequency 208 9418 40155 159320 41044
Proportion 0.001 0.038 0.161 0.637 0.164

POPH
n missing distinct
250145 0 5

Value 1 2 3 4 5
Frequency 111423 25442 1499 110511 1270
Proportion 0.445 0.102 0.006 0.442 0.005

EDUH
n missing distinct
250145 0 12

Value 0 1 10 11 2 3 4 5 6 7 8
Frequency 28497 12454 4228 28926 927 7 36 42560 18759 55897 54818
Proportion 0.114 0.050 0.017 0.116 0.004 0.000 0.000 0.170 0.075 0.223 0.219

Value 9
Frequency 3036
Proportion 0.012

ES1H
n missing distinct
250145 0 2

Value 1 2
Frequency 173421 76724
Proportion 0.693 0.307

We note that in the following analysis concerning classification trees by applying the package rpart in R, we will give only the essential statistical outputs.

First step (step1 analysis): detecting multidimensional poverty patterns

In this investigation, the decision trees-based capabilities analysis will be based on the overall deprivation index we will built. First of all, we define a deprivation line for each of the 17 variables. Deprivation line is the criteria of distinguishing who is deprived or not regarding the variable. As we have already mentioned in GHS 2011 case, the utility of each variable representing a different functioning is determined according to universal human rights chart. According to the deprivation threshold, if a person experiences deprivation regarding the variable then he or she takes the score 1, and 0 otherwise for that variable. Taking in consideration all variables simultaneously, we count the number of deprivations the person suffers to obtain her/his overall individual deprivation score. Then, as a function of the distribution of all person's overall deprivation index scores in the data, we create several deprivation groups. At last stage, we will proceed with classification tress method to discover the most influential capabilities patterns emerging in the data distinguishing deprivation groups among them. The table below gives information on deprivation line for each variable used in the index. The table contains 4 columns: first column gives the code name of the variable given that each variable represents a functioning, second column mentions deprivation line regarding the variable, third column indicates the utility of the functioning for being capable of living a decent life according to universal human rights chart, and last column contains remarks about the variable.

Table 80. Deprivation lines and composition of variables for step 1: overall deprivation index and multidimensional poverty patterns, data: Brazil PNAD 2011

Variable (each variable represents a functioning)	**Deprivation line[208]**	**The utility of functionings for having the capability of living a decent life according to Universal Human Rights Chart[209]**	**Remark**
WAL_D	Deprived (WAL_D=1) if WAL is 2,3,4,5	Article 25: functioning of living with an adequate housing.	Rigged wood, uncoated wall of mud, harnessed wood, and straw are considered as not appropriate wall materials compared to masonry.
ROF_D	Deprived (ROF_D=1) if ROF is 3,4,5,6	Article 25: functioning of living with an adequate housing.	Rigged wood, zinc, harnessed wood, and straw are considered as not appropriate roof materials

[208] For the codes and categories, see the table provided in section 4.2.

[209] Please refer to the link below for further information (accessed 12 February 2020) : https://www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/eng.pdf

			compared to tile, and stone slab.
RO_D	Deprived if the number of rooms available in the dwelling divided by the household size is inferior than 1	Article 25: functioning of living with an adequate housing.	Excluding kitchen, bathroom, garden, and entranceway, in normal conditions every single person should have at least 1 room for her/his privacy.
WATD_D	Deprived (WATD_D=1) if piped water system is not inside the house or the water comes from well or spring or other source outside the dwelling	Article 25: functioning of living with an adequate housing.	This variable is composed of the variables PW1 and PW2 (see section 4.2). A household is considered not deprived if the piped water system is available for at least one room in the house/dwelling and comes from the general distribution network.
SAN2_D	Deprived (SAN2_D=1) if there is no toilet/bathroom in the dwelling or if it is shared by other households in the dwelling	Article 25: functioning of living with an adequate housing.	
RUB_D	Deprived (RUB_D=1) if RUB is 2,3,4,5	Article 25: functioning of living with an adequate housing.	
ENG_L_D	Deprived (ENG_L_D=1) if ENG_L is 3	Article 25: functioning of living with an adequate housing.	Canister oil, kerosene, gas are considered as not appropriate energy sources for lighting compared to electric.
ENG_C_D	Deprived (ENG_C_D=1) if there is no one-burner stove neither two-burner stove in the dwelling, or there is a two-burner stove but the energy source is a gas cylinder or firewood or coal in the dwelling, or there is a two-burner stove but the energy source is a gas cylinder or firewood or coal in the dwelling	Article 25: functioning of living with an adequate housing.	The variable is composed of the two variables namely ENG_C1 and ENG_C2.

TEL_D	Deprived (TEL_D=1) if the household does not have a landline telephone	Article 25: functioning of living with an adequate housing in terms of material and social service regarding availability of landline network in the neighborhood where the person lives. It is not necessary that telephone, internet networks are provided by the State, but at least State should make sure that every single citizen has access to such services.	
MPHO_D	Deprived (MPHO_D=1) if the household does not have a mobile phone	The same argument mentioned for TEL_D.	
RAD_D	Deprived (RAD_D=1) if the household does not have a radio	Article 25: functioning of living with an adequate housing in terms of material.	
TV_D	Deprived (TV_D=1) if the household does not have a television	Article 25: functioning of living with an adequate housing in terms of material.	
REF_D	Deprived (REF_D=1) if the household does not have a refrigerator	Article 25: functioning of living with an adequate housing in terms of material.	
WM_D	Deprived (WM_D=1) if the household does not have a washing machine	Article 25: functioning of living with an adequate housing in terms of material.	

CAR_D	Deprived (CAR_D=1) if the household does not have a car	Article 25: functioning of living with an adequate housing in terms of material.	
COMPINT_D	Deprived (COMPINT_D=1) if there is no computer in the dwelling	Article 25: functioning of living with an adequate housing in terms of material.	
EDU_D	Deprived (EDU_D=1) if EDU is 0,1,2,3,4,5,6,7,9 (less than high school level diploma)	Article 26: functioning of having an educational attainment[210].	

Regarding the overall deprivation index 'DEP' we built based on 17 variables above, 0.30 per cent of the target population (18+) has no deprivation whilst 0.02 percent of them has deprivation in 17 items. The most frequent groups are those having deprivation in 3,4,5, or 6 items, which represent 14.23, 14.12, 14.19, 14.60 per cent of the population, respectively. We drop 752 persons having 0 deprivation (the most prosperous group) in the data. After that, there are in total 249,393 persons in the data. The summary statistics of DEP is as follows:

```
                    Total Deprivation
-------------------------------------------------------------
      Percentiles      Smallest
 1%             1              1
 5%             2              1
10%             2              1       Obs                 249393
25%             3              1       Sum of Wgt.         249393

50%             5                      Mean           5.367657
                        Largest        Std. Dev.      2.509228
75%             7             17
90%             9             17       Variance       6.296225
95%            10             17       Skewness       .7230904
99%            13             17       Kurtosis       3.711054
```

As we did in previous country analysis, here also we apply the same methodology and we built a deprivation groups variable namely DEP_GRS regrouping the population by 20th, 40th, 60th, 80th, and 100th percentiles. However, from the summary statistics above we can see that, the division of the population into less than 5 different deprivation groups will be also convenient given that diversity is less striking as the median corresponds to 5 deprivations and the upper 25 per cent starts at 7 deprivations. In that vein, we will create 3 groups as follows: 1st group

[210] Article 26 mentions that "Everyone has the right to education. Education shall be free, at least in the elementary and fundamental stages. Elementary education shall be compulsory. Technical and professional education shall be made generally available and higher education shall be equally accessible to all on the basis of merit". However, we argue that at least a general high school diploma or its equivalent in terms of vocational technical high school is a fundamental necessary for people to obtain opportunities for higher professional education and specific job skills.

(DEP_GRS1) includes people having deprivation in 1 to 4 items; 2nd group (DEP_GRS2) includes people having deprivation in 5 to 6 items: 3rd group (DEP_GRS3) includes people having deprivation in 7 to 17 items. When we take into account simultaneously 17 deprivation variables (e.g., ROF_D, WAL_D-see the table of deprivation lines above) to find out the capabilities profiles of 249,393 individuals in the data (the variable named PATH17), the results show that there are 3,037 different profiles, which is relatively high amount. For instance, with the most frequent 5 profiles (with at least 100 observations each of them), we can only capture 26.16 per cent of the population. But also, in terms of profiles with at least 100 observations, by employing 50 of 3,037 profiles we can obtain information for 67.84 per cent of the population as follows:

PATH17	Freq.	Percent	%<=
00100001000000000	21256	8.52	8.52
00100001110000111	16608	6.66	15.18
00100001100000000	13514	5.42	20.60
00100001010000111	7444	2.98	23.59
00100001010000000	6422	2.58	26.16
00100001110000011	5585	2.24	28.40
00100001100000010	4872	1.95	30.35
00100001000000010	4816	1.93	32.29
00100001110000000	4140	1.66	33.95
00110101110000111	3863	1.55	35.49
00100001100000111	3501	1.40	36.90
00100001110000110	3333	1.34	38.23
00100001010000010	3252	1.30	39.54
00100001010000011	3247	1.30	40.84
00100001110000010	3198	1.28	42.12
00100001100000011	3095	1.24	43.36
00100001110000101	3086	1.24	44.60
00100001100000001	3044	1.22	45.82
00100001010000110	2890	1.16	46.98
00100001111000111	2861	1.15	48.13
00100101110000111	2814	1.13	49.26
00100001110000001	2660	1.07	50.32

00110001110000111	2620	1.05	51.37
00100001000000100	2342	0.94	52.31
00100001100000110	2171	0.87	53.18
00100001010000001	2110	0.85	54.03
00000001000000000	2078	0.83	54.86
00100001100000100	2068	0.83	55.69
00100001010000100	2043	0.82	56.51
00100001000000110	1985	0.80	57.31
00100001000000001	1911	0.77	58.07
00100001110000100	1878	0.75	58.83
00100001010000101	1831	0.73	59.56
00000001110000111	1802	0.72	60.28
00100001000000111	1760	0.71	60.99
00100000000000000	1733	0.69	61.68
00100001000000011	1626	0.65	62.33
00110101111000111	1446	0.58	62.91
10100001110000111	1416	0.57	63.48
00110101110000101	1319	0.53	64.01
00100001100000101	1276	0.51	64.52
00000001111000111	1028	0.41	64.94
00110001010000111	984	0.39	65.33
00100101010000111	955	0.38	65.71
10110101110000111	945	0.38	66.09
00110001000000000	914	0.37	66.46
00100101111000111	898	0.36	66.82
10100001110000011	880	0.35	67.17
00000001010000111	870	0.35	67.52
00000001100000000	802	0.32	67.84

The table below shows the number of persons in the group, unique capabilities profiles based on the 17 variables and the ratio of unique profiles by the number of persons in the group.

Table 81. Ratio of unique profiles to the number of persons in the deprivation group (based on the variable DEP_GRS and 17 variables), data: Brazil PNAD 2011

Group	Unique capabilities profiles	Number of persons in the group	Ratio (Unique profiles/Number of persons in the group)
DEP_GRS1 (1-4 deprivations)	400	100,907	0.00396
DEP_GRS2 (5-6 deprivations)	721	71,810	0.01004
DEP_GRS3 (7-17 deprivations)	1,916	76,676	0.024
Total (whole sample)	**3,037**	**249,393**	**0.012**

As we can see from the data there is a larger variety of living conditions for people in the third deprivation group, compared to the 2 others. In fact, as the ratio numbers show, compared to the population of the groups the number of unique capabilities is quite low; however, it is still very complicated to interpret these capabilities profiles. Within first group, the most frequent 10 profiles represent 65 per cent of the group but still in total 107 profiles reach to explain 97 per cent of the 1st group; within second group, the most frequent 10 profiles capture 53 per cent of the group (and 111 profiles explain 94 per cent of the 2nd group); within third group, the most frequent 10 profiles provide information for 45.16 per cent of the group (and 159 profiles explain 85.3 per cent of the 3rd group). The analysis of capabilities is here less complicated than South Africa, Turkey, and Mexico cases, probably due to the less items in the overall deprivation index; nevertheless, it remains still complicated to interpret the combination 17 variables through at least 50 profiles for obtaining a synthetic information on poverty across Brazil. For that reason, in next paragraphs, we will show how to deal with such complexity by employing classification trees method. The goal is to find out the most informative capabilities patterns distinguishing the 3 deprivation groups. More specifically, we will do in total 3 classification tree analysis predicting how the deprivation groups distinguish among them by 17 deprivation variables of actual capabilities: {DEP_GRS1 vs. DEP_GRS2, DEP_GRS1 vs. DEP_GRS3, DEP_GRS2 vs. DEP_GRS3}.

Table 82. Per centage of groups in the sample of classification target groups, data: Brazil PNAD 2011

Target group in classification	Per centage of groups in the sample
DEP_GRS1 vs. DEP_GRS2	{58.4; 41.6}
DEP_GRS1 vs. DEP_GRS3	{56.8; 43.2}
DEP_GRS2 vs. DEP_GRS3	{48.3; 51.7}

As the table above shows we have a relatively balanced data but still we don't have a 100 per cent equal distribution between two groups.

CAP111: GR1 vs. GR2 analysis

In this analysis, we focus on DEP_GRS1 vs. DEP_GRS2. We use the data containing only people from the deprivation group 1 and the deprivation group 2. The outcome/target response variable is dichotomous, namely CAP111, taking the value 1 for the group 1 and the value 2 for the group 2 (so that people identified in the group 1 take the value 1 and people identified in the group 2 take the value 2 according to CAP111).

Table 83. Training vs. test samples in CAP111 analysis, data: Brazil PNAD 2011

	GR1	GR2
Training sample	67,271	47,873
Test sample	33,636	23,937

As the table above shows, there are 115,144 individuals in the training data (corresponding to 67 per cent of the initial data) and 57,573 individuals in the test sample. Training and test data are obtained by randomly partitioning the initial data/total sample into two subsets by 70 per cent vs. 30 per cent respectively. The data is quite balanced with 58.4 per cent for GR1 and 41.6 per cent for GR2 in both training and test samples (in the initial data, the proportion is also the same).

Concerning the rpart application we have set cp at 0.025 and we used gini index as split criterion, and we kept other parameters at their default values. Classification rules are discovered on the training data and they will be tested on the test data for evaluating predictive accuracy. The classification tree is called CAP111, and the model is named dtreeCAP111. The results in the box below show that the initial saturated tree given the parameters we used contains 34 splits (i.e., 35 terminal nodes) with a cross-validation error rate around 0.11. This tree is quite big and therefore difficult for interpretation, so we will prune the tree to around 10-11 splits to obtain an optimal interpretation, which yields also a relatively low level of cross-validation error rate (around 0.24).

```
> printcp(dtreeCAP111)

Classification tree:
rpart(formula = CAP111 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
 [1] CAR_D     COMPINT_D EDU_D     MPHO_D     RO_D      RUB_D     TEL_D
 [8] WAL_D     WATD_D    WM_D

Root node error: 47873/115144 = 0.41577

n= 115144

          CP nsplit rel error  xerror      xstd
1  0.4834458      0   1.00000 1.00000 0.0034934
2  0.0563157      1   0.51655 0.51655 0.0029108
3  0.0291187      3   0.40392 0.40392 0.0026496
4  0.0178876      5   0.34569 0.34569 0.0024866
5  0.0166273      8   0.29202 0.29202 0.0023150
6  0.0144758     10   0.25877 0.25877 0.0021963
7  0.0122094     11   0.24429 0.24429 0.0021412
8  0.0110918     13   0.21987 0.21987 0.0020428
9  0.0110814     14   0.20878 0.20300 0.0019704
10 0.0090448     16   0.18662 0.18662 0.0018962
11 0.0071718     17   0.17757 0.17757 0.0018535
```

```
12 0.0057235     20   0.15606 0.15606 0.0017459
13 0.0052848     21   0.15034 0.15119 0.0017204
14 0.0034675     22   0.14505 0.14505 0.0016874
15 0.0031403     23   0.14158 0.14158 0.0016683
16 0.0029174     26   0.13216 0.13204 0.0016145
17 0.0028617     29   0.12341 0.12441 0.0015698
18 0.0027155     30   0.12055 0.12203 0.0015555
19 0.0026737     31   0.11783 0.11748 0.0015278
20 0.0026320     32   0.11516 0.11693 0.0015244
21 0.0025000     34   0.10989 0.11365 0.0015040
```

The CAP111 pruned tree, is composed of 12 terminal nodes providing 6 patterns of GR1, and 6 patterns of GR2. As we can see from the pruned tree below, the discovered patterns from the left to the right side of the tree are as follows: (1) no deprivation regarding having a computer (COMPINT), a washing machine (WM), and a landline telephone (TEL) in the household's dwelling; (2) no deprivation regarding COMPINT, WM, and educational attainment (EDU), but deprivation regarding TEL; (3) no deprivation regarding COMPINT, WM, and having a car (CAR) in the household, but deprivation regarding TEL, and EDU; (4) no deprivation regarding COMPINT, and WM, but deprivation regarding TEL, EDU, and CAR; (5) no deprivation regarding COMPINT, CAR, and EDU, but deprivation regarding WM; (6) no deprivation regarding COMPINT, CAR, and TEL, but deprivation regarding WM, and EDU; (7) no deprivation regarding COMPINT, and CAR, but deprivation regarding WM, EDU, and TEL; (8) no deprivation regarding COMPINT, but deprivation regarding WM, and CAR; (9) deprivation regarding COMPINT, and CAR; (10) no deprivation regarding CAR, but deprivation regarding COMPINT, and TEL; (11) no deprivation regarding CAR, TEL, and WM, but deprivation regarding COMPINT; (12) no deprivation regarding CAR, and TEL, but deprivation regarding COMPINT, and WM.

The patterns of GR1 are 1st, 2nd, 11th, 5th, 3rd, and 6th with 58, 12.8, 6.3, 5.5, 4.6, and 2.2 per centage of coverage respectively (see the table below). For instance, with the 1st pattern, 58 per cent of people from GR1 are detected. On the other hand, the patterns of GR2 are 9th, 10th, 8th, 4th, 7th, and 12th with 48, 15, 15, 6, 3.5, and 2.8 per centage of coverage respectively (see the table below). For instance, with the 9th pattern, 48 per cent of people from GR2 are captured. In the training data, by using the CAP111 pruned tree classification patterns, we lose 11 per cent of information for GR1 and 10 per cent of information for GR2.

Table 84. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP111 pruned tree, data: Brazil PNAD 2011

Pattern no	**Coverage of pattern (in %) among the total population of GR1**	**Cumulative coverage (in %)**
1	58	58
2	12.8	70.8
11	6.3	77.1
5	5.5	82.6
3	4.6	87.2
6	2.2	89.4~**89**

Table 85. Coverage of GR2's patterns among the total population of GR2 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP111 pruned tree, data: Brazil PNAD 2011

Pattern no	Coverage of pattern (in %) among the total population of GR2	Cumulative coverage (in %)
9	48	48
10	15	63
8	15	78
4	6	84
7	3.5	87.5
12	2.8	90.3~**90**

We can also say that through the 6 patterns of GR1, deprivation regarding TEL, WM, EDU, and COMPINT concern 17.4, 7.7, 6.8, and 6.3 per cent of the entire population of GR1 (among 67,271 persons) respectively. On the other hand, concerning GR2, the 6 patterns show that the variables CAR, COMPINT, TEL, WM, and EDU affect 69, 66, 24.5, 21, and 9.5 per cent of the entire population of GR2 (among 47,873 persons). The 4 deprivation variables, COMPINT, TEL, WM, and EDU are the common in terms of deficits for both GR1 and GR2; however, the magnitude of COMPINT is almost 10 times higher, and those of TEL and EDU are 1.4 times higher, and those of WM is 2.7 times higher in terms of per centage points for GR2, compared to GR1.

```
       Confusion Matrix and Statistics

          Reference
Prediction   GR1   GR2
       GR1 29996  2262
       GR2  3640 21675

               Accuracy : 0.8975
                 95% CI : (0.895, 0.9)
    No Information Rate : 0.5842
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7907

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.8918
            Specificity : 0.9055
         Pos Pred Value : 0.9299
         Neg Pred Value : 0.8562
             Prevalence : 0.5842
         Detection Rate : 0.5210
   Detection Prevalence : 0.5603
      Balanced Accuracy : 0.8986

       'Positive' Class : GR1
```

Regarding confusion matrix statistics, this model has a relatively good overall accuracy of around 90 per cent, and the rate of sensitivity is 89 per cent, and the rate of specificity is 90 per cent, which are relatively good as well.

