Université de Neuchâtel
Faculté de droit et des sciences économiques

Estimation de la variance de l'indice suisse des prix à la consommation à l'aide de techniques empiriques, sous l'hypothèse d'échantillons aléatoires

Thèse
présentée à la Faculté de droit et des sciences économiques
pour obtenir le grade de docteur ès sciences économiques

par
Sandrine König

Neuchâtel
1995

Imprimerie de l'Evole SA
Neuchâtel

Madame Sandrine König est autorisée à imprimer sa thèse de doctorat ès sciences économiques intitulée:

"Estimation de la variance de l'indice suisse des prix à la consommation à l'aide de techniques empiriques, sous l'hypothèse d'échantillons aléatoires".

Elle assume seule la responsabilité des opinions énoncées.

Neuchâtel, le 20 février 1995

Le Doyen
de la Faculté de droit
et des sciences économiques

Daniel Haag

A mes parents

A Jean-Claude

Préface

Les économistes analysent la santé économique d'un pays notamment au travers des trois indicateurs économiques suivants: le produit national brut, le taux de chômage et le taux d'inflation. Ce dernier se mesure grâce à l'indice des prix à la consommation et il est aussi utilisé pour l'indexation des salaires et des rentes (par exemple, en Suisse, les rentes AVS) et pour déflater d'autres statistiques économiques. Il s'avère donc essentiel que la méthode de construction et de calcul de l'indice des prix à la consommation soit fiable d'un point de vue statistique et qu'au niveau politique il bénéficie d'un large consensus et soit accepté par tous les partenaires sociaux d'un pays. L'indice suisse des prix à la consommation (ISPC) est calculé mensuellement par l'Office fédéral de la statistique (OFS).

L'ISPC est calculé selon l'indice de Laspeyres. Celui-ci mesure, par rapport à une période de base, l'évolution du niveau général des prix des biens et services pour une population de référence. L'indice tient compte de l'importance des différents articles dans le budget des ménages. Le panier de la ménagère, ou panier-type, est composé des biens et services consommés par la population de référence durant la période de base. Nous remarquons que pour construire et calculer un IPC, il faut connaître deux types d'informations. Premièrement, il faut savoir quelle est la part du budget des ménages qui est consacrée à la consommation des différents articles. Deuxièmement, on doit connaître le prix des biens et services pour la période de base et pour la période courante. Or, nous constatons qu'il n'est pratiquement pas possible d'examiner le prix de tous les biens et services, ni d'observer la consommation de tous les ménages d'un pays. Ces deux éléments doivent donc être obtenus en utilisant des échantillons. Le plan d'échantillonnage est ainsi composé d'un échantillon de ménages, de zones géographiques, de points de vente, d'articles et d'habitations. On retrouve ce plan pour tous les IPC calculés dans les pays industrialisés. L'échantillon de ménages participe à une enquête sur la consommation dont les résultats servent à estimer les coefficients budgétaires des articles. L'échantillon de zones géographiques définit les régions à l'intérieur desquelles un échantillon de points de vente est sélectionné. Au sein de ces derniers, on échantillonne les articles pour lesquels les prix sont relevés. Enfin, pour estimer l'indice du loyer du logement, il est nécessaire de disposer d'un échantillon d'habitations.

Lorsque le statisticien établit une statistique basée sur un échantillon, il veut pouvoir estimer la précision de l'estimateur calculé. Ainsi, on mesure l'erreur due à l'échantillonnage par l'estimation de la variance. Cependant, cette dernière ne peut être calculée que si l'échantillon a été sélectionné avec les méthodes d'échantillonnage aléatoires.
Il faut noter que les échantillons entrant dans la construction de l'indice suisse des prix à la consommation ne sont pas tous sélectionnés aléatoirement. De ce fait, il n'est pas possible de calculer la variance de l'ISPC. Remarquons que les Etat-Unis sont, à l'heure actuelle, le seul pays où la construction de l'indice des prix à la consommation repose sur un ensemble d'échantillons aléatoires et où l'on estime la variance de l'IPC. Au niveau international, le cas américain est considéré comme le modèle de référence.

Les objectifs principaux de cette thèse sont de poser les principes permettant de déboucher sur un plan d'échantillonnage aléatoire pour l'ISPC et de proposer une méthode, adaptée au cas suisse, qui permette l'estimation de la précision de l'ISPC. Par précision de l'ISPC, nous comprenons deux éléments: premièrement le calcul de la variance de l'ISPC et deuxièmement le calcul de la variance du changement de prix de l'indice. Dans le premier cas, nous voulons connaître l'intervalle de confiance à l'intérieur duquel la vraie valeur de l'IPC se situe pour une probabilité donnée. Et dans le second cas, notre objectif est de savoir si la variation de l'IPC entre deux périodes (comme deux mois consécutifs, ou deux mêmes mois à un an d'intervalle) est significative ou non. C'est-à-dire si la variation de l'IPC est due à une variation réelle des prix ou si cette variation n'est attribuable qu'aux aléas de l'échantillonnage.

Enfin, nous nous proposons d'apporter quelques réflexions sur l'utilisation de la médiane dans le cadre de la statistique des prix.

Dans le premier chapitre, nous abordons les concepts de base qui sont nécessaires à la compréhension des éléments développés ultérieurement au cours de cette thèse. Nous examinons l'historique, la définition et les propriétés des indices statistiques. Nous présentons aussi les différentes techniques d'échantillonnage, comme les méthodes à choix raisonné et les méthodes aléatoires. Le deuxième chapitre est consacré à l'étude descriptive de l'indice suisse des prix à la consommation. Puis, l'IPC américain est présenté au chapitre 3. Nous avons examiné avec attention le cas américain puisqu'il représente le modèle de référence selon les recommandations internationales. Cette étude nous a permis d'élaborer les propositions que nous faisons pour l'ISPC. Les suggestions quant à la sélection aléatoire des échantillons sont exposées dans le quatrième chapitre. Puis, au chapitre suivant, nous présentons notre méthodologie pour l'estimation de la variance. Celle-ci a été testée sur les données de Genève. Enfin, au chapitre 6, nous apportons quelques réflexions sur l'usage de la médiane dans le calcul de l'indice des prix à la consommation et ceci afin de suggérer une alternative possible à l'utilisation de la moyenne arithmétique.

Relevons que l'Office fédéral de la statistique (1993) ainsi que la Commission de statistique conjoncturelle et sociale (1992) considèrent la mise sur pied d'un plan d'échantillonnage aléatoire ainsi que la mesure des erreurs dues à l'échantillonnage comme des thèmes qu'il conviendra d'étudier lors des futures révisions de l'indice suisse des prix à la consommation.

Remerciements:
Nous adressons nos remerciements au Fonds national suisse de la recherche scientifique qui, dans le cadre du projet no 12-27837.89 *"Mesure de la variance de l'indice des prix à la consommation en Suisse"*, nous a accordé un statut de doctorante.

Pour le soutien et les conseils qui nous ont permis d'élaborer cette thèse, nous exprimons nos remerciements et notre gratitude à Monsieur le Professeur Yadolah Dodge directeur de thèse, à Monsieur le Professeur Claude Jeanrenaud corapporteur, ainsi qu'à Monsieur Dr. Fahrad Mebran deuxième corapporteur, chef du Bureau des statistiques du Bureau International du Travail (BIT) à Genève.

Nos remerciement s'adressent aussi au Service Cantonal de Statistique de Genève qui nous a fourni les données numériques, à Monsieur Rietschin responsable du secteur économie et à Madame Davaudet pour leur précieuse collaboration.

L'Office statistique de la ville de Zurich et Madame Pelli sont également remerciés pour les données qui nous ont été fournies.

Nous témoignons notre reconnaissance à Madame Séverine Pfaff pour sa collaboration sur le projet du FNRS, à Madame Nicole Vauthier et Monsieur Michel Chapuis pour leur travail de relecture.

Table des matières

Chapitre 1

Concepts de base

1.1 Introduction

Le but de ce chapitre est d'exposer les concepts de base qui sont utilisés tout au long de ce travail. L'indice des prix à la consommation (IPC) est un indice statistique. On expose dans la section suivante l'historique des indices, puis, dans la section 1.3, leurs définitions et propriétés. La construction d'un IPC nécessite de sélectionner plusieurs échantillons. Ces derniers peuvent être choisis par différentes méthodes de sondage, comme la méthode par choix raisonné ou la méthode aléatoire; nous les décrivons dans la section 1.4. A la suite de tout sondage, la statistique estimée fournit un résultat différent de celui obtenu si l'on avait pu observer toute la population. Cette différence comprend deux types d'erreurs que nous expliquons dans la section 1.5. Enfin, la dernière section est consacrée à l'erreur quadratique moyenne, cette dernière permettant de mesurer l'erreur totale d'une statistique.

1.2 Historique des indices statistiques

Commençons cette section par un historique des indices statistiques qui est basé sur un article de M.G. Kendall (1969). Cela nous permettra

de comprendre le contexte dans lequel les indices se sont développés.
Dès le XVIIIème siècle, on s'est intéressé à étudier l'évolution des prix, voir W. Fleetwood (1707), à l'évolution de la monnaie, voir C. Dutot (1738) et G.R. Carli (1764). Ces savants, en étudiant le changement relatif de variables (le prix de diverses marchandises) entre deux périodes de temps, ont établi la notion d'indice statistique.
W. Fleetwood (1707), dans son ouvrage "Chronicon Preciosum", étudia le cas suivant. Un collège anglais, fondé entre 1440 et 1460, inscrivit dans ses statuts une clause qui prévoyait que toute personne admise comme membre du collège devait jurer de le quitter si sa fortune personnelle annuelle dépassait £5. W. Fleetwood se demanda si en l'an 1700 on pouvait encore prêter le même serment, étant donné la perte de valeur de la monnaie durant la période écoulée. Il essaya de déterminer quelle somme de monnaie était nécessaire en 1700 pour acheter la même quantité de blé, de viande, de boissons et de tissus que l'on pouvait acquérir en 1440-1460 avec £5. Sa conclusion fut qu'il fallait disposer d'environ £30 en 1700 pour obtenir les mêmes quantités de produits qu'en 1440-1460.
C. Dutot (1738) étudia la diminution de la valeur de la monnaie au travers du revenu de deux souverains, Louis XV et Louis XII. Les revenus de ces rois s'établissaient respectivement à £100'000'000 en 1735 et £7'650'000 en 1515. Dutot se demanda alors lequel des deux rois disposait, en termes réels, du plus grand revenu. Afin de répondre à cette interrogation, il releva les prix, pour ces deux époques, d'un certain nombre de biens et de services; notamment le prix d'une oie, d'un poulet, d'un lapin, d'un journalier. Il additionna ainsi les prix observés pour chaque ensemble de données et divisa ces deux groupes. Il en conclut que Louis XII fut le plus riche en termes réels.
Gian Rinaldo Carli (1764), qui fut professeur d'astronomie à Padoue, antiquaire, historien et économiste à Milan, s'intéressa plus particulièrement à la dépréciation de la valeur de la monnaie depuis la découverte de l'Amérique. Il releva les prix des grains, du vin et de l'huile en 1500 et en 1750. Puis, il construisit la mesure suivante:

$$I = \frac{1}{n} \cdot \sum \frac{P_1}{P_0}$$

L'indice ainsi obtenu est une moyenne arithmétique des prix relatifs de chaque produit observé.

Bien d'autres auteurs contribuèrent au développement de la théorie des indices statistiques, tel Sir George Shuckburgh Evelyn (1798), ou encore J. Lowe (1822) qui s'intéressa aux fluctuations de la valeur de la monnaie, et fut le premier à considérer dans son analyse l'importance de chaque bien dans la consommation domestique; cette importance relative est aujourd'hui appelée, dans la théorie des indices statistiques, le coefficient de pondération.

Au cours du XIXème siècle, plusieurs chercheurs allemands, comme E. Laspeyres (1864, 1871), professeur de sciences économiques à l'Université de Giessen, H. Paasche (1874) et M.W. Drobisch (1871), firent entre autres des travaux sur l'évolution du niveau des prix à Hambourg; ces trois auteurs rejetèrent les méthodes qui ne tenaient pas compte de l'importance relative de chaque bien. I. Fisher (1936), au début de ce siècle, réalisa également des études sur les indices statistiques. Ces auteurs influencèrent la théorie des indices statistiques; en effet, aujourd'hui encore, les formules qu'ils ont proposées sont celles qui sont principalement utilisées. Notons que l'indice des prix à la consommation est calculé avec l'indice de Laspeyres, et cela non seulement en Suisse mais aussi dans les autres pays. Pour terminer cette section consacrée à l'histoire des indices statistiques, nous pouvons citer une phrase d'Edgeworth (1925) qui est devenue classique dans la définition des indices:

"I proposed to define an index-number as a number adapted by its variations to indicate the increase or decrease of a magnitude not susceptible of accurate measurement." 1925, "The Quality of Index Numbers", Economic Journal, 35, pp.379-388.

1.3 Les indices statistiques

Un indice, au sens le plus large, est une mesure du changement relatif d'une variable étudiée entre deux situations. Ces situations peuvent être deux périodes de temps, tels que deux mois, deux années, ou deux points de l'espace, comme deux régions d'un pays, deux communes. Au cours de notre étude, on se concentrera sur les indices portant sur deux périodes de temps. Nous verrons qu'il existe deux catégories d'indices statistiques. La première mesure la variation relative d'une variable entre deux situations; on parle alors d'indice statistique élémentaire. La seconde catégorie permet de résumer, en un indice unique, un ensemble d'indices statistiques élémentaires. Cet indicateur est un indice statistique synthétique. Commençons par examiner les indices statistiques élémentaires.

1.3.1 Les indices statistiques élémentaires

Les indices statistiques élémentaires mesurent la variation relative d'une variable entre deux situations.
Considérons une variable X qui prend différentes valeurs

$$x^0, x^1, \ldots, x^t$$

à différentes dates ou périodes successives $0, 1, \ldots, t$. Alors x^j dénote la valeur prise par la variable X à la période $j, j \in (0, 1, \ldots, t)$. On peut définir un indice statistique élémentaire de la manière suivante:

Définition 1.1 On appelle indice statistique élémentaire d'une variable X entre les périodes 0 et t, noté $i^{t/0}$, le ratio suivant:

$$i^{t/0} = \frac{x^t}{x^0}$$

où t représente la date ou période courante et 0 la date ou période de base ou de référence.

L'indice élémentaire peut être exprimé en pourcentage, on dit alors que l'indice $i^{t/0}$ est exprimé base 100 à la date de référence 0:

$$I^{t/0} = i^{t/0} \cdot 100 = \frac{x^t}{x^0} \cdot 100$$

Examinons maintenant deux propriétés importantes des indices statistiques élémentaires. Il s'agit des propriétés de circularité et de réversibilité.

• **La circularité**

Cette propriété permet d'affirmer qu'étant donné deux périodes 0 et t, l'indice calculé pour le temps t sur la base 0 peut également être obtenu par l'intermédiaire d'indices portant sur d'autres périodes que t et 0. Soit trois situations 0, t', t, on a:

$$i^{t/0} = i^{t/t'} \cdot i^{t'/0}$$

La démonstration vient de la définition même des indices élémentaires:

$$\frac{x^t}{x^0} = \frac{x^t}{x^{t'}} \cdot \frac{x^{t'}}{x^0}$$

Ainsi on peut comparer la valeur des indices élémentaires aux dates t et t' grâce au quotient des indices $i^{t/0}$ et $i^{t'/0}$, mais aussi sur la valeur des variables elles-mêmes, car

$$i^{t/t'} = \frac{i^{t/0}}{i^{t'/0}} = \frac{x^t}{x^0} \cdot \frac{x^0}{x^{t'}} = \frac{x^t}{x^{t'}}$$

On peut généraliser la propriété de circularité pour une suite d'indices:

$$i^{t/0} = i^{t/t-1} \cdot i^{t-1/t-2} \cdot \ldots \cdot i^{2/1} \cdot i^{1/0}$$

On forme ainsi une chaîne d'indices. La valeur $i^{t/0}$ est identique à celle qui serait obtenue par un calcul direct

$$i^{t/0} = \frac{x^t}{x^0}$$

• **La réversibilité**

Cette propriété permet d'affirmer que si on inverse la période de base et la période courante, le nouvel indice a la valeur réciproque de l'ancien. Soit $i^{0/t}$ un indice élémentaire, base de référence t et date courante 0, et $i^{t/0}$ un indice élémentaire base de référence 0 et période courante t. On a alors:

$$i^{0/t} = \frac{1}{i^{t/0}}$$

En effet,

$$\frac{x^0}{x^t} = \frac{1}{\frac{x^t}{x_0}}$$

La réversibilité garantit l'absence de distorsions puisque l'on peut retourner à la valeur de base depuis n'importe quelle période de temps.

1.3.2 Les indices statistiques synthétiques

Jusqu'à présent, seuls les indices statistiques élémentaires, qui mesurent l'évolution relative entre deux situations d'une variable, ont été présentés. Cependant, on souhaite souvent obtenir un seul indicateur qui reflète la variation relative de plusieurs variables qui peuvent être des périodes de temps ou des lieux géographiques.
Imaginons par exemple que l'on veuille mesurer l'évolution du niveau général des prix. On peut considérer ce dernier comme un ensemble de variables dont chacune représente le prix d'un article déterminé. A partir des valeurs de chaque variable, pour deux situations, on peut calculer les indices élémentaires. On dispose d'un indice élémentaire pour chaque variable. Le but est alors de les résumer en un seul indicateur que l'on appelle indice synthétique. En fonction de la méthode d'agrégation utilisée, on obtient différents types d'indices synthétiques. Dans cette section, on va étudier les trois types d'indices synthétiques les plus utilisés en pratique. Il s'agit de l'indice de Laspeyres, de l'indice de Paasche et de l'indice de Fisher.On peut définir les indices synthétiques de la façon suivante:

Définition 1.2 Soit X un ensemble de variables, chaque variable X_j pouvant prendre les valeurs:

$$(x_1^0, \ldots, x_1^t), \ldots, (x_j^0, \ldots, x_j^t), \ldots, (x_n^0, \ldots x_n^t)$$

à différentes dates ou périodes successives.
Les indices élémentaires de chaque x_j se calculent en faisant le rapport entre la valeur x_j au temps t et la valeur de x_j au temps 0. On a:

$$i_j^{t/0} = \frac{x_j^t}{x_j^0}$$

Pour obtenir un indice synthétique, il faut résumer les différents indices élémentaires $i_1^{t/0}, \ldots, i_n^{t/0}$ en un seul indice. Ci-dessous, on verra qu'il existe différentes méthodes pour résumer les indices élémentaires en un indice synthétique.