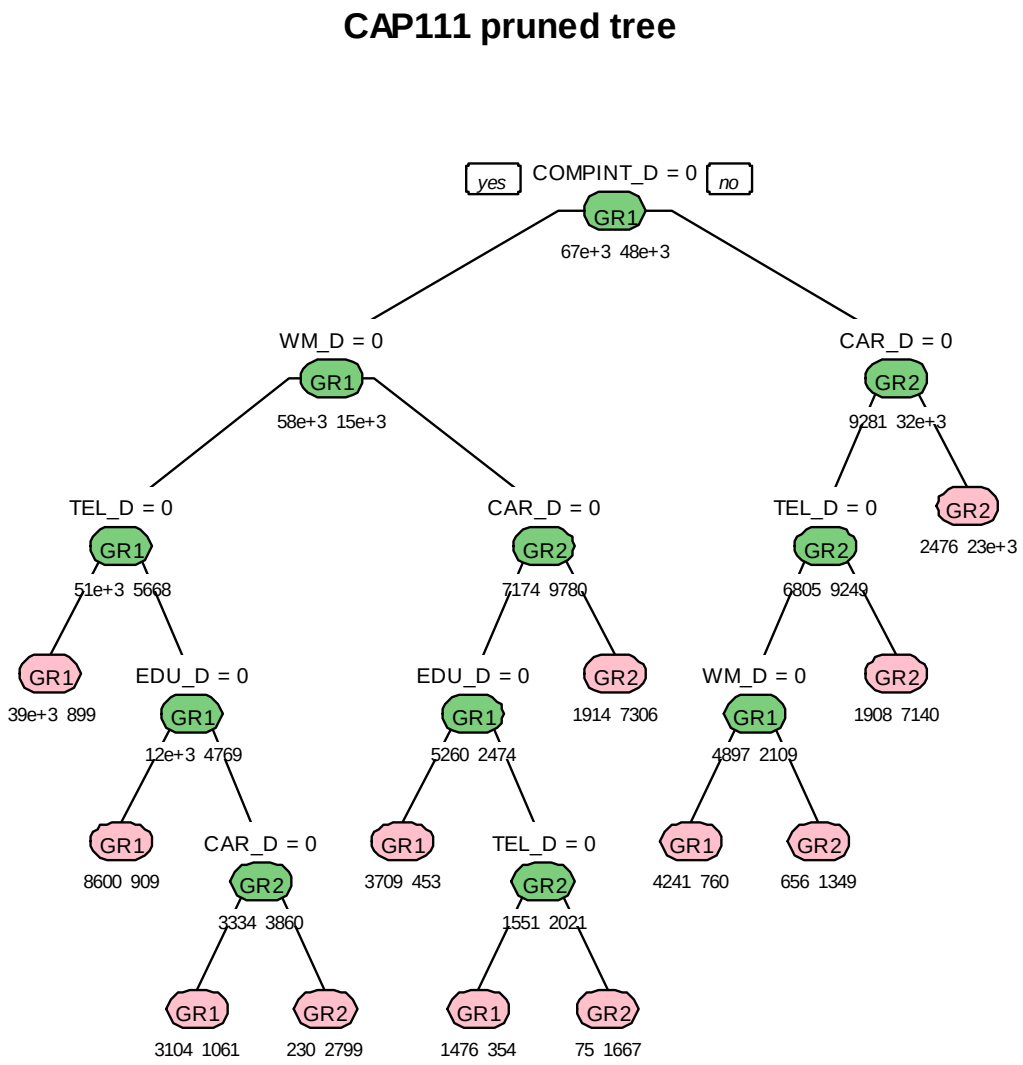


Figure 33. CAP111 pruned tree, data: Brazil PNAD 2011

CAP112: GR1 vs. GR3 analysis

In this analysis, we focus on DEP_GRS1 vs. DEP_GRS3. We use the data containing only people from the deprivation group 1 and the deprivation group 3. The outcome/target response variable is dichotomous, namely CAP112, taking the value 1 for the group 1 and the value 3 for the group 3 (so that people identified in the group 1 take the value 1 and people identified in the group 3 take the value 3 according to CAP112).

Table 86. Training vs. test samples in CAP112 analysis, data: Brazil PNAD 2011

	GR1	GR3
Training sample	67,271	51,117
Test sample	33,636	25,559

As the table above shows, there are 72,388 individuals in the training data (corresponding to 67 per cent of the initial data) and 59,195 individuals in the test sample. Training and test data are obtained by randomly partitioning the initial data/total sample into two subsets by 70 per cent vs. 30 per cent respectively. The data is very balanced with 56.8 per cent for GR1 and 43.2 per cent for GR3 in both training and test samples (in the initial data, the proportion is the same as well).

In first attempt of fitting the model based on the cp equal to 0.025, we obtained a very good tree with 10 splits generating almost zero cross-validated error rate. Given the results, we won't proceed with pruning phase. The box below provides information on dtreeCAP112 model.

```
Classification tree:
rpart(formula = CAP112 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
[1] COMPINT_D EDU_D     TEL_D     WATD_D    WM_D

Root node error: 51117/118388 = 0.43178

n= 118388

         CP nsplit rel error   xerror       xstd
1 0.7784494      0  1.000000 1.000000 0.00333409
2 0.0940196      1  0.221551 0.221551 0.00197979
3 0.0240429      2  0.127531 0.127531 0.00153542
4 0.0138897      3  0.103488 0.103488 0.00139071
5 0.0118552      5  0.075709 0.075709 0.00119694
6 0.0085197      6  0.063854 0.063854 0.00110215
7 0.0062797      8  0.046814 0.046814 0.00094727
8 0.0025000      9  0.040534 0.040534 0.00088266
```

The CAP112 pruned tree is composed of 10 terminal nodes providing 5 patterns of GR1, and 5 patterns of GR3. As we can see from the pruned tree below, the discovered patterns from the left to the right side of the tree are as follows: (1) no deprivation regarding having a computer (COMPINT) in the household, and a piped water from general distribution network (WATD) inside the dwelling where the household lives; (2) no deprivation regarding COMPINT, and having a washing machine (WM) in the household, but deprivation regarding WATD; (3) no deprivation regarding COMPINT, but deprivation regarding WATD, and WM; (4) no deprivation regarding WM, and having a landline telephone (TEL) in the household, but deprivation regarding COMPINT; (5) no deprivation regarding TEL, and educational attainment (EDU), but deprivation regarding COMPINT, and WM; (6) no deprivation regarding TEL, but deprivation regarding COMPINT, WM, and EDU; (7) deprivation regarding COMPINT, TEL, and WM; (8) no deprivation regarding WM, but deprivation regarding COMPINT, TEL, and EDU; (9) no deprivation regarding WM, EDU, and WATD, but deprivation regarding COMPINT, and TEL; (10) no deprivation regarding WM, and EDU, but deprivation regarding COMPINT, TEL, and WATD.

The patterns of GR1 are 1st, 4th, 2nd, 9th, and 5th with 83, 9.3, 2.9, 2.8, and 1 per centage of coverage respectively (see the table below). For instance, with the 1st pattern, 83 per cent of people from GR1 are detected. On the other hand, the patterns of GR3 are 7th, 8th, 6th, 3rd, and 10th with 82, 8, 3.8, 2, and 0.6 per centage of coverage respectively (see the table below). For instance, with the 7th pattern, 82 per cent of people from GR3 are captured.

Table 87. Coverage of GR1's patterns among the total population of GR1 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP112 pruned tree, data: Brazil PNAD 2011

Pattern no	Coverage of pattern (in %) among the total population of GR1	Cumulative coverage (in %)
1	83	83
4	9.3	92.3
2	2.9	95.2
9	2.8	98
5	1	**99**

Table 88. Coverage of GR3's patterns among the total population of GR3 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP112 pruned tree, data: Brazil PNAD 2011

Pattern no	Coverage of pattern (in %) among the total population of GR3	Cumulative coverage (in %)
7	82	82
8	8	90
6	3.8	93.8
3	2	95.8
10	0.6	96.4~**96**

In the training data, by using the CAP112 pruned tree classification patterns, we lose only 1 per cent of information for GR1 and 4 per cent of information for GR3.

We can also say that through the 5 patterns of GR1, deprivation regarding COMPINT, WATD, TEL, and WM concern 13.1, 2.9, 2.8, and 1 per cent of the entire population of GR1 (among 67,271 persons) respectively. On the other hand, concerning GR3, the 5 patterns show that the variables COMPINT, TEL, WM, EDU, and WATD affect 94.4, 90.6, 87.8, 11.8, and 2.6 per cent of the entire population of GR3 (among 51,117 persons). The 4 deprivation variables, COMPINT, TEL, WM, and WATD are the common in terms of deficits for both GR1 and GR3; however, the magnitude of COMPINT is 7 times higher, those of TEL is 32 times, and those of WM is 88 times higher in terms of per centage points for GR3 compared to GR1.

Regarding the overall accuracy of the mode, as we can see from the confusion matrix statistics, prediction on the test data is very good generating around 98 per cent of accuracy, and sensitivity and specificity rates vary between around 97.5 and 99 per cent, and misclassification rates are very small.

```
Confusion Matrix and Statistics

          Reference
Prediction   GR1   GR3
       GR1 33337   705
       GR3   299 24854

               Accuracy : 0.983
                 95% CI : (0.982, 0.9841)
    No Information Rate : 0.5682
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9654

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.9911
            Specificity : 0.9724
         Pos Pred Value : 0.9793
         Neg Pred Value : 0.9881
             Prevalence : 0.5682
         Detection Rate : 0.5632
   Detection Prevalence : 0.5751
      Balanced Accuracy : 0.9818

       'Positive' Class : GR1
```

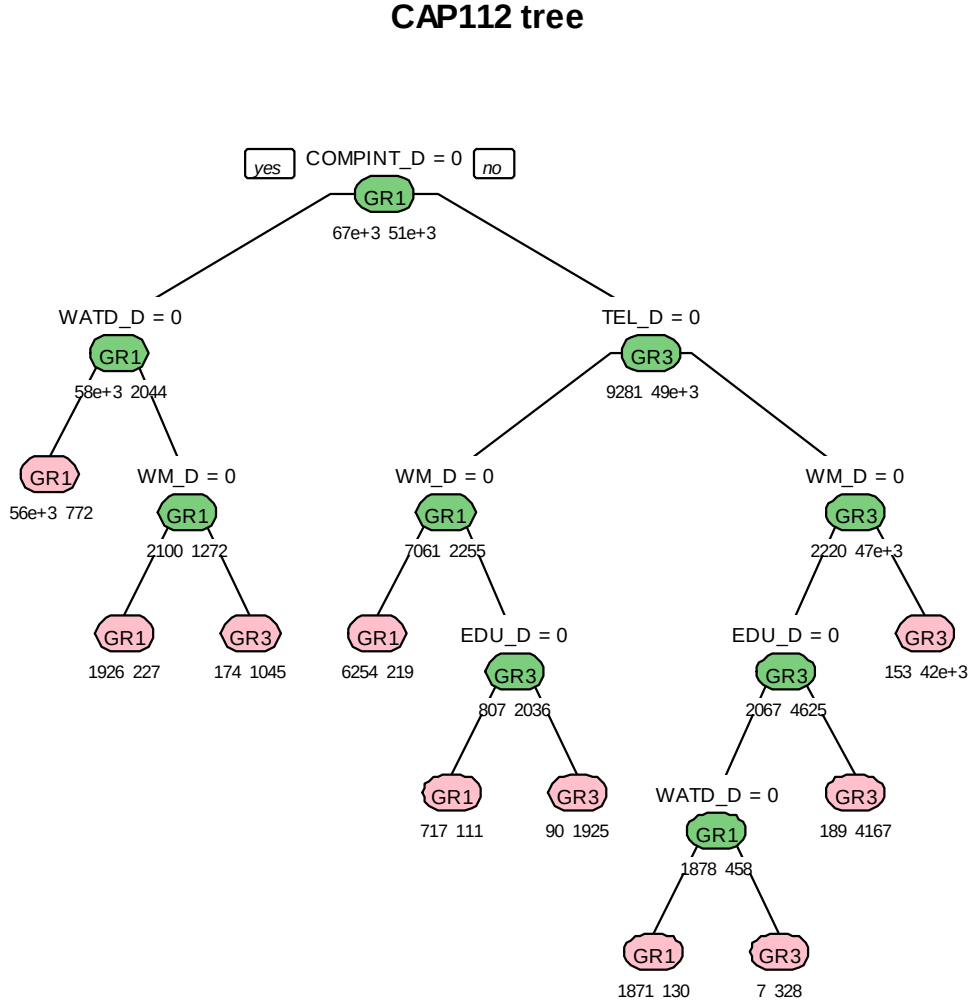


Figure 34. CAP112 tree, data: Brazil PNAD 2011

CAP115: GR2 vs. GR3 analysis

In this analysis, we focus on DEP_GRS2 vs. DEP_GRS3. We use the data containing only people from the deprivation group 2 and the deprivation group 3. The outcome/target response variable is dichotomous, namely CAP115, taking the value 2 for the group 2 and the value 3 for the group 3 (so that people identified in the group 2 take the value 2 and people identified in the group 3 take the value 3 according to CAP115).

Table 89. Training vs. test samples in CAP115 analysis, data: Brazil PNAD 2011

	GR2	GR3
Training sample	47,873	51,117
Test sample	23,937	25,559

As the table above shows, there are 98,990 individuals in the training data (corresponding to 67 per cent of the initial data) and 49,496 individuals in the test sample. Training and test data are obtained by randomly partitioning the initial data/total sample into two subsets by 70 per cent vs. 30 per cent respectively. The data is relatively balanced with 48.3 per cent for GR2 and 51.7 per cent for GR3 in both training and test samples (in the initial data, the proportion is also the same).

Fitting this model with the initial default parameters, and a cp equal to 0.025 has produced a saturated tree with 32 splits with a cross-validated error rate around 0.12. Thus, we will proceed with the pruning phase in order to obtain a tree with around 10-11 splits. The box below contains detailed information on the results of the model.

```
> printcp(dtreeCAP115)

Classification tree:
rpart(formula = CAP115 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
 [1] CAR_D     COMPINT_D EDU_D     MPHO_D     RO_D      ROF_D     RUB_D
 [8] TEL_D     WAL_D     WATD_D    WM_D

Root node error: 47873/98990 = 0.48361

n= 98990

          CP nsplit rel error  xerror      xstd
1  0.3144152      0   1.00000 1.00000 0.0032843
2  0.1096965      1   0.68558 0.68558 0.0030940
3  0.0594072      4   0.35617 0.35617 0.0024816
4  0.0309987      5   0.29676 0.29676 0.0023042
5  0.0211602      6   0.26577 0.26577 0.0021995
6  0.0086897      7   0.24461 0.24461 0.0021225
7  0.0078854      8   0.23592 0.23592 0.0020894
8  0.0061204     10   0.22014 0.22014 0.0020271
9  0.0055564     11   0.21402 0.21526 0.0020071
10 0.0054728     12   0.20847 0.20887 0.0019805
11 0.0053057     13   0.20300 0.20124 0.0019480
12 0.0045119     16   0.18681 0.18681 0.0018840
13 0.0043588     18   0.17778 0.18006 0.0018530
14 0.0035928     22   0.16028 0.16028 0.0017574
15 0.0035302     23   0.15669 0.15756 0.0017437
16 0.0034048     24   0.15316 0.15242 0.0017173
17 0.0031855     26   0.14635 0.14618 0.0016845
18 0.0031751     28   0.13997 0.14340 0.0016696
19 0.0029035     29   0.13680 0.13893 0.0016453
20 0.0027051     30   0.13390 0.13682 0.0016337
21 0.0025000     32   0.12849 0.12993 0.0015948
```

The CAP115 pruned tree is composed of 11 terminal nodes providing 7 patterns of GR2, and 4 patterns of GR3. As we can see from the tree below, the discovered patterns from the left to the right side of the tree are as follows: (1) no deprivation regarding rubbish collection of the household (RUB), and having a washing machine (WM) in the household; (2) no deprivation regarding RUB, educational attainment (EDU), and having a piped water from the general distribution network inside the dwelling where the household lives (WATD), but deprivation regarding WM; (3) no deprivation regarding RUB, EDU, and having a computer (COMPINT) in the dwelling, but deprivation regarding WM, and WATD; (4) no deprivation regarding RUB, and EDU, but deprivation regarding WM, WATD, and COMPINT; (5) no deprivation regarding RUB, and COMPINT, but deprivation regarding WM, and EDU; (6) no deprivation regarding RUB, and having a landline telephone (TEL), but deprivation regarding WM, EDU, and COMPINT; (7) no deprivation regarding RUB, WATD, and having a car (CAR) in the household, but deprivation regarding WM, EDU, COMPINT, and TEL; (8) no deprivation regarding RUB, and CAR, but deprivation regarding WM, EDU, COMPINT, TEL, and WATD; (9) no deprivation regarding RUB, but deprivation regarding WM, EDU, COMPINT, TEL, and

CAR; (10) no deprivation regarding COMPINT, but deprivation regarding RUB; (11) deprivation regarding RUB, and COMPINT.

The patterns of GR2 are 1st, 2nd, 5th, 6th, 7th, 10th, and 3rd with 39.7, 23, 12.5, 8.8, 4.8, 4.4, and 1 per centage of coverage respectively (see the table below). For instance, with the 1st pattern, 39.7 per cent of people from GR2 are detected. On the other hand, the patterns of GR3 are 11th, 9th, 4th, and 8th with 41, 41, 2.2, and 0.85 per centage of coverage respectively (see the table below). For instance, with the 11th pattern, 41 per cent of people from GR3 are captured.

Table 90. Coverage of GR2's patterns among the total population of GR2 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP115 tree, data: Brazil PNAD 2011

Pattern no	**Coverage of pattern (in %) among the total population of GR2**	**Cumulative coverage (in %)**
1	39.7	39.7
2	23	62.7
5	12.5	75.2
6	8.8	84
7	4.8	88.8
10	4.4	93.2
3	1	94.2~**94**

Table 91. Coverage of GR3's patterns among the total population of GR3 (in per centage) and their cumulative frequencies in per centage from the most to the least frequent patterns of CAP115 tree, data: Brazil PNAD 2011

Pattern no	**Coverage of pattern (in %) among the total population of GR3**	**Cumulative coverage (in %)**
11	41	41
9	41	82
4	2.2	84.2
8	0.85	85.05~**85**

In the training data, by using the CAP115 tree classification patterns, we lose 7 per cent of information for GR2 and 18 per cent of information for GR3.

We can also say that through the 7 patterns of GR2, deprivation regarding WM, EDU, COMPINT, TEL, RUB, and WATD concern 50.1, 26.1, 13.6, 4.8, 4.4, and 1 per cent of the entire population of GR2 (among 47,873 persons). On the other hand, concerning GR3, the 4 patterns show that the variables COMPINT, WM, EDU, TEL, CAR, RUB, and WATD affect 85, 44, 41.85, 41.85, 41, 41, and 3 per cent of the entire population of GR3 (among 51,117 persons). The deprivation variables, COMPINT, WM, EDU, TEL, RUB, and WATD are the common in terms of deficits for both GR2 and GR3; however, the magnitude of COMPINT is

6.25 times higher, EDU is 1.6 times higher, those of TEL is 8.7 times higher, those of RUB is 9 times higher, and those of WATD is 3 times higher in terms of per centage points for GR3 compared to GR2.

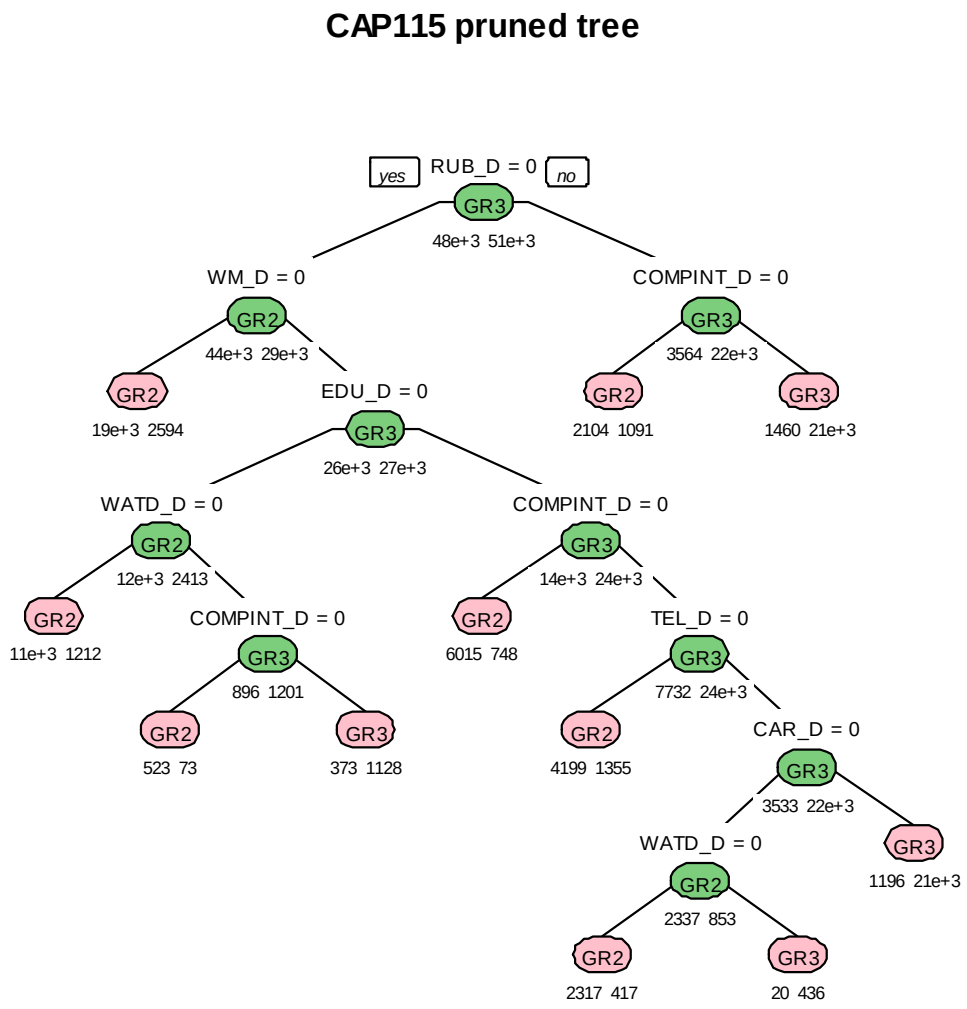


Figure 35. CAP115 tree, data: Brazil PNAD 2011

Regarding the overall accuracy of the model, as we can see from the confusion matrix statistics, prediction on the test data is relatively good generating 89 per cent of accuracy, nevertheless; there is a discrepancy between the sensitivity and specificity rates. The sensitivity rate equals 93 and the specificity rate is 85 per cent, which means that the prediction for GR2 is better than GR3.

```
 Confusion Matrix and Statistics

          Reference
Prediction   GR2   GR3
       GR2 22391  3767
       GR3  1546 21792

               Accuracy : 0.8927
                 95% CI : (0.8899, 0.8954)
    No Information Rate : 0.5164
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7857
```

```
Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.9354
            Specificity : 0.8526
         Pos Pred Value : 0.8560
         Neg Pred Value : 0.9338
             Prevalence : 0.4836
         Detection Rate : 0.4524
   Detection Prevalence : 0.5285
      Balanced Accuracy : 0.8940

       'Positive' Class : GR2
```

Comparative summary

After presented in detail the results of classification tress through 3 different deprivation groups, we will provide in this sub-section a summary of the discovered patterns regarding content and predictive accuracy.

CAP111

Based on the training data of GR1 and GR2 in this model, we can see that people living in a household lacking a computer and a car are at most risk of being in GR2 compared to GR1. Subsequently, people living in a household not lacking a car but lacking a computer and a landline telephone are at risk of being in GR2 compared to GR1. Moreover, people living in a household not lacking a computer but lacking a washing machine and a car, and people living in a household not lacking a computer and a washing machine but lacking a landline telephone and a car and not having at least a grade 12 educational level are also likely to be in GR2 compared to GR1. On the other hand, people living in a household having a computer, a washing machine, and a landline telephone are most likely to be part of the least poor population. From the results, we can see that GR2 is mainly distinguished from GR1 in terms of material deprivation including car, computer, landline telephone, washing machine respectively, but also of educational attainment. There are also some people of GR1 being in situation of deprived regarding some material such as computer, washing machine, landline telephone; however, the magnitude of such deprivations is considerably higher for GR2 compared to GR1.