• Les formules d'indices

Les indices synthétiques les plus utilisés en pratique sont l'indice de Laspeyres, l'indice de Paasche et l'indice de Fisher; ceux-ci sont exposés dans les paragraphes suivants.
Avant de passer à la présentation de ces derniers, il faut tout d'abord expliquer la notion de coefficient de pondération. Ces derniers sont utilisés lors du calcul des trois indices synthétiques mentionnés ci-dessus, car chacune de ces trois méthodes est une moyenne pondérée. Pour construire ces moyennes, on associe à chaque individu x_j un poids correspondant à son importance relative. Celle-ci, appelée coefficient de pondération, est exprimée par w_j; w_j^0 représente l'importance relative de l'individu x_j à la date de référence 0 et w_j^t l'importance relative de x_j à la date courante t.
Si X représente le niveau général des prix et x_j celui du lait, alors w_j^0 représente la part du budget des ménages consacrée au lait à la période de référence 0. Notons que la somme des coefficients de pondération est égale à 1:

$$\sum_{j=1}^{n} w_j = 1$$

• La formule de l'indice de Laspeyres

Définition 1.3 L'indice de Laspeyres (IL) est la moyenne arithmétique des indices élémentaires pondérés par leur coefficient w_j^0 de la date de référence.

$$IL^{t/0} = \frac{\sum_{j=1}^{n} w_j^0 \cdot I_j^{t/0}}{\sum_{j=1}^{n} w_j^0} = \sum_{j=1}^{n} w_j^0 \cdot \frac{x_j^t}{x_j^0} \cdot 100$$

• La formule de l'indice de Paasche

Définition 1.4 L'indice de Paasche (IP) se définit comme la moyenne harmonique des indices élémentaires pondérés par les coefficient w_j^t de la période courante.

$$\begin{aligned} IP^{t/0} &= \frac{1}{\sum_{j=1}^{n} \frac{w_j^t}{i_j^{t/0}}} \cdot 100 \\ &= \frac{1}{\sum_{j=1}^{n} w_j^t \cdot \frac{x_j^0}{x_j^t}} \cdot 100 \end{aligned}$$

Remarquons que les coefficients de pondération sont ceux de la période courante et qu'ils varient lors de chaque changement de période.

• La formule de l'indice de Fisher

Définition 1.5 L'indice de Fisher (IF) est la moyenne géométrique des indices de Laspeyres et de Paasche.

$$IF^{t/0} = \sqrt{IL^{t/0} \cdot IP^{t/0}}$$

Les formules de Laspeyres et de Paasche aboutissent à des résultats différents. Selon B. Grais (1979), aucune des deux formules n'est plus vraie que l'autre, d'où l'idée de calculer une moyenne entre les deux indices exposés précédemment.

• Les types d'indices

Avec un indice synthétique, on peut observer trois types de variations: des variations de valeur, de prix ou de quantité. Dans le cadre de ce travail, on se penchera uniquement sur l'indice de prix, puisqu'il concerne directement l'indice des prix à la consommation. Un indice de prix peut être calculé avec l'une des trois formules exposées ci-dessus, à savoir celle de Laspeyres, de Paasche ou de Fisher.
Examinons successivement l'indice de Laspeyres des prix, l'indice de Paasche des prix et l'indice de Fisher des prix.

•L'indice de Laspeyres des prix (IL_p)

Nous avons vu que la formule d'un indice de Laspeyres est:

$$IL^{t/0} = \sum_{j=1}^{n} w_j^0 \cdot \frac{x_j^t}{x_j^0} \cdot 100$$

Pour le cas d'un indice de Laspeyres des prix, examinons ce que signifie chacun des éléments de la formule et comment on peut réécrire l'IL des prix.
Les coefficients de pondération w_j^0, appelés, dans le cas d'un indice des prix, coefficients budgétaires, sont déterminés par la part de la dépense totale des ménages qui est consacrée à la consommation des
biens et services au cours de la période de base (ou période de référence).
Les coefficients budgétaires peuvent donc s'exprimer par:

$$w_j^0 = \frac{q_j^0 \cdot p_j^0}{\sum_{j=1}^{n} q_j^0 \cdot p_j^0}$$

où p_j^0 est le prix de l'article j au temps 0 et q_j^0 est la quantité consommée de l'article j au temps 0.

On peut reformuler l'IL_p comme étant le rapport entre la dépense totale de la période de référence estimée au prix de la période courante et la dépense totale de la période de base:

$$\begin{aligned} IL_p^{t/0} &= \frac{\sum_{j=1}^{n} q_j^0 \cdot p_j^0 \cdot \frac{p_j^t}{p_j^0}}{\sum_{j=1}^{n} q_j^0 \cdot p_j^0} \\ &= \frac{\sum_{j=1}^{n} q_j^0 \cdot p_j^t}{\sum_{j=1}^{n} q_j^0 \cdot p_j^0} \end{aligned}$$

•L'indice de Paasche des prix (IP_p)

Nous connaissons la formule d'indice de Paasche:

$$IP^{t/0} = \frac{1}{\sum_{j=1}^{n} \frac{w_j^t}{i_j^{t/0}}} \cdot 100$$

$$= \frac{1}{\sum_{j=1}^{n} w_j^t \frac{x_j^0}{x_j^t}} \cdot 100$$

Voyons la signification des divers éléments pour un indice de Paasche des prix. Les coefficients budgétaires sont déterminés par la part de la dépense totale des ménages qui est consacrée à la consommation des différents articles durant la période courante:

$$w_j^t = \frac{q_j^t \cdot p_j^t}{\sum_{j=1}^{n} q_j^t \cdot p_j^t}$$

où p_j^t est le prix de l'article j au temps t et q_j^t est la quantité consommée de l'article j au temps t.
Nous pouvons formuler l'IP_p comme étant le rapport entre la dépense totale de la période courante et la dépense totale de la période courante estimée au prix de l'année de base.

$$\begin{aligned} IP_p^{t/0} &= \frac{\sum_{j=1}^{n} q_j^t \cdot p_j^t}{\sum_{j=1}^{n} q_j^t \cdot p_j^t \cdot \frac{p_j^0}{p_j^t}} \\ &= \frac{\sum_{j=1}^{n} q_j^t \cdot p_j^t}{\sum_{j=1}^{n} q_j^t \cdot p_j^0} \end{aligned}$$

•L'indice de Fisher des prix (IF_p)

Pour l'indice de Fisher des prix, IF_p, il s'agit de multiplier l'indice de Laspeyres des prix IL_p et l'indice de Paasche des prix IP_p, puis d'en prendre la racine carrée:

$$IF_p^{t/0} = \sqrt{IL_p^{t/0} \cdot IP_p^{t/0}}$$

• **Les raccords d'indices et les indices-chaînes**

En pratique, un indice est renouvelé régulièrement. Lors de ces changements, on se trouve confronté au problème de l'estimation de l'évolution de l'indice sur une longue période où plusieurs indices, avec des bases différentes, se sont succédés.
Soit un indice I, base 100 à la période t_0, calculé jusqu'à la période t_i, période où il est remplacé par un indice I', base 100 à la période t_e. Pour juger de l'évolution de l'indice entre t_0 et t_j, période postérieure à t_i, on est conduit à estimer la valeur qu'aurait prise l'indice I à la période t_j en faisant un raccord d'indices:

$$I^*_{j/0} = I'_{j/e} \cdot \frac{I_{g/0}}{I'_{g/e}}, \mathrm{e} \leq g \leq i$$

Notons que cette formule repose sur la propriété de circularité de l'indice; nous avons vu ci-dessus que cette propriété n'est pas vérifiée pour les indices de Laspeyres et de Paasche; il s'agit donc d'une estimation. Le rapport $\frac{I_{g/0}}{I'_{g/e}}$ correspond au coefficient de raccordement. Il s'agit du coefficient par lequel il faut multiplier le nouvel indice pour obtenir une estimation de la valeur qu'aurait atteinte l'ancien indice si on avait continué de le calculer. Le coefficient de raccordement peut être calculé à partir de n'importe quelle période g pour laquelle l'ancien et le nouvel indice ont été calculés. On remarquera cependant que c'est le plus souvent la dernière période, pour laquelle on a encore le dernier indice, qui sert de base. Si la période de raccordement est identique à celle de la période de base du nouvel indice, le raccord d'indices s'écrit alors:

$$I^*_{j/0} = I'_{j/e} \cdot \frac{I_{e/0}}{100}$$

Les indices-chaînes sont le résultat de l'opération de raccord de deux indices. Un des inconvénients des indices à base fixe est de vieillir relativement rapidement. En effet, les produits sélectionnés pour représenter l'évolution des catégories d'articles ne sont plus adaptés et les coefficients de pondération de la période de base et de la période courante deviennent trop différents pour que la comparaison demeure valable. On propose donc de calculer des indices dont la base varie. Pour faire une comparaison entre deux périodes, on calcule le raccord des indices. Notons $I_{t/(t-1)}$ l'indice à la date t par rapport à la date $t-1$, l'indice-chaîne à la date 2 par rapport à la date 0 est donné par:

$$C_{2/0} = I_{2/1} \cdot I_{1/0}$$

L'indice-chaîne à la date t par rapport à la date 0 est donné par:

$$C_{t/0} = I_{t/(t-1)} \cdot C_{(t-1)/0}$$

et par récurrence:

$$C_{t/0} = I_{t/(t-1)} \cdot I_{(t-1)/(t-2)} \cdot \ldots \cdot I_{1/0}$$

Notons que chaque maillon de la chaîne peut être du type de Laspeyres ou de Paasche, mais que leur produit n'est ni un indice de Laspeyres ni un indice de Paasche. Chaque maillon offre un meilleur suivi qu'avec les indices de Laspeyres et de Paasche de la variation de l'indice entre deux périodes successives. Chaque élément de la chaîne est un indice dont la base est la période immédiatement précédente, cela permet de comparer deux périodes successives.

• **Remarques**

Un indice de prix peut être calculé avec l'une des trois méthodes exposées précédemment. Cependant en pratique, c'est l'indice de Laspeyres qui est utilisé. L'une des difficultés d'un indice de prix réside dans le fait qu'il faille établir des coefficients de pondération. Dans le cas d'un indice de prix à la consommation, il faut effectuer une enquête auprès des ménages sur leurs dépenses de consommation. Les résultats de cette enquête permettent alors d'obtenir les coefficients budgétaires.
Un indice de Laspeyres nécessite la détermination de coefficients budgétaires pour la période de base. Ces poids sont conservés jusqu'à ce qu'une révision de l'indice soit entreprise. Pour un indice de Paasche, le calcul des coefficients de pondération doit être effectué pour chaque période courante. Nous remarquons donc que l'indice de Laspeyres est utilisé en pratique car l'indice de Paasche se révèle être très coûteux.
L'objectif d'un indice de Laspeyres des prix à la consommation peut être défini de la façon suivante. **Un indice de Laspeyres des prix à la consommation a pour but de mesurer l'évolution dans le temps, par rapport à une période de base, du niveau général des prix des biens et services, compte tenu de l'importance**

relative qu'ils ont pour la population de référence. Le panier de la ménagère, composé des biens et services consommés par la population de référence durant la période de base, est fixe et invariant.

L'indice peut être calculé pour plusieurs niveaux, tels les indices partiels obtenus par régions, par groupes d'articles (par exemple: l'alimentation, l'habillement,...), par canaux de distribution, etc. L'indice général des prix à la consommation est le résultat de l'agrégation des indices des niveaux inférieurs.
Pour pouvoir construire et calculer un IPC nous avons vu qu'il faut connaître deux types d'informations; premièrement les coefficients budgétaires et deuxièmement les prix des biens et services, à la période de base et de référence. Comme il n'est possible ni d'examiner le prix de tous les biens et services dans tous les points de vente d'un pays, ni d'observer la consommation de tous les ménages d'un pays, ces deux éléments doivent être obtenus en utilisant des données issues de cinq échantillons, qui sont:

- Un échantillon des ménages, qui participe à une enquête sur la consommation et dont les résultats servent à estimer les coefficients budgétaires (w).

- Un échantillon d'aires géographiques, par exemple des communes.

- Un échantillon de points de vente, choisis à l'intérieur de chaque aire géographique sélectionnée.

- Un échantillon d'articles, sélectionnés à l'intérieur des points de vente choisis, et dont on relève les prix.

- Un échantillon d'habitations, qui sert à estimer le niveau moyen des loyers.

La sélection de ces échantillons peut se faire par plusieurs méthodes que nous présentons dans la section suivante. Les techniques d'échantillonnage sont exposées, notamment, chez Y. Dodge (1993) et chez Y. Dodge , F. Mehran et M. Rousson (1990).

1.4 Les techniques d'échantillonnage

Supposons que l'on veuille connaître la part du revenu des ménages résidant en Suisse qui est consacrée au loyer. Pour déterminer cette part, on peut observer chaque ménage et déterminer le pourcentage du revenu consacré au loyer. En observant chaque individu d'une population, on procède ainsi à un recensement. Une autre méthode consisterait à choisir un certain nombre de ménages en Suisse et à observer la part des dépenses que chacun de ces ménages consacre au loyer, et en fonction de ces résultats, à tirer des conclusions pour l'ensemble des ménages résidant en Suisse. En ne choisissant qu'une fraction de la population, appelée échantillon, on procède à un échantillonnage.
Le problème principal réside dans le choix de l'échantillon, qui doit être représentatif de la population, afin que les conclusions établies sur la base des résultats de l'échantillon soient applicables à la population.
L'échantillonnage présente les avantages suivants:

1. L'échantillonnage est plus économique que le recensement; en effet, on réduit le coût de collecte.

2. En cas d'échantillonnage, le dépouillement des données peut être réalisé dans un temps plus court. Il y a donc gain de temps.

3. Une analyse exhaustive de la population n'est pas toujours réalisable. L' échantillonnage permet alors de couvrir des cas qui ne pourraient l'être par le recensement.

4. Souvent, une meilleure qualité des résultats est obtenue par l'échantillonnage. Il y a plus de ressources disponibles pour la formation des enquêteurs; le nombre restreint d'observations permet un meilleur contrôle de celles-ci.

Nous exposerons dans les sous-sections suivantes les deux grandes familles de techniques d'échantillonnage, à savoir les méthodes d'échantillonnage par choix raisonné et les méthodes d'échantillonnage aléatoires ou probabilistes. Cependant, avant d'examiner ces techniques, nous proposons un historique de l'échantillonnage, dont la source se trouve chez J. Mairesse (1988).

Très tôt dans l'histoire des hommes, les dirigeants, tels les Princes et les Rois, voulurent connaître leur puissance. Pour cela, il leur fallait connaître certains éléments susceptibles de représenter leur force, comme la population, les richesses, etc. Pour atteindre ce but, ils observèrent systématiquement et d'une manière exhaustive ces éléments; ainsi, l'idée du recensement apparut dans tous les berceaux de la Civilisation, comme en Mésopotamie, en Egypte, en Chine, etc.

En Occident, dès le XVIIème siècle, on trouve les premières tentatives d'extrapolations, notamment en Angleterre avec la fondation par John Graunt et William Petty de l'école d'arithmétique politique anglaise. L'un des buts de cette école est de quantifier et de rechercher des constantes dans le comportement humain qui pourraient permettre des estimations et des prévisions, comme par exemple le nombre d'enfants par femme, le nombre d'habitants par maison. Au XVIIIème siècle, en France, Colbert fait mener une enquête sur l'état des Provinces. Quant à Vauban, il s'intéresse aux données chiffrées, qu'elles proviennent de recensements ou d'échantillons. Dans son écrit "Méthode généralle et facille pour faire le dénombrement des peuples" (1686), il préconise l'utilisation d'échantillons de terres arables dans chaque province pour estimer au mieux les capacités agricoles du pays. L'Abbé de Saint-Pierre souhaite que soit créé "un bureau pour recueillir les divers dénombrements". Pierre-Simon Laplace, en 1783, propose de calculer ce qu'il appelle "l'erreur à craindre", c'est-à-dire de calculer l'incertitude de l'estimateur.

Jusqu'au XIXème siècle, il y a domination de l'exhaustivité et des recensements. C'est à la fin du XIXème que l'Institut International de Statistique (IIS) (créé à Londres en juin 1885) sera le lieu des premiers débats sur la représentativité. Présenté au congrès de l'IIS à Berne en 1895, l'article "Observations et expériences concernant des dénombrements représentatifs" d'Anders Nicolaï Kiaer va déclencher des réactions négatives. Kiaer est alors directeur du Bureau Central de Statistique du Royaume de Norvège et émet l'idée d'une investigation partielle (l'échantillon) basée sur une méthode qu'il appelle "méthode représentative". Le but de sa méthode est de choisir un échantillon de manière qu'il soit une miniature de la population. Un des courants d'opposition les plus forts fut celui mené par George Von Mayr, Université de Mūnich, pour qui une investigation partielle ne peut jamais remplacer un recensement. En 1897, à la conférence d'IIS à Saint-Petersburg et à une conférence des statisticiens scandinaves qui eut lieu à Stockholm, Kiaer exposa à nouveau ses idées sur le sujet et elles furent acceptées. En 1901 et 1903 aux réunions d'IIS à Budapest et Berlin, il continue de défendre son point de vue. En 1901, C.D. Wright, fondateur du Bureau des Statistiques américaines, prit parti en faveur de Kiaer, de même que A.L. Bowley.