CAP112

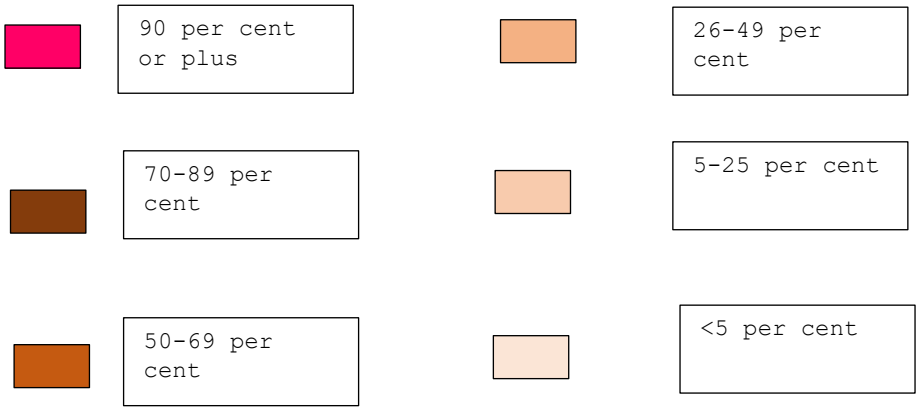
In this model, the capabilities patterns of the least poor group, GR1, and the poorest group, GR2, are discovered. First of all, people living in a household lacking a computer, a washing machine, and a landline telephone are at most risk of being in GR3 compared to GR1. This is the most important pattern distinguishing GR1 and GR3, affecting 82 per cent of GR3 people on the training data. We can see that here also as it was in the case of CAP111 model, the combination of some types of material deprivation is appeared as the most influential configuration of poverty separating GR3 from GR1. Subsequently, people living in a household not lacking a washing machine but lacking a computer and a landline telephone, and not having at least a 12-grade degree are likely to be part of GR3 compared to GR1.

CAP115

In this model comparing GR2 and GR3 based on the training data, we find out that people living in a household lacking a computer and an appropriate rubbish removal system, and people living in a household not lacking an appropriate rubbish removal system but not having a washing machine, a computer, a landline telephone, a car, and at least a 12 grade degree are

mostly are risk of being in GR3 compared to GR2. Similar to previous models, material deprivation and deficit in educational attainment are found the most influential functionings playing crucial role in the most striking configurations of poverty experienced by GR3 compared to GR2.

Figure 36 below is a kind of snapshot view on the overall importance of variables in terms of deprivation people experience regarding them on the training data across all employed models. There are 19 columns labelled as model, group, and variables-from WAL to EDU, respectively. The rows in the figure include the name of 3 models (e.g., CAP111) and the groups existing in the model/target groups (e.g., GR1 vs. GR2). The 6 different criteria, corresponding each of them to a different interval of per centage of population in the training data experiencing deprivation regarding the variable, are arbitrarily set as follows: 90 per cent +, 70-89 per cent, 50-69 per cent, 26-49 per cent, 5-25 per cent, and < 5 per cent. For instance, in CAP112 model, per centage of GR3 population in the training data suffering in terms of deprivation regarding the variable INT is in the interval of 70-89 per cent. Each criterion is presented in a different color as follows:



From the information above, we can see that the variables WAL, ROF, RO1, SAN, ENG_L, ENG_C, MPHO, RAD, TV, and REF are not used in any selected tree to distinguish the deprivation groups. It is important to mention that the variables appearing in the selected trees are the most influential ones in terms of information gain to distinguish the two groups of the target response variable. However, it does not mean that the groups in the model are not deprived regarding other variables that do not appear in the tree.

Based on the results displayed in Figure 36, we can observe that the poorest group (GR3) distinguishes greatly from the least poor group (GR1) in terms of computer, landline telephone, and washing machine, and from the middle poor group (GR2) in terms of computer. In fact, the deprivation regarding the items of access to piped water provided by general distribution network inside the dwelling, landline telephone, washing machine, computer, and educational attainment appear the influential poverty elements of the poorest group, GR3, compared to both GR1 and GR2. However, the magnitude of the deprivation regarding car, washing machine, and computer become more striking in model CAP115 then CAP112; this is due to the fact that deprivation difference for each of these 3 items between GR1 and GR3 is pronounced than between GR2 and GR3. Moreover, the item of rubbish removal surfaces only in CAP115 model, both in the patterns of GR2 and GR3, with a higher deprivation among GR3. The deprivation regarding the item car appears in CAP115 affecting GR3 and not GR2; however, GR2 appears suffering deprivation in this item in CAP111 compared to GR1- this can be explained by the

fact that there is a relatively important share of GR2 emerging deprived in terms of washing machine in CAP115; thus, people in such situation may also include those in deprivation in terms of car, which leads to no-apparition of car in the patterns of GR2 in CAP115. It is very important to note that despite GR1 is the least poor group experiencing a smaller number of deprivations compared to GR2 and GR3, include also some people lacking a piped water provided by the general distribution network inside the dwelling. On the other hand, in CAP111 model, we see that GR2 is distinguishes greatly from GR1 in terms of car, and computer. But there are also some people in GR1 experiencing deprivation regarding landline telephone, washing machine, and educational attainment. It is possible that landline telephone is now preferred to a mobile phone use from 2000s, and this is why some GR1 people do not possess it, but we don't have information on that. We used 17 variables in the overall deprivation index, and only 7 of 17 items appear in the patterns of GR1, GR2, and GR3 across CAP111, CAP112, and CAP115 models, and we can remark that material deprivation in terms of computer, washing machine, landline telephone, car, and educational attainment dominate the poverty configurations among these 3 groups. In second model, we will more specifically focus on dwelling, material situations, and we will check also deprivations regarding the other items that did not appear in the patterns of CAP111, CAP112, and CAP115.

In Figure 37, patterns of groups are shown per model. The green color means no deprivation and the red color means deprivation. Rows are the models, and columns include the title of model (Mod), groups (Gr), pattern (Pat), frequency of pattern (Fre) or in other words the coverage of pattern among the training population in per centage, and variables from WATD to EDU, respectively (including only variables emerging in the patterns). For instance, the first pattern (P1) of GR1 in CAP111 concerns 58 per cent of GR1 people in the training data. This pattern includes no deprivation regarding TEL, WM, and COMPINT at the same time. Concerning GR2, from CAP111, we can see that the combination of deprivation regarding CAR and COMPINT is the most important, then the combination of deprivation regarding CAR and no deprivation regarding COMPINT, and TEL, and the combination of deprivation regarding WM, CAR, and no deprivation regarding COMPINT appear. To eliminate inequality between GR1 and GR2, material deprivation regarding these items should be considered. Concerning GR3, from CAP112, we can observe that the combination of deprivations regarding landline telephone, washing machine, and computer is the most important poverty configuration experienced by 82 per cent of GR3 population on the training data compared to the least poor group; then, those having a washing machine in their dwelling but lacking a landline telephone, a computer, and at least a 12 grade degree are coming as the second most likely poverty profile. In addition, 3.8 per cent of GR3 on the training data in this model have a landline telephone in their dwelling, but lack a washing machine, a computer, and a 12-grade degree. Thus, to eliminate discrepancy between the least poor and the poorest group in the country, educational attainment as well as material deprivation regarding the 3 items should take priority. Furthermore, CAP115 model shows that the combination of deprivations regarding rubbish removal and computer, and the combination of deprivation regarding landline telephone, washing machine, computer, car, educational attainment, and non-deprivation regarding rubbish removal are the most relevant poverty patterns for GR3 compared to GR2, capturing in total 82 per cent of GR3 on the training data.

Table 93 below shows the main confusion matrix statistics clarifying the predictive accuracy of our 3 models. In the table, columns are the models and rows correspond to accuracy (acc),

sensitivity (sens), specificity (spec), positive predictive value (pos pred value), negative predictive value (neg pred value), balanced accuracy (balanced acc) computed on the test data, and misclassification rate (misc rate) computed in the training data, respectively. From the results, we can see that the 3 models have a relatively good performance for predicting the groups, notably CAP112 model. In this line, we will use the most relevant patterns of **CAP112** for *GR1*, **CAP115** for *GR2*, and **CAP112** for *GR3* for further analysis.

Model	Group	WAL	ROF	RO1	WATD	SAN	RUB	ENG_L	ENG_C	TEL	MPHO	RAD	TV	REF	WM	CAR	COMPINT	EDU
CAP111	GR1																	
	GR2																	
CAP112	GR1																	
	GR3																	
CAP115	GR2																	
	GR3																	

Figure 36. Deprivation per centages represented by different colors in terms of variables emerging in the patterns by group within each model, data: Brazil PNAD 2011

Mod	Gr	Pat	Freq	WATD	RUB	TEL	WM	CAR	COMPINT	EDU
CAP111	1	P1	58							
		P2	12.8							
		P11	6.3							
		P5	5.5							
		P3	4.6							
	2	P9	48							
		P10	15							
		P8	15							
		P4	6							
CAP112	1	P1	83							
		P4	9.3							
	3	P7	82							
		P8	8							
		P6	3.8							

Mod	Gr	Pat	Freq	WATD	RUB	TEL	WM	CAR	COMPINT	EDU
CAP115	2	P1	39.7							
		P2	23							
		P5	12.5							
		P6	8.8							
		P7	4.8							
		P10	4.4							
	3	P11	41							
		P9	41							

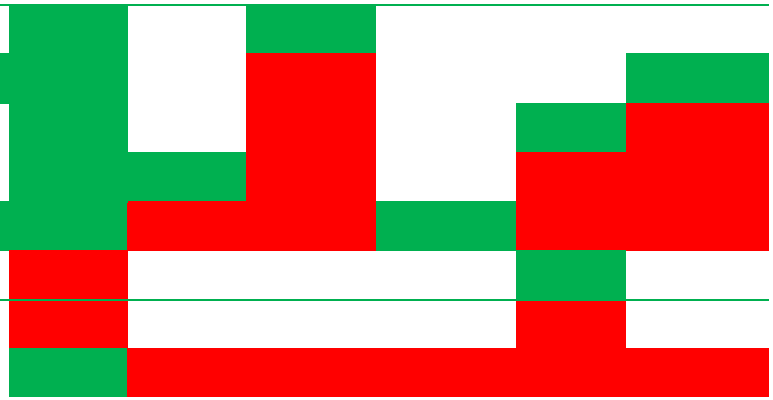


Figure 37. Frequency and composition of classification patterns per model, data: Brazil PNAD 2011

Table 92. Main confusion matrix statistics per model, data: Brazil PNAD 2011

	CAP111 (GR1 vs. GR2)	CAP112 (GR1 vs. GR3)	CAP115 (GR2 vs. GR3)
Acc	0.897	0.983	0.892
Sens	0.891	0.991	0.935
Spec	0.905	0.972	0.852
Pos pred value	0.929	0.979	0.856
Neg pred value	0.856	0.988	0.933
Balanced acc	0.898	0.981	0.894
No-Misc rate (test data)	89 90.5	99 97	93.5 85

Second step (step2 analysis): detecting dimensional poverty profiles of poverty groups

In this sub-section, we will in particular focus on GR2, and GR3 in terms of their dimensional profiles. By taking in consideration the performance of models (i.e., predictive accuracy), we can see that for GR2 the model CAP115, and for GR3 the model CAP112, bring the most efficient information concerning their most relevant capabilities patterns. Inside each of these three models, by selecting the most influential patterns, we will attempt to find out the variety of profiles of the three groups regarding dwelling, material, and education aspects respectively. The aim of doing this is to detect more information on living conditions of these three most deprived groups.

The dimension of *dwelling conditions* is separated into two sub-sets: (i) DWE1: WAL, ROF, and RO; (ii) DWE2: WATD, SAN, RUB, ENG_L, and ENG_C. The first covers deprivation regarding the main wall material, the main roof material, the number of available rooms per person in the dwelling where the household lives. The second includes deprivation in terms of whether the dwelling has an access to a piped water provided by the general distribution network (inside the house), whether the dwelling has a toilet/bathroom, whether the rubbish of the household is collected directly, whether the main source of energy used in the dwelling for lighting is electric or not, and whether the main source of energy used in the dwelling for cooking is piped gas, electric or not or whether there is no stove for cooking, respectively. The main reason behind this separation is profiles become easily complex to interpret when many variables are included, as we have seen in the beginning of the analysis. The dimension of *material* (MAT) is also divided into two subsets: (i) MAT1: REF, and WM; (ii) MAT2: RAD, TV, TEL, MPHO, COMPINT, and CAR. The first group comprises deprivation regarding whether there is a refrigerator, and a washing machine in the dwelling. The second concerns deprivation regarding whether there is a radio, a television, a landline telephone, a mobile phone, a computer, and a car in the dwelling. The dimension of *education* (EDUC) includes only deprivation regarding educational attainment.

We start off by detecting the population of GR2, and GR3 conforming to the most informative patterns (see below) in the full data (training plus test data), and then we find out their dimensional profiles according to the variables indicated above. This exercise is performed in statistical software STATA, for each group separately.

GR2 dimensional profiles based on CAP115 model

With this model, we saw that the following 6 patterns capture 93 per cent of population of GR2 in the training data: (i) no deprivation regarding RUB, and WM; (ii) no deprivation regarding RUB, EDU, and WATD, but deprivation regarding WM; (iii) no deprivation regarding RUB, and COMPINT, but deprivation regarding WM, and EDU; (iv) no deprivation regarding RUB, and COMPINT, but deprivation regarding WM, EDU, and TEL; (v) no deprivation regarding RUB, CAR, and WATD, but deprivation regarding WM, EDU, COMPINT and TEL; and (vi) no deprivation regarding COMPINT, but deprivation regarding RUB. When we applied these 6 classification rules in the full data, we lose 7.5 per cent of information reaching to 92.5 per cent (as in the case of training data), which shows that the model predicts very well GR2.

The table below illustrates the dimensional profiles of GR2 based on the extracted 6 patterns touching each of them at least 3 per cent of GR2. The table has 5 columns representing pattern number, and the most substantial dwelling conditions (DWE1, and DWE2), material (MAT1, and MAT2), and educational attainment (EDUC) profiles per each pattern. In first column, next

to the pattern number, the frequency of the pattern is mentioned in parenthesis in terms of per centage. For instance, 1st pattern concern 39.7 per cent of population of GR2 in the entire data. Profiles are composed of 0 and 1 where 0 means no deprivation and 1 means deprivation regarding the variable. The profiles are composed in the following way:

DWE1= WAL_D+ROF_D+RO_D.

DWE2= WATD_D+SAN_D+RUB_D+ENG_L_D+ENG_C_D.

MAT1=REF_D+WM_D.

MAT2=RAD_D+TV_D+TEL_D+MPHO_D+COMPINT_D+CAR_D.

EDU=EDU_D.

In table, next to the profiles, their frequency (i.e., Freq) within the pattern population is mentioned in parenthesis in terms of per centage. For instance, the profile 001 in DWE1 concerns 81.5 per cent of population of pattern 1. The sum of frequencies of dimensional profiles in each column for each pattern equals maximum 100 per cent. In case of DWE1, for the 1st pattern, there are 3 profiles are selected as the most substantial ones with at least 100 observations. The last row in the table indicates the total information captured in each dimension for the ensemble of GR3.

Concerning DWE1 profile throughout the 6 patterns, the extracted profiles shown in the table capture 90.4 per cent of GR2 (given that the 6 patterns explain 93 per cent of GR2). There are in total 16 DWE1 profiles. Among them, the combination of non-deprivation regarding wall, roof quality of the dwelling where the person lives, and of deprivation regarding the number of available rooms per person in the dwelling is the most relevant profile covering 79 per cent of GR2. After that, being deprived regarding the 3 items of DWE1 at the same time comes concerning 6.6 per cent of GR2, and then the deprivation regarding the wall quality of the dwelling and the available number of rooms per person in the dwelling with no deprivation regarding the quality of roof of the dwelling comes covering 5.16 per cent of GR2.

Concerning DWE2 profile based on the 6 patterns and the profiles extracted in the table, we can say that the profile of being deprived regarding the main source of energy used for cooking in the dwelling and not being deprived regarding the other 4 items of DWE2 flourishes as the most important DWE2 pattern affecting 82 per cent of GR2. Subsequently, the profile of being deprived regarding the main source of energy used for cooking in the dwelling and the access to piped water provided by the general distribution network within the dwelling, and of not being deprived regarding the other 3 items of DWE2 surfaces as the 2nd important DWE2 pattern concerning 4.9 per cent of GR2.

When it comes to materials, in terms of MAT1 dimension, via the extracted 7 profiles based on the 6 patterns shown in the table below we detect 91 per cent of GR2. The most relevant profile is not having a washing machine but having a refrigerator in the dwelling where the household lives, affecting 51 per cent of GR2. Then, the profile of non-deprivation regarding both these 2 items of MAT1 comes, which concerns 42 per cent of GR2.

On the other hand, regarding the MAT2 dimension, there is more complexity compared to MAT1. In total, 23 profiles based on the 6 patterns give information on 87 per cent of GR2. Among these 23 profiles, the most relevant is the profile of living in a household lacking a

landline telephone, a computer, a car but not lacking a radio, a television, and a mobile phone. This profile concerns 27 per cent of GR2. After that, the combination of deprivation regarding telephone and car and non-deprivation regarding the other 4 items of MAT2 comes affecting 17 per cent of GR2. Besides, the combination of deprivation regarding computer and car and non-deprivation regarding the other 5 items of MAT2 covers 14 per cent of GR2. Moreover, the combination of deprivation regarding telephone and computer, and non-deprivation regarding the other 5 items of MAT2 emerges concerning 9.5 per cent of GR2. Throughout the 23 profiles, we can say that 70.5 per cent of GR2 live in a household lacking a car, 63 per cent of GR2 live in a household lacking a telephone, and 59 per cent of GR2 live in a household lacking a computer. It is important to note that 63 per cent of GR2 live in a household lacking a telephone but not lacking a mobile phone vs. only 2.45 per cent live in a household lacking a mobile phone but not lacking a landline telephone. Even we cannot see from the survey due to the way the question was asked according to the survey questionnaire, we ca think that probably due to the use of mobile phone, households may would prefer to not use a landline telephone for communication. On the other hand, 7 per cent of GR2 live in a household having both a radio and a television, and the rate of households lacking a mobile phone is quite low.

If we look at the educational attainment of GR2, we can observe that based on the 6 patterns, almost 60 per cent of GR2 don't have at least a 12-grade degree (and 63 per cent of the entire population of GR2 is deprived in educational attainment). More specifically, 31.75 per cent of GR2 are high school graduates and only 3.82 per cent are university graduates vs. 25.33 per cent are elementary school graduates, 14.2 per cent are primary school graduates, 10.10 per cent are literate but they do not have a formal school diploma, 6.4 per cent are illiterate, and 5.65 per cent are secondary school graduates.

As a result of these discovered profiles and the patterns of GR2 from the CAP115 analysis, we conclude that the following items should be considered to increase capabilities of GR2: educational attainment, dwelling conditions in terms of number of rooms available per person, and availability of at least one stove, and energy source used for cooking in the dwelling, and household material situation in terms of washing machine, landline telephone, computer, and car. Nevertheless, landline telephone might be preferred by mobile phone which is often available in households according to the results; but also, instead of car, the amelioration of public transports in terms of security, and availability might be preferred for environmental issues such as air pollution, car traffic. In CAP111 mode, when GR1 and GR2 were compared, we also found out that the items of car, computer, telephone, washing machine, and educational attainment are the most relevant to increase capabilities of GR2 compared to GR1. The items of available room per person, and energy source for cooking did not appear in CAP111, it is probably due to the fact that GR1 also suffers considerably in these 2 items. We will check this point in the supplementary analysis, in the case of GR1 examination. On the other hand, based on the results of CAP115, we can say that GR2 is in better position compared to GR3 notably regarding the item of washing machine, educational attainment, and computer.

Table 93. Dimensional profiles of GR2 based on the CAP115 model, data: Brazil PNAD 2011

Pattern number of GR2 (Freq in % in full data)	DWE1 profile (Freq in % in full data)	DWE2 profile (Freq in % in full data)	MAT1 profile (Freq in % in full data)	MAT2 profile (Freq in % in full data)	EDUC profile (Freq in % in full data)
1 (39.7)	001 (81.5), 101 (9.5), 000 (6)	00001 (87), 10001 (12.3)	00 (99.4)	001011 (38), 001001 (15.7), 000011 (14.2), 001010 (13.9), 000111 (4.6), 001000 (3.8)	0 (21.5), 1 (78.5)
2 (23)	001 (88), 000 (7), 101 (3.3)	00001 (99.5)	01 (99.17)	001011 (51.6), 001001 (19), 001010 (12), 000011 (11.4)	0 (100), 1 (0)
3 (12.5)	001 (91.5), 101 (3.6), 000 (3.6)	00001 (92)	01 (99.5)	001001 (40.4), 000001 (27.6), 001000 (25.4), 000000 (3.8)	0 (0), 1 (100)
4 (8.8)	001 (81.2), 000 (16.9)	00001 (97)	01 (99.87)	000011 (65.2), 000010 (23.6), 000111 (7.2), 000110 (3.6)	0 (0), 1 (100)
5 (4.8)	001 (89.6), 000 (9)	00001 (99.9)	01 (99.6)	001010 (99)	0 (0), 1(100)
6 (4.4)	001 (88.7), 000 (5.16), 101 (4.16)	00101 (66), 10101 (33.4)	00 (59.9), 01 (39.8)	001000 (35.3), 001001 (27.5), 000001 (25), 000000 (9.4)	0 (45.6), 1 (54.4)
Total info/GR2 population coverage in %	(39.7*0.97) + (23*0.953) + (12.5*0.987) + (8.8*0.981) + (4.8*0.986) + (4.4*0.980) = **90.4**	(39.7*0.993) + (23*0.995) + (12.5*0.92) + (8.8*0.97) + (4.8*0.999) + (4.4*0.994) =**91.5**	(39.7*0.994) + (23*0.9917) + (12.5*0.995) + (8.8*0.9987) + (4.8*0.996) + (4.4*0.638) =**91**	(39.7*0.902) + (23*0.94) + (12.5*0.972) + (8.8*0.996) + (4.8*0.99) + (4.4*0.972) =**87**	**93**

GR3 dimensional profiles based on CAP112 model

With this model, we saw that the following 3 patterns capture 94 per cent of population of GR3 in the training data: (i) deprivation regarding COMPINT, TEL, and WM; (ii) no deprivation regarding WM, but deprivation regarding COMPINT, TEL, and EDU; (iii) no deprivation regarding TEL, but deprivation regarding COMPINT, WM, and EDU. When we applied these 3 classification rules in the full data, we lose only 5.5 per cent of information reaching to 94.5 per cent (very close to what we obtained on the training data), which shows that the model predicts very well GR3.