Jusqu'en 1925, date de la réunion d'IIS à Rome, il n'y eut plus d'autre discussion sur la méthode représentative. Dès 1925, cette méthode est acceptée et le débat portera dès lors sur la technique à utiliser pour sélectionner un échantillon. Une commission, nommée par l'IIS en 1924, étudia l'application de la méthode représentative. Le rapport de cette commission aboutit à une résolution, adoptée à Rome en 1925, qui acceptait certaines méthodes d'échantillonnage aléatoires ou raisonnées.

Jerzy Neyman fut l'un des pionniers de la théorie de l'échantillonnage. L'un de ses articles ,"On the two different aspects of the representative methods: the method of stratified sampling and the method of purposive selection" (1934), fut reconnu comme une contribution importante dans le domaine des statistiques. L'une des raisons du succès de Jerzy Neyman est qu'il put fournir des raisons théoriques ainsi que des exemples pratiques pour expliquer pourquoi les méthodes aléatoires donnent un meilleur résultat que la sélection par choix raisonné.

Les statisticiens russes, avant la Révolution de 1917, avaient déjà développé des méthodes fondées sur une vision partielle de la population. C'est dans le cadre des "Zemstvos", qui étaient des gouvernements locaux au sein desquels il existait des départements de statistiques, que se développèrent les idées de l'échantillonnage. Les Zemstvos furent créés en 1864 par A.I. Tchuprov (1842-1908). Etant donné l'immensité du pays, on utilisa très tôt des techniques d'échantillonnage car il y avait impossibilité d'observer exhaustivement tous les territoires. A.I. Tchuprov, présent au congrès d'IIS à Saint-Petersburg, joua un rôle important pour répandre les idées de Kiaer chez les statisticiens russes. Dès 1918, A.A. Tchuprov (fils de A.I. Tchuprov) étudie l'échantillonnage aléatoire; il l'utilise pour mesurer la précision des estimateurs et il fait référence au tirage en grappes et à l'échantillonnage stratifié. A.G. Kovalsvsky, en 1924, a déjà traité l'échantillonnage par stratifiation et l'allocation optimale par strate. Ainsi, ces deux statisticiens ont obtenu des résultats qui seront découverts et répandus une dizaine d'années plus tard en Europe Occidentale par J. Neyman. On peut se demander pour quelles raisons les résultats de Tchuprov ne se sont pas répandus. Certains auteurs notent que, premièrement, après la Révolution russe, Tchuprov s'établit à Berlin puis à Dresden où il vécut en solitaire et que, deuxièmement, le style de Tchuprov est moins lisible que celui de Neyman.

Quant à J. Neyman, il donna lecture des résultats de ses recherches devant le "Royal Statistical Society" de Grande-Bretagne. Puis, invité à Washington par W.E. Deming, il donna une série de conférences sur l'échantillonnage aléatoire, ce qui influença profondément les statisticiens américains.

Citons encore comme chercheurs H. Hansen et W. Hurwitz qui travaillèrent pour le bureau du recensement américain, ou encore P.C. Mahalanobis, fondateur de l'Indian Statistical Institute, qui s'intéressa à l'échantillonnage dès 1932.

1.4.1 Les méthodes d'échantillonnage à choix raisonné

Parmi les méthodes à choix raisonné, la technique la plus usitée est la méthode des quotas. Cette dernière peut se résumer comme suit. A partir d'informations connues concernant la population, sur la répartition selon certaines caractéristiques comme l'âge, le sexe, la catégorie socio-professionnelle, etc, on construit l'échantillon de manière à y retrouver les mêmes proportions que celles observées dans la population. Les caractéristiques ainsi utilisées sont appelées variables de contrôle. De cette façon, on essaie d'assurer la représentativité de l'échantillon. Parmi les variables de contrôle les plus utilisées, mentionnons pour un échantillon de personnes: le sexe et l'âge; pour un échantillon de ménages: la catégorie socio-professionnelle, l'effectif ou la région et pour un échantillon de points de vente: le type de commerce, le nombre de salariés et la nature de l'activité.

Un contrôle sur l'enquêteur est difficile à effectuer puisqu'il choisit librement l'individu; un certain nombre de directives peuvent cependant lui être données. On y trouve l'interdiction de sélectionner les individus à partir de listes (par exemple, une liste des membres d'associations), d'opérer dans la rue ou encore d'interroger deux fois le même individu. La méthode des itinéraires ou méthode de Politz impose à l'enquêteur un itinéraire fixe avec des points d'arrêt prédéterminés. On dispose ainsi d'un moyen de contrôler le travail effectué. Les avantages de ce type de méthodes sont leur faible coût et leur rapidité, et le fait qu'il ne faille pas de base de sondage. Mais leurs inconvénients sont de deux ordres. Premièrement, la méthode par choix raisonné n'a pas de fondement théorique suffisant. Elle repose sur l'hypothèse qu'une distribution correcte des variables de contrôle assure la représentativité de la distribution des caractères étudiés. Deuxièmement, ces méthodes ne permettent pas d'évaluer la précision des estimateurs. En effet, faute de pouvoir contrôler la probabilité d'inclusion des individus choisis, il n'est pas possible d'utiliser la théorie des probabilités. Cette dernière est utilisée dans le cas d'un échantillonnage aléatoire et elle permet d'associer à chaque estimation une mesure de l'erreur qui peut avoir été commise.

1.4.2 Les méthodes d'échantillonnage aléatoires

Avec ce type de méthodes, chaque individu de la population a une probabilité, connue et non nulle, de faire partie de l'échantillon. Pour pouvoir sélectionner les individus à partir de la population, on doit disposer d'une base de sondage, c'est-à-dire d'une liste composée de tous les individus de la population sans omission et sans répétition. Avec les méthodes aléatoires, la théorie statistique peut être appliquée afin de dériver les propriétés des estimateurs des échantillons.
Il faut tout d'abord distinguer la sélection avec remise ou sans remise. Lorsqu'il y a remise, l'individu sélectionné est replacé dans la population. Un même individu peut ainsi être sélectionné plusieurs fois. Lorsque le tirage est effectué sans remise, l'individu choisi n'est pas replacé dans la population.
Les variables observées sur l'échantillon sont des variables aléatoires. A partir de ces dernières, on peut estimer les grandeurs correspondantes de la population et calculer l'erreur due à l'échantillonnage. Les principales méthodes d'échantillonnage probabiliste sont l'échantillonnage aléatoire simple, l'échantillonnage en grappes et l'échantillonnage stratifié. Chez W.G. Cochran (1963), chez M.H. Hansen, W.N. Hurwitz,

W.G. Madow (1960), chez Ch. Gourieroux (1981) et chez A.M. Dussaix, J.M. Grosbras (1993), on trouve une description très complète des différentes techniques d'échantillonnage.

• **L'échantillonnage aléatoire simple.**

En pratique, les plans de sondage sont plus complexes qu'un échantillonnage aléatoire simple, mais les principes fondamentaux restent les mêmes; c'est pourquoi nous commençons par examiner ce cas. Nous distinguerons le cas avec remise et celui sans remise.

On est en présence d'une procédure d'échantillonnage aléatoire simple à probabilités égales si tous les échantillons de taille n que l'on peut former à partir de la population de taille N ont la même chance ou probabilité d'être choisis. Pour pouvoir pratiquer une procédure d'échantillonnage aléatoire simple, il est nécessaire de disposer d'une base de sondage parfaite, c'est-à-dire sans omission, ni double enregistrement ou erreur, et cette condition n'est pas souvent réalisée en pratique.
On souhaite sélectionner un échantillon de taille n issu d'une population de taille N et sur lequel des mesures sont réalisées. Un élément de la population est désigné par i, on a donc $i = 1, \ldots, N$ individus au sein de la population.
Si X représente la caractéristique que l'on veut mesurer (par exemple le revenu par habitant, le nombre d'enfants par famille, etc), alors X_i est la valeur de la caractéristique pour l'individu i. Lorsqu'un échantillon de taille n est tiré de la population, il y a $i = 1, \ldots, n$ individus dans l'échantillon, et x_i représente la valeur de la caractéristique X pour le iième individu de l'échantillon. Notons que les lettres capitales se réfèrent aux caractéristiques de la population et les lettres minuscules aux caractéristiques de l'échantillon.

La moyenne de la population s'exprime par:

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

La moyenne de l'échantillon s'exprime par:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Pour mesurer la variabilité, on utilise comme mesure de dispersion la variance (σ^2) qui se calcule comme suit dans le cas de la population:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N}$$

σ représente l'écart-type de la population.
La variance de l'échantillon (s^2), qui est un estimateur non biaisé, se définit par:

$$\begin{aligned} s^2 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)} \\ &= \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{(n-1)} \end{aligned}$$

s représente l'écart-type de la population.
Le rapport $n/N = f$ est appelé le taux de sondage.

L'échantillonnage aléatoire simple avec remise:

Les étapes suivantes sont à effectuer pour obtenir un échantillon aléatoire simple à probabilités égales avec remise. Tout d'abord, chaque individu de la population de référence est numéroté de 1 à N. Puis on choisit au hasard un nombre entre 1 et N; traditionnellement on utilise une table de nombres aléatoires; un individu est identifié, puis il est remis dans l'échantillon. Enfin, on recommence n fois l'opération afin d'obtenir un échantillon de taille n. Notons qu'une même unité peut ainsi apparaître plusieurs fois dans l'échantillon.

L'échantillonnage aléatoire simple sans remise:

Le principe est le même que celui exposé pour l'échantillonnage aléatoire simple à probabilités égales sans remise, mais l'individu sélectionné est retiré de l'échantillon. Les variables aléatoires x_i ne sont plus indépendantes.

$$\begin{aligned} V(\bar{x}) &= \frac{N-n}{Nn}s^2 \\ &= (1-f)\frac{s^2}{n} \end{aligned}$$

$$= \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$$

• **L'échantillonnage en grappes.**

Illustrons le principe de l'échantillonnage en grappes à l'aide de l'exemple suivant. Imaginons que l'on veuille calculer le loyer moyen des habitations d'une ville. Si l'on veut utiliser l'échantillonnage aléatoire simple, il faut disposer d'une base de sondage contenant toutes les habitations de la ville. Puis on doit sélectionner un échantillon de logements de taille n. Une fois les habitations choisies, il faut observer chacune d'entre elles. Notons que celles-ci peuvent être très dispersées sur le territoire de la ville et que cela engendrera notamment des frais de déplacements, une perte de temps pour se déplacer. Une autre alternative est de disposer d'une carte détaillée de la ville et de découper celle-ci en quartiers. Puis on peut sélectionner quelques quartiers et, pour chacun d'eux, on observera tous les loyers des habitations. Les avantages de cette sélection sont le regroupement géographique des unités et le fait d'observer uniquement les habitations des quartiers sélectionnés. Avec ce type de sélection, on procède à un échantillonnage en grappes. Le principe de l'échantillonnage en grappes peut donc être défini comme suit. Les unités de la population de référence (appelées unités élémentaires) sur lesquelles on désire effectuer des mesures, sont réunies en groupes ou grappes. Ceux-ci servent d'unités primaires d'échantillonnage (UP). Ensuite, on sélectionne un échantillon d'unités primaires et tous les individus appartenant aux UP choisies sont inclus dans l'échantillon. Cela correspond à un échantillonnage à un degré; notons qu'il peut y avoir des sondages à plusieurs degrés. Parmi les avantages de l'échantillonnage en grappes, mentionnons que la collecte des données pour des individus voisins géographiquement est plus facile, plus rapide et moins coûteuse que dans le cas d'individus dispersés. En effet, il y a économie de frais de déplacement, de contacts, d'identification. Ce type de sélection est aussi utile lorsqu'une base de sondage ne peut être établie. Avant de présenter les estimateurs issus de l'échantillonnage en grappes, nous exposons la notation utilisée.

Posons:

N : nombre de grappes dans la population

n : nombre de grappes dans l'échantillon

M : nombre d'unités élémentaires dans la grappe
x_{ij}: valeur de la caractéristique étudiée pour la jième unité ($j = 1, \ldots, M$) pour la iième grappe ($i = 1, \ldots, N$).
La moyenne de la iième grappe s'exprime par:

$$\bar{x}_{i.} = \sum_{j=1}^{M} x_{ij}/M$$

La moyenne des moyennes des grappes dans un échantillon de n grappes est:

$$\bar{x}_n = \sum_{i=1}^{n} \bar{x}_{i.}/n$$

La moyenne des moyennes des grappes dans la population est donnée par:

$$\bar{X}_N = \sum_{i=1}^{N} \bar{x}_{i.}/N$$

La moyenne dans la population est:

$$\bar{X} = \sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}/NM$$

La variance entre unités à l'intérieur de la iième grappe ($i = 1, \ldots, N$) s'exprime par:

$$S_i^2 = \sum_{j=1}^{M}(x_{ij} - \bar{x}_{i.})^2/(M-1)$$

La variance à l'intérieur des grappes est donnée par:

$$S_w^2 = \sum_{i=1}^{N} S_i^2/N$$

La variance entre les moyennes des grappes dans la population est:

$$S_b^2 = \sum_{i=1}^{N}(\bar{x}_{i.} - \bar{X})^2/(N-1).$$

Pour un échantillonnage aléatoire simple sans remise de n grappes, contenant chacune M unités d'une population de N grappes de moyenne échantillonnale $\bar{x}_n$, sa variance est donnée par:

$$
\begin{aligned}
V(\bar{x}_n) &= \frac{1-f}{n} S_b^2 \\
&\cong \frac{1-f}{nM} S^2(1+(M-1)\rho)
\end{aligned}
$$

où $f = n/N$ représente le taux de sondage et ρ le coefficient de corrélation intra-grappes, ce dernier s'exprime par:

$$
\rho = \frac{\sum_{i=1}^{N}\sum_{j\neq k}^{M}(x_{ij}-\bar{X})(x_{ik}-\bar{X})}{(M-1)(NM-1)S^2}
$$

L'estimateur de la variance de $\bar{x}_n$, pour un échantillon aléatoire simple sans remise, est donné par:

$$
\hat{V}(\bar{x}_n) = \frac{1-f}{n} s_b^2
$$

où:

$$
s_b^2 = \frac{\sum_{i=1}^{n}(\bar{x}_{i.}-\bar{x}_n)^2}{(n-1)}
$$

Pour un échantillon de n grappes, la variance totale peut être décomposée en tenant compte de la variance entre les grappes et de la variance à l'intérieur des grappes.

La variance entre les grappes s'exprime par:

$$
s_b^2 = \sum_{i=1}^{n} \frac{(\bar{x}_{i.}-\bar{x}_n)^2}{n-1}
$$

La variance à l'intérieur des grappes est donnée par:

$$
s_w^2 = \sum_{i=1}^{n}\sum_{j=1}^{M} \frac{(x_{ij}-\bar{x}_{i.})^2}{n(M-1)}
$$

La variance totale devient alors:

$$\begin{aligned} s^2 &= \sum_{i=1}^{n}\sum_{j=1}^{M} \frac{(x_{ij} - \bar{x}_n)^2}{(nM-1)} \\ &= \frac{(n-1)s_b^2 + n(M-1)s_w^2}{(nM-1)} \end{aligned}$$

s_b^2 et s_w^2 sont des estimateurs non biaisés de S_b^2 et S_w^2. s^2 est un estimateur biaisé de S^2, car un échantillon de nM unités n'est pas choisi aléatoirement parmi la population de NM éléments. Un estimateur non biaisé est donné par:

$$\hat{S}^2 = \frac{(N-1)s_b^2 + N(M-1)s_w^2}{(NM-1)}.$$

• **L'échantillonnage aléatoire simple stratifié.**

Dans l'échantillonnage stratifié, la population est divisée en plusieurs sous-populations. Ces sous-populations, appelées strates, sont mutuellement exclusives et exhaustives. Un échantillon est tiré pour chacune des strates. Les tirages sont indépendants pour chacune d'entre elles. La stratification a plusieurs avantages d'une point de vue pratique; en voici quelques-uns:

1. Une stratification réalisée en fonction de caractéristiques naturelles peut améliorer la plan d'échantillonnage. Par exemple, lors d'une enquête sur les revenus, on a tout intérêt à répartir les ménages dans des strates en fonction du niveau du revenu.

2. La stratification est efficace lorsqu'il y a des valeurs extrêmes dans la population qui peuvent être réparties en différentes strates. On obtient ainsi des strates homogènes, ce qui réduit la variabilité à l'intérieur de la strate. On peut alors obtenir un estimateur pour chaque strate et ceux-ci peuvent être combinés de façon à donner un estimateur précis pour l'ensemble de la population.

3. La stratification rend possible l'utilisation de plans d'échantillonnage différents pour chaque strate. En pratique, l'information

dont on dispose n'est pas uniforme pour toutes les unités; la population peut alors être subdivisée en fonction de l'information disponible.

Pour ces diverses raisons, la stratification est une technique qui est presque toujours utilisée en pratique.

La population considérée est divisée en k sous-ensembles homogènes appelés les strates. N représente le nombre total d'unités élémentaires de la population et N_h représente l'effectif de la strate h. Ainsi

$$N = \sum_{h=1}^{k} N_h = N_1 + N_2 + \ldots + N_k$$

Dans ce qui suit, on va considérer que la sélection des échantillons à l'intérieur des strates se fait par un plan aléatoire simple, et à probabilités égales. Ainsi, la taille de l'échantillon tiré de la strate h sera désignée par n_h et

$$n = \sum_{h=1}^{k} n_h$$

est le nombre total des éléments tirés dans toutes les strates.

La valeur de la caractéristique X de la iième unité de l'échantillon dans la strate h est désignée par X_{hi}.

Le total des caractéristiques de chaque unité dans une même strate se note de la manière suivante:

$$X_h = \sum_{i=1}^{N_h} X_{hi}$$

On note par x_{hi} la valeur de la caractéristique de la iième unité de l'échantillon dans la strate h. On obtient donc les résultats suivants:

$$x_h = \sum_{i=1}^{n_h} x_{hi}$$

De même

$$X = \sum_{h=1}^{k} X_h = \sum_{h=1}^{k} \sum_{i=1}^{N_h} X_{hi}$$

représente la somme des valeurs des caractéristiques de toutes les unités élémentaires de la population et

$$x = \sum_{h=1}^{k} x_h = \sum_{h=1}^{k} \sum_{i=1}^{N_h} x_{hi}$$

est la somme totale de la caractéristique en considération pour toutes les unités de l'échantillon.

La moyenne de toute la population est désignée par:

$$\bar{X} = \frac{X}{N} = \frac{1}{N} \sum_{h=1}^{k} X_h = \frac{1}{N} \sum_{h}^{k} \sum_{i}^{N_h} X_{hi}$$

La notation $\bar{X}_h$ représente la moyenne à l'intérieur de la strate h:

$$\bar{X}_h = \frac{X_h}{N_h} = \frac{1}{N_h} \sum_{i=1}^{N_h} X_{hi}$$

La moyenne de l'échantillon de taille n_h de cette strate vaut:

$$\bar{x}_h = \frac{x_h}{n_h} = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$$

La variance de la caractéristique entre les unités élémentaires à l'intérieur d'une même strate h s'écrit de la manière suivante:

$$\sigma_h^2 = \frac{\sum_{i=1}^{N_h} (X_{hi} - \bar{X}_h)^2}{N_h}$$

et

$$S_h^2 = \frac{N_h}{N_h - 1} \sigma_h^2 = \frac{\sum_{i=1}^{N_h} (X_{hi} - \bar{X}_h)^2}{N - 1}$$

Ainsi la variance pour toute la population, sans égard pour la strate, est:

$$\sigma^2 = \frac{\sum_{h=1}^{k}\sum_{i=1}^{N_h}(X_{hi}-\bar{X})^2}{N}$$

où:

$$s^2 = \frac{\sum_{h=1}^{k}\sum_{i=1}^{N_h}(X_{hi}-\bar{X})^2}{N-1}.$$

1.5 Erreurs consécutives à l'échantillonnage

A la suite de tout échantillonnage, il existe une erreur entre l'estimateur, calculé sur la base de l'échantillon, et la vraie valeur de la population. Cette erreur est composée de deux éléments: le premier est l'erreur d'échantillonnage et le second l'erreur de mesure.

1.5.1 L'erreur d'échantillonnage

L'erreur dite d'échantillonnage apparaît lorsque l'on travaille sur un échantillon et non sur la population. En effet, l'échantillon sélectionné n'est qu'un représentant de l'ensemble de tous les échantillons possibles de même taille qui auraient pu être sélectionnés à partir de la population. Les estimateurs, que l'on obtiendrait pour chacun des échantillons possibles, seraient différents les uns des autres. L'erreur d'échantillonnage est donc une mesure de la variation parmi les estimateurs issus de tous les échantillons possibles. Cette erreur dépend du plan d'échantillonnage choisi (plan de sondage stratifié, à plusieurs degrés) et de la forme de l'estimateur. Notons le cas particulier du recensement où l'erreur d'échantillonnage est inexistante, puisque l'on observe tous les individus de la population.

1.5.2 L'erreur de mesure

Le deuxième type d'erreur est l'erreur dite de mesure. Ce type d'erreur se retrouve aussi bien dans l'échantillonnage que dans le recensement. Plusieurs sources d'erreurs existent; voyons-en quelques-unes.

- Une des premières sources d'erreur de mesure est constituée par les erreurs de couverture. Celles-ci sont dues au fait que la population couverte ne correspond pas à l'ensemble de la population cible, par exemple les données que l'on a sur la population peuvent être incomplètes, périmées ou fausses.

- Les non réponses constituent également une erreur de mesure lorsqu'un individu sélectionné ne peut pas être atteint, ou que l'individu choisi et contacté refuse de répondre.

- L'enquêteur, lors d'entretiens, peut influencer la réponse du répondant. Par exemple, en ne lisantpas la question d'une façon correcte, il influence le répondant lorsqu'il pose la question, ou encore, il enregistre incorrectement la réponse.

- Un autre type d'erreur est dû à la rédaction du questionnaire. Il est important de faire attention aux termes utilisés. En effet, si ces derniers sont savants, ou s'il s'agit d'abréviations, le répondant peut ne pas les comprendre correctement et sa réponse sera incorrecte. La présentation esthétique du questionnaire est également importante ainsi que l'ordre dans lequel sont posées les questions. Des modifications apparemment anodines de formulation des questions peuvent entraîner de grands changements dans les réponses.

- La psychologie du répondant peut l'entraîner à répondre volontairement d'une manière erronnée.

- La manière dont est réalisée l'enquête (par téléphone, par visites personnelles) influence les réponses.

- Des erreurs peuvent apparaître lors de la saisie ou du dépouillement des données (erreurs de codage, de transfert).

1.6 Erreur quadratique moyenne

Dans la présente section nous examinons l'erreur totale d'une statistique que l'on appelle aussi l'erreur quadratique moyenne (EQM).

A partir des observations effectuées sur l'échantillon, il faut estimer avec un maximum d'efficacité la valeur du paramètre de la population et apprécier la précision de cette estimation.

Estimer un paramètre de la population, à partir d'un échantillon aléatoire, contient deux difficultés. La première concerne le choix de l'estimateur et la seconde consiste à déterminer la précision de l'estimation. Un bon estimateur doit notamment être sans biais et être efficace.

- **Un estimateur sans biais:**

Un estimateur, $\hat{\theta}$, est dit sans biais si l'espérance mathématique de $\hat{\theta}$ est égale à la vraie valeur, θ, du paramètre de la population:

$$E(\hat{\theta}) = \theta$$

le biais, $B(\hat{\theta})$, est mesuré par:

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta$$

- **Un estimateur efficace:**

Plus sa dispersion est faible, plus un estimateur, $\hat{\theta}$, est efficace. La variabilité de l'estimateur se mesure en calculant sa variance:

$$Var(\hat{\theta}) = E(\hat{\theta} - E(\hat{\theta}))^2$$

Notons que si l'on a deux estimateurs sans biais $\hat{\theta_1}$ et $\hat{\theta}_2$, le plus efficace des deux est celui qui a la dispersion la plus faible.

On peut montrer que l'EQM est la somme de la variance et du carré du biais.

Soient $X_1, \ldots, X_n$ un échantillon de taille n et $\hat{\theta}$ un estimateur du paramètre θ, l'écart au carré de l'estimateur est donné par $(\theta - \hat{\theta})^2$. L'EQM est défini par l'espérance mathématique de ce dernier.

$$EQM(\hat{\theta}) \quad = \quad Var(\hat{\theta}) + (B(\hat{\theta}))^2.$$

1.6.1 Erreur quadratique moyenne, concept appliqué à l'IPC

Le concept de l'EQM, appliqué à l'IPC, fut introduit dans la littérature par différents auteurs, par exemple Bailar (1983), L. Biggeri et A. Giommi (1983), C. Anderson, G. Forsman et J. Wretman (1987). En suivant leurs propositions, posons que:

I^* - l'indice idéal
I - l'indice exact (défini)
$\hat{I}$ - l'indice estimé

Alors l'EQM s'exprime par:

$$\begin{aligned} EQM(\hat{I}) &= E(\hat{I} - E(\hat{I}))^2 + (E(\hat{I}) - I)^2 \\ &\quad + (I - I^*)^2 + 2(E(\hat{I}) - I)(I - I^*) \end{aligned}$$

- **Variance**

$E(\hat{I} - E(\hat{I}))^2$: ce terme représente la part de l'erreur totale due aux erreurs d'échantillonnage.

Rappelons que pour construire un IPC, on doit prélever un échantillon de ménages, d'aires géographiques, de points de vente, de biens et services et d'habitations.

- **Biais**

$E((\hat{I}) - I)^2$: ce terme représente la part de l'erreur totale due à l'erreur de mesure issue de la mise sur pied des différents plans d'échantillonnage.

Les auteurs susmentionnés distinguent différentes sources d'erreurs de mesure propres au cas de l'IPC:

- Les erreurs de couverture dues à des listes inadéquates des produits, des ménages, des points de vente.

- L'omission de catégories de biens, par exemple les repas pris hors du domicile, les dépenses des vacances.

- Les erreurs d'homogénéité, les biens observés ne sont pas les mêmes que ceux de la période de référence.
- les non réponses qui touchent particulièrement l'enquête réalisée auprès des ménages.
- L'inexactitude des données pour les prix relevés ou pour les données relatives aux dépenses des ménages.

$(I - I^*)^2$: cette part d'erreur est due au fait que la méthode de calcul utilisée n'est pas adéquate par rapport à l'indice théorique idéal.

Ce type d'erreur peut provenir de plusieurs sources, notamment parce qu'on choisit un indice qui ne correspond pas à l'indice théorique idéal. Par exemple, on utilise un indice de Laspeyres alors qu'un autre indice serait meilleur d'un point de vue théorique. Une autre source d'erreur peut être le système des coefficients de pondération qui est utilisé; celui-ci ne doit pas être obsolète, ce qui est rapidement le cas avec un système de pondération fixe et invariant. Cependant en pratique, il s'avère peu réalisable de tenir à jour le système des coefficents budgétaires. En effet, cela entraînerait, par exemple, la construction annuelle d'une nouvelle base, et ce travail d'un point de vue pratique serait très difficile à mettre en place.

Les types d'erreurs énumérés ci-dessus montrent combien le problème de l'estimation des erreurs de l'IPC est complexe.

Chapitre 2

L'indice suisse des prix à la consommation

2.1 Introduction

Ce chapitre a pour but de décrire le plan d'échantillonnage et la méthode de calcul de l'indice suisse des prix à la consommation (ISPC). Les sources de ce chapitre se trouvent dans les documents de l'Office fédéral des arts et métiers et du travail (OFIAMT) [1] et de l'Office fédéral de la statistique (OFS) [2]. Notons que jusqu'en mars 1987, l'Office fédéral de l'industrie, des arts et métiers et du travail était chargé de calculer l'ISPC. Depuis cette date, c'est l'Office fédéral de la statistique qui en est responsable.

Note:

1 *L'indice suisse des prix à la consommation.* (1977)

2 *La révision de l'indice suisse des prix à la consommation. Concept de base.* (1990)
Le nouvel indice suisse des prix à la consommation: mai 1993. Aperçu des méthodes. (1993)
Révision de l'indice suisse des prix à la consommation. Conception du nouvel indice suisse des prix à la consommation. (1993)

2.2 Le plan d'échantillonnage

En pratique, il n'est pas possible d'observer les prix de toutes les transactions de consommation réalisées en Suisse. L'indice des prix à la consommation de la Suisse est donc estimé, par l'Office fédéral de la statistique (OFS), en se fondant sur des données issues de cinq échantillons. Tout d'abord, des ménages sont sélectionnés et participent à l'enquête sur les budgets des ménages, dont les résultats servent à déterminer les coefficients budgétaires des articles. Puis un échantillon de communes est choisi, et à l'intérieur de celles-ci, des points de vente sont à leur tour sélectionnés. Dans chaque point de vente, un échantillon d'articles est ensuite tiré et ce sont les prix de ces articles qui sont relevés. Enfin, le dernier échantillon qu'il faut sélectionner est l'échantillon d'habitations locatives qui sert au calcul de l'indice du loyer. Notons que les habitations en propriété ne sont pas, actuellement, inclues dans le calcul de l'indice du loyer.
Dans les sections suivantes, nous décrivons les plans d'échantillonnage en vigueur jusqu'à fin avril 1993 ainsi que ceux qui font suite à la dernière révision générale de l'IPC. Le nouvel indice, issu de cette révision complète, est entré en vigueur en mai 1993.

2.2.1 L'échantillon des ménages

La première enquête sur le budget des ménages date de 1912, puis il y a eu des enquêtes en 1919-1922, 1936-1937 et 1937-1938. Dès 1943, on procéda chaque année à une enquête sur le budget des ménages. Cependant, jusqu'en 1975, seuls les ménages familiaux de salariés étaient pris en compte. A partir de cette date, on étendit le champ de l'enquête aux ménages de salariés d'une ou deux personnes.

L'enquête sur les budgets des ménages de salariés de 1975

Les résultats de l'enquête sur la consommation des ménages de 1975 servirent de base à l'établissement de la structure et des coefficients de pondération du panier de la ménagère (ou panier-type) lors de la révision complète de l'ISPC, base 100 en 1977. Cette enquête s'est déroulée auprès de 980 ménages. Ces derniers ont été choisis parmi les quelques 2500 ménages qui s'étaient portés volontaires auprès de l'OFIAMT. L'échantillon a été sélectionné selon la méthode des quotas.

Les résultats du recensement fédéral de la population de 1970 servirent de base à l'élaboration de l'échantillon. Trois variables de contrôle furent retenues et, pour chacune d'entre elles, l'OFIAMT s'efforça de retrouver au sein de l'échantillon le même pourcentage que celui établi par le recensement de la population. Les variables de contrôle choisies furent la taille du ménage, l'appartenance régionale (divisée par région linguistique, agglomérations et canton) et le revenu brut du chef du ménage. Certaines catégories de la population furent exclues; il s'agissait des personnes actives indépendantes, de la population agricole et des retraités et rentiers.

Les enquêtes sur les budgets des ménages de salariés et de retraités de 1981 et 1986

Lors de ces deux enquêtes, on désirait étendre la population de référence à l'ensemble des ménages privés. L'échantillon des ménages a comme base les résultats du recensement fédéral de la population de 1980. 442 ménages de salariés, qui s'étaient volontairement inscrits auprès de L'OFIAMT, furent soumis à l'enquête. Ils participèrent bénévolement, moyennent une rémunération symbolique de Frs. 300.-. A partir des résultats de cette enquête, on détermina le panier-type et les coefficients de pondération qui servirent à la révision partielle de 1982. L'indice a été calculé sur une base 100 en décembre de 1982 jusqu'en avril 1993. Dans les cantons de Bâle-Ville, Bâle-Campagne, du Tessin, de Vaud et Genève et dans les villes de Berne et Zurich, les données furent collectées et partiellement dépouillées. Pour toutes les autres régions de Suisse, l'OFIAMT réalisa la collecte et le dépouillement des données. En 1986, un nouvel échantillon de 484 ménages fut soumis à l'enquête.
Au cours de ces enquêtes, les dépenses de consommation se définissent comme la part du revenu que les ménages affectent à l'achat de marchandises et de services à des fins de consommation directe. Certaines dépenses, comme les impôts, les taxes et les assurances, ne sont donc pas considérées comme des dépenses de consommation. Notons toutefois que les assurances liées à la possession d'un véhicule à moteur sont assimilées à des dépenses de consommation.

L'enquête sur la consommation des ménages de 1990

Cette enquête a permis de calculer les coefficients budgétaires utilisés pour le nouvel ISPC, base 100 en mai 1993. Jusqu'alors l'enquête était effectuée auprès des ménages de salariés (sauf travailleurs agricoles) et des retraités. La Commission de statistique conjoncturelle et sociale (CSCS) releva la nécessité de définir la base de l'enquête sur la population résidante permanente et de constituer un échantillon représentatif.

L'enquête sur la consommation 1990 s'est composée en fait de deux enquêtes (annuelle et mensuelle) complémentaires et qui furent menées en parallèle. Lors de l'enquête annuelle, les ménages sélectionnés furent interrogés durant toute l'année 1990 sur leurs grandes catégories de dépenses et recettes ainsi que sur des données relatives à la structure du ménage (catégorie socio-professionnelle, situation du logement, etc). Au cours de l'enquête mensuelle, les ménages furent questionnés, pour un mois, sur leurs dépenses et recettes détaillées. Le plan de sondage fut à deux degrés: le premier degré correspondant à un échantillon aléatoire de communes et le second degré à un échantillon aléatoire simple de ménages. Les unités primaires d'échantillonnage, représentées par les communes, avaient un plan de sondage stratifié avec remise et étaient sélectionnées avec une probabilité proportionnelle à leur taille. La stratification s'est faite selon les trois régions linguistiques (allemand y compris romanche, français et italien) et selon la taille des communes, représentée par le nombre d'abonnés au téléphone. Puis dans chacune des strates, une commune fut choisie. Les unités secondaires d'échantillonnage, représentées par les ménages, ont été tirées par échantillonnage systématique; le point de départ fut sélectionné aléatoirement. La taille de l'échantillon de ménages dans chaque commune était proportionnelle à la taille de la strate. La base de sondage fut le registre des abonnés au téléphone. L'échantillon sélectionné pour l'enquête annuelle est resté le même tout au long de l'enquête. Quelque 2000 ménages ont participé à cette enquête. Quant à l'enquête mensuelle, douze échantillons de 1000 ménages chacun ont été sélectionnés. En effet, l'échantillon de ménages fut renouvelé tous les mois; il s'agissait donc d'un échantillonnage par rotation. Les unités primaires d'échantillonnage furent ainsi renouvelées tout au long de l'année, d'une manière partielle. Chaque unité primaire restait au moins deux mois, mais au plus trois ou six mois, et seules les grandes villes restèrent douze mois dans l'échantillon. Les taux de participation aux deux enquêtes furent de 25% à 30% pour l'enquête annuelle et de 40% à 50% pour l'enquête mensuelle. Pour chacune des enquêtes, une première prise de contact eut lieu par téléphone. Ce court entretien permettait de connaître les caractéristiques des ménages (par exemple la taille, la catégorie socio-professionnelle).

2.2.2 L'échantillon des communes

Dans cette section, nous commençons par présenter les communes qui ont participé aux relevés des prix jusqu'en avril 1993. Puis nous exposerons la variante retenue par l'OFS dans le cadre de la dernière révision générale.

Jusqu'en avril 1993, quarante-huit communes ont participé aux relevés des prix de l'alimentation, du mazout et de l'essence. La collaboration des communes fut volontaire et ces dernières supportèrent les coûts engendrés par les relevés de prix. Les communes participantes sont énumérées ci-après.