Based on the 3 extracted patterns of GR3 in CAP112, we find out that with the 8 DWE1 profiles shown in the table below, 88 per cent of GR3 can be captured. The most relevant DWE1 profile is the combination of deprivation regarding the available number of rooms per person, and non-deprivation regarding the wall, and roof of the dwelling where the person lives. This profile concerns 67.5 per cent of GR3. Besides, the combination of deprivation regarding the wall of

the dwelling, and the available number of rooms per person in the dwelling, and non-deprivation regarding the roof of the dwelling where the person lives covers 16 per cent of GR3.

Regarding DWE2 dimension, by employing the extracted 15 profiles based on the 3 patterns, 87.6 per cent of GR3 are detected. The most relevant profile is the combination of deprivation regarding energy source used for cooking in the dwelling where the person lives and of non-deprivation regarding the other 4 items of DWE2. This profile concerns 39 per cent of GR2. Subsequently, the profile of deprivation regarding access to piped water provided by the general distribution network inside the dwelling (WATD), rubbish removal (RUB) of the dwelling, and energy source used for cooking (ENG_C) in the dwelling where the person lives, and of non-deprivation regarding the other 2 items of DWE2 comes affecting 17.7 per cent of GR2. Moreover, the profile of deprivation regarding WATD and ENG_C and of non-deprivation regarding the other 3 items of DWE2 emerges affecting 12 per cent of GR2, and the profile of deprivation regarding RUB, and ENG_C, and of non-deprivation regarding the other 3 items of DWE2 surfaces concerning 11 per cent of GR2. Across the extracted 15 profiles, we can say that: (i) 87.6 per cent of GR3 don't experience deprivation regarding energy source used for lighting, and a very few of them experience deprivation regarding sanitation facilities in the dwelling where the person lives; (ii) 87.6 per cent of GR3 live in a dwelling lacking a proper energy source for cooking, or a stove for cooking; (iii) 37 per cent of GR3 live in a dwelling not having an access to a piped water provided by the general distribution network; and (iv) 36.5 per cent of GR3 live in a dwelling where the rubbish is not directly collected.

In terms of material deprivation, by employing 6 profiles of MAT1 based on the 3 patterns, we capture 93.7 per cent of GR3. The most frequent profile is living in a dwelling lacking a washing machine, which concerns 75 per cent of GR3. Besides, 10.5 per cent of GR3 are found living in a dwelling lacking both a refrigerator and a washing machine. On the other hand, only 7.6 per cent of GR3 are found living in a dwelling possessing both a refrigerator and a washing machine. When the 6 profiles are taken in consideration, we can observe that 86 per cent of GR3 are deprived regarding washing machine, and 11 per cent are deprived regarding refrigerator.

On the other side, concerning MAT2 dimension, by employing 12 profiles of MAT2 based on the 3 patterns, we detect 87 per cent of GR3. The most frequent profile is living in a dwelling lacking a landline telephone (TEL), a computer (COMPINT), and a car (CAR), which affects 49 per cent of GR3. Moreover, 16.3 per cent of GR3 are found living in a dwelling lacking a telephone, a computer, a car, and a mobile phone. And, 5.7 per cent of GR3 live in a dwelling lacking a telephone, and a computer. When we take in consideration the 12 profiles, we find out that 82 per cent of GR3 are deprived regarding TEL, 88 per cent are deprived regarding COMPINT, and 81 per cent are deprived regarding CAR. But also, 23 per cent of GR3 are found living in a dwelling deprived regarding both telephone and mobile phone.

Concerning educational attainment, based on the 3 patterns we find out that 85 per cent of GR3 are deprived vs. only 9 per cent are not deprived. When the entire data of GR3 is taken in consideration, 89 per cent are found deprived vs. 11 per cent are not deprived. More specifically, 37.8 per cent of GR3 are elementary school graduates, 21.8 per cent are illiterate, 15 per cent are primary school graduates, 10 per cent are high school graduates, 9 per cent are literate but with no diploma, and only 0.72 per cent are university graduates.

We can summarize that for GR3, in order to increase their capabilities, policies should refer to amelioration of dwelling conditions where the person lives in terms of available rooms per person in the dwelling, main energy source used for cooking and existence of at least one stove for cooking, of material situation of the household in terms of washing machine, telephone, computer, and car, and of educational attainment. Nevertheless, it becomes also striking deprivation regarding rubbish removal, access to piped water provided by the general network distribution, refrigerator, and mobile phone, compared to GR2.

Table 94. Dimensional profiles of GR3 based on the CAP112 model, data: Brazil PNAD 2011

Pattern number of GR4 (Freq in % in full data)	DWE1 profile (Freq in % in full data)	DWE2 profile (Freq in % in full data)	MAT1 profile (Freq in % in full data)	MAT2 profile (Freq in % in full data)	EDUC profile (Freq in % in full data)
1 (82)	001 (73.5), 101 (15.3), 000 (5.5)	00001 (43), 10101 (21.2), 00101 (11.3), 10001 (11.3), 11101 (6.45)	01 (87.3), 11 (12.65)	001011 (59), 001111 (18.2), 001010 (5.55), 101011 (4.2), 101111 (3.2), 011111 (3)	0 (11.4), 1 (88.6)
2 (8)	001 (55.3), 101 (36.4)	10101 (30.8), 00001 (29.6), 10001 (21.3), 00101 (12.7)	00 (95.3), 10 (4.7)	001011 (56.2), 001111 (18), 001010 (14.8),	0 (0), 1 (100)
3 (3.8)	001 (75.3), 101 (16.5), 000 (4)	00001 (43.4), 10001 (22.5), 00101 (19), 10101 (10.3)	01 (95.6), 11 (4.4)	000011 (44.5), 000111 (42.4), 000010 (3.65)	0 (0), 1 (100)
Total info/GR3 population coverage in %	(82*0.943) + (8*0.917) + (3.8*0.958) =**88**	(82*0.932) + (8*0.944) + (3.8*0.952) =**87.6**	(82*0.999) + (8*1) + (3.8*1) =**93.7**	(82*0.931) + (8*0.89) + (3.8*0.905) =87	94

The results show that the poorest group, GR3, experience (i) a much higher deprivation in terms of washing machine, computer, telephone, car, and educational attainment; (ii) relatively at same level deprivation in terms of available rooms per person in the dwelling, and energy source for cooking including existence of at least one stove in the dwelling where the person lives, compared to GR2. But also, even the deprivation remains at relatively low levels, lack of a refrigerator and/or a mobile phone in the dwelling becomes striking among GR3 compared to GR2.

Third step (step3 analysis): detecting various profiles of poverty groups

In this sub-section, we will proceed with sub-analysis in order to examine in more detail the extent of capabilities of GR2, and GR3 in terms of geographic situation, and economic activity status. Concerning individual characteristics, and household composition, a substantial number of profiles emerge in both GR2 and GR3 cases, which makes not possible to have a synthetic information. For that reason, classification tree model is built including the variables, namely, GEN, AGE, POP, HS, GH, AH, POPH, EDUH, ES1H, OGH, and ES3H. However, for GR2, based on the CAP115 model, an accurate prediction could not be obtained; so that we will only provide the result for this point for GR3. We also note that the distribution of educational attainment across GR2, and GR3 are already shown in second model section.

GR2 sub-group profiles based on CAP115 model

When we look at the distribution of geographical situation composed of the 2 variables namely region (REG) and setting (SET1) of the entire GR2, we find out that 94.65 per cent of the group live in urban areas whilst only 5.35 per cent live in rural areas. In terms of region, 30.44 per cent live in northeast, 28.65 per cent live in southeast, 15.72 per cent live in south, 13.58 per cent live in north, and 11.62 per cent live in central west. When it comes to the combination of REG and SET1, the results show that 29 per cent of GR2 live in urban areas of northeast, 27.4 per cent live in urban areas of southeast (where Rio de Janeiro, and Sao Paulo are located), 14 per cent live in urban areas of south, 13 per cent live in urban areas of north, 11 per cent live in urban areas of central west, 1.8 per cent live in rural areas of south, 1.4 per cent live in rural areas of northeast, and 1.3 per cent live in rural areas of southeast. As we can see the majority of GR2 live in urban areas, and east coast of the country.

In terms of economic activity, there are 7 main variables, namely, ES1, SC, WH, ES3, JOB, OG, and IND. These variables represent employed vs. not employed, working hours, economic activity status, number of jobs performed, occupational group, and industry group of the main work during the reference week, respectively. When we combine these 6 variables, an extensive number of profiles emerge for GR2 based on its 6 patterns from CAP115 model. Notably, economic activity status, occupational group, and industry group variables have several modalities, which diversifies the economic activity profiles. For that reason, we only combined the 4 variables, ES1, SC, WH, and JOB. When we take in consideration the 6 patterns of GR2 from CAP115 model, the following profiles are obtained (with at least 100 observations and 5 per cent of frequency): (1) not economically active; (2) contributing to social security, working between 40 and 44 hours per week, and performing one job; (3) not contributing to social security, working between 40 and 44 hours per week, and performing one job; (4) contributing to social security, working between 45 and 48 hours per week, and performing one job; (5) not contributing to social security, working between 15 and 39 hours per week, and performing one job; (6) contributing to social security, working at least 49 hours per week, and performing one job; (7) contributing to social security, working between 15 and 39 hours per week, and performing one job; (8) not contributing to social security, working at least 49 hours per week, and performing one job. These 8 profiles concern 36 per cent, 18.7 per cent, 7 per cent, 6.5 per cent, 5.8 per cent, 3.85 per cent, 3.8 per cent, and 3.4 per cent of the entire GR2, respectively.

When we take in consideration not economically active population, and the economically active population based on these 8 profiles, we reach to capture 85 per cent of the entire GR2. Moreover, we can also say that (i) 13.75 per cent of the entire GR2 work long hours, at least 45 hours per week, 25.7 per cent are working between 40 and 44 hours per week, which is an upper threshold as more than 40 hours per week is problematic, and 9.6 per cent work between 15 and 39 hours per week; (ii) 16.2 per cent of the entire GR2 do not contribute to social security; and (iii) 49 per cent of the entire GR2 perform one job during the reference week. Moreover, it is important to remark that 10.4 per cent of the entire GR2 work at least 40 hours per week and do not contribute to social security.

Regarding economic activity status, 43.7 per cent of economically active population of GR2 based on its 6 patterns from CAP115 are employees with a signed card work, 21 per cent are own account worker, 15 per cent are employees without a signed work card, 6.3 per cent are domestic worker without a signed work card, and 5.7 per cent are statutory civil servants. If we look at the GR2 economically active population in terms of no contribution to social security

and working hours per week (WH=3, 4, 5) , we find that 46.35 per cent are own account workers, 35.27 per cent are employees without a signed work card, and 10.84 per cent are domestic workers without a signed work card. Moreover, in terms of occupational group, 31.72 per cent of GR2 based on its 6 patterns from CAP115 are goods and service production and repair and maintenance workers, 27 per cent are service workers, 13.72 per cent are trade vendors and service providers, 8.5 per cent are administrative workers, 6.6 per cent are agricultural workers, and 5.6 per cent are mid-level technicians. On the other hand, the per centages of managers, science and art professionals are considerably low. And in terms of industry, 20.66 per cent are working in trade and repair sector, 13.95 per cent are working in manufacture, 11.22 per cent are working in construction, and 9.62 per cent are working in domestic services.

GR3 sub-group profiles based on CAP112 model

When we look at the distribution of geographical situation composed of the 2 variables namely region (REG) and setting (SET1) of the entire GR3 we find out that 36.26 per cent of GR3 live in rural areas vs. 63.74 per cent live in urban areas. In terms of region, 42.23 per cent of GR3 live in northeast, 24 per cent live in north, 15.9 per cent live in southeast, 10 per cent live in south, and 7.7 per cent live in central west. When we combine REG and SET1, the results show that 25.6 per cent of GR3 live in urban areas of northeast, 16.5 per cent live in rural areas of northeast, 15.6 per cent live in urban areas of north, 11 per cent live in urban areas of southeast, 8.4 per cent live in rural areas of north, 6 per cent live in urban areas of south, 5.4 per cent live in urban areas of central west, 4.8 per cent live in rural areas of southeast, 4 per cent live in rural areas of south, and 2.3 per cent live in rural areas of central west. Compared to GR2, we can see that (i) there is much more heterogeneity-94 per cent of GR2 live across 5 regions vs. 94 per cent of GR3 live across 8 regions; (ii) the per centage of people living in rural areas are quite higher among GR3 population; and (iii) 57.7 per cent of GR3 live in north part of the country including north and northeast regions.

In terms of economic activity, when we take in consideration the 3 patterns of GR3 from CAP112 model, the following profiles are obtained (with at least 100 observations and 5 per cent of frequency): (1) economically not active; (2) not contributing to social security, working between 15 and 39 hours per week, and performing one job; (3) not contributing to social security, working between 40 and 44 hours per week, and performing one job; (4) contributing to social security, working between 40 and 44 hours per week, and performing one job; (5) not contributing to social security, working at least 49 hours per week, and performing one job; (6) not contributing to social security, working between 45 and 48 hours per week, and performing one job; (7) not contributing to social security, working between 1 and 14 hours per week, and performing one job; and (8) contributing to social security, working between 45 and 48 hours per week, and performing one job. These 8 profiles concern 37.7 per cent, 12 per cent, 10.4 per cent, 8.2 per cent, 5.6 per cent, 4.8 per cent, 4.2 per cent, and 3.5 per cent of the entire GR3 population, respectively. Herewith, we can say that (i) 13.9 per cent of the entire GR3 work long hours, at least 45 hours per week-this per centage is almost the same compared to those of GR2; 18.6 per cent are working between 40 and 44 hours per week, which is an upper threshold as more than 40 hours per week is problematic; 12 per cent work between 15 and 39 hours, and 4.2 per cent work between 1 and 14 hours per week; (ii) 37 per cent of the entire GR3 do not contribute to social security-in terms of per centage points, this score is 2.3 times higher than those of GR2; and (iii) 49 per cent of the entire GR3 perform one job during the reference week. Moreover, it is important to remark that 20.8 per cent of the entire GR3 work at least 40 hours

per week and do not contribute to social security-the per centage is double of those of GR2. We can, therefore, point out that compared to GR2, the per centage of people working between 40 and 44 hours per week is 1.38 times lower in terms of per centage points among GR3; however, even if weekly hours of work on average is lower among GR3 compared to GR2 the share of GR3 working without contributing to social security is significantly higher than GR2. This last issue may due to the fact that GR3 population find more precarious jobs requiring ay be less hours of work but most importantly with no registration to social security.

Regarding economic activity status, 29 per cent of economically active population of GR3 based on its 3 patterns from CAP112 are own account workers, 20.2 per cent are employees without a signed work card, 21 per cent are employees with a signed work card, 10.4 per cent are workers in production for own consumption, 8 per cent are domestic workers without a signed work card, and 5.6 per cent are unpaid workers. If we look at the GR3 economically active population in terms of no contribution to social security and working hours per week: (i) for those are working at least 40 hours per week in a job without contributing to social security, we find that 43.2 per cent are own account workers, 35.65 per cent are employees without a signed work card, and 8.8 per cent are domestic workers without a signed work card; and (ii) for those are working between 1 and 39 hours per week in a job without contributing to social security, we find that 33 per cent are own account workers, 25.85 per cent are workers in production for own consumption, 14.35 per cent are domestic workers without signed work card, 13.45 per cent are employees without a signed work card, and 12.46 per cent are unpaid workers. We can see that people working less than 40 hours per week in a job without contributing to social security are in more precarious positions in terms of economic status (unpaid job or working without a signed work card) compared to those are working at least 40 hours per week in a job without contributing to social security.

Moreover, in terms of occupational group, 40.5 per cent of GR3 based on its 3 patterns from CAP115 are agricultural workers, 24.75 per cent are goods and service production and repair and maintenance workers, 21.4 per cent are service workers, and 7.37 per cent are trade vendors and service providers, 8.5 per cent are administrative workers, 6.6 per cent are agricultural workers, and 5.6 per cent are mid-level technicians. Here, we can observe that the per centage of agricultural workers is quite remarkable compared to GR2; but also, the per centages of managers, science and art professionals are together equal to around only 2 per cent, the per centage of administrative workers is around 2 per cent, and the per centage of mid-level technicians is also around 2 per cent. And in terms of industry, 40.7 per cent are working in agriculture, 12 per cent are working in construction, 10.7 per cent are working in trade and repair sector, 9.93 per cent are working in domestic services, and 8.43 per cent are working in manufacturing. Compared to GR2, there is a considerable part of GR3 economically active population performing in the sector of agriculture.

Lastly, we will focus on the result of the classification tree we built ‘dtreeCAP112DEC’ for GR3 including the variables, namely, GEN, AGE, POP, HS, GH, AH, POPH, EDUH, ES1H, OGH, and ES3H in as the regressors to predict CAP112 (DEP_GRS==1 vs. DEP_GRS_GR==3). We use the default values, and a cp equal to 0.025 as we employed in all previous classification tree models. The box below contains the classification tree results of the model ‘dtreeCAP112DEC’

```
> printcp(dtreeCAP112DEC)

Classification tree:
rpart(formula = CAP112 ~ ., data = train.set, method = "class",
    parms = list(split = "gini"), control = temp)

Variables actually used in tree construction:
[1] OGH  REG  SET1

Root node error: 51097/118345 = 0.43176

n= 118345

        CP nsplit rel error  xerror      xstd
1 0.345187      0   1.00000 1.00000 0.0033348
2 0.093792      1   0.65481 0.65481 0.0030318
3 0.002500      3   0.46723 0.46723 0.0027017
```

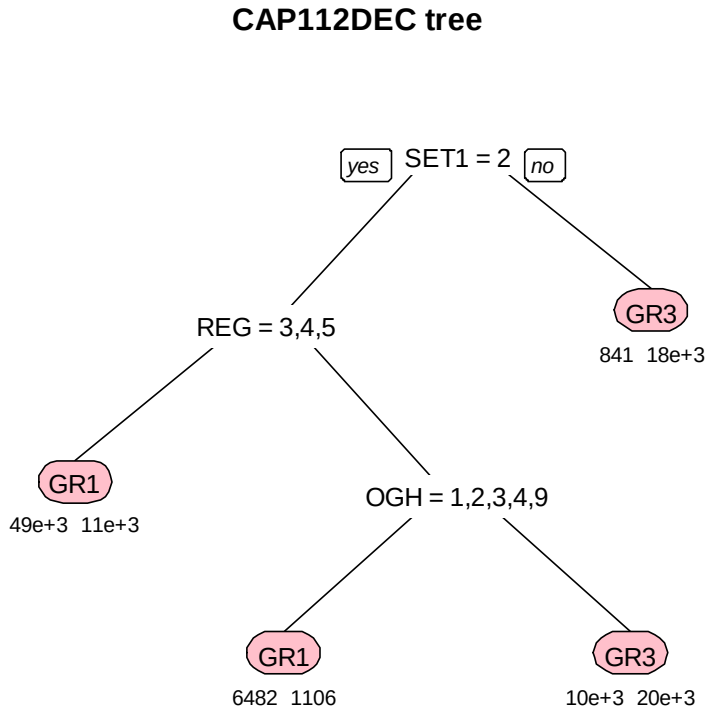


Figure 38. CAP112DEC tree model, data: Brazil PNAD 2011

The CAP112DEC model on the training data including 67,248 individuals from GR1 and 51,097 individuals from GR3 results in a tree with 4 terminal nodes, generating 0.467 cross-validation error rate for a cp equal to 0.025[211]. We don't proceed with the pruning phase. The 4 discovered patterns are: (i) living in urban areas of the southeast, south, central-west regions-73 per cent of GR1; (ii) living in urban areas of north and northeast regions and working as managers, science and art professionals, mid-level technicians, administrative workers, and armed and auxiliary members-9.6 per cent of GR1; (iii) living in urban areas of north and northeast regions and working as service workers, trade vendors and service providers, agricultural workers, and goods and service production and repair and maintenance workers-31 per cent of GR3; and (iv) living in rural areas-35 per cent of GR3. The first two patterns are of

[211] 64 individuals are dropped due to missing values in occupational group of the head of the household (OGH).

GR1 vs. the last two models are of GR3. Thus, the first two patterns show the most important features of the least poor group compared to the poorest group, and the last two patterns show the most important features of the poorest group compared to the least poor group in the country. We can clearly see that the poorest group is characterized by living in rural areas compared to GR1, and also by 4 socio-professional groups. We can also point out that the majority of GR1 don't live in north parts of the country. As the confusion matrix statistics show (see the box below), the predictive accuracy of the tree is 79 per cent with 83 per cent of non-misclassification for GR1 and 75 per cent of non-misclassification for GR3 in the test data. Thus, the performance of the tree is relatively good but notably the performance is higher by 8 per centage points for GR1 compared to GR3.

```
Confusion Matrix and Statistics

          Reference
Prediction   GR1   GR3
       GR1 27854  6336
       GR3  5771 19213

               Accuracy : 0.7954
                 95% CI : (0.7921, 0.7986)
    No Information Rate : 0.5682
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.5819

 Mcnemar's Test P-Value : 2.963e-07

            Sensitivity : 0.8284
            Specificity : 0.7520
         Pos Pred Value : 0.8147
         Neg Pred Value : 0.7690
             Prevalence : 0.5682
         Detection Rate : 0.4707
   Detection Prevalence : 0.5778
      Balanced Accuracy : 0.7902

       'Positive' Class : GR1
```

Fourth step (step4 analysis): detecting multidimensional poverty profiles of specific groups

To finalize the analysis, we will briefly focus on the case of elderly, GR1 population respectively, and after that we will produce some cross-tabulations regarding some additional variables.