Aarau	Genève	Saint-Imier
Altdorf	Glaris	Saint-Moritz
Arbon	Granges	Sarnen
Appenzell	Hérisau	Schaffhouse
Baden	Langenthal	Schwyz
Bâle	Lausanne	Sion
Bellionze	Liestal	Soleure
Berne	Locarno	Stans
Bienne	Le Locle	Thoune
Berthoud	Lugano	Vevey
Brigue	Lucerne	Wil (SG)
La Chaux-de-Fonds	Neuchâtel	Winterthour
Coire	Olten	Wohlen (AG)
Delémont	Porrentruy	Yverdon
Frauenfeld	Rorschach	Zoug
Fribourg	Saint-Gall	Zurich

Les offices de statistiques de Bâle-Ville, Berne, Genève et Zurich relevaient les prix de tous les biens et services entrant dans le calcul de l'indice suisse des prix à la consommation, exception faite des relevés centralisés (tarifs CFF, tarifs postaux et téléphoniques, prix des prestations médicales et médico-dentaires, etc). En effet, les prix de ceux-ci étaient relevés d'une manière centralisée par l'Office fédéral de la statistique. Puis chacune de ces quatre villes calculait un indice complet en se référant aux directives fournies par l'OFS. Ces indices communaux étaient ensuite pondérés et intégrés lors du calcul de l'ISPC. Les coefficients de pondération des communes furent de 10% pour Bâle-Ville, 10% pour Berne, 10% pour Genève, 20% pour Zurich et de 50% pour toutes les autres communes de Suisse.

Avant d'exposer la solution retenue par l'OFS pour le nouvel indice, nous présentons les variantes qui avaient été envisagées au cours de cette révision. Dans un de ses documents techniques, l'Office fédéral de la statistique (1990) envisageait trois variantes pour constituer l'échantillon de communes. Ces variantes étaient les suivantes.

Pour la première variante, un échantillon de communes, sélectionné par choix raisonné, aurait participé à l'enquête, suivant le modèle en vigueur jusqu'en avril 1993. On aurait trouvé dans ces communes participantes les chefs-lieux des cantons, les centres d'agglomération et éventuellement quelques centres régionaux. La taille de l'échantillon se serait composée d'une cinquantaine de communes. L'Office fédéral de la statistique fit remarquer que ce système était déjà en place et donc opérationnel, et que l'on garantissait la continuité avec les enquêtes précédentes. L'OFS nota que le système de pondération des communes devrait être affiné.
Dans la deuxième variante, l'Office fédéral de la statistique proposait une sélection, par choix raisonné, d'une vingtaine de communes, représentant quelque quinze cantons. Parmi les avantages, on énuméra la décharge des informateurs et des services d'enquêtes, celle-ci pouvant être mise à profit pour améliorer l'observation des prix. Ce nombre restreint aurait toutefois permis de garantir une représentativité suffisante au niveau national. Les désavantages cités par l'OFS étaient d'ordre politique et psychologique; en effet, tous les cantons n'auraient pas été représentés.

La troisième variante envisagée consistait à sélectionner un échantillon aléatoire de communes. Cet échantillon stratifié comprenait des communes de plus de 10'000 habitants, avec une probabilité de sélection proportionnelle à la taille de la commune, et les cinq plus grandes villes retenues d'office dans l'échantillon. La taille de l'échantillon aurait été d'au maximum 50 communes. Les désavantages de cette variante résidaient, selon l'OFS, dans la disparition des avantages des deux précédentes variantes.

Les communes rurales étaient exclues de l'échantillon du précédent ISPC. L'OFS envisagea donc dans son concept de base d'inclure un échantillon aléatoire de petites communes en option aux trois variantes.

La participation actuelle.
L'OFS a finalement choisi une option intermédiaire. Tout d'abord notons que l'OFS voit dans la commune une bonne unité d'enquête, et que 24 communes ont été choisies. Pour établir une répartition régionale optimale, l'OFS s'est fondé sur la subdivision de la Suisse en macrorégions de mobilité spatiale, dites macrorégions MS. La Suisse est découpée en 16 macrorégions MS; ces dernières sont axées sur des centres (grands ou moyens) et sont liées à la notion de bassins d'emplois. Ainsi, dans les 24 communes participant aux relevés des prix, on trouve les 16 communes centres des macrorégions, quatre communes périphériques et quatre communes rurales. L'importance des communes sera pondérée en fonction de la population des régions. Les communes participantes sont:

Centres des macrorégions MS:
Aarau, Bâle, Berne, Bienne, Coire, Fribourg, Genève, Lausanne, Lucerne, Lugano, Neuchâtel, Saint-Gall, Sion, Thoune, Winterthour, Zurich.

Centres périphériques:
Biasca, Loèche, Porrentruy, Schwanden.

Communes rurales:
Beggingen (SH), Kestenholz (SO), Wassen (UR), Zuoz (GR).

2.2.3 L'échantillon des points de vente

La situation jusqu'en avril 1993
Les prix étaient relevés chez les vendeurs. Le libre choix de la sélection des points de vente était laissé à l'organisme chargé de relever les prix. Quelques instructions étaient néanmoins données, comme observer les prix dans le plus grand nombre de magasins possible, répartir ces points de vente de façon équilibrée au sein de la commune et choisir le point de vente au vu de son importance économique locale. Compte tenu de la structure du commerce de détail en Suisse, l'OFS calculait, pour le groupe de l'alimentation, une moyenne pondérée des prix "Coop" (15%), "Migros" (20%) et "Autres magasins" (65%).
Les prix étaient relevés par correspondance par l'Office fédéral de la statistique pour les relevés centralisés, tels que les tarifs CFF et PTT, les prix auprès d'associations professionnelles, de grands magasins ou de services administratifs.

La situation dès mai 1993
Le choix des points de vente à l'intérieur d'une commune est effectué avec la collaboration des organismes communaux responsables de l'enquête. Les points d'observation des prix sont choisis en fonction des types de magasins, des formes de vente et de l'emplacement au sein de la commune. De deux à six points de vente par commune et par type de produits sont sélectionnés. Certains prix peuvent être observés d'une manière centralisée, notamment dans le cas de distributeurs nationaux où les prix sont fixes pour toutes les régions du pays.
Dans le nouvel indice, l'OFS tient compte des canaux de distribution d'une manière plus affinée. En effet, il prévoit des coefficients de pondération pour chaque groupe d'articles. Les canaux de distribution se composent dorénavant de "Coop", "Migros", "Denner", de supermarchés, de grands magasins et succursales et des magasins locaux spécialisés.

2.2.4 L'échantillon des articles

La situation jusqu'en avril 1993
Le panier-type comprenait les neufs groupes principaux suivants: l'alimentation, les boissons et tabacs, l'habillement, le loyer du logement, le

chauffage et éclairage, l'aménagement et entretien du logement, les transports et communications, la santé et les soins personnels, l'instruction et les loisirs. Ces groupes principaux étaient subdivisés en 413 rubriques, appelées aussi postes de dépenses. Chaque rubrique avait un coefficent de pondération issu de l'enquête sur la consommation des ménages. Envrion 1350 variétés composaient les rubriques. Au niveau des variétés, il n'y avait pas de coefficients de pondération, et elles constituaient en fait un échantillon de tous les produits existant pour une rubrique. Les variétés représentaient les articles ou groupes d'articles dont le prix devait être relevé. L'indice général reposait sur un ensemble de quelque 300'000 relevés de prix; parmi ceux-ci, environ 75'000 relevés de prix concernaient l'alimentation et 100'000 le loyer du logement.

Trois différentes méthodes étaient appliquées par les fonctionnaires communaux pour relever les prix des articles. La première consistait à relever personnellement les prix des articles, la deuxième à distribuer et rassembler les formulaires d'enquête, les prix étant notés par les informateurs du point de vente, et la troisième à envoyer les formulaires par voie postale.

Les prix de la plupart des denrées alimentaires, du chauffage, de l'éclairage et de l'essence étaient relevés mensuellement pendant la première quinzaine du mois dans les 48 communes. Les prix des autres groupes principaux étaient relevés trimestriellement et les loyers du logement semestriellement.

Pour obtenir le détail des groupes principaux, des rubiques et des postes de dépenses ainsi que leurs coefficients de pondération, le lecteur peut consulter la publication de l'OFIAMT (1983) *"Nouvelle méthode de calcul de l'indice suisse des prix à la consommation. Base décembre 1982=100"*.

La situation dès mai 1993

Le nouveau panier-type se fonde sur la nomenclature internationale appelée "Nomenclature de l'utilisation de la consommation finale des ménages" (Nomenclature SNA). Elaborée par l'ONU, elle est adoptée par un grand nombre d'Etats, notamment les pays de la CEE. Cette nomenclature subdivise les biens et services en huit groupes. Chacun d'eux est divisé en quatre niveaux: les groupes de produits; les agrégations de postes de dépenses ou autres groupes de produits; les

postes de dépenses, qui sont les dernières rubriques pondérées et pour lesquels on publie encore des séries d'indices; enfin, les variétés qui constituent le niveau le plus bas du panier-type. Celles-ci correspondent aux biens et services dont les prix sont relevés. Ces assortiments peuvent être modifiés selon l'offre et la demande. Notons que le nombre de rubriques du nouvel indice est de 276. Le détail de ces rubiques et leurs poids de pondération se trouvent dans le document de l'OFS(1993) *"Le nouvel indice suisse des prix à la consommation; mai 1993=100"*.
Pour calculer l'IPC, l'OFS retient, comme dans la pratique internationale, le concept de consommation. Ce dernier, équivalant au concept de la consommation en comptabilité nationale, exclut les impôts directs, les assurances sociales, les cotisations versées aux assurances privées et les transferts (comme les pensions).
Les relevés de prix se font tous les mois pour les produits alimentaires, tous les six mois pour les vêtements et tous les trois mois pour les autres biens et services. Remarquons que le relevé des prix est toujours effectué au cours des huit premiers jours du mois. Les communes relèvent dorénavant le prix de tous les biens et services. Parmi les nouveaux postes, nous pouvons citer le chauffage à distance, le do-it-yourself et outils de jardinage, les appareils médicaux (lunettes, appareils auditifs, etc), les prestations paramédicales, la location de garages et de places de parc, les ordinateurs domestiques y compris les logiciels et les accessoires, les articles personnels (montres, sacs, valises, etc). Pour les groupes de produits qui ont des prix uniques, l'OFS procède à des relevés centralisés par courrier. Il s'agit notamment des médicaments, des livres, des tarifs CFF et PTT, des concessions radio et télévision. Les prix déterminants à relever sont les prix payés à l'étalage par le consommateur. Les prix des biens soldés, des invendus, des fins de série ne sont pas pris en compte; par contre, les actions et les rabais le sont.

2.2.5 L'échantillon des logements

La situation jusqu'en avril 1993
Dès 1925, le groupe loyer du logement donne l'évolution moyenne des loyers pour la Suisse, pour diverses grandeurs d'habitation et selon leur date de construction.
Le plan d'échantillonnage (non aléatoire) en vigueur jusqu'en avril 1993 date de la révision de l'ISPC de 1977. Le plan d'échantillonnage pré-

voyait initialement de relever 120'000 loyers dans 159 communes; cela aurait représenté le 12,5% des appartements. L'OFS distingua cinq catégories d'appartements, correspondant aux habitations d'une à cinq chambres, et les loyers de quelque 100'000 loyers étaient relevés semestriellement (en mai et en novembre). Quatre-vingt-cinq communes participaient à cette enquête et elles supportèrent les frais engendrés par ces relevés. Les appartements sélectionnés représentaient environ le 10% du nombre total des habitations locatives. Les habitations étaient réparties en trois catégories selon leur date de construction, à savoir les appartements anciens (construits avant 1947), les nouveaux et les récents (mis sur le marché depuis le dernier relevé de prix, (nouvelle construction)). Le coefficient de pondération du groupe principal loyer du logement était alors d'environ 20%. La méthode de calcul de l'indice des loyers du logement fut la suivante. Tout d'abord, on considérait le loyer (sans les charges pour le chauffage et l'eau chaude) pour les appartements sélectionnés. Les appartements étaient ensuite regroupés selon le nombre de chambres; on formait ainsi les types d'appartement, puis chaque type était divisé par catégories d'appartements (anciens, nouveaux et récents). Le loyer moyen se calculait pour chaque commune et pour chaque type et catégorie d'appartement. Avec les indices obtenus par commune, l'OFS calculait ensuite l'indice moyen des loyers pour l'ensemble de la Suisse. Les indices communaux étaient pondérés proportionnellement au nombre d'appartements de la commune; cette pondération fut adaptée aux résultats du recensement fédéral des logements de 1980. La grandeur du logement fut déterminée par le nombre de pièces habitables, y compris les mansardes et les autres pièces séparées du corps du logement. Les pièces sans cuisine ni cuisinette furent aussi considérées comme des logements. A partir du recensement, aucun nouvel échantillon ne fut mis sur pied et cela pour deux raisons majeures. Tout d'abord, les adresses sur les formules d'enquêtes n'étaient pas informatisées et la sélection manuelle d'un échantillon de 100'000 appartements aurait entraîné des difficultés considérables. La seconde raison concerne la protection des données. En effet, le Service de la protection des données de l'Office fédéral de la justice refusa que le recensement des logements et les adresses puissent servir de base de sondage au tirage d'un échantillon. L'OFS dut détruire les formules d'enquêtes.

La situation dès mai 1993

Les logements couverts par le champ de l'indice sont les logements non meublés et les maisons familiales loués en Suisse et ils doivent être habités durablement. Les logements exclus du champ de l'indice sont les habitations de plus de cinq chambres, les logements de vacances et les logements utilisés à d'autres fins que l'habitation, par exemple les bureaux.

L'OFS tire aléatoirement les adresses du registre téléphonique des PTT, puis les locataires contactés doivent fournir le nom et l'adresse du propriétaire ou de la gérance. L'OFS effectue les relevés des loyers chez les propriétaires ou gérants et cela chaque trimestre. On interroge les bailleurs non seulement sur le loyer mais aussi sur d'autres caractéristiques permettant de connaître la structure du parc des logements. L'OFS distingue trois types de caractéristiques supplémentaires. Premièrement, les caractéristiques dites d'identification qui regroupent l'adresse du logement, l'étage et le nom du locataire. Deuxièmement, les caractéristiques dites du logement, à savoir le genre du bâtiment, le nombre de chambres, la surface habitable en m^2, l'année de construction, l'année et le genre de la dernière rénovation, le genre du logement (attique, duplex). Enfin, le dernier type de caractéristiques est composé de celles dites auxiliaires, tels que le nom et l'adresse du propriétaire ou de la gérance, l'annonce du changement de locataire, le motif de la variation du loyer ou encore le statut d'occupation (conciergerie, logement de service, logement subventionné). Parmi les inconvénients de la base de sondage utilisée par l'OFS, on peut relever notamment la nécessité d'éliminer un certain nombre d'adresses inutilisables (par exemple les logements de vacances, les bureaux, etc) et le fait que les logements non raccordés au réseau téléphonique ne sont pas pris en considération. Cependant, il semble qu'à l'heure actuelle ce soit la meilleure base de sondage disponible. Les logements sont répartis en vingt catégories, par taille (les logements de une à cinq chambres) et par classe d'âge (de 0-5 ans, de 6-10 ans, de 11-20 ans, 20 ans et plus). L'échantillon net est composé de quelque 5000 logements. L'OFS avait prévu lors de la conception du nouvel indice d'inclure d'une manière explicite les logements occupés par leur propriétaire (LOP) dans l'indice des prix à la consommation, et cela pour deux raisons. Premièrement, les LOP ont une importance croissante et deuxièmement, l'EC 90 est élargi à la population résidante. Cependant, l'inclusion explicite d'un indice LOP est restée en suspens; elle sera à l'ordre du jour de la prochaine révision.

2.3 La méthode de calcul

La méthode de calcul de l'ISPC qui est exposée dans cette section fut en vigueur de 1982 (base 100) à avril 1993. Les différentes étapes du calcul sont exposées dans les paragraphes suivants, puis un exemple numérique simplifié illustre cette méthode.

2.3.1 La méthode des indices élémentaires

Examinons tout d'abord la métbode des indices élémentaires appliquée à la méthode de calcul de l'indice suisse des prix à la consommation.

La première étape correspond à la détermination des indices élémentaires des observations individuelles de prix.

L'indice élémentaire d'une observation de prix à l'époque t, noté (I^t_{hije}) est défini comme le rapport prix-quantité de l'observation entre la période courante et la période de référence, multiplié par 100. On l'exprime par:

$$\begin{aligned} I^t_{hije} &= \frac{p^t_{hije}}{q^t_{hije}} : \frac{p^0_{hije}}{q^0_{hije}} \cdot 100 \\ &= \frac{p^t_{hije} \cdot q^0_{hije}}{p^0_{hije} \cdot q^t_{hije}} \cdot 100 \end{aligned}$$

où t désigne le mois sous revue, 0 la période de base, p le prix relevé et q la quantité relevée. h correspond au numéro de la rubrique de l'indice dans l'indice général, i au canal de distribution avec $i = 1, \ldots, 3$ (1=Coop, 2=Migros et 3=autres points de vente). j désigne le numéro du groupe de communes, avec $j = 1, \ldots, 5$ (1= Zurich, 2=Bâle, 3=Berne, 4=Genève et 5=autres communes), enfin e représente le numéro de l'observation de prix à l'intérieur de h, i et j.

Par exemple, I^t_{hije}, peut représenter l'indice élémentaire en mai 1992 (t) du *lait entier upérisé, 1l* ($= e$) à Genève ($j = 3$) pour la Migros ($i = 1$) de la rubrique *lait entier en emballage* ($= h$).

2.3.2 La méthode de calcul

Calcul de l'indice en cas d'assortiment constant et de structures constantes des groupes de communes et des canaux de distribution

A partir des indices élémentaires, on calcule, avec l'opération suivante, un indice (I^t_{hij}) par groupe de communes j à l'intérieur d'un canal de distribution i et d'une rubrique de l'indice h (I^t_{hij}). I^t_{hij} est exprimé par:

$$I^t_{hij} = \frac{\sum_{e=1}^{n^t_{hij}} I^t_{hije}}{n^t_{hij}}$$

où n^t_{hij} indique le nombre d'indices élémentaires utilisés pour la rubrique h, i et j. Par exemple I^t_{hij} peut représenter l'indice en mai 1992 (t) de la rubrique *lait entier en emballage* (h) pour la Migros ($i = 2$) à Genève ($j = 4$).

Les indices des groupes de communes (I^t_{hij}), déterminés selon la méthode ci-dessus, sont ensuite agrégés avec les coefficients de pondération des groupes de communes pour former l'indice par canal de distribution à l'intérieur d'une rubrique de l'indice (I^t_{hi}). On calcule I^t_{hi} comme suit:

$$I^t_{hi} = \frac{\sum_{j=1}^{5} g_j \cdot I^t_{hij}}{\sum_{j=1}^{5} g_j}$$

où g_j correspond aux coefficients de pondération des communes. Avec $g_1 = 0{,}2$ pour Zurich, $g_2 = 0{,}1$ pour Bâle, $g_3 = 0{,}1$ pour Berne, $g_4 = 0{,}1$ pour Genève, et $g_5 = 0{,}5$ pour autres communes. Par exemple I^t_{hi} peut représenter l'indice en mai 1992 (t) de la rubrique *lait entier en emballage* (h) pour le canal de distribution Migros (i).

Lorsqu'il n'y a pas, dans un groupe de communes, d'observation de prix pour le canal de distribution correspondant, le coefficient de pondération de ce groupe de communes est alors réparti proportionnellement sur les autres groupes de communes. Pour cela on divise la somme des indices pondérés agrégés par la somme des coefficients de pondération des groupes de communes restantes.

Les indices nationaux par canal de distribution (I_{hi}^t) sont, dans l'étape suivante, pondérés et agrégés avec les coefficients de pondération des canaux de distribution pour former l'indice par rubrique de l'indice général (I_h^t).

$$I_h^t = \frac{\sum_{i=1}^{3} a_i \cdot I_{hi}^t}{\sum_{i=1}^{3} a_i}$$

où a_i représente les coefficients de pondération des canaux de distribution, $a_1 = 0{,}15$ pour Coop, $a_2 = 0{,}20$ pour Migros et $a_3 = 0{,}65$ pour autres points de vente. Par exemple I_h^t peut représenter l'indice au mois de mai 1992 (t) de la rubrique *lait entier en emballage h* au niveau national.

Si, dans un canal de distribution, il n'y a pas d'observation de prix, alors le coefficient de pondération est réparti proportionnellement sur les autres canaux de distribution. La manière de procéder est analogue à celle appliquée à la pondération des groupes de communes.

Le calcul de l'indice en cas de disparition d'observations de prix, de groupes de communes ou de canaux de distribution

Si une observation de prix, dont l'indice élémentaire était au mois précédent supérieur ou inférieur à la moyenne, n'est plus faite au mois sous revue, il faut alors appliquer le calcul suivant. On établit un rapport des moyennes des indices élémentaires du mois courant et du mois précédent, dans lequel n'entrent que les observations de prix relevées pour les deux périodes, et ensuite on enchaîne ce rapport à l'indice par groupe de communes et canal de distribution déterminé au mois précédent.

$$I_{hij}^{t} = \frac{\sum\limits_{e=1}^{n_{hij}^{t}} I_{hije}^{t}}{\sum\limits_{e=1}^{n_{hij}^{t}} I_{hije}^{t-1}} \cdot I_{hij}^{t-1}$$

Pour calculer l'indice par canal de distribution, on procède de la même manière que ci-dessus:

$$I_{hi}^{t} = \frac{\sum\limits_{j=1}^{5} g_j \cdot I_{hij}^{t}}{\sum\limits_{j=1}^{5} g_j \cdot I_{hij}^{t-1}} \cdot I_{hi}^{t-1}$$

Pour tous les j où $I_{hi}^{t-1} > 0$.

Pour déterminer l'indice par rubrique de l'indice, on appliquera la formule suivante pour les canaux de distribution ayant disparu :

$$I_{h}^{t} = \frac{\sum\limits_{i=1}^{3} a_i \cdot I_{hi}^{t}}{\sum\limits_{i=1}^{3} a_i \cdot I_{hi}^{t-1}} \cdot I_{h}^{t-1}$$

Le calcul de l'indice en cas d'introduction de nouvelles observations de prix, de nouveaux groupes de communes ou de nouveaux canaux de distribution

Lorsqu'une nouvelle observation de prix r est introduite pour le mois sous revue, on lui attribue un indice élémentaire correspondant au niveau de l'indice moyen de l'agrégat qui lui est immédiatement supérieur:

$$I_{hijr}^{t} \equiv I_{hij}^{t}$$

On pose comme hypothèse que la nouvelle observation de prix aurait accusé dans le passé la même évolution de prix que l'agrégat supérieur, c'est-à-dire la moyenne des observations faites jusqu'alors dans le groupe de communes concerné à l'intérieur d'un canal de distribution.

De même, on attribue au nouveau groupe de communes ou au nouveau canal de distribution le niveau correspondant à l'agrégat qui lui est immédiatemment supérieur. Pour le nouveau canal de distribution, on pose que I_{hi}^t est égal à I_h^t et pour le nouveau groupe de communes, que I_{hij}^t est équivalent à I_{hi}^t.

La détermination des indices pour les sous-groupes et les groupes de marchandises ou de services et détermination de l'indice général

Les indices des diverses rubriques de l'indice calculés sont agrégés et pondérés pour obtenir les indices pour les sous-groupes et groupes de marchandises ou de services (I_f^t) ainsi que pour le niveau de l'indice général (I^t). Rappelons que les coefficients budgétaires sont déterminés à partir des résultats de l'enquête sur la consommation des ménages.

Les indices pour les sous-groupes et groupes de marchandises ou de services s'expriment ainsi:

$$I_f^t = \frac{\sum_{k=1}^{n_f} w_{fk} \cdot I_{fk}^t}{w_f}$$

$$\sum_{k=1}^{n_f} w_{fk} = w_f$$

où f désigne un sous-groupe ou un groupe de marchandises ou de services, k représente le numéro de la rubrique de l'indice à l'intérieur d'un sous-groupe f de marchandises ou de services, n_f indique le nombre de rubriques de l'indice dans le sous-groupe f de marchandises et w_{fk} représente le coefficient de pondération de la rubrique k de l'indice (parts des dépenses selon l'échantillon de marchandises et de services). Et l'indice général se calcule selon la formule suivante:

$$I^t = \sum_{h=1}^{n} w_h \cdot I_h^t;$$

$$\sum_{h=1}^{n} w_h = 1$$

où w_h représente le coefficient budgétaire de de la rubrique h.

2.3.3 La méthode des indices élémentaires dans la pratique

Dans la pratique, il y a quelques changements par rapport à la méthode théorique exposée ci-dessus.
Premièrement, l'indice des loyers est déterminé au moyen d'une méthode spécifique de pondération et de calcul. Deuxièmement, la pondération par canal de distribution n'est prise en compte que pour le groupe de l'alimentation. Pour ce qui est des autres groupes de marchandises ou de services, on calcule des moyennes non pondérées de tous les indices élémentaires par groupe de communes, et ces moyennes sont agrégées directement pour former l'indice par rubrique de l'indice:

$$I_h^t \equiv I_{hi.}^t$$

Et troisièmement, pour les prix relevés d'une manière centralisée, la pondération par groupe de communes tombe, de même que celle par canal de distribution. On utilisera alors la méthode suivante:

$$I_h^t \equiv I_{hij.}^t$$

2.3.4 Exemples numériques

Pour illustrer la méthode de calcul exposée ci-dessus, considérons l'exemple numérique fictif suivant. Supposons que l'on connaisse le prix de trois articles de la rubrique *lait entier en emballage* à la Coop à Berne. Pour chacun des trois articles (e_1, e_2 et e_3), on connaît leur prix à la période de base (t_0) et à la période courante (t_1). Les prix sont donnés dans le tableau suivant:

	t_0	t_1
e_1	1.75	1.80
e_2	1.70	1.70
e_3	1.80	1.90

Les indices élémentaires des observations individuelles de prix (I^t_{hije}) se calculent comme suit:

$$I^t_{hij1} = \frac{1.80}{1.75} \cdot 100 = 102.85$$

$$I^t_{hij2} = \frac{1.70}{1.70} \cdot 100 = 100.00$$

$$I^t_{hij3} = \frac{1.90}{1.80} \cdot 100 = 105.55$$

La deuxième étape correspond au calcul de l'indice de la rubrique *lait entier en emballage* pour le canal de distribution Coop à Berne:

$$I^t_{hij} = \frac{102.85 + 100.00 + 105.55}{3} = 102.80$$

Si l'on dispose pour les différents groupes de communes, pour le canal de distribution Coop, les indices pour la rubrique *lait entier en emballage*, on peut alors calculer l'indice national pour la Coop pour la rubrique *lait entier en emballage.*

groupe de communes	coefficients de pondération	I^t_{hij}
g_1 (Zurich)	0.2	105.00
g_2 (Bâle)	0.1	103.95
g_3 (Berne)	0.1	102.80
g_4 (Genève)	0.1	104.20
g_5 (autres)	0.5	102.30

$$I^t_{hi} \;=\; \frac{\sum_{j=1}^{5} g_j \cdot I^t_{hij}}{\sum_{j=1}^{5} g_j} = 103.24$$

Si l'on dispose de l'indice national de la rubrique *lait entier en emballage* par canal de distribution, on peut alors calculer l'indice national pour la rubrique *lait entier en emballage.*

canaux de distribution	coefficients de pondération	I_{hi}^t
a_1 (Migros)	0.2	104.25
a_2 (Coop)	0.15	103.24
a_3 (autres)	0.65	105.30

$$I_h^t = \frac{\sum_{i=1}^{3} a_i \cdot I_{hi}^t}{\sum_{i=1}^{3} a_i} = 104.78$$

L'indice général national est ensuite obtenu par agrégation des différents indices d'une rubrique de l'indice (I_h^t) à l'aide des coefficients budgétaires correspondants. Rappelons que ces coefficients budgétaires sont issus de l'enquête sur la consommation qui est réalisée auprès des ménages. Pour l'exemple, supposons que nous ayons trois groupes de produits: *lait entier en emballage* (I_1^t), *pain* (I_2^t) et *viande* (I_3^t). Les coefficients budgétaires et les indices sont donnés dans le tableau ci-dessous:

coefficients de pondération	I_h^t
0.31	$I_1 = 104.78$
0.40	$I_2 = 106.30$
0.29	$I_3 = 105.75$

$$I^t = \sum_{h=1}^{n} w_h \cdot I_h = 105.67$$

L'indice national s'établit ainsi à 105.67. Cet exemple numérique fictif, bien que simplifié, permet de comprendre la méthode pratique de calcul de l'indice suisse des prix à la consommation.

Nous avons vu précédemment que si l'on est en présence du cas où il y disparition d'observations de prix, il faut alors établir un rapport des moyennes des indices élémentaires du mois courant et du mois précédent, dans lequel n'entrent que les observations de prix relevées pour les deux périodes, et ensuite on enchaîne ce rapport à l'indice

par groupe de communes et canal de distribution déterminé au mois précédent.
Illustrons ce principe à l'aide de l'exemple numérique suivant.

Supposons que l'on connaisse le prix de quatre articles (e_1 à e_4). Pour chacun d'entre eux, on connait leur prix à la période de base (t_0), à la période t_1 et à la période t_2. L'observation du prix de l'article e_1 disparait à la période t_2. Exposons ci-après le calcul qui doit être fait pour calculer l'indice de la rubrique par groupe de communes et canal de distribution.
Les prix sont donnés dans le tableau suivant:

	t_0	t_1	t_2
e_1	5.-	7.-	–
e_2	6.-	8.-	7.-
e_3	9.-	9.-	8.-
e_4	6.-	8.-	8.-

Les indices élémentaires des prix au temps t_0 et t_1 se présente comme suit:

	t_0	t_1
e_1	100	140
e_2	100	133
e_3	100	100
e_4	100	133

L'indice par groupe de communes à l'intérieur d'un canal de distribution et d'une rubrique de l'indice se calcule comme suit:

$$I_{hi1} = (140 + 133 + 100 + 135) : 4 = 126.50$$

L'observation du prix de l'article e_1 disparait au temps t_2. En conséquence, seules les observations de prix de e_2 à e_4 subsistent. L'indice par groupe de communes à l'intérieur d'un canal de distribution et d'une rubrique de l'indice se calcule comme suit.

$$I_{hij2} = 7 : 8 \cdot 133 = 115.71$$

$$I_{hij3} = 8 : 9 \cdot 100 = 88.00$$

$$I_{hij4} = 8 : 8 \cdot 133 = 133.00$$

$$I_{hi1} = \frac{(115.71 + 88.00 + 133.00)}{(133.00 + 100.00 + 133.00)} \cdot 126.50 = 115.11$$

Le principe de calcul est le même en cas de disparition de canal de distribution ou de communes de l'indice.

2.3.5 La proposition générale de l'OFS pour le nouvel indice

La méthode de calcul reste fondamentalement la même que celle qui a prévalu jusqu'en avril 1993, à savoir que le calcul repose sur l'indice de Laspeyres des prix avec la méthode des indices élémentaires pour l'agrégation des prix individuels. Les principales nouveautés sont qu'il y a pondération non plus pour les quatre grandes villes seulement (Bâle, Berne, Genève et Zurich) et le reste du pays, mais que chaque commune participante se verra attribuer un poids de pondération. Ce dernier étant fondé sur la population de la macrorégion MS d'où provient la commune.

Quant aux pondérations des canaux de distribution, l'OFS prévoit des coefficients pour tout le panier-type avec une détermination variable de la pondération selon les groupes du panier-type. Et on ne considère plus seulement Migros et Coop, mais l'OFS tiendra compte des autres types de canaux de distribution. Les coefficients de pondération seront établis selon les données relatives à la structure du marché (données provenant d'analyses du marché, d'informations sur les branches économiques).

Chapitre 3

L'indice américain des prix à la consommation

3.1 Introduction

A travers ce chapitre, nous nous proposons de décrire le cas de l'indice des prix à la consommation des Etats-Unis. Rappelons que l'indice américain est le seul IPC à se composer d'un ensemble d' échantillons aléatoires, et pour lequel on calcule la variance. Il était donc essentiel que l'on examine en détail le plan d'échantillonnage utilisé ainsi que la technique d'estimation de la variance qui a été développée.

Nous présentons le plan d'échantillonnage, la méthode de calcul, ainsi que la méthode de l'estimation de la variance tels qu'ils sont appliqués par le Bureau of Labor Statistics (BLS). Cet office est responsable du calcul de l'IPC américain. Les informations contenues dans ce chapitre trouvent leur source dans la publication du BLS *"BLS Handbook of Methods"* (1988).

3.2 Le plan d'échantillonnage

Cette section est consacrée à l'étude descriptive des échantillons utilisés par le Bureau of Labor Statistics (BLS) des Etats-Unis pour construire l'indice des prix à la consommation.
Les échantillons sélectionnés permettent de construire l'IPC représentatif des prix payés pour les biens et services achetés par les consommateurs urbains américains. Comme dans le cas de la Suisse, on trouve les cinq échantillons suivants: un échantillon aréolaire, un échantillon de ménages, sélectionné dans chacune des aires choisies, un échantillon de points de vente où les ménages sélectionnés font leurs achats, un échantillon des biens et services achetés par les consommateurs, un échantillon d'habitations dans chaque aire urbaine choisie.

3.2.1 L'échantillon aréolaire

Pour des considérations statistiques, les Etats-Unis sont partagés en 1088 unités primaires d'échantillonnage ("Primary Sampling Units" ou PSU). Une PSU est définie comme étant un comté ("county") ou un groupe de comtés contigus. Si une PSU peut être considérée comme représentante des autres PSU, alors on parle de PSU non-auto-représentative ("non-self-representing PSU" ou NSR PSU). Si une PSU ne représente qu'elle-même, on est alors en présence d'une PSU auto-représentative ("self-representing PSU" ou SR PSU).
Les 1088 PSU sont combinées selon les sept caractéristiques suivantes, qui ont une étroite corrélation avec les mouvements de prix, à savoir:

1. la région, la taille, et s'il s'agit d'une aire métropolitaine ou non,
2. le revenu moyen en dividende et intérêts par unité d'habitation,
3. le gain et salaire moyen par unité d'habitation,
4. le pourcentage d'habitations chauffées à l'électricité,
5. le pourcentage d'habitations chauffées au mazout,
6. le pourcentage de population noire,
7. le pourcentage de retraités.

Le BLS a ainsi divisé les USA en 91 strates géographiques dont 31 strates auto-représentatives et 60 strates non-auto-représentatives. Les 31 strates auto-représentatives sont constituées d'aires métropolitaines de plus de 1,2 million d'habitants. Dans les 60 strates non-auto-représentatives, on distingue 22 PSU de taille moyenne, 24 PSU de petite taille et 14 PSU d'aires urbaines non métropolitaines. C'est le nombre d'habitants qui détermine les grandes et les petites PSU. Ce critère de taille varie selon les quatres grandes régions des USA: pour la région Nord-Est 500'000 habitants, pour la région Centre-Ouest 360'000 habitants, pour la région Sud 450'000 habitants et pour la région Ouest 300'000 habitants. Notons que les strates rurales ne sont pas prises en compte. Les 31 strates auto-représentatives sont composées de 34 SR PSU, c'est-à-dire 1 PSU par strate, plus 2 PSU supplémentaires pour New York et 1 PSU supplémentaire pour Los Angeles. Dans chacune des 60 strates non-auto-représentatives, on sélectionne une PSU. On obtient ainsi au total 94 PSU.