Elderly

In terms of the 3 deprivation groups, 38 per cent of elderly population live in the poorest group, 30.87 per cent live in GR2, and 31 per cent live in GR1. Besides, 34.4 per cent of them live in a household composed of 2 persons, 21.4 per cent live in a household of 3 persons, 15.7 per cent live alone, 15 per cent live in a household f at least 5 persons, and 13 per cent live in a household of 4 persons. As a function of the poverty features discovered through classification tree, dimensional, and decomposition analysis, the situation of elderly can be revised for increasing their living conditions. For this population, it becomes extremely important the health issues, and their monthly retirement or other earnings to afford their general and specific needs due to their age. For that reason, specific questions on health, and grades can help a lot for capabilities examination of elderly. From the survey results, we cannot know how the household income is distributed among the members, and how it is used for affording the needs

of elderly in case they live in households with at least 2 people. Moreover, elderly living alone should be on specific focus, notably in terms of whether they receive or not grades and whether they are able or not to reach medical help when they need it.

GR1

We will briefly see the profiles of GR1 in terms of geographic situation, educational attainment, and economic activity respectively. Regarding the first aspect, 98.7 per cent of GR1 live in urban areas vs. only 1.28 per cent live in rural areas. In other words, 41 per cent of GR1 live in urban areas of southeast region (where the most famous cities of the country, Rio de Janeiro and Sao Paulo are located), 20.7 per cent live in urban areas of south region, 17,4 percent live in urban areas of northeast region, 11.6 per cent live in urban areas of central-west region, 7.8 per cent live in urban areas of north region. In terms of educational attainment, 37.68 per cent of GR1 are high school graduates, 24.73 per cent are university graduates, 10.85 per cent are literate without a diploma, 9.92 per cent are elementary school graduates, 7 per cent are primary school graduates, and 5.7 per cent are secondary school graduates. Hence, the most striking two points regarding education are among the least poor group, the per centage of university graduates is significantly higher, and the per centage of illiterates is significantly lower compared to those of GR2, and GR3. However, still 35 per cent of GR1 are deprived regarding educational attainment; such rate can be considered a relatively important number for a less poor group.

When it comes to economic activity, 69.5 per cent are economically active vs. 30.5 per cent are not. Compared to GR2, and GR3, the per centage of economically active population is around 20 per centage points superior for GR1. Regarding the combination of the variables, SC, WH, and JOB respectively, we find out that 29.36 per cent of GR1 contributes to social security, work between 40 and 44 hours per week, and perform one job; 7 per cent contributes to social security, work between 45 and 48 hours per week, and perform one job; 7 per cent contributes to social security, work between 15 and 39 hours per week, and perform one job; 6 per cent contributes to social security, work at least 49 hours per week, and perform one job; 5.6 per cent do not contribute to social security, work between 40 and 44 hours per week, and perform one job; and 4.4 per cent do not contribute to social security, work between 15 and 39 hours per week, and perform one job. These 6 profiles are the most relevant constructing 59 per cent of GR1. There are also 12 other profiles concerning 10.5 per cent of GR1, with a relatively small per centage each of them. Across 18 profiles, we can conclude that 13.45 per cent of GR1 are active in the market but they do not contribute to social security; such number still stands problematic given it is the least poor group pf the country. But also, 20 per cent of GR1 work long hours, at least 45 hours per week, and 35.4 per cent work between 40 and 44 hours per week. It is also important to note that compared to GR2 and GR3, the per centage of people working at least 45 hours per week and not contributing to social security is considerably low, around 5 per cent of GR1.

Besides, concerning economic activity status, 32.8 per cent of GR1 are employees with a signed work card, 12.5 per cent are own account workers, 8.4 per cent are statutory civil servants, 7.5 per cent are employees without a signed work card, and 4 per cent are self-employers. In terms of occupational group, 13.74 per cent of GR1 are goods and service production and repair and maintenance workers, 11.31 per cent are science and art professionals, 9.45 per cent are administrative workers, 8.75 per cent are trade vendors and service providers, 7.78 per cent are mid-level technicians, and 5.90 per cent are managers. Compared to GR2 and GR3, the per

centages of science and art professionals, and managers, and administrative workers are higher for GR1. However, it is important to note that across the country the per centage of managers is 2.93 per cent, and those of science and art professionals is 5.58 per cent, which are considerably low. So that, the majority of people working as managers and science and art professionals are from GR1. Finally, in terms of industrial activity, 14.62 per cent of GR1 are working in trade and repair sector, 10.34 per cent are working in education, health and social services, 9 per cent are working in manufacturing, and 6 per cent are working in public administration. We can see that compared to the poorest group, the per centage of people working in agricultural sector (1.34 per cent) is considerably low.

Additional variables

We will finalize the Brazil PNAD 2011 analysis by doing some cross-tables regarding the variables BEN_1 to BEN_5, UPW, CNPJ, INV, PAYC, DWE, WFIL, DR, and hh_child respectively.

The variables, namely, BEN_1, BEN_2, BEN_3, BEN_4, BEN_5, UPW, CNPJ, INV, and PAYC focus on some specific target groups in the labour market. The variables BEN_1, BEN_2, BEN_3, BEN_4, and BEN_5 concern the benefits received in the main job during the reference week for employees, and domestic workers (i.e., ES3 is 1,2,3,4,5,6). More specifically, from BEN_1 to BEN_5 refer to housing, food, transport, education or nursery, and health or rehabilitation benefits, respectively. The results show that, among economically active during the reference week employees, and domestic workers, 4.22 per cent receive housing benefits, 41.55 per cent receive food benefits, 40.83 per cent receive transport benefits, 2.28 per cent receive education/nursery benefits, and 17.68 per cent receive health or rehabilitation benefits from their main job. When we look at the combination of these 5 variables as the profile of benefits received in main job, we find out that the following 6 profiles arrive to explain 90 per cent of the target group: (i) no benefits received (42 per cent); (ii) receiving food and transport benefits and not receiving the other 3 benefits (17 per cent): (iii) receiving transport benefit and not receiving the other 4 benefits (10.45 per cent); (iv) receiving food, transport, and health or rehabilitation benefits and not receiving the other 2 benefits (8.9 per cent); (v) receiving food benefit and not receiving the other 4 benefits (8.6 per cent); and (vi) receiving food and health or rehabilitation benefits and not receiving the other 3 benefits (2.96 per cent). On the other hand, if we look at the distribution of benefits profile among employees and domestic workers in GR2, we find out that the following 5 profiles are the most relevant capturing in total 90.75 per cent of this group: (i) no benefits received (44.49 per cent); (ii) receiving food and transport benefits and not receiving the other 3 benefits (18.3 per cent); (iii) receiving transport benefits and not receiving the other 4 benefits (12.22 per cent); (iv) receiving food benefit and not receiving the other 4 benefits (8.44 per cent); and (v) receiving food, transport, health or rehabilitation benefits and not receiving the other 3 benefits (7.29 per cent). Besides, among employees and domestic workers in GR3, the most relevant 5 profiles detecting 90 per cent of the group are: (i) no benefits received (56 per cent); (ii) receiving food and transport benefits and not receiving the other 3 benefits (11.46 per cent); (iii) receiving food benefits and not receiving the other 4 benefits (10.5 per cent); (iv) receiving transport benefits and not receiving the other 4 benefits (7.9 per cent); (v) receiving housing benefit and not receiving the other 4 benefits (4.11 per cent). And, concerning GR1, the following 7 profiles explain 91.67 per cent of GR1: (i) no benefits received (33.74 per cent); (ii) receiving food and transport benefits and not receiving the other 3 benefits (18.88 per cent); (iii) receiving food, transport, and health or rehabilitation benefits and not receiving the other 2 benefits (12.94 per

cent); (iv) receiving transport benefit and not receiving the other 4 benefits (10.56 per cent); (v) receiving food benefit and not receiving the other 4 benefits (7.83 per cent); (vi) receiving food and health or rehabilitation benefits and not receiving the other 3 benefits (4.68 per cent); and (vii) receiving health or rehabilitation benefit and not receiving the other 4 benefits (3 per cent).

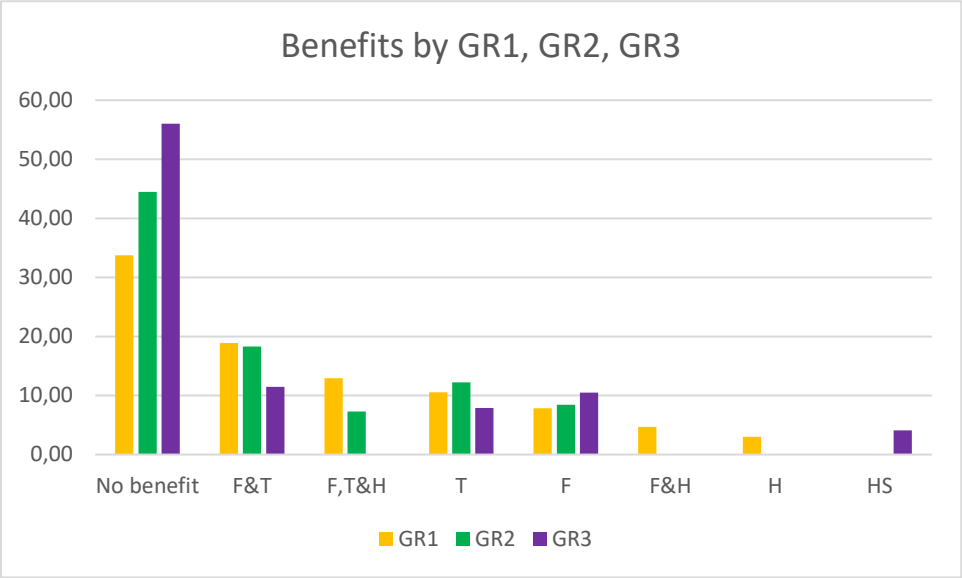


Figure 39. Benefits received by employees and domestic workers by GR1, GR2, GR3. Note: F=food, T=transport, H=health, HS=housing. Column numbers are in per centage, data: Brazil PNAD 2011

As we can see from Figure 39, GR3 receives less benefits but receive slightly more housing benefit compared to GR1 and GR2. Economically active population in GR3, the percentages of those do not contribute to social security or those do not have a signed work card are higher compared to GR1 and GR2; related to that, we can intuitively say that the per centage of people receiving benefits from their job may be lower as well among GR3.

The variable, UPW, is based on the question addressed to own account workers and employers to know whether there is at least one unpaid work in their workplaces. 12.18 per cent of them answered yes vs. 87.82 per cent answered no. There are in total 4,054 unpaid workers in our 18+ economically active population in the reference week (corresponding to 1.63 per cent). 63.20 per cent of unpaid workers are in GR3, 16.77 per cent are in GR2, and 20 per cent are in GR1. We can see that the per centage of unpaid workers is much higher in GR1, but the per centages emerge relatively important for GR1 and GR2 as well.

Moreover, the variable, CNPJ, helps to find out the frequency of enterprises not registered to National Register of Legal Entities (CNPJ). Enterprises not registered to CNPJ may be seen as the part of informal sector. The related question is asked to employees in private sector, own account workers and employers except domestic workers. Among GR1, 76.2 per cent said their workplace is registered to CNPJ, 20. 48 per cent said no, and there are 3.31 per cent missing values; among GR2, 61 per cent said yes, 32.56 per cent said no, and there are 6.38 per cent missing values; and among GR3, 32.47 per cent said yes, 58.68 per cent said no, and there are 8.85 per cent missing values. Based on these numbers, we can deduce that people in the poorest group are more likely working in informal sector as employees or own account workers. When we look at the distribution of the variable, INV (whether the enterprise where the person works provide invoices to the clients) among deprivation groups, we observe that 32.36 per cent of

private sector employers in GR3 said no, 13.19 per cent of those in GR2 said no, and 7.96 per cent of those in GR1 said no. Besides, 44.24 per cent of private sector employees in GR1 said the enterprise where s/e works do not provide paychecks to the employees, 21.11 per cent of those in GR2, and 12.75 per cent of those in GR1 gave also the same answer. These two last variables support also the possibility of higher frequency of employees in informal sector among the poorer groups.

Now we will look at the distribution of 3 dwelling conditions variables, namely, DWE (type pf main dwelling), WFIL (water filter), DR (drainage of bathroom or toilet). 80.5 per cent of GR1 live in a house and 19.39 per cent live in an apartment; 94.14 per cent of GR2 live in a house and 5.5 per cent live in an apartment; 98 per cent of GR3 live in a house and 1.36 per cent live in an apartment. Besides, 37.41 per cent of GR1, 50.13 per cent of GR2, and 61 per cent of GR3 do not have a water filter in their dwelling. And concerning drainage of bathroom/toilet, 68.5 per cent of GR1, 50.65 of GR2, and 21.86 per cent of GR3 live in a dwelling having a sewer or pluvial collection network; 12.5 per cent of GR1, 10.4 per cent of GR2, and 5.19 per cent of GR3 live in a dwelling using a sceptic tank connected to the collecting network sewage or rain; 11.65 per cent of GR1, 20.24 per cent of GR2, and 22.60 per cent of GR3 live in a dwelling using a sceptic tank not connected to the collecting network sewage or rain; 6.39 per cent of GR1, 16.14 per cent of GR2, and 34.97 per cent of GR3 live in a dwelling using a rudimentary pit for the drainage of bathroom/toilet. We can observe that the use of sewer or pluvial collection network, which is the best way for drainage is higher among GR1 but still its frequency is quite low for a least poor group. The poorest group (GR3) mostly uses a rudimentary pit. But also, 3.19 per cent of GR3 live in a dwelling using a ditch, and 3.13 per cent of GR3 live in a dwelling using river, lake or sea for the drainage. The deprivation regarding drainage becomes worst from the least poor to the poorest group. The situation should be checked for GR2 and GR3 and seen as one of the priorities for increasing sanitation facilities. However, it is important to remark that for GR1 also the situation is not perfect.

We will finish the analysis by a household composition variable, hh_child (number of children in the household). 50 per cent of individuals from GR1, 45.33 per cent of individuals from GR2, and 39.67 per cent of individuals from GR3 live in a household with no child. The per centage of people living in a household with at least 4 children is about 1 per cent for GR1 whilst it is 3.18 per cent for GR2, and 7.64 per cent for GR3. The per centage of people living in a household with 3 children is 3.86 for GR1, 6.20 for GR2, and 9.43 for GR3. For GR1, 29. 35 per cent live in a household with 1 child and 15.41 per cent live in a household with 2 children; for GR2, 28.39 per cent live in a household with 1 child and 16.91 per cent live in a household with 2 children; and for GR3, 25 per cent live in a household with1 child and 18.22 per cent live in a household with 2 children. We can observe that the per centage of people living in a household with no children is higher among GR1 compared to GR3 by 10.63 per centage points, and the per centage of people living in a household with at least 4 children is higher among GR3 compared to GR1 by 6.64 per centage points.

9.7 Additional comparative analysis

Here, we provide detailed information on comparisons between South Africa, Turkey, Mexico, and Brazil. Before entering into detail on comparisons, the table below shows the availability of indicators and dimension per country and whether the indicator is used for the overall index or decomposition analysis in case studies.

Table 95. Indicators used in overall deprivation index and decomposition analysis (except the main variables used in Fourth Model, individual characteristics and household composition related variables) of 4 country cases: South Africa GHS 2011, Turkey SILC 2011, Mexico MCS 2012, and Brazil PNAD 2011.

Indicators	***Dimension***	***Overall index (IND) or Decomposition analysis (DEC)***	**South Africa GHS 2011**	**Turkey SILC 2011**	**Mexico MCS 2012**	**Brazil PNAD 2011**
DWE_D	DWS	IND	Yes	*No*	*No*	*No*
ROF_D	DWS	IND	Yes	*No*	Yes	Yes
WAL_D	DWS	IND	Yes	*No*	Yes	Yes
ROFC_D	DWS	IND	Yes	Yes	*No*	*No*
WALC_D	DWS	IND	Yes	Yes	*No*	*No*
RO_D	DWS	IND	Yes	Yes	*No*	Yes
KIT_D	DWS	IND	Yes	Yes	Yes	*No*
SAN_D	DWS	IND	Yes	Yes	Yes	Yes
WATD_D	DWS	IND	Yes	Yes	Yes	Yes
ELC_D	DWS	IND	Yes	*No*	Yes	*No*
ENG_C_D	DWS	IND	Yes	*No*	Yes	Yes
ENG_L_D	DWS	IND	Yes	*No*	*No*	Yes
ENG_H_D	DWS	IND	Yes	Yes	*No*	*No*
TV_D	MS	IND	Yes	Yes	Yes	Yes
RAD_D	MS	IND	Yes	*No*	Yes	Yes
REF_D	MS	IND	Yes	Yes	Yes	Yes
TEL_D	MS	IND	Yes	Yes	Yes	Yes
INT_D	MS	IND	Yes	*No*	Yes	*No*
EDU_D	EDUC	IND	Yes	Yes	Yes	Yes
RUB_D	ENV/DWS	IND	Yes	*No*	Yes	Yes
LIT_D	ENV	IND	Yes	*No*	*No*	*No*
WP_D	ENV	IND	Yes	*No*	*No*	*No*
AP_D	ENV	IND	Yes	*No*	*No*	*No*
LD_D	ENV	IND	Yes	*No*	*No*	*No*
NOI_D	ENV	IND	Yes	Yes	*No*	*No*
POST_D	ENV	IND	Yes	*No*	*No*	*No*
MC_D	HEALTH	IND	Yes	*No*	Yes	*No*
ES1	ECOA	DEC	Yes	Yes	Yes	Yes
CHN_D	NUT	IND	Yes	*No*	*No*	*No*
ADN_D	NUT	IND	Yes	*No*	*No*	*No*
NU1_D	NUT	IND	Yes	*No*	*No*	*No*
NU2_D	NUT	IND	Yes	*No*	*No*	*No*
NU3_D	NUT	IND	Yes	*No*	*No*	*No*
NU4_D	NUT	IND	Yes	*No*	*No*	*No*
HWS_D	DWS	IND	*No*	Yes	Yes	*No*
HEATS_D	DWS	IND	*No*	Yes	Yes	*No*
INS_D	DWS	IND	*No*	Yes	*No*	*No*
LIGHT_D	DWS	IND	*No*	Yes	*No*	*No*
AIRC_D	MS	IND	*No*	Yes	Yes	*No*
WM_D	MS	IND	*No*	Yes	Yes	Yes
CAR_D	MS	IND	*No*	Yes	Yes	Yes
DISHW_D	MS	IND	*No*	Yes	*No*	*No*

COMPINT_D	MS	IND	*No*	Yes	Yes	Yes
MPHO_D	MS	IND	Yes	Yes	Yes	Yes
POL_D	ENV	IND	*No*	Yes	*No*	*No*
CRI_D	ENV	IND	*No*	Yes	*No*	*No*
ES3	ECOA	DEC	*No*	Yes	*No*	Yes
WH	ECOA	DEC	*No*	Yes	*No*	Yes
SC	ECOA	DEC	*No*	Yes	*No*	Yes
FS1_D	FS	IND	*No*	Yes	*No*	*No*
FS2	FS	DEC	*No*	Yes	*No*	*No*
FS3	FS	DEC	*No*	Yes	*No*	*No*
FS4	FS	DEC	*No*	Yes	*No*	*No*
FS5	FS	DEC	*No*	Yes	*No*	*No*
FS6	FS	DEC	*No*	Yes	*No*	*No*
FS7	FS	DEC	*No*	Yes	*No*	*No*
FS8	FS	DEC	*No*	Yes	*No*	*No*
FS9	FS	DEC	*No*	Yes	*No*	*No*
FS10	FS	DEC	*No*	Yes	*No*	*No*
FS11	FS	DEC	*No*	Yes	*No*	*No*
FS12	FS	DEC	*No*	Yes	*No*	*No*
FLO_D	DWS	IND	*No*	*No*	Yes	*No*
SINK_D	DWS	IND	*No*	*No*	Yes	*No*
DR_D	DWS	IND	*No*	*No*	Yes	Yes
STOVE_D	MS	IND	*No*	*No*	Yes	*No*
IRON_D	MS	IND	*No*	*No*	Yes	*No*
VC_D	MS	IND	*No*	*No*	Yes	*No*
JOB	ECOA	DEC	*No*	*No*	Yes	Yes
AF1_D	NUT	IND	*No*	*No*	Yes	*No*
AF2_D	NUT	IND	*No*	*No*	Yes	*No*
AF3_D	NUT	IND	*No*	*No*	Yes	*No*
MED_D	HEALTH	IND	*No*	Yes	*No*	*No*
DEN_D	HEATH	IND	*No*	Yes	*No*	*No*

Comparison between South Africa, Turkey and Brazil

Second comparison model includes the three countries, South Africa, Turkey, and Brazil. Taking in consideration the common variables (see above), the essential capabilities profile is the combination of the variables WATD, RO, EDU, respectively (ESCAP_C2= {WATD, RO, EDU}). On the other hand, the material situation profile is the combination of the variables TV, REF, TEL, MPHO, respectively (MSCAP_C2= {TV, REF, TEL, MPHO}). And, the complete profile is composed of WATD, RO, EDU, TV, REF, TEL, MPHO, respectively, (CAP_C2= {WATD, RO, EDU, TV, REF, TEL, MPHO}). Each variable represents a functioning such as WATD means access to drinking water. The outputs of functionings are *deprivation* or *non-deprivation*. The situation of deprivation is coded 1 and the situation of non-deprivation is coded 0. For instance, for ESCAP_C2, the profile {111} corresponds to deprivation regarding WATD, RO, EDU at the same time.