Avant de présenter un tableau récapitulatif, il est nécessaire d'expliquer à ce stade de l'exposé la notion de réplique, car cette notion est utilisée par le BLS lors du calcul de la variance de l'IPC. En effet, le plan d'échantillonnage américain est aléatoire à tous les niveaux et il permet donc d'estimer la précision de l'IPC en calculant sa variance. Pour pouvoir effectuer ce calcul, le BLS procède à au moins deux sélections, appelées répliques, d'articles et de points de vente à l'intérieur des SR PSU. Le BLS dispose d'au moins deux échantillons de prix pour chaque SR PSU et d'un échantillon pour chaque NSR PSU. Le BLS applique une technique d'estimation de la variance qui est basée sur l'utilisation de ces répliques; cette méthode sera décrite dans la section 3.4.

Le BLS constitue donc, à partir des 94 PSU, un ensemble de 134 répliques qu'il partage en deux groupes appelés "répliques paires" et "répliques impaires".

Une aire géographique peut être composée soit d'une seule SR PSU soit de plusieurs NSR PSU.

Dans le cas où l'aire n'est composée que d'une seule SR PSU, alors cette dernière n'est représentative que d'elle-même. Le calcul de la variance à l'intérieur de cette aire nécessite de constituer deux ou quatre répliques à partir de la SR PSU. Ces répliques sont ensuite partagées aléatoirement en un groupe de répliques appelé "répliques paires" et en

un second groupe appelé "répliques impaires".
Dans le cas où l'aire est composée de plusieurs NSR PSU, alors chaque NSR PSU est représentative de toutes les autres NSR PSU. On considère chaque NSR PSU comme une réplique en elle-même et elle doit alors être regroupée avec d'autres NSR PSU pour procéder au calcul de la variance.
Le tableau suivant résume les PSU incluses dans l'IPC ainsi que le nombre de répliques par PSU.

Strate	Nb. de PSU	Nb. de répliques
New York centre	1	4
New York, New Jersey banlieues	2	4
Los Angeles centre	1	4
Los Angeles banlieue	1	2
Chicago	1	4
Philadelphia	1	2
San Fransisco, Detroit	2	4
Grands SR PSU	9	18
Petits SR PSU	16	32
NSR PSU	60	60
TOTAL	94	134

3.2.2 L'échantillon des ménages

Dans le cadre de l'enquête sur les dépenses des consommateurs (Consumer Expenditure Survey = CES), deux échantillons de ménages sont sélectionnés. La population de référence correspond à la population civile non institutionnelle. C'est-à-dire que le BLS exclut la population qui a une institution comme chef de ménage, tel qu'un hôpital, un home ou encore une prison. La base de sondage de la population de référence est donnée par le recensement de la population.
Pour le CES, les Etats-Unis sont représentés par 109 PSU dont 91 correspondent aux PSU de l'indice des prix à la consommation et 18 PSU correspondant à des aires rurales non métropolitaines.
Le CES est divisé en deux enquêtes distinctes: l'enquête par carnet de compte et l'enquête par entretien. Pour chacune de ces deux enquêtes, un échantillon est sélectionné. Examinons tout d'abord l'enquête par carnet de compte.

L'enquête par carnet de compte

Cette enquête, où environ 8000 ménages sont contactés, se compose de deux parties. La première partie correspond à un questionnaire sur les caractéristiques démographiques et socio-économiques du ménage. Dans la seconde partie, on demande aux ménages de relever quotidiennement toutes leurs dépenses pour les articles de consommation courante et fréquemment achetés. Cette partie se déroule pendant deux semaines consécutives.

L'enquête par entretien

Le but de l'enquête par entretien est de collecter des données sur les types de dépenses effectuées dans un intervalle de trois mois ou plus. Il s'agit d'observer des dépenses, tel l'achat d'une habitation, d'une automobile, et les dépenses intervenant à intervalle régulier, tel le loyer, les factures des prestations des services publics, les primes d'assurances.

Lors de l'enquête par entretien, environ 9000 ménages sont contactés. Chaque ménage est interrogé une fois par trimestre pendant cinq trimestres consécutifs. Après le cinquième trimestre, le ménage est retiré de l'échantillon et est remplacé par un autre ménage. Ainsi, chaque trimestre, quelque 20% des ménages sont renouvelés. Notons que cette technique d'échantillonnage est appelée échantillonnage par rotation.

Lors du premier trimestre, un entretien initial a lieu et il porte sur les caractéristiques démographiques et socio-économiques des familles. Un inventaire des biens durables les plus importants est établi et on s'informe sur les dépenses du ménage. Lors des entretiens ultérieurs, les ménages sont interrogés sur les composantes majeures de leurs dépenses, sur les points de vente où ils font leurs achats et sur leurs gains et salaires.

Dans le contexte de l'indice des prix à la consommation, les résultats issus de ces deux enquêtes servent à sélectionner le nouveau panier de la ménagère, à déterminer l'importance relative des différents articles et, ainsi, à calculer les coefficients budgétaires pour le panier de la ménagère.

3.2.3 L'échantillon des articles

Les articles inclus dans l'indice des prix à la consommation sont regroupés en huit groupes majeurs (y compris habitation). Ceux-ci sont les suivants: alimentation et boissons, combustible et prestations des services publics, services et aménagement de l'habitation, habillement et entretien, transports, soins médicaux, loisirs, autres biens et services.

Le BLS a établi quatre niveaux de classification pour les articles. Le premier niveau correspond à celui des groupes majeurs. Le deuxième niveau est celui des classes de dépenses, ces dernières étant au nombre de 69 ("expenditure classes" (EC)). Le troisième niveau se compose des 207 strates d'articles ("item strata" (IS)). Enfin, le dernier niveau est formé de 364 articles au niveau des relevés ("entry level item" (ELI)). L'exemple suivant permet de mieux illustrer ces niveaux de classification:

- Alimentation et boissons
    - EC: 01 Céréales et produits céréaliers
        - IS: 0101 Farine et mélanges de farines préparés
            - ELI: 01011 Farine
            - ELI: 01012 Mélanges de farines préparés
        - IS: 0102 Céréales
            - ELI: 01021 Céréales
        - IS: 0103 Riz, pâtes et semoule
            - ELI: 01031 Riz
            - ELI: 01032 Macaronis, produits similaires et semoule

Chaque année, dans 20% des PSU, de nouveaux échantillons d'articles sont sélectionnés. Il y a ainsi un échantillonnage par rotation et, en cinq ans, une rotation complète est réalisée. Chaque année, quatre listes exhaustives, une par grande région, des biens et services consommés sont dressées à partir des données issues des deux dernières enquêtes sur les dépenses des consommateurs. Puis, à partir de ces quatre listes, un échantillon indépendant d'ELI est tiré pour chacune des quatre régions.

L'échantillon est tiré pour chaque strate d'articles et pour chacune des répliques-PSU soumise à la rotation. Ainsi, sur une période de cinq ans, 134 échantillons d'ELI sont sélectionnés au niveau national. L'échantillon d'ELI est sélectionné selon une procédure d'échantillonnage systématique. Chaque ELI, à l'intérieur d'une strate d'articles, a une probabilité de sélection proportionnelle aux dépenses relatives pour l'ELI à l'intérieur de la strate d'articles pour la région concernée. Le BLS a établi le nombre de sélections par strate d'articles, par réplique-PSU comme suit: alimentation et boissons (73), combustible et prestations des services publics (12), services et aménagement de l'habitation (66), habillement et entretien (47), transports (34), soins médicaux (18), loisirs (27), autres biens et services (21).

Les prix sont relevés mensuellement pour les catégories suivantes: alimentation, énergie, loyer, équivalence-locative. Pour les autres articles, les prix sont collectés tous les deux mois. Notons que dans les cinq plus grandes zones géographqiues, les prix sont collectés mensuellement pour tous les biens et services. Les plus grandes zones sont: New York, Los Angeles, Chicago, Philadelphia et San Francisco. Chaque réplique-PSU est attribuée à un mois pair ou impair pour les relevés de prix.

3.2.4 L'échantillon des points de vente

Une enquête sur les points de vente (Point-of-Purchase Survey = POPS) est réalisée chaque année sur une période de quatre à six semaines dans environ 20% des PSU. Le but de cette enquête est de disposer de données sur les points de vente où les ménages urbains effectuent leurs achats pour des groupes définis de biens et services. La population de référence pour cette enquête est la même que celle participant à l'enquête sur les dépenses des consommateurs. C'est à partir des données issues de cette enquête que le plan d'échantillonnage des points de vente sera déterminé. Et c'est au sein de ces points de vente que seront relevés les prix des biens et services. Cette enquête fournit le plan d'échantillonnage pour environ deux tiers des articles de l'indice des prix à la consommation. Les articles qui ne sont pas couverts par l'enquête sont le loyer, l'équivalence-locative, le gaz naturel et l'électricité.

Depuis 1985, les ménages sont choisis selon un plan d'échantillonnage identique à celui utilisé pour l'enquête sur les dépenses des consommateurs. Le recensement de la population sert de base de sondage.

L'entretien auprès des ménages se déroule en deux parties. La première partie consiste à recueillir des informations sur les caractéristiques socio-économiques et démographiques des ménages. Cela permet de déterminer le nombre d'unités de consommation qui résident dans l'habitation et qui doivent être interrogées. Une unité de consommation réunit tous les membres d'une habitation qui sont soit liés par un lien de parenté soit qui ne sont pas parents mais financièrement dépendants l'un de l'autre pour les dépenses majeures, tels que l'alimentation ou le loyer. Dans la seconde partie, on cherche à savoir si les unités de consommation achètent les catégories de biens et services à l'intérieur d'une période donnée. Cette période est appelée période de rappel. Cette dernière varie d'une semaine à cinq ans. La période de rappel d'une ou deux semaines concerne des articles de consommation courante, comme l'alimentation, le tabac, la benzine; la période de rappel de cinq ans touche des produits tels que l'automobile, le revêtement des sols.

Les biens et services sont regroupés en catégories d'échantillonnage. Les articles sont regroupés en une seule catégorie s'ils sont vendus dans un même point de vente. Ces catégories, appelées catégories-POPS, consistent donc en un ELI ou en une combinaison d'ELI. Par exemple, la catégorie-POPS 107 *Produits laitiers, y compris oeufs*, est composée de: *oeufs, lait entier frais, autres laits entiers frais et crème, beurre, autres produits laitiers, fromages, crème glacée et produits similaires.* Il existe deux listes différentes de catégories-POPS qui sont utilisées lors des entretiens. Chacune des listes est utilisée dans la moitié des ménages interrogés. Chaque liste compte 147 catégories-POPS. La plupart des catégories sont incluses dans les deux listes. Quelques catégories ayant une courte période de rappel ne sont incluses que dans une seule liste. Le répondant doit indiquer pour chacune des catégories-POPS de la liste si ses achats sont faits durant la période de rappel, et si c'est le cas, il doit donner les noms et lieux d'achat ainsi que le montant de la dépense.

Chaque année, un nouvel échantillon de points de vente est sélectionné dans 20% des PSU. Durant l'année suivant l'enquête, le BLS initie les nouveaux points de vente. Le nombre de points de vente sélectionnés pour chaque catégorie-POPS, pour chaque groupe majeur et pour chacun des groupes PSU est donné dans le tableau suivant:

Points de vente par catégories-POPS	1	2	3	4	5	6	7	8	9	10
Alimentation ...	6	7	6	6	8	9	9	4	2	6
Combustible ...	7	8	4	4	7	8	6	4	2	6
Service et ...	1	1	1	1	1	1	1	1	1	1
Habillement ...	2	2	2	2	2	3	3	2	1	2
Transport	2	4	3	3	3	4	4	3	1	3
Soins médicaux	3	3	3	3	3	5	3	3	1	4
Loisirs	1	1	1	1	1	1	1	1	1	1
Autres biens ...	1	1	1	1	1	1	1	1	1	1

Le BLS a établi 173 catégories-POPS, ces dernières se répartissent entre les différents groupes majeurs comme suit:

Alimentation et boissons	17
Combustible et prestation..	8
Services et aménagement..	38
Habillement et entretien	31
Transport	27
Soins médicaux	8
Loisirs	23
Autres biens et services	21

On arrive à un total de 34'790 points de vente visités. Cependant, chaque mois ce sont en réalité 25'000 points de vente environ qui sont visités. La différence s'explique par le fait que la plupart des PSU ne participent au relevé qu'un mois sur deux.

Les procédures d'échantillonnage des points de vente

Lorsqu'un échantillon d'EL1 est sélectionné, cela permet d'identifier une catégorie-POPS spécifique pour la sélection d'un point de vente. Par exemple, l'EC "Volailles" (Poultry) est subdivisé en IS et en ELI et, pour les différents ELI, nous trouvons différentes catégories-POPS.

Illustrons cela par le schéma suivant:
EC: 06 Volailles
IS: 0601 Poulet entier frais
ELI: 06011 Poulet entier frais
IS: 0602 Morceaux de poulet frais ou surgelés
ELI: 06021 Morceaux de poulet frais ou surgelés
IS: 0603 Autres volailles
ELI: 06031 Autres volailles
Les catégories-POPS correspondantes pour ces ELI sont:
POPS: 106 Viandes et volailles
06011 Poulet entier frais
06021 Morceaux de poulets frais ou surgelés
06031 Autres volailles

Pour les SR PSU, les ménages de l'échantillon pour le POPS sont divisés en au moins deux groupes indépendants. Le nombre de groupes correspond au nombre de répliques des SR PSU. Cela fournit au minimum deux bases de sondage pour les points de vente. Une réplique-PSU est constituée par une sélection d'ELI et de points de vente pour toutes les strates d'articles au sein d'un PSU. Les échantillons des points de vente sont sélectionnés d'une manière indépendante pour chaque PSU, réplique-PSU et catégorie-POPS en utilisant une procédure d'échantillonnage systématique. Chaque point de vente de la base de sondage a une probabilité de sélection proportionnelle au montant des dépenses reportées dans le POPS. Tous les ELI sélectionnés lors de la procédure d'échantillonnage des articles, pour une réplique-PSU donnée, doivent voir leurs prix relevés. Les prix doivent être relevés dans chaque point de vente. Si un point de vente est sélectionné plusieurs fois, une augmentation comparable est faite quant au nombre de prix relevés dans le point de vente.

Les procédures de sélection à l'intérieur des points de vente
Pour chacun des ELI désignés pour le relevé de prix dans un point de vente de l'échantillon, un représentant du BLS doit sélectionner un article spécifique à l'intérieur du point de vente. Pour cela, il utilise une technique de sélection probabiliste à plusieurs degrés.

Le représentant va suivre une démarche composée de plusieurs étapes. La première étape consiste à identifier tous les articles inclus dans la définition EL1 et qui sont en vente. Lors de la deuxième étape, le repésentant du BLS va regrouper les articles par caractéristiques communes, comme la taille, le type d'emballage, etc. A l'étape suivante, le représentant du BLS, avec l'assistance du répondant du point de vente, va attribuer des probabilités de sélection à chacun des groupes. Ces probabilités de sélection sont proportionnelles aux ventes des articles inclus dans chaque groupe. Pour estimer la proportion des ventes, il dispose de quatre alternatives. Ces dernières sont, par ordre de préférence: obtenir directement les proportions des ventes par le répondant du point de vente, ou bien ranger les groupes par importance des ventes avec l'aide du répondant et estimer les proportions, ou bien utiliser l'espace de vente pour estimer les proportions, ou encore utiliser des probabilités de sélection égales.

Après avoir attribué des probabilités de sélection, le représentant du BLS utilise une table de nombres aléatoires pour sélectionner un groupe. Une fois un groupe sélectionné, il faut alors identifier tous ses articles puis former des sous-groupes sur la base des caractéristiques communes des articles. Le représentant du BLS va alors attribuer des probabilités de sélection à chaque sous-groupe. Il utilise ensuite une table de nombres aléatoires afin de sélectionner un sous-groupe. Ce processus est répété jusqu'à l'identification d'un seul article.

Cette procédure peut être illustrée par l'exemple suivant. Supposons que le repésentant du BLS doive sélectionner un canapé dans un des points de vente. Il va tout d'abord énumérer tous les articles correspondant à l'ELI-sofas en vente. Puis il va les regrouper selon leurs caractéristiques communes.

L'exemple fictif suivant illustre ce principe. Posons que le représentant regroupe les canapés en deux groupes: *canapé cuir*, *canapé tissu*. A chacun des groupes, il va attribuer une probabilité de sélection. Supposons que cette dernière soit proportionnelle au montant des ventes. Il va alors sélectionner aléatoirement un groupe, par exemple le groupe *canapé cuir*. Le représentant énumérera alors tous les canapés cuirs en vente et formera de nouveaux groupes, tel que: *canapé cuir 3 places*, *canapé cuir 2 places*. Une probabilité de sélection sera attribuée aux groupes nouvellement formés et l'un d'entre eux sera sélectionné. Si le *canapé cuir 2 places* est choisi, et supposons qu'il n'y ait qu'un seul modèle, alors le représentant du BLS va relever son prix. Durant une période de cinq ans, ce sera le prix de cet article qui sera observé pour l'ELI-sofas.

3.2.5 L'échantillon des habitations

L'échantillon d'habitations va servir à calculer deux indices liés à l'habitation, à savoir l'indice relatif aux babitations louées, appelé indice du loyer, et l'indice des habitations occupées par leur propriétaire, appelé indice de l'équivalence-locative.
L'échantillon est stratifié en grappes et sélectionné selon une procédure d'échantillonnage systématique. L'échantillon est composé de quelque 40'000 habitations louées et 20'000 habitations occupées par leur propriétaire. La base de sondage est constituée par les données issues du recensement des habitations. En ce qui concerne les habitations construites après le recensement, elles sont prises en compte par l'intermédiaire des permis de construction accordés depuis cette date.

La stratification

Deux variables sont utilisées afin d'établir la stratification. Il s'agit du niveau moyen du loyer, qui compte trois niveaux, et du pourcentage d'habitations louées par rapport au nombre total d'habitations; cette variable compte six niveaux. Les deux variables choisies pour la stratification sont liées aux variations que subissent les loyers. Dix-huit strates sont ainsi formées.
Le BLS stratifie au niveau le plus détaillé des aires publiées par le Bureau du recensement à l'intérieur des 94 PSU. Le Bureau du recensement fournit des données par groupes de quartiers et par quartiers.