Table 96. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, and BR PNAD 2011

	South Africa (SA GHS 2011) 6% mv.	Turkey (TR SILC 2011)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C2)	111 (30.44), 000 (13.99), 101 (13.53), 011 (12.65), 001 (9.15), 110 (7.75), 010 (7.74), 100 (4.76) Total info=100%	001 (51.2), 011 (18.7), 000 (16.8), 010 (10.6), 101 (2.3) Total info=100%	011 (40.26), 010 (31.31), 111 (12.11), 001 (5.84), 000 (4.6), 110 (4) Total info=94%
Material situation profiles (MSCAP_C2)	0010 (58.27), 0000 (14.85), 1110 (10.57), 0110 (6.5), 1010 (3.21), 1111 (2.5), 0011 (2.18) Total info=98%	0000 (51.4), 0010 (42), 0001 (3.13) Total info=96.61%	0010 (48), 0000 (38.14), 0011 (6.1), 0001 (2.26) Total info=95%
Complete profiles (CAP_C2)	1110010 (16.74), 0110010 (8.77), 1010010 (7.25), 0000000 (7.03), 0000010 (6.33), 1111110 (6.22), 1100010 (5.67), 0100010 (5.33), 0010010 (4.77), 1110110 (3.58), 1000010 (3.42), 0010000 (2.83), 1011110 (2.36), 0110000 (2.04) Total info=82.33%	0010010 (25.7), 0010000 (23), 0000000 (10.8), 0110000 (10.3), 0100000 (7), 0000010 (5.88), 0110010 (5.47), 0100010 (3.4), 0110001 (2.11) Total info=94%	0110010 (20.25), 0100000 (16.49), 0110000 (15), 0100010 (13.55), 1110010 (6.94), 1100010 (2.48), 0000000 (2.36), 0110011 (2.34), 0010010 (2.2) Total info=82%

Regarding ESCAP profiles, we can say that: (i) in terms of deprivation in EDU and not in other items, Turkey has the worst score concerning 51.2 per cent of 18+ population, which is 5.6 times higher than South Africa, and 8.8 times higher than Brazil; (ii) in terms of deprivation in RO and EDU, and not in other items, Brazil has the worst score concerning 40 per cent of 18+ population, which is 2.15 times higher than Turkey and 3.2 times higher than South Africa; (iii) in terms of no deprivation in all items at the same time, Turkey has the best performance concerning 16.8 per cent of 18+ population, which is 1.2 times higher than Turkey and 3 times higher than Brazil; (iv) in terms of deprivation in RO and not in other items, Brazil has the worst performance concerning 31.3 per cent of 18+ population, which is 3 times higher than Turkey and 4 times higher than South Africa. In terms of total prosperity. Turkey has the highest score (16.8), then South Africa (13.99), and Brazil (4.6). In terms of total poverty, South Africa (30.44) has the highest score, then Brazil (12.11), and Turkey (0) come. In terms of intermediate poverty, Turkey has the highest score (82.2), then Brazil (83.29), and South Africa (55.57) come. If we take into account intensity in terms of number of deprivations for intermediate poverty case, 2 deprivations should be considered the heaviest, and 1 deprivation as the lowest one. So that, in terms of intermediate poverty with 2 deprivations, Brazil appears as the worst

and then South Africa and Turkey come; in terms of 1 deprivation, Turkey, Brazil, and South Africa come respectively.

Regarding MSCAP profiles, we can observe that; (i) in terms of no deprivation in all items at the same time, Turkey has the best performance concerning 51.4 per cent of 18+ population, which is 1.35 times higher than Brazil and 3.5 times higher than South Africa; (ii) in terms of TEL deprivation and not in other items, South Africa has the worst performance concerning 58.3 per cent of 18+ population, which is 1.2 times higher than Brazil, and 1.4 times higher than South Africa. In terms of total prosperity, Turkey has a highest score (51.4), then Brazil (38.14), and South Africa (14.85) come. In terms of total poverty, South Africa has a score under 3 per cent, and in case of Turkey and Brazil it is zero or near zero. The ranking of intermediate poverty is the same of total prosperity. If we take into account intensity in terms of number of deprivations for intermediate poverty case, 3 deprivations should be considered the heaviest, 2 deprivations as the middle level, and 1 deprivation as the lowest one. So that, in terms of intermediate poverty with 3 deprivations, South Africa appears as the worst; in terms of 2 deprivations, South Africa as first and then Brazil emerges; in terms of 1 deprivation, South Africa, Brazil, and Turkey come.

Regarding CAP profiles, the situation appears very complex and a proper comparison becomes impossible due to heterogeneity of profiles in South Africa and Brazil cases and there are no relevant common profiles between countries.

At an approximative level, we can conclude that Turkey has the worst situation regarding EDU, Brazil has the worst situation regarding RO, and South Africa has the worst situation regarding WATD and material situation (tv, refrigerator, telephone). And, at overall aspect, Brazil is in the middle situation between Turkey and South Africa.

Comparison between South Africa, Turkey and Mexico

First comparison model includes the three countries, South Africa, Turkey, and Mexico. Taking in consideration the common variables (see above), the essential capabilities profile is the combination of the variables WATD, SAN, KIT, EDU, respectively (ESCAP_C3= {WATD, SAN, KIT, EDU}). On the other hand, the material situation profile is the combination of the variables TV, REF, TEL, MPHO, respectively (MSCAP_C3= {TV, REF, TEL, MPHO}). And, the complete profile is composed of WATD, SAN, KIT, EDU, TV, REF, TEL, MPHO, respectively (CAP_C3= {WATD, SAN, KIT, EDU, TV, REF, TEL, MPHO}). Each variable represents a functioning such as WATD means access to drinking water. The outputs of functionings are *deprivation* or *non-deprivation*. The situation of deprivation is coded 1 and the situation of non-deprivation is coded 0. For instance, for ESCAP_C3, the profile {1001} corresponds to deprivation regarding WATD, non-deprivation regarding SAN, non-deprivation regarding KIT, and deprivation regarding EDU.

Table 97. Essential capabilities and material situation profiles across SA GHS 2011, TR SILC 2011, and MX MCS 2012

	South Africa (SA GHS 2011) 7.3% mv.	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)
Essential capabilities profiles (ESCAP_C3)	1001 (32), 0000 (20.78), 0001 (19.82), 1000 (9.65), 1011 (7.96), 1101 (2.68),	0001 (60.88), 0000 (26.8), 0101 (6.4) Total info=94%	0001 (35.59), 0000 (31), 1001 (18.95), 1000 (4.25), 1011 (3.49) Total info=93%

	1010 (2.46), 0011 (2.03) Total info=97.37%		
Material situation profiles (MSCAP_C3)	0010 (58.31), 0000 (14.96), 1110 (10.51), 0110 (6.49), 1010 (3.16), 1111 (2.48), 0011 (2.19) Total info=98%	0000 (51.4), 0010 (42), 0001 (3.13) Total info=96.6%	0010 (33.9), 0000 (27.9), 0011 (10.85), 0001 (10.75), 0111 (4.77), 0110 (4.76), 1111 (2.9) Total info=96%
Complete profiles (CAP_C3)	10010010 (19.96), 00010010 (12.43), 00000010 (10.99), 00000000 (8.94), 10000010 (7.47), 00010000 (4.82), 10011110 (4.7), 10110010 (3.19), 10010110 (2.97), 10111110 (2.2) Total info=77.68%	00010000 (30.74), 00010010 (26.17), 00000000 (17.54), 00000010 (9), 01010010 (3.67), 00010001 (2.42), Total info=89.5%	00000000 (16.29), 00010010 (13.2), 00010000 (9.4), 00000010 (9.3), 10010010 (6.38), 00010001 (5.49), 00010011 (4), 10010011 (3.72), 00000001 (3), 10010111 (2.2) Total info=73%

Regarding ESCAP profiles, we can say that: (i) Turkey has the worst performance in terms of deprivation only in EDU concerning 60.88 per cent of 18+ population, which is 1.7 times higher than Mexico, and 3 times higher than South Africa; (ii) in terms of no deprivation in all items, Mexico has the best performance concerning 31 per cent of 18+ population, which is 1.5 times higher than South Africa and has the similar performance with Turkey.

Regarding MSCAP profiles, we can observe that: (i) in terms of no deprivation in all items, Turkey has the best performance concerning 51.4 per cent of 18+ population, which is 1.84 times higher than Mexico, and 3.4 times higher than South Africa; (ii) in terms of deprivation in TEL and not in other items, South Africa has the worst performance concerning 58.3 per cent of 18+ population, which is 1.38 times higher than Turkey, and 1.7 times higher than Mexico. In terms of total prosperity, Turkey has the highest score (51.4), then Mexico (27.9) and South Africa (14.96) come. In terms of total poverty, Mexico and South Africa have a score under 3 per cent, and such case was not found in Turkey. In terms of intermediate poverty with 3 deprivations, South Africa has the highest figure (10.5), then Mexico comes (4.77). Concerning intermediate poverty with 2 deprivations, Mexico has the highest figure (15.6), then South Africa comes (11.8). On the other hand, South Africa has the highest score for 1 deprivation (58), then Mexico comes (44.65), then Turkey emerges (45).

Regarding CAP profiles, we can remark that: (i) in terms of deprivation in EDU and no deprivation in other items, Turkey has the worst performance concerning 30.74 per cent of 18+ population, which is 6.37 times higher than South Africa and 3.27 times higher than Mexico; (ii) in terms of deprivation in EDU and TEL and not in other items, Turkey has the worst performance concerning 26.17 per cent of 18+ population, which is two times higher than Mexico and South Africa; (iii) in terms of no deprivation in all items, Mexico has the best performance concerning 16 per cent of 18+ population, then Turkey comes with 17.54 per cent, and South Africa with almost 2 times worser performance compared to Mexico and Turkey; (iv) in term of deprivation in TEL and not in other items, the three countries have the similar performance.

At an approximative level, we can conclude that South Africa has the worst situation regarding WATD and material deprivation (TV, REF, TEL), Turkey has the worst situation regarding EDU and SAN, but also best situation regarding material deprivation, and Brazil has an intermediate level in suffering also deprivation regarding EDU, TEL, and WATD.

Comparison between South Africa, Mexico and Brazil

First comparison model includes the three countries, South Africa, Mexico, and Brazil. Taking in consideration the common variables (see above), the essential capabilities profile is the combination of the variables WATD and EDU, respectively (ESCAP_C4= {WATD, EDU}). On the other hand, the material situation profile is the combination of the variables TV, REF, TEL, MPHO, RAD, respectively (MSCAP_C4= {TV, REF, TEL, MPHO, RAD}). And, the complete profile is composed of WATD, EDU, TV, REF, TEL, MPHO, RAD, respectively (CAP_C4= {WATD, EDU, TV, REF, TEL, MPHO, RAD}). Each variable represents a functioning such as WATD means access to drinking water. The outputs of functionings are *deprivation* or *non-deprivation*. The situation of deprivation is coded 1 and the situation of non-deprivation is coded 0. For instance, for ESCAP_C4, the profile {11} corresponds to deprivation regarding both WATD and EDU.

Table 98. Essential capabilities and material situation profiles across SA GHS 2011, MX MCS 2012, and BR PNAD 2011

	South Africa (SA GHS 2011) 6.5% mv.	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C4)	11 (44), 01 (21.7), 00 (21.65), 10 (12.56) Total info=100%	01 (37.8), 00 (31.9), 11 (25.11), 10 (5.14) Total info=100%	01 (46.11), 00 (35.92), 11 (13.58), 10 (4.39) Total info=100%
Material situation profiles (MSCAP_C4)	00100 (50.41), 00000 (13.94), 00101 (7.8), 11101 (5.35), 11100 (5.27), 01100 (4.28), 01101 (2.25), 10100 (2.17) Total info=91.48%	00101 (28.9), 00001 (22.5), 00111 (8.89), 00011 (8.5), 00000 (5.39), 00100 (5), 01101 (4), 01111 (3.84), 11111 (2.3), 00010 (2.24) Total info=91.6%	00100 (46.98), 00000 (38), 00110 (5.47), 00010 (2.26) Total info=93%
Complete profiles (CAP_C4)	1100100 (20), 0100100 (11.8), 0000100 (10.56), 0000000 (8.52), 1000100 (8.04), 0100000 (4.5), 1111101 (4.34), 1111100 (4.28), 1100101 (4), 1101100 (3) Total info=79%	0000001 (13.26), 0100101 (11.84), 0000101 (8.39), 0100001 (7.74), 1100101 (6.68), 0100011 (4.38), 1100111 (3.92), 0100111 (3.5), 0000000 (3.25), 1101111 (2.6), 0000011 (2.46), 0100100 (2.19), 1000101 (2) Total info=72%	0100100 (22.23), 0000000 (18.83), 0100000 (16.86), 0000100 (15.39), 1100100 (6.77), 0100110 (2.94), 1000100 (2.59) Total info=86%

Regarding ESCAP profiles, we can say that: (i) in terms of deprivation in WATD and EDU, South Africa has the worst performance concerning 44 per cent of 18+ population, which is 1.75 times higher than Mexico, and 3.24 times higher than Brazil; (ii) in terms of deprivation in WATD and not in EDU, South Africa has the worst performance concerning 12.56 per cent of 18+ population, which is 7.4 times higher than Mexico, and 2.86 times higher than Brazil; (iii) in terms of deprivation in EDU, and not in WATD, Brazil has the worst performance concerning 46 per cent of 18+ population, which is 1.2 times higher than Mexico, and 2.1 times higher than South Africa; (iv) in terms of non-deprivation in all items, Brazil has the best score concerning 36 per cent of 18+ population, which is 1.12 times higher than Mexico and 1.65 times higher than South Africa. In terms of total prosperity, Brazil has the highest score (35.92), then Mexico (31.9) and South Africa (21.65) come. In terms of total poverty, South Africa has the highest score (44), then Mexico (25.11) and Brazil (13.58) come. In terms of intermediate poverty, Brazil has the highest score (50.5), then Mexico (42.94) and South Africa (34.26) come.

Regarding MSCAP profiles, we can observe that: (i) in terms of deprivation in TEL and not in other items, South Africa has the worst performance concerning 50.4 per cent of 18+ population, which is 1.07 times higher than Brazil, and 10 times higher than Mexico; (ii) in terms of non-deprivation in all items, Brazil has the best performance concerning 38 per cent of 18+ population, which is 2.7 times higher than South Africa and 7 times higher than Mexico.

Regarding CAP profiles, due to heterogeneity there is no possibility of doing much comparison except we can say that (i) in terms of deprivation in EDU and TEL, Brazil has the worst performance concerning 22.2 per cent of 18+ population, which is 1.88 times higher than South Africa, and 10 times higher than Mexico; (ii) in terms of non-deprivation in all items, Brazil has the best performance concerning 18.8 per cent of 18+ population, which is 2.2 times higher than South Africa and 5.8 times higher than Mexico.

At an approximative level, we can conclude that South Africa has the worst situation regarding WATD, EDU, TV, REF, TEL, Mexico has slightly worst situation compared to Brazil.

Comparison between Turkey, Mexico and Brazil

Fifth comparison model includes the three countries, Turkey, Mexico, and Brazil. Taking in consideration the common variables (see above), the essential capabilities profile is the combination of the variables WATD, EDU, respectively (ESCAP_C5= {WATD, EDU}). On the other hand, the material situation profile is the combination of the variables TV, REF, TEL, MPHO, WM, CAR, COMPINT, respectively (MSCAP_C5= {TV, REF, TEL, MPHO, WM, CAR, COMPINT}). And, the complete profile is composed of WATD, EDU, TV, REF, TEL, MPHO, WM, CAR, COMPINT, respectively (CAP_C5= {WATD, EDU, TV, REF, TEL, MPHO, WM, CAR, COMPINT). Each variable represents a functioning such as WATD means access to drinking water. The outputs of functionings are *deprivation* or *non-deprivation*. The situation of deprivation is coded 1 and the situation of non-deprivation is coded 0. For instance, for ESCAP_C5, the profile {11} corresponds to deprivation regarding both WATD and EDU.

Table 99. Essential capabilities and material situation profiles across TR SILC 2011, MX MCS 2012, and BR PNAD 2011

	Turkey (TR SILC 2011) 0.77% mv.	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C5)	01 (70) 00 (27.4), 11 (2.5) Total info:100%	01 (37.83), 00 (31.92), 11 (25.11), 10 (5.14) Total info=100%	01 (46.11), 00 (35.92), 11 (13.58), 10 (4.39) Total info=100%
Material situation profiles (MSCAP_C5)	0010011 (22.57), 0000000 (16.06), 0000010 (14.5), 0000011 (13.94), 0010010 (6.22), 0000001 (6), 0010001 (5.52), 0010000 (4.5), 0010111 (2.85), 0001011 (2.5) Total info: 95%	0010011 (14.23), 0000000 (11.44), 0010111 (7.79), 0000010 (6.63), 0011111 (5), 0000011 (4.77), 0011011 (4.64), 0001011 (4.56), 0111111 (3.8), 0010001 (3.7), 0010010 (3.35), 0110111 (2.9), 1111111 (2.8), 0000001 (2.63), 0010000 (2.56) Total info:80.86%	0010111 (19), 0000000 (18.3), 0010011 (6.1), 0010000 (6), 0000010 (5.34), 0011111 (4.65), 0010101 (3.82), 0010110 (3.79), 0010010 (3.78), 0010001 (3.17), 0000111 (3.16), 0000011 (2.94), 0000001 (2.71), 0010100 (2.32), 0000100 (2.31), 0000110 (2.26) Total info: 90%
Complete profiles (CAP_C5)	010010011 (19), 010000011 (11.7), 010000010 (8.8), 010000000 (7.5), 000000000 (8.55), 000000010 (5.7), 010000001 (4.7), 010010001 (4.4), 010010010 (3.6), 000010011 (3), 000010010 (2.6), 010001011 (2.4), 000010000 (2.4), 000000011 (2.22), 010010000 (2) Total info: 89%	000000000 (8.88), 010010011 (6.76), 000000010 (3.95), 110010011 (3.75), 010010111 (2.99), 110010111 (2.95), 010001011 (2.89), 010000011 (2.79), 000010011 (2.77), 110111111 (2.69), 110011111 (2.65), 010000000 (2.34), 111111111 (2.3), 010000010 (2.3), 010011011 (2.12) Total info=52%	-

Regarding ESCAP profiles, Turkey has the worst situation in terms of EDU but the best in terms of WATD. Brazil has slightly a better position compared to Mexico in terms of EDU, and WATD.

Regarding MSCAP profiles, we can observe that: (i) in terms of deprivation in terms of TEL, CAR, COMPINT, and not in other items, Turkey has the best performance concerning 22.6 per cent of 18+ population, which is 1.2 times higher than Brazil and 1.6 times higher than Mexico; (ii) in terms of non-deprivation in all items, Brazil has the best performance concerning 18 per cent of 18+ population, which is 1.14 times higher than Turkey and 1.6 times higher than Mexico; (iii) in terms of deprivation in CAR but not in other items, Turkey has the worst performance concerning 14 per cent of 18+ population, which is 2.2 times higher than Mexico and 2.7 times higher than Brazil. Moreover, in terms of washing machine, Brazil has a worse

performance compared to Turkey and Mexico but in terms of CAR and COMPINT, Brazil has a better performance. And, Turkey and Mexico have quite similar performance regarding CAR and COMPINT.

Regarding CAP profiles, it is not possible to do a comparison as we could not obtain a relevant per cent of information for Brazil due to heterogeneity of profiles.

Comparison between South Africa and Turkey

Sixth comparison model includes the two countries, South Africa, and Turkey. Taking in consideration the common variables (see above), the essential capabilities profile is the combination of the variables WATD, SAN, RO, KIT, EDU, NOI, respectively (ESCAP_C6= {WATD, SAN, RO, KIT, EDU, NOI}). On the other hand, the material situation profile is the combination of the variables TV, REF, TEL, MPHO, respectively (MSCAP_C6= {TV, REF, TEL, MPHO}). And, the complete profile is composed of WATD, SAN, RO, KIT, EDU, NOI, TV, REF, TEL, MPHO, respectively (CAP_C6= {WATD, SAN, RO, KIT, EDU, NOI, TV, REF, TEL, MPHO}). Each variable represents a functioning such as WATD means access to drinking water. The outputs of functionings are *deprivation* or *non-deprivation*. The situation of deprivation is coded 1 and the situation of non-deprivation is coded 0. For instance, for ESCAP_C6, the profile {101010} corresponds to deprivation regarding WATD, non-deprivation regarding SAN, deprivation regarding RO, non-deprivation regarding KIT, deprivation regarding EDU, and non-deprivation regarding NOI.

Table 100. Essential capabilities and material situation profiles across SA GHS 2011, and TR SILC 2011

	South Africa (SA GHS 2011) 7.7% mv.	Turkey (TR SILC 2011)
Essential capabilities profiles (ESCAP_C6)	101010 (18.48), 000000 (12.15), 001010 (9.15), 100010 (8.78), 000010 (7.2), 001000 (5.93), 101000 (4.97), 101110 (4.55), 100000 (3.18), 101011 (3.16), 001011 (2.17) Total info=80%	000010 (36.37), 001010 (15.2), 000000 (13.13), 001000 (8.5), 000011 (6.9), 010010 (5), 000001 (3.1), 001011 (2.37), 001001 (2.07) Total info=93%
Material situation profiles (MSCAP_C6)	0010 (58.25), 0000 (14.98), 1110 (10.52), 0110 (6.49), 1010 (3.17), 1111 (2.48), 0011 (2.19) Total info=98%	0000 (51.4), 0010 (42.08), 0001 (3.13) Total info=96.61%
Complete profiles (CAP_C6)	1010100010 (11.33), 0010100010 (6.4), 0000000000 (6.38), 1000100010 (5.45), 0000000010 (5.3), 0010000010 (4), 1010000010 (3.85), 0000100010 (3.72), 1010101110 (3.09), 0000100000 (2.48),	0000100010 (17.6), 0000100000 (17.44), 0000000000 (8.6), 0010100000 (8.5), 0010000000 (5.8), 0000000010 (4.4), 0010100010 (4.4), 0000110010 (3.45), 0000110000 (3.34), 0100100010 (3),

	1000000010 (2.45), 1010110010 (2.11), 1011100010 (2.05) Total info=59%	0010000010 (2.66) Total info=79%

Regarding ESCAP profiles, we can say that: (i) in terms of deprivation in EDU, and not in other items, Turkey has the worst performance concerning 36.37 per cent of 18+ population, which is 5 times higher than South Africa; (ii) in terms of non-deprivation in all items, the three countries have similar performance; and (iii) in terms of deprivation in RO, and not in other items, in South Africa it concerns 6 per cent and in Turkey it concerns 8.5 per cent of 18+ population respectively. As we have found before, Turkey has a worse situation in terms of deprivation in EDU and South Africa has a worse situation in terms of deprivation in WATD.

Regarding MSCAP profiles, we can observe that: (i) in terms of deprivation in TEL, South Africa has the worst performance concerning 58 per cent of 18+ population, which is 1.4 times higher than Turkey; and (ii) in terms of non-deprivation in all items, Turkey has the best performance concerning 51 per cent of 18+ population, which is 3.4 times higher than South Africa. And, notably for South Africa, material deprivation in terms of TV and REF, has a considerably worser position compared to Turkey (based on 97 per cent of 18+ Turkey population, there is no deprivation regarding TV and/or REF whilst in South Africa, around 15 per cent of 18+ population suffer deprivation in these two items).

Regarding CAP profiles, there is no relevant results notably due to heterogeneity in South Africa case.

At an approximative level, we can conclude that South Africa has the worst situation regarding WATD, RO, TV, REF, TEL compared to Turkey. On the other hand, Turkey has a considerably worser situation regarding EDU, but also Turkey has a slightly worser situation regarding SAN compared to South Africa.

Comparison between South Africa and Brazil

Seventh comparison model includes the two countries, South Africa, and Brazil. Taking in consideration the common variables (see above), the essential capabilities profile is the combination of the variables WATD, RO, EDU, respectively (ESCAP_C7= {WATD, RO, EDU}). On the other hand, the material situation profile is the combination of the variables TV, REF, TEL, respectively (MSCAP_C7= {TV, REF, TEL}). And, the complete profile is composed of WATD, RO, EDU, TV, REF, TEL, respectively (CAP_C7= {WATD, RO, EDU, TV, REF, TEL}). Each variable represents a functioning such as WATD means access to drinking water. The outputs of functionings are *deprivation* or *non-deprivation*. The situation of deprivation is coded 1 and the situation of non-deprivation is coded 0. For instance, for ESCAP_C7, the profile {111} corresponds to deprivation regarding WATD, RO, EDU at the same time.

Table 101. Essential capabilities and material situation profiles across SA GHS 2011, and BR PNAD 2011

	South Africa (SA GHS 2011) 6.15% mv.	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C7)	111 (30.44), 000 (14), 101 (13.5), 011 (12.65), 001 (9.15),	011 (40.26), 010 (31.31), 111 (12.11), 001 (5.84), 000 (4.6),

	110 (7.75), 010 (7.74), 100 (4.76), Total info=100%	110 (4) Total info=98%
Material situation profiles (MSCAP_C7)	001 (60.45), 000 (15.5), 111 (13), 011 (7.1), 101 (3.5) Total info=99.7%	001 (54.17), 000 (40.4), 011 (2.44) Total info=97%
Complete profiles (CAP_C7)	111001 (17.2), 011001 (9.3), 101001 (7.7), 111111 (7.4), 000000 (7.2), 000001 (6.4), 110001 (5.7), 010001 (5.4), 001001 (5.1), 111011 (3.8), 100001 (3.5), 101111 (3.36), 001000 (3), 011000 (2.15), 010000 (2) Total info=89.44%	011001 (22.6), 010000 (16.8), 011000 (16), 010001 (14), 111001 (8.85), 001001 (2.96), 110001 (2.7), 000000 (2.48), 001000 (2.43), 000001 (2) Total info=91%

Regarding ESCAP profiles, we can say that: (i) in terms of deprivation in WATD, RO, and EDU, South Africa has the worst performance concerning 30.44 per cent of 18+ population, which is 2.5 times higher than Brazil; (ii) in terms of deprivation in RO and EDU, and not in WATD, Brazil has the worst performance concerning 40.26 per cent of 18+ population, which is 3.2 times higher than South Africa; (iii) in terms of deprivation in RO, and not in other items, Brazil has the worst performance concerning 31.3 per cent of 18+ population, which is 4 times higher than South Africa; and (iv) in terms of non-deprivation in all items, South Africa has the best performance concerning 14 per cent of 18+ population, which is 3 times higher than Brazil. In terms of total prosperity, South Africa has the highest score (14) and after Brazil (4.6) come. In terms of 3 deprivations, South Africa has the highest figure (30.44), and Brazil (12.11) comes. In terms of 2 deprivations, Brazil has the highest score (44.26), then South Africa (34) comes. And in terms of 1 deprivation, Brazil has the highest score (37.15), then South Africa comes (21.65).

Regarding MSCAP profiles, we can observe that: (i) in terms of deprivation in TEL and not in other items, South Africa has the worst performance concerning 60.5 per cent of 18+ population, which is 1.1 times higher than Brazil; (ii) in terms of deprivation in REF and TEL, and not in TV, South Africa has the worst performance concerning 7 per cent of 18+ population, which is 2.9 times higher than Brazil; and (iii) in terms of non-deprivation in all items, Brazil has the best performance concerning 40.4 per cent of 18+ population, which is 2.6 times higher than South Africa. In terms of total prosperity, Brazil has the highest score (40.4), then South Africa (15.65) come. In terms of total poverty, South Africa has the highest score (12.98). Besides, in terms of 1 and 2 deprivations cases, South Africa has the highest scores compared to Brazil.

Regarding CAP profiles, we can remark that: (i) in terms of deprivation in WATD, RO, EDU, TEL, and not in other items, South Africa has the worst performance concerning 17.2 per cent of 18+ population, which is 1.94 times higher than Brazil; (ii) in terms of deprivation in RO, EDU, TEL, and not in other items, Brazil has the worst performance concerning 22.6 per cent of 18+ population, which is 2.4 times higher than South Africa; (iii) in terms of deprivation in RO, and not in other items, Brazil has the worst performance concerning 16.8 per cent of 18+ population, which is 5.6 times higher than South Africa; (iv) in terms of deprivation in RO and EDU, Brazil has the worst performance concerning 16 per cent of 18+ population, which is 7.4 time higher than South Africa; (v) in terms of deprivation in RO and TEL, and not in other items, Brazil has the worst performance concerning 14 per cent of 18+ population, which is 2.6 times higher than South Africa.

At an approximative level, we can conclude that South Africa suffer more regarding WATD EDU, TV, REF, and TEL compared to Brazil. On the other hand, Brazil suffers more regarding RO compared to South Africa.

Comparison between South Africa and Mexico

Eight comparison model includes the two countries, South Africa, and Mexico. Taking in consideration the common variables (see above), the essential capabilities profile is the combination of the variables WATD, SAN, KIT, ELC, EDU, MC, NUT, respectively (ESCAP_C8= {WATD, SAN, KIT, ELC, EDU, MC, NUT}). On the other hand, the material situation profile is the combination of the variables TV, RAD, REF, TEL, MPHO, INT, respectively (MSCAP_C8= {TV, RAD, REF, TEL, MPHO, INT}). And, the complete profile is composed of WATD, RO, EDU, TV, REF, TEL, respectively (CAP_C8= {WATD, SAN, KIT, ELC, EDU, MC, NUT, TV, RAD, REF, TEL, MPHO, INT}). Each variable represents a functioning such as WATD means access to drinking water. The outputs of functionings are *deprivation* or *non-deprivation*. The situation of deprivation is coded 1 and the situation of non-deprivation is coded 0.

Table 102. Essential capabilities and material situation profiles across SA GHS 2011, and MX MCS 2012

	South Africa (SA GHS 2011) 8.4% mv.	Mexico (MX MCS 2012)
Essential capabilities profiles (ESCAP_C8)	1000110 (18.72), 0000110 (12.6), 0000000 (9.8), 0000010 (8.67), 1000111 (7.98), 1000010 (5.92), 0000111 (4.05), 1010110 (3.33), 1001110 (3.02), 0000100 (2.59) Total info=76.76%	0000100 (24.15), 0000000 (22.69), 1000100 (11.43), 0000010 (6.36), 0000110 (6.28), 0000101 (3.84), 1000110 (3.17), 1000101 (3.09), 1000000 (2.78) Total info=83.8%
Material situation profiles (MSCAP_C8)	000101 (48.29), 000001 (9), 010101 (7.64), 111101 (5.29), 101101 (5.23), 000000 (5), 001101 (4.2), 000100 (2.27), 011101 (2.26),	000101 (26.32), 000000 (14.89), 000111 (8.77), 000001 (7.63), 000011 (6), 010101 (4.56), 001101 (3.9), 001111 (3.8), 010000 (3.56),

	100101 (2.11) Total info=91.3%	000100 (2.59), 000010 (2.42), 101111 (2.3) Total info=86.85%
Complete profiles (CAP_C8)	-	-

Regarding ESCAP profiles, we can say that: (i) in terms of deprivation in EDU, and not in other items, Mexico has the worst performance concerning 24 per cent of 18+ population, which is 9.3 times higher than South Africa; (ii) in terms of deprivation in WATD, EDU, MC, and not in other items, South Africa has the worst performance concerning 18.7 per cent of 18+ population, which is 5.9 times higher than Mexico; (iii) in terms of deprivation in EDU and MC, and not in other items, South Africa has the worst performance concerning 12.6 per cent of 18+ population, which is 2 times higher than Mexico; (iv) in terms of non-deprivation in all items, Mexico has the best performance concerning 22. per cent of 18+ population, which is 2.3 times higher than South Africa.

Regarding MSCAP profiles, we can observe that: (i) in terms of deprivation in TEL and INT, and not in other items, South Africa has the worst performance concerning 48.3 per cent of 18+ population, which is 1.8 times higher than Mexico; (ii) in terms of deprivation in INT and not in other items, the performance of South Africa is 1.18 times worser than Mexico; and (iii) in terms of non-deprivation in all items, Mexico has the best performance concerning 15 per cent of 18+ population, which is 3 times higher than South Africa.

At an approximative level, we can conclude that South Africa has a worser situation regarding WATD, MC, NUT, TV, RAD, REF, TEL, and INT compared to Mexico.

Comparison between Turkey and Brazil

Ninth comparison model includes the two countries, Turkey, and Brazil. Taking in consideration the common variables (see above), the essential capabilities profile is the combination of the variables WATD, RO, EDU, respectively (ESCAP_C9= {WATD, RO, EDU}). On the other hand, the material situation profile is the combination of the variables TV, REF, TEL, MPHO, WM, CAR respectively (MSCAP_C9= {TV, REF, TEL, MPHO, WM, CAR}). And, the complete profile is composed of WATD, RO, EDU, TV, REF, TEL, MPHO, WM, CAR, respectively (CAP_C9= {WATD, RO, EDU, TV, REF, TEL, MPHO, WM, CAR}). Each variable represents a functioning such as WATD means access to drinking water. The outputs of functionings are *deprivation* or *non-deprivation*. The situation of deprivation is coded 1 and the situation of non-deprivation is coded 0.

Table 103. Essential capabilities and material situation profiles across TR SILC 2011, and BR PNAD 2011

	Turkey (TR SILC 2011)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C9)	001 (51.2), 011 (18.7), 000 (16.8), 010 (10.6), 101 (2.3) Total info:100%	011 (40.26), 010 (31.31), 111 (12.11), 001 (5.84), 000 (4.61), 110 (4) Total info=98%

Material situation profiles (MSCAP_C9)	0010011 (22.57), 0000000 (16), 0000010 (14.5), 0000011 (14), 0010010 (6.22), 0000001 (6), 0010001 (5.5), 0010000 (4.5), 0010111 (2.85), 0001011 (2.5) Total info: 94.68%	0010111 (19), 0000000 (18.32), 0010011 (6.1), 0010000 (6.06), 0000010 (5.34), 0011111 (4.65), 0010101 (3.82), 0010110 (3.79), 0010010 (3.78), 0010001 (3.17), 0000111 (3.16), 0000011 (2.94), 0000001 (2.71), 0010100 (2.32), 0000100 (2.31), 0000110 (2.26) Total info=90%
Complete profiles (CAP_C9)	0010010011 (15.6), 0010000011 (7.11), 0010000010 (7.10), 0010000000 (5.8), 0000000000 (4.9), 0110000011 (4.6), 0000000010 (3.7), 0100000000 (3.6), 0010010001 (3.5), 0110010011 (3.36), 0010010010 (2.97), 0010000001 (2.58), 0110000001 (2.13), 0000010011 (2.09), 0100000010 (2) Total info=71%	-

Regarding ESCAP profiles, we can say that: (i) in terms of deprivation in EDU, and not in other items, Turkey has the worst performance concerning 51.2 per cent of 18+ population, which is 8.76 times higher than Brazil; (ii) in terms of deprivation in RO, and not in other items, Brazil has the worst performance concerning 31 per cent of 18+ population, which is 2.96 times higher than Turkey; (iii) in terms of deprivation in EDU and RO, and not in WATD, Brazil has the worst performance concerning 40 per cent of 18+ population, which is 2.15 times higher than Turkey; (iv) in terms of deprivation in all items, Brazil has the worst performance concerning 12 per cent of 18+ population and such profiles does not exist in Turkey; and (v) in terms of non-deprivation in all items, Turkey has the best performance concerning 16.8 per cent of 18+ population, which is 3.64 times higher than Brazil. In terms of total prosperity, Turkey has the highest score compared to Brazil. In terms of total poverty, Brazil has a score of 12 per cent whilst such case was not found in Turkey. In terms of 1 deprivation, Turkey has the highest score (61.8) compared to Brazil (37.15). In terms of 2 deprivations, Brazil has the highest score (44. 26) compared to Turkey (21).

Regarding MSCAP profiles, we can observe that: (i) in terms of deprivation in TEL, CAR, and COMPINT, and not in other items, Turkey has the worst performance concerning 22.6 per cent of 18+ population, which is 1.2 times higher than Brazil; (ii) in terms of deprivation in CAR,

and not in other items, Turkey has the worst performance concerning 14.5 per cent of 18+ population, which is 2.7 times higher than Brazil; (iii) in terms of deprivation in CAR and COMPINT, and not in other items, Turkey has the worst performance concerning 14 per cent of 18+ population, which is 4.8 times higher than Brazil; and (iv) in terms of non-deprivation in all items, Brazil has the best performance concerning 18 per cent of 18+ population, which is 1.14 times higher than Turkey.

At an approximative level, we can conclude that Brazil has worser performance in WATD, RO, TEL, WM. On the other hand, Turkey has worser performance in EDU, CAR, and COMPINT.

Comparison between Turkey and Mexico

Tenth comparison model includes the two countries, Turkey, and Mexico. Taking in consideration the common variables (see above), the essential capabilities profile is the combination of the variables WATD, SAN1 (bathroom), SAN2 (toilet), KIT, EDU, respectively (ESCAP_C10= {WATD, SAN1, SAN2, KIT, EDU}). On the other hand, the material situation profile is the combination of the variables TV, REF, TEL, MPHO, WM, CAR, COMPINT, AIRC, respectively (MSCAP_C10= {TV, REF, TEL, MPHO, WM, CAR, COMPINT, AIRC}). And, the complete profile is composed of WATD, RO, EDU, TV, REF, TEL, respectively (CAP_C10= {WATD, SAN1, SAN2, KIT, EDU, TV, REF, TEL, MPHO, WM, CAR, COMPINT, AIRC}). Each variable represents a functioning such as WATD means access to drinking water. The outputs of functionings are *deprivation* or *non-deprivation*. The situation of deprivation is coded 1 and the situation of non-deprivation is coded 0.

Table 104. Essential capabilities and material situation profiles across TR SILC 2011, and MX MCS 2012

	Turkey (TR SILC 2011)	Mexico (MX MCS 2012)
Essential capabilities profiles (ESCAP_C10)	00001 (60.56), 00000 (26.8), 00101 (5) Total info:92%	00000 (29.34), 00001 (29), 11001 (14.39), 01001 (6.49), 10001 (4.57), 11011 (2.94), 11000 (2.63) Total info=89.45%
Material situation profiles (MSCAP_C10)	00100111 (20.64), 00000111 (12.58), 00000101 (12.19), 00000001 (11.82), 00100101 (5.18), 00000011 (5.09), 00100011 (4.81), 00000000 (4.24), 00100001 (3.38), 00101111 (2.75), 00010111 (2.33), 00000100 (2.31) Total info: 87%	00100111 (12.65), 00101111 (7.34), 00000001 (7.22), 00111111 (4.93), 00000101 (4.78), 00110111 (4.42), 00000000 (4.22), 00010111 (4), 00000111 (4), 01111111 (3.8), 00100011 (2.9), 01101111 (2.86), 11111111 (2.79), 00100101 (2.64) Total info=68.73%
Complete profiles (CAP_C10)	0000100100111 (14.12), 0000100000111 (9.2), 0000100000101 (7.4), 0000000000001 (5.88), 0000100000001 (5.7), 0000000000101 (4.4),	-

	0000100000011 (3.7), 0000100100011 (3.45), 0000100100101 (3), 0000000000000 (2.6), 0000000100111 (2.5), 0000000100101 (2.04), 0010100100111 (2) Total info=66%	

Regarding ESCAP profiles, we can say that: (i) in terms of deprivation in EDU and not in other items, Turkey has the worst performance concerning 60.56 per cent of 18+ population, which is 2 times higher than Mexico; (ii) in terms of deprivation in WATD, SAN1, and EDU, and not in other items, Mexico has the worst performance concerning 14.4 per cent of 18+ population given that such profile was not found in Turkey; and (iii) in terms of non-deprivation in all items, Mexico has the best performance concerning 29 per cent of 18+ population, which is not very far from the performance of Turkey (26.8 per cent).

Regarding MSCAP profiles, we can observe that: (i) in terms of deprivation in TEL, CAR, COMPINT, and AIRC, and not in other items, Turkey has the worst performance concerning 20.6 per cent of 18+ population, which is 1.6 times higher than Mexico; (iii) in terms of deprivation in CAR, COMPINT, and AIRC, and not in other items, Turkey has the worst performance concerning 12.6 per cent of 18+ population, which is 3.1 times higher than Mexico; (iv) in terms of deprivation in CAR and COMPINT, and not in other items, Turkey has the worst performance concerning 12 per cent of 18+ population, which is 2.55 times higher than Mexico; and (v) in terms of deprivation in AIRC and not in other items, Turkey has the worst performance concerning 11.8 per cent of 18+ population, which is 1.6 times higher than Mexico.

At an approximative level, we can conclude that Turkey has worser performance in EDU, CAR, COMPINT, and AIRC. On the other hand, Mexico has worser performance in WATD, SAN1, WM, REF, and TEL.

Comparison between Mexico and Brazil

Eleventh comparison model includes the two countries, Mexico, and Brazil. Taking in consideration the common variables (see above), the essential capabilities profile is the combination of the variables WATD, EDU, respectively (ESCAP_C11= {WATD, EDU}). On the other hand, the material situation profile is the combination of the variables TV, REF, TEL, respectively (MSCAP_C11= {TV, REF, TEL, MPHO, RAD, WM, CAR, COMPINT}). And, the complete profile is composed of WATD, EDU, TV, REF, TEL, MPHO, RAD, WM, CAR, COMPINT, respectively (CAP_C11= {WATD, EDU, TV, REF, TEL, MPHO, RAD, WM, CAR, COMPINT}). Each variable represents a functioning such as WATD means access to drinking water. The outputs of functionings are *deprivation* or *non-deprivation*. The situation of deprivation is coded 1 and the situation of non-deprivation is coded 0.

Table 105. Essential capabilities and material situation profiles across MX MCS 2012, and BR PNAD 2011

	Mexico (MX MCS 2012)	Brazil (BR PNAD 2011)
Essential capabilities profiles (ESCAP_C11)	01 (37.83), 00 (31.92), 11 (25.11), 10 (5.14) Total info=100%	01 (46.11), 00 (35.92), 11 (13.58), 10 (4.39) Total info=100%

Material situation profiles (MSCAP_C11)	00100011 (12), 00000000 (9), 00100111 (6.66), 00000010 (5.4), 00110111 (4), 00000011 (3.87), 00110011 (3.8), 00010011 (3.5), 00100001 (3.22), 01110111 (3), 00100010 (2.86), 01100111 (2.48), 00001000 (2.4), 11110111 (2.2), 00100000 (2.16), 00101011 (2.15), 00000001 (2.13) Total info=71%	00000000 (18.3), 00100111 (18.14), 00100000 (6), 00100011 (6), 00000010 (5.33), 00110111 (4), 00100101 (3.77), 00100010 (3.76), 00100110 (3.76), 00100001 (3.16), 00000111 (3.14), 00000011 (2.93), 00000001 (2.7), 00100100 (2.3), 00000100 (2.3), 00000110 (2.25) Total info=88%
Complete profiles (CAP_C11)	-	-

Regarding ESCAP profiles, we can say that: (i) in terms of deprivation in EDU and not in WATD, Brazil has the worst performance concerning 46 per cent of 18+ population, which is 1.2 times higher than Mexico; (ii) in terms of deprivation in WATD and EDU, Mexico has the worst performance concerning 25 per cent of 18+ population, which is 1.85 times higher than Brazil; and (iii) in terms of non-deprivation in WATD and EDU, Brazil has slightly better performance concerning 36 per cent of 18+ population, which is 1.125 times higher than Mexico. In terms of total prosperity, Brazil has the highest score (35.92) compared to Mexico (31.92). In terms of total poverty, Mexico has the highest score (25.11) compared to Brazil (13.58). And in terms of 1 deprivation, Brazil has the highest score (50.5) compared to Mexico (42.97).

Regarding MSCAP profiles, we can observe that: (i) in terms of non-deprivation in all items, Brazil has the best performance concerning 18 per cent of 18+ population, which is two times higher than Mexico; (ii) in terms of TEL, CAR, COMPINT, and not in others, Mexico has the worst performance concerning 2 per cent of 18+ population, which is two times higher than Brazil; and (iii) in terms of deprivation in TEL, WM, CAR, and COMPINT, Brazil has the worst performance concerning 18 per cent of 18+ population, which is 3 times higher than Mexico.

At an approximative level, we can conclude that Mexico has worser performance in WATD, EDU, MPHO, CAR, COMPINT. On the other hand, Brazil has worser performance in TEL and WM.

9.8 List of national statistical agencies of select upper-middle-income countries

Table 106. List of national statistical agencies of selected upper-middle income countries (accessed 8 January 2020)

Country	National Statistical Agencies	Web-sites
Argentina	Instituto Nacional de Estadística y Censos	https://www.indec.gob.ar/ftp/cuadros/sociedad/EPH_metodologia_22_pobreza.pdf https://www.indec.gob.ar/indec/web/Nivel3-Tema-4-46 https://www.indec.gob.ar/indec/web/Nivel4-Tema-4-47-156
Azerbaijan	The State Statistical Committee of the Republic of Azerbaijan	https://www.stat.gov.az/source/budget_households/?lang=en
Belize	Statistical Institute of Belize	http://sib.org.bz/wp-content/uploads/2017/05/Poverty_Assessment_Report_2009.pdf
Botswana	Statistics Botswana	http://www.statsbots.org.bw/sites/default/files/publications/BMTHS%20POVERTY%20STATS%20BRIEF%202018.pdf http://www.statsbots.org.bw/poverty http://www.statsbots.org.bw/sites/default/files/publications/BMTHS%20POVERTY%20STATS%20BRIEF%202018.pdf
Brazil	Instituto Brasileiro de Geografia e Estatística (IBGE)	https://www.ibge.gov.br/en/statistics/multi-domain/living-conditions-poverty-and-inequality.html https://sidra.ibge.gov.br/tabela/6682 https://odsbrasil.gov.br/relatorio/sintese
Colombia	DANE	https://www.dane.gov.co/index.php/en/statistics-by-topic/poverty-and-life-conditions https://www.dane.gov.co/index.php/en/statistics-by-topic-1/poverty-and-life-conditions/necesidades-basicas-insatisfechas-nbi
Costa Rica	Instituto Nacional de Estadística y Censos	http://www.inec.go.cr/noticia/pobreza-por-ingresos-se-mantiene-en-210-respecto-ano-anterior http://www.inec.go.cr/sites/default/files/documentos/pobreza_y_presupuesto_de_hogares/pobreza/metodologias/mepobrezaenaho2015-01.pdf http://www.inec.go.cr/objetivos-de-desarrollo-sostenible
Dominican Republic	Oficina Nacional de Estadistica	https://www.one.gob.do/sociales/pobreza-asistencia-social-y-condiciones-de-vida https://www.one.gob.do/publicaciones?ID=209
Ecuador	Instituto Nacional de Estadística y Censos	https://www.ecuadorencifras.gob.ec/pobreza2/ https://www.ecuadorencifras.gob.ec/pobreza-junio-2019/ https://www.ecuadorencifras.gob.ec/objetivos-de-desarrollo-sostenible/
Georgia	National Statistics office of Georgia	https://www.geostat.ge/en/modules/categories/192/living-conditions http://sdg.gov.ge/intro
Jordan	Department of Statistics	http://dosweb.dos.gov.jo/population/poverty/
Lebanon	Central Administrati	http://www.cas.gov.lb/index.php/mdg-en http://www.cas.gov.lb/index.php/demographic-and-social-en/poverty-en

	on for Statistics	
Malaysia	Department of Statistics	https://www.dosm.gov.my/v1/index.php?r=column/cglossary2&menu_id=eWd2VFdIZ2xpdzBmT2Y0a0pweDcwQT09 https://www.dosm.gov.my/v1/index.php?r=column/cone&menu_id=Ukh4bnhIbjl6T1BNQlNDTk9MMGwyUT09
Mexico	Instituto Nacional de Estadística, Geografía (INEGI)	http://www.inegi.org.mx/ https://www.coneval.org.mx/Medicion/MP/Paginas/Pobreza-2018.aspx https://www.coneval.org.mx/Medicion/Paginas/English-version.aspx https://www.coneval.org.mx/Medicion/MP/Paginas/Lineas-de-bienestar-y-canasta-basica.aspx
Paraguay	Direccion general de estadistica, encuestas y censos	https://www.dgeec.gov.py/Publicaciones/Biblioteca/POBREZA-MONETARIA%20-2018/Pobreza%20Monetaria_Boletin.pdf https://www.dgeec.gov.py/Publicaciones/Biblioteca/POBREZA-MONETARIA%20-2018/Pobreza%20Monetaria_presentacion.pdf
Peru	Instituto Nacional de Estadística e Informática (INEI)	https://www.inei.gob.pe/estadisticas/indice-tematico/sociales/ https://www.inei.gob.pe/media/MenuRecursivo/metodologias/pobreza01.pdf
South Africa	Statistics South Africa	http://www.statssa.gov.za/?p=12075 http://www.statssa.gov.za/publications/P03101/P031012019.pdf http://www.statssa.gov.za/publications/Report-03-10-08/Report-03-10-082014.pdf http://www.statssa.gov.za/?page_id=739&id=5
Thailand	National statistical office	http://statbbi.nso.go.th/staticreport/page/sector/en/08.aspx
Turkey	Turkiye Istatistik Kurumu	http://www.turkstat.gov.tr/PreTablo.do?alt_id=1013 http://www.turkstat.gov.tr/PreTablo.do?alt_id=1097

9.9 Quality report of variables, South Africa GHS 2011

Quality report of variables

```
#SA GHS 2011 Data Analysis for PhD thesis
#Author: A. Tugce Beycan (tugce.beycan@unine.ch), Department of Sociology,
University of Neuchatel
#Updated: 14 September 2019

#Open data
library(foreign)

SA2011derived=read.dta("E:\\PhD_Research\\Datas_May2017\\SA_GHS_2011\\Stata
\\DerivedBasedata_SA_GHS2011_11Sept2019_TB.dta")

#Quality report of main and some decomposition analysis variables

library (Hmisc)
describe (SA2011derived [,1:90])

(response.percentage=round(100*prop.table(table(SA2011derived$PRO)),
digits=1))
mylabels=paste("PRO",names(response.percentage),"\n",  response.percentage,
"%", sep=" ")
pie(response.percentage, labels=mylabels)
```

```
boxplot(SA2011derived$Age)
```

Descriptive statistics outputs:

SA2011derived[, 1:90]

90 Variables 47117 Observations

---

hid (household id)

n	missing	distinct
47117	0	20041

lowest: 101000190000001601 101000190000005601 101000190000007601 101000190000011601 101000240000001401

highest: 986010670000010901 986010670000012701 986010670000014501 986010670000016301 986010670000018101

---

id (personal id)

n	missing	distinct
47117	0	47117

lowest: 10100019000000160101 10100019000000160102 10100019000000560101 10100019000000560102 10100019000000560103

highest: 98601067000001270103 98601067000001450101 98601067000001630101 98601067000001810101 98601067000001810102

---

serial (serial number)

n	missing	distinct
47117	0	19

Value	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	19	20
Frequency	19825	14464	7059	3325	1439	556	241	101	45	27	13	8	2	3	2	2	1	3
Proportion	0.421	0.307	0.150	0.071	0.031	0.012	0.005	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

PRO

n	missing	distinct
47117	0	9

Value	1	2	3	4	5	6	7	8	9
Frequency	6032	5619	2671	4290	7595	4359	7178	4502	4871
Proportion	0.128	0.119	0.057	0.091	0.161	0.093	0.152	0.096	0.103

SET

n	missing	distinct
47117	0	4

Value	1	2	4	5
Frequency	27215	2933	15462	1507
Proportion	0.578	0.062	0.328	0.032

GEN

n	missing	distinct
47117	0	2

```
Value         1     2
Frequency  25988 21129
Proportion 0.552 0.448
```

```
POP
      n        missing distinct
  47117        0        4

Value         1     2     3     4
Frequency  36235  5603  1370  3909
Proportion 0.769 0.119 0.029 0.083
```

```
AGE
      n        missing distinct
  47117        0        6

Value         1     2     3     4     5     6
Frequency  10297 10881  8688  7449  5156  4646
Proportion 0.219 0.231 0.184 0.158 0.109 0.099
```

```
Age
      n        missing distinct     Info     Mean      Gmd      .05      .10      .25      .50      .75      .90      .95
      47117        0       88        1    40.05    18.88       19       21       26       37       51       64       72

lowest:  18  19  20  21  22, highest: 102 104 106 110 113
```

```
HS
      n        missing distinct
  47117        0        5

Value         1     2     3     4     5
Frequency  3457  6276  7318  8156 21910
Proportion 0.073 0.133 0.155 0.173 0.465
```

```
hh_child
      n        missing distinct
  47117        0        6

Value         1     2     3     4     5     6
Frequency  15165 10501  9329  5732  3139  3251
Proportion 0.322 0.223 0.198 0.122 0.067 0.069
```

```
GH
      n        missing distinct
  47117        0        2

Value         1     2
Frequency  19924 27193
Proportion 0.423 0.577
```

```
AH
      n        missing distinct
  47117        0        5

Value         1     2     3     4     5
Frequency    11  1195  5211 30956  9744
Proportion 0.000 0.025 0.111 0.657 0.207
```

POPH
n missing distinct
47117 0 4

Value 1 2 3 4
Frequency 36202 5601 1376 3938
Proportion 0.768 0.119 0.029 0.084

HH_NCHSA
n missing distinct
47117 0 2

Value 0 1
Frequency 45014 2103
Proportion 0.955 0.045

INDG
n missing distinct
47117 0 4

Value 1 2 3 99
Frequency 4544 40058 1727 788
Proportion 0.096 0.850 0.037 0.017

DWELS
n missing distinct
47117 0 3

Value 1 2 99
Frequency 8182 38846 89
Proportion 0.174 0.824 0.002

DWE
n missing distinct
47117 0 11

Value 1 2 3 4 5 6 7 8 9 10 11
Frequency 35700 4643 1357 93 252 816 676 1055 2009 503 13
Proportion 0.758 0.099 0.029 0.002 0.005 0.017 0.014 0.022 0.043 0.011 0.000

ROF
n missing distinct
47117 0 12

Value 1 2 3 4 5 6 7 8 9 10 11 12
Frequency 177 667 28346 9 61 28 34 16 9482 9 1425 6863
Proportion 0.004 0.014 0.602 0.000 0.001 0.001 0.001 0.000 0.201 0.000 0.030 0.146

WAL
n missing distinct
47117 0 12

Value 1 2 3 4 5 6 7 8 9 10 11 12
Frequency 32455 6697 2804 648 1 9 1268 203 8 2961 2 61
Proportion 0.689 0.142 0.060 0.014 0.000 0.000 0.027 0.004 0.000 0.063 0.000 0.001

ROFC
n missing distinct
47117 0 5

```
Value         1     2     3     4     5
Frequency  2724  5114 11064 19578  8637
Proportion 0.058 0.109 0.235 0.416 0.183
```

```
WALC
       n  missing distinct
   47117        0        5

Value          1     2     3     4     5
Frequency   2640  5025 10056 20629  8767
Proportion 0.056 0.107 0.213 0.438 0.186
```

```
RO
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25      .50      .75      .90      .95
   47117        0      107    0.993   0.9541   0.7193   0.2222   0.3333   0.5000   0.7500   1.1250   2.0000   2.5000

lowest:  0.00000000  0.07142857  0.08333334  0.08695652  0.09090909, highest:  7.00000000  8.00000000  9.0
0000000 10.00000000 14.00000000
```

```
KIT
       n  missing distinct
   47117        0        2

Value         0     1
Frequency  6398 40719
Proportion 0.136 0.864
```

```
SAN
       n  missing distinct
   47117        0        7

Value         1     2     3     4     5     6     7
Frequency  27181  1609   303  6662  9207   264  1891
Proportion 0.577 0.034 0.006 0.141 0.195 0.006 0.040
```

```
SAN_SH
       n  missing distinct
   47117        0        3

Value         1     2     3
Frequency  4797 40429  1891
Proportion 0.102 0.858 0.040
```

```
SAN_LOC
       n  missing distinct
   47117        0        6

Value         1     2     3     4     5     6
Frequency  20460 23953   720    88     5  1891
Proportion 0.434 0.508 0.015 0.002 0.000 0.040
```

```
WATD
       n  missing distinct
   47117        0       12

Value         1     2     3     4     5     6     7     8     9    10    11    12
Frequency  21677 14025   724   304  1059  5775   402   544  1435   185   319   668
Proportion 0.460 0.298 0.015 0.006 0.022 0.123 0.009 0.012 0.030 0.004 0.007 0.014
```

```
WATDIST
```

```
      n        missing distinct
  47117        0           5

Value          1          2       3        4     5
Frequency   5068  3353  1302   664 36730
Proportion 0.108 0.071 0.028 0.014 0.780
```

```
SW
      n        missing distinct
  47117        0           2

Value          1          2
Frequency  43739  3378
Proportion 0.928 0.072
```

```
CW
      n        missing distinct
  47117        0           2

Value          1          2
Frequency  43504  3613
Proportion 0.923 0.077
```

```
GTW
      n        missing distinct
  47117        0           2

Value          1          2
Frequency  43198  3919
Proportion 0.917 0.083
```

```
FFBSW
      n        missing distinct
  47117        0           2

Value          1          2
Frequency  44265  2852
Proportion 0.939 0.061
```

```
WAT_T
      n        missing distinct
  47117        0           3

Value          1          2       3
Frequency   1766  2539 42812
Proportion 0.037 0.054 0.909
```

```
WATI
      n        missing distinct
  47117        0           3

Value          1          2       3
Frequency  18554 21841  6722
Proportion 0.394 0.464 0.143
```

```
WATI_D
      n        missing distinct
  47117        0           4
```

```
Value         1     2     3     4
Frequency  10450  8104 21841  6722
Proportion 0.222 0.172 0.464 0.143
```

ELC

```
      n        missing distinct
  47117        0           3

Value         1     2     3
Frequency  31287  4772 11058
Proportion 0.664 0.101 0.235
```

ELCCUT

```
      n     missing distinct
  47117        0        8

Value         0     1     2     3     4     5     6     7
Frequency  4772 21661  9274  1988   521   138  7557  1206
Proportion 0.101 0.460 0.197 0.042 0.011 0.003 0.160 0.026
```

ENG_C

```
      n        missing distinct
  47117        0           9

Value         0     1     2     3     4     5     6     7     8
Frequency    44 35961     7  1021  2014  7332   625     7   106
Proportion 0.001 0.763 0.000 0.022 0.043 0.156 0.013 0.000 0.002
```

ENG_H

```
      n  missing distinct
  47117        0       10

Value         0     1     2     3     4     5     6     7     8     9
Frequency  9797 19483     7   939  5197 10025  1546    16    96    11
Proportion 0.208 0.414 0.000 0.020 0.110 0.213 0.033 0.000 0.002 0.000
```

ENG_L

```
      n  missing distinct
  47117        0       10

Value         0     1     2     3     4     5     6     7     8     9
Frequency    17 42767    35     2   836    25    17  3374     5    39
Proportion 0.000 0.908 0.001 0.000 0.018 0.001 0.000 0.072 0.000 0.001
```

ENVP1

```
      n  missing distinct
  47117        0        2

Value         1     2
Frequency  8469 38648
Proportion 0.18  0.82
```

RUB

```
      n  missing distinct
  47117        0        9

Value         1     2     3     4     5     6     7     8     9
Frequency  27231  1086  1020    99   138    13   522 15419  1589
Proportion 0.578 0.023 0.022 0.002 0.003 0.000 0.011 0.327 0.034
```

LIT
n missing distinct
47117 0 2

Value 1 2
Frequency 14149 32968
Proportion 0.3 0.7

WP
n missing distinct
47117 0 2

Value 1 2
Frequency 6243 40874
Proportion 0.132 0.868

AP
n missing distinct
47117 0 2

Value 1 2
Frequency 7788 39329
Proportion 0.165 0.835

LD
n missing distinct
47117 0 2

Value 1 2
Frequency 14626 32491
Proportion 0.31 0.69

NOI
n missing distinct
47117 0 2

Value 1 2
Frequency 7436 39681
Proportion 0.158 0.842

POST
n missing distinct
47117 0 8

Value 1 2 3 4 5 6 7 8
Frequency 24609 7229 2720 1482 2641 1135 417 6884
Proportion 0.522 0.153 0.058 0.031 0.056 0.024 0.009 0.146

TEL
n missing distinct
47117 0 7

Value 1 2 3 4 5 6 7
Frequency 375 36619 7544 1745 649 125 60
Proportion 0.008 0.777 0.160 0.037 0.014 0.003 0.001

INT_1
n missing distinct

```
  47117       0       2

Value          1     2
Frequency   3869 43248
Proportion  0.082 0.918
---------------------------------------------------------------------------------------------------------------------
INT_2
      n  missing distinct
  47117       0       2

Value          1     2
Frequency   1561 45556
Proportion  0.033 0.967
---------------------------------------------------------------------------------------------------------------------
INT_3
      n  missing distinct
  47117       0       2

Value          1     2
Frequency   2544 44573
Proportion  0.054 0.946
---------------------------------------------------------------------------------------------------------------------
INT_4
      n  missing distinct
  47117       0       2

Value          1     2
Frequency   7110 40007
Proportion  0.151 0.849
---------------------------------------------------------------------------------------------------------------------
INT_5
      n  missing distinct
  47117       0       2

Value          1     2
Frequency   1817 45300
Proportion  0.039 0.961
---------------------------------------------------------------------------------------------------------------------
INT_6
      n  missing distinct
  47117       0       2

Value          1     2
Frequency   1406 45711
Proportion  0.03  0.97
---------------------------------------------------------------------------------------------------------------------
INT_7
      n  missing distinct
  47117       0       2

Value          1     2
Frequency   9927 37190
Proportion  0.211 0.789
---------------------------------------------------------------------------------------------------------------------
TV
      n  missing distinct
  47117       0       2

Value          1     2
Frequency  39633  7484
```

Proportion 0.841 0.159

RAD

n	missing	distinct
47117	0	2

Value	1	2
Frequency	37972	9145
Proportion	0.806	0.194

REF

n	missing	distinct
47117	0	2

Value	1	2
Frequency	38139	8978
Proportion	0.809	0.191

ADN

n	missing	distinct
47117	0	6

Value	1	2	3	4	5	6
Frequency	38728	2569	4335	1017	356	112
Proportion	0.822	0.055	0.092	0.022	0.008	0.002

CHN

n	missing	distinct
47117	0	6

Value	1	2	3	4	5	6
Frequency	27134	1856	3054	661	249	14163
Proportion	0.576	0.039	0.065	0.014	0.005	0.301

NU1

n	missing	distinct
47117	0	2

Value	1	2
Frequency	11148	35969
Proportion	0.237	0.763

NU2

n	missing	distinct
47117	0	2

Value	1	2
Frequency	10023	37094
Proportion	0.213	0.787

NU3

n	missing	distinct
47117	0	2

Value	1	2
Frequency	8263	38854
Proportion	0.175	0.825

NU4

n	missing	distinct

```
  47117       0       2

Value         1           2
Frequency  10303 36814
Proportion   0.219 0.781
---------------------------------------------------------------------------------------------------------------------------
EDU
      n  missing distinct
  47117       0      14

Value           0    1      2      3      4        5       6    7        8      9    10        11    12     13
Frequency     542   49  4313  2762  7350 10818   123   135   232   203 14135   539  1446  4470
Proportion  0.012 0.001 0.092 0.059 0.156 0.230 0.003 0.003 0.005 0.004 0.300 0.011 0.031 0.095
---------------------------------------------------------------------------------------------------------------------------
MC
      n  missing distinct
  47117       0       2

Value         1           2
Frequency   7997 39120
Proportion  0.17  0.83
---------------------------------------------------------------------------------------------------------------------------
DIS1
      n  missing distinct
  47117       0       3

Value         1           2       99
Frequency   1568 45324   225
Proportion   0.033 0.962 0.005
---------------------------------------------------------------------------------------------------------------------------


DIS2
      n  missing distinct
  47117       0       5

Value         1           2       3      4        99
Frequency   1806   964  4140 39580   627
Proportion   0.038 0.020 0.088 0.840 0.013
---------------------------------------------------------------------------------------------------------------------------
HWC
      n  missing distinct
   5415    41702       7

Value          0        1      2     3       4      5    99
Frequency   4180    55    14   279   795     6    86
Proportion   0.772 0.010 0.003 0.052 0.147 0.001 0.016
---------------------------------------------------------------------------------------------------------------------------
CILM_1
      n  missing distinct
   1345    45772       3

Value         1           2      99
Frequency   1166   135    44
Proportion   0.867 0.100 0.033
---------------------------------------------------------------------------------------------------------------------------
CILM_2
      n  missing distinct
   2029    45088       3
```

```
Value         1     2    99
Frequency  1913    80    36
Proportion 0.943 0.039 0.018
```

CILM_3

```
      n  missing distinct
    195    46922        3

Value         1     2    99
Frequency   157    35     3
Proportion  0.805 0.179 0.015
```

CILM_4

```
      n  missing distinct
    948    46169        3

Value         1     2    99
Frequency   816    96    36
Proportion  0.861 0.101 0.038
```

CILM_5

```
      n  missing distinct
   6276    40841        3

Value         1     2     99
Frequency  5875   272   129
Proportion 0.936 0.043 0.021
```

CILM_6

```
      n  missing distinct
   1826    45291        3

Value         1     2     99
Frequency  1543   223    60
Proportion 0.845 0.122 0.033
```

TIMEH2

```
      n  missing distinct
  47117        0       13

Value         1     10    11    12     2     3     4     5     6     7     8     9    99
Frequency 38019     42    82   307  7669    18   153    35    35   198    73     9   477
Proportion 0.807 0.001 0.002 0.007 0.163 0.000 0.003 0.001 0.001 0.004 0.002 0.000 0.010
```

DISG

```
      n  missing distinct
  47117        0        4

Value         1     2     3    99
Frequency  1630  5551 39877    59
Proportion 0.035 0.118 0.846 0.001
```

OLDG

```
      n  missing distinct
  47117        0        4

Value         1     2     3    99
Frequency  5283  1885 39877    72
Proportion 0.112 0.040 0.846 0.002
```

WARG
n missing distinct
47117 0 4

Value 1 2 3 99
Frequency 11 7173 39877 56
Proportion 0.000 0.152 0.846 0.001

FOSG
n missing distinct
47117 0 4

Value 1 2 3 99
Frequency 119 7065 39877 56
Proportion 0.003 0.150 0.846 0.001

DEPG
n missing distinct
47117 0 4

Value 1 2 3 99
Frequency 40 7144 39877 56
Proportion 0.001 0.152 0.846 0.001

CHSG
n missing distinct
47117 0 3

Value 2 3 99
Frequency 7183 39877 57
Proportion 0.152 0.846 0.001

AIDG
n missing distinct
47117 0 4

Value 1 2 3 99
Frequency 6 7177 39877 57
Proportion 0.000 0.152 0.846 0.001

SRG
n missing distinct
47117 0 4

Value 1 2 3 99
Frequency 10 7169 39877 61
Proportion 0.000 0.152 0.846 0.001

ES1
n missing distinct
47117 0 8

Value 1 2 3 4 5 6 7 99
Frequency 18676 5589 183 5244 2266 13124 707 1328
Proportion 0.396 0.119 0.004 0.111 0.048 0.279 0.015 0.028

SEC
n missing distinct
47117 0 4

Value	0	1	2	99
Frequency	28441	15019	3480	177
Proportion	0.604	0.319	0.074	0.004