La sélection de l'échantillon à l'intérieur des strates

Le BLS choisit un échantillon de grappes à l'intérieur de chaque strate à l'aide d'une méthode de sélection systématique et avec une probabilité proportionnelle à la taille de la grappe. Une fois une grappe choisie, le BLS la fractionne à son tour en un certain nombre de grappes (ce nombre dépend de la taille de la grappe) et sélectionne à nouveau une grappe aléatoirement.

Une procédure de sélection sur le terrain détermine le choix final des habitations de l'échantillon. Le représentant du BLS qui travaille sur le terrain doit énumérer toutes les habitations de la grappe choisie.

Puis, à l'aide de feuilles de sélection qui sont préparées par le BLS, le représentant va sélectionner les habitations à soumettre à l'enquête et leur type d'occupation désiré. Notons que les feuilles de sélection sont préparées par le BLS pour chaque grappe, en tenant compte de la proportion des habitations louées et des habitations occupées par leur propriétaire. Cette proportion est issue du recensement des logements. Cela permet de déterminer combien d'habitations sont à observer dans la grappe et quel est le pourcentage des habitations louées et des habitations occupées par leur propriétaire. Avant que le représentant ne contacte les répondants des babitations pour la première fois, il les range dans l'une des quatre catégories suivantes:

1. Soumettre à la procédure de sélection mais initier seulement si l'habitation est louée.

2. Soumettre à la procédure de sélection mais initier seulement si l'habitation est occupée par le propriétaire.

3. Soumettre à la procédure de sélection et initier quel que soit le type d'occupation de l'habitation.

4. Ne pas contacter, ni pour la procédure de sélection ni pour l'initiation.

Le représentant prend contact avec un répondant pour chaque habitation soumise à la procédure de sélection, obtient son type d'habitation actuelle et détermine si l'habitation peut être retenue dans l'échantillon. Si l'habitation est retenue, alors le représentant procède à l'initation; cela signifie qu'un premier entretien avec le répondant de l'habitation est réalisé.

3.3 La méthode de calcul

3.3.1 La méthode pratique du calcul de l'IPC

En pratique, le BLS utilise la formule suivante pour calculer l'indice des prix à la consommation des Etats-Unis:

$$I_{t,0} = \frac{\sum_{H}\sum_{Z} A_{HZ} I_{HZt,0}}{A}$$

où A est la somme de tous les poids d'agrégation pour tous les articles et toutes les aires géographiques de l'IPC. $I_{HZt,0}$ représente l'indice à la période t pour un agrégat d'aires H et un agrégat d'articles Z. Il peut s'agir, par exemple, de l'indice du groupe majeur "Alimentation et boissons" pour la région ouest des USA en mai 1992.
Au niveau inférieur, on calcule l'indice $I_{HZt,0}$ de la manière suivante:

$$I_{HZt,0} = \frac{\sum_{h}\sum_{z} A_{hz} I_{hzt,0}}{A_{HZ}}$$

où A_{hz} représente le poids d'agrégation pour l'aire h et la strate d'articles z et A_{HZ} représente le poids d'agrégation pour l'aire agrégée H et l'agrégat d'articles Z et $I_{hzt,0}$ correspond à l'indice de l'aire h et la strate d'articles z. Illustrons cela à l'aide de l'exemple suivant: $I_{hzt,0}$ peut représenter l'indice de la strate d'articles "céréales et produits céréaliers" de l'aire de San Fransisco en mai 1992.

Au niveau z et h, l'IPC est calculé par enchaînement. L'indice du mois précédent ($I_{hzt-1,0}$), pour chaque aire (h) et chaque strate d'articles (z), est multiplié par un estimateur ($R_{hzt,t-1}$) du changement de prix relatif entre le mois courant et le mois précédent. Ceci permet d'obtenir l'indice du mois courant (I_{hzt}) par aire et par strate d'articles. On peut écrire cet indice de la façon suivante:

$$I_{hzt,0} = I_{hzt-1,0} \cdot R_{hzt,t-1}$$

Par exemple, on obtient $I_{hzt,0}$ qui est, supposons, l'indice de la strate d'articles "céréales et produits céréaliers" pour l'aire h au temps t en multipliant l'indice pour la strate d'articles "céréales et produits céréaliers" pour l'aire h au temps 0 par le changement du prix de la strate d'articles "céréales et produits céréaliers" pour l'aire h au cours du mois t.

La forme de l'estimateur $R_{hzt,t-1}$ dépend des procédures utilisées pour sélectionner les échantillons des points de vente et d'articles. Si les échantillons sont sélectionnés avec une probabilité proportionnelle à la quantité pour chacune des unités, alors l'estimateur $R_{hzt,t-1}$ est donné par:

$$R_{hzt,t-1} = \frac{\sum_{i \in z} W_{hi} P_{hi,t}}{\sum_{i \in z} W_{hi} P_{hi,t-1}}$$

où W_{hi} représente l'inverse de la probabilité de sélection de l'article soumis au relevé, $P_{hi,t}$ le prix de l'article soumis au relevé au temps t et $P_{hi,t-1}$ le prix de l'article soumis au relevé au temps $t-1$.

Si l'échantillon de points de vente et les articles sont sélectionnés avec des probabilités proportionnelles aux dépenses, l'estimateur de $R_{hzt,t-1}$ s'exprime par:

$$R_{hzt,t-1} = \frac{\sum_{i \in z} W_{hi} P_{hi,t} / P_{hia}}{\sum_{i \in z} W_{hi} P_{hi,t-1} / P_{hia}}$$

où P_{hia} est un estimateur du prix de l'article sélectionné pour la période a, correspondant aux dépenses faites au sein des points de vente de l'échantillon.

3.3.2 Les indices du loyer et de l'équivalence-locative

Les indices du loyer et de l'équivalence-locative mesurent le changement du coût de l'habitation pour les locataires et les propriétaires. Les données sont issues de l'enquête sur les logements.

L'indice du loyer

Mensuellement, le BLS estime l'indice du loyer. Pour ce faire, il utilise l'indice du loyer du mois précédent et l'indice du loyer établi six mois plus tôt; le BLS utilise également une mesure du changement du loyer sur 1 mois et sur 6 mois estimée à partir de l'enquête sur les habitations.

Mesure du changement du loyer sur 1 mois: Cette mesure s'estime par le rapport de la somme des loyers du mois courant, ajustés pour le vieillissement subi durant un mois, et la somme des loyers ajustés pour le mois précédent.

Mesure du changement du loyer sur 6 mois: Cette mesure s'estime par le rapport de la somme des loyers du mois courant, ajustés pour le viellissement subi pendant les six derniers mois, et la somme des loyers ajustés pour les six derniers mois. L'indice du loyer du mois courant est une moyenne pondérée de l'indice du loyer du mois précédent, ajusté par la mesure du changement du loyer sur 1 mois, et l'indice du loyer pour le sixième mois antérieur ajusté par la mesure du changement du loyer sur 6 mois.

Avant de donner les estimateurs du changement du loyer pour 1 mois et 6 mois, posons: S_1 est l'ensemble des unités d'habitation louées qui ont été interrogées au temps t et qui sont comparables à celles qui ont été interrogées au temps $t-1$, et S_6 est l'ensemble des unités d'habitation louées qui ont été interrogées au temps t et $t-1$. r_{it} est le loyer de l'unité i au temps t et a_i est le facteur lié à la diminution de la qualité du logement à cause du vieillissement entre le temps t et $t-1$.
W_{i1} et W_{i6} représentent les probabilités de sélection de l'unité i.
Alors nous pouvons écrire les estimations du changement du loyer comme étant:

$$\begin{aligned} R_{t,t-1} &= \frac{\sum_{i\in S_1}(r_{it}+a_i r_{i,t-1})W_{i1}}{\sum_{i\in s_6}(r_{i,t-1}W_{i1})} \\ R_{t,t-6} &= \frac{\sum_{i\in S_6}(r_{it}+6a_i r_{i,t-6})W_{i6}}{\sum_{i\in s_6}(r_{i,t-6}W_{i6})} \end{aligned}$$

On peut exprimer l'indice mensuel du loyer I_t de la manière suivante:

$$I_t = A(I_{t-1}R_{t,t-1}) + (1-A)(I_{t-6}R_{t,t-6})$$

où $A = 0.65$; ce chiffre est issu d'études de simulation réalisées par le BLS et qui ont montré qu'il permettait de minimiser l'écart quadratique moyen. I_{t-1} est l'indice du loyer du mois précédent et I_{t-6} est l'indice du loyer semestriel.

L'indice des logements occupés par leur propriétaire ou l'indice de l'équivalence-locative

Le BLS calcule l'équivalence-locative en estimant le loyer implicite (m_j) des propriétaires pour chaque unité j de l'échantillon. La valeur initiale est estimée à partir de la réponse obtenue à la question suivante: " Si cette habitation (maison ou logement) était louée, quel serait le montant de la location, y compris l'entretien mais sans les charges (mazout, gaz, électricité,...) et sans l'aménagement du logement (réfrigérateur, revêtement,...) ?".

A cette question, soit le propriétaire est capable de répondre soit il ne le peut pas, et c'est alors le BLS qui estime le loyer par comparaison avec d'autres habitations du même type. Le BLS attribue à chaque unité-propriétaire j un ensemble de locataires Q_j; cet ensemble est constitué en se basant sur l'emplacement géographique à l'intérieur d'une PSU, sur le type de structure et caractéristiques communes. Le BLS essaie tout d'abord de trouver des locataires (Q_j) qui concordent parfaitement avec l'unité-propriétaire (j). Si le BLS ne trouve pas d'ensemble de locataires correspondant à l'unité-propriétaire, les contraintes sont assouplies jusqu'à l'obtention d'un ensemble d'unités-locataires.

Le loyer implicite (m_{jt}) pour le propriétaire j au temps t est estimé à partir du loyer implicite au temps $t-6$ et du changement moyen dans le loyer des unités Q_{j6}, où Q_{j6} sont des sous-ensembles de locataires dans Q_j qui ont des loyers comparables au temps t et $t-6$ et Q_{j1} a la même signification mais au temps t et $t-1$.

On peut exprimer $m_{j,t}$ de la façon suivante:

$$m_{j,t} = m_{j,t-6} \sum_{j \in Q_{j6}} \left(\frac{P_{i,t}}{P_{i,t-6}} + 6a_i \right) / n_{j6}$$

où $P_{i,t}$ est le loyer de l'unité-locataire i au temps t, a_i est le facteur d'ajustement du vieillissement et n_{j6} est le nombre d'unités-location dans Q_{j6}.

Le loyer implicite au temps $t-1$, $m_{j,t-1}$, peut être exprimé de la même façon, à savoir:

$$m_{j,t-1} = m_{j,t} \sum_{j \in Q_{j1}} \left(\frac{P_{i,t-1}}{P_{i,t} + a_i P_{i,t-1}} \right) / n_{j1}$$

Une fois ces deux loyers implicites calculés, le BLS estime l'indice de l'équivalence-locative du mois courant. Le BLS calcule des estimateurs du changement de prix des habitations des propriétaires pour 1 mois et 6 mois. Posons encore que S_1 est l'ensemble des unités-propriétaires ayant un loyer implicite au temps t et $t-1$ et S_6 est l'ensemble des unités-propriétaires ayant un loyer implicite au temps t et $t-6$.
Alors l'estimateur du changement de prix sur 1 mois et sur 6 mois peut être défini par:

$$R_{i,t-1} = \frac{\sum_{i \in S_1} m_{j,t} W_{j1}}{\sum_{i \in S_1} m_{j,t-1} W_{j1}}$$

$$R_{i,t-6} = \frac{\sum_{i \in S_6} m_{j,t} W_{j6}}{\sum_{i \in S_6} m_{j,t-1} W_{j6}}$$

où W_{i1} et W_{i6} représentent les probabilités de sélection des unités-propriétaires.
Enfin, l'estimateur final de l'indice de l'équivalence-locative au mois t (I_t) est donné par:

$$I_t = A(I_{t-1}R_{t,t-1}) + (1-A)(I_{t-6}R_{t,t-6})$$

où $A = 0.65$ et I_{t-1} représente l'indice de l'équivalence-locative au mois $t-6$.

3.4 L'estimation de la variance

3.4.1 Introduction

Au cours du premier chapitre, nous avons vu que l'un des avantages importants de l'échantillonnage aléatoire est que l'on peut mesurer l'erreur d'échantillonnage. Dans le cas des USA, le plan d'échantillonnage aléatoire de l'IPC permet de calculer l'erreur; pour ce faire, le BLS procède à au moins deux sélections (les répliques) d'articles et de points de vente à l'intérieur d'une aire géographique de l'indice.
Au moins deux relevés de prix dans chaque PSU auto-représentative et un relevé dans chaque PSU non-auto-représentative doivent être effectués. Une fois ces données disponibles, qui représentent tous les niveaux du plan d'échantillonnage, des techniques d'estimation de la variance basées sur des méthodes de réplication sont utilisées.

3.4.2 L'estimation de la variance de l'IPC américain jusqu'en 1986

Dès 1978, il existe une structure de réplique au niveau des articles dont le prix est relevé. Pour chaque aire géographique, deux estimateurs de l'indice des prix à la consommation étaient construits, en utilisant les répliques.
La différence au carré de ces indices fournissait un estimateur de la variance de l'indice. Il faut noter que ces estimateurs de la variance sont conditionnels aux valeurs des dépenses des ménages pour la période de base, c'est-à-dire aux coefficients budgétaires.
Pour les biens et services, chaque relevé de prix était sélectionné indépendamment pour chaque réplique.

Pour les habitations, chacune d'entre elles était attribuée soit à la réplique paire (A) soit à la réplique impaire (B) dans le cas des PSU-AR. Pour les PSU-NAR, l'indice PSU déterminait la réplique pour chacune des unités.
L'indice avant 1987 était calculé par enchaînement. Un estimateur des dépenses pour le mois précédent, pour chacune des strates d'articles, était multiplié par le prix relatif pour fournir un estimateur des dépenses du mois courant pour la strate d'articles.
Les valeurs des dépenses pour les strates d'articles, C_{zt}, étaient alors agrégées et comparées au total des dépenses. Pour une seule strate d'articles, C_{zt} se calculait comme suit:

$$C_{zt} = C_{zt-1} \cdot R_{zt,t-1}$$

et

$$I_{zt} = \frac{C_{zt}}{C_{z0}} \cdot 100$$

Pour chacune des répliques, le calcul se faisait d'une manière identique. Les estimateurs de la variance conditionnels aux valeurs de dépenses pour les strates d'articles de la période de référence, pour tous les articles (ou une partie des articles) au niveau national (ou local) étaient calculés somme suit:

$$U_{ht,0} = \frac{\sum_{z=1}^{Z} C_{zt}}{\sum_{z=1}^{z} C_{t0}}$$

où le numérateur correspond à la somme des C_{zt} pour les strates d'articles considérées pour l'aire h au temps t et le dénominateur correspond à la somme des C_{t0} pour les strates d'articles considérés pour l'aire h au temps t. $U^A_{ht,0}$ et $U^B_{ht,0}$ se calculent de la même manière que $U_{ht,0}$ et correspondent à la réplique paire (A) et à la réplique impaire (B).
Une formule de la variance pour l'indice $I_{t,0}$ est donnée par:

$$\sigma^2(I_{t,0}) = \sum_h \frac{100^2}{2}((U^A_{ht,0} - U_{ht,0})^2 + (U^B_{ht,0} - U_{ht,0})^2)$$

Le changement de prix entre t et $t-m$ est donné par:

$$I_{t,t-m} = \frac{I_{t,0}}{I_{t-m,0}}$$

L'estimateur de la variance de $I_{t,t-m}$ est exprimé par:

$$\begin{aligned}\sigma^2(I_{t,t-m}) &= \frac{1}{(I_{t-m,0})^2}(\sigma^2(I_{t,0}) + (I_{t,t-m})^2\sigma^2(I_{t-m,0}) \\ &\quad -2I_{t,t-m}Cov(I_{t,0}, I_{t-m,0}))\end{aligned}$$

La covariance entre I_t et I_{t-m} se calcule comme suit:

$$\begin{aligned}Cov(I_{t,0}, I_{t-m,0}) &= \sum_h \frac{100^2}{2}((U^A_{ht,0} - U_{ht,0})(U^A_{ht-m,0} - U_{ht-m,0}) \\ &\quad +(U^B_{ht,0} - U_{ht,0})(U^B_{ht-m,0} - U_{ht-m,0})\end{aligned}$$

3.4.3 L'estimation de la variance de l'IPC américain dès 1987

A partir de 1987, les estimateurs incorporent directement la contribution à la variance due aux dépenses des consommateurs (poids de pondération). Dès 1987, les coefficients budgétaires utilisés dans le calcul de l'indice sont dérivés de l'enquête sur la consommation de 1982-1984. Les estimateurs de la variance dépendent de la structure d'agrégation et de la structure de réplique de l'indice. L'utilisation de répliques donne un moyen de mesurer la variation totale de l'indice à partir des indices calculés sur des sous-échantillons de l'échantillon complet. L'échantillon complet, pour chaque indice, comprend au moins deux répliques, une moitié étant désignée par réplique impaire et l'autre moitié par réplique paire.

Les estimateurs de la variance de l'indice

Pour estimer la variance d'un indice, considérons l'indice, IX, pour un agrégat d'articles, I, pour un agrégat d'aires géographiques, M, au mois t, et qui s'exprime par:

$$IX(I,M,t) = IX(I,M,p) \cdot \frac{WI(I,M,t)}{WI(I,M,p)}$$

où $IX(I,M,p)$ est l'indice pivot, qui habituellement vaut 100 et joue comme une constante de normalisation.
$WI(I,M,t)$ représente un indice pondéré. Ce dernier est égal à la somme sur l'agrégat d'aires M du produit entre poids d'agrégation des aires et leur indice.

$$WI(I,m,f,t) = AW(I,m,f) \cdot IX(I,m,f,t)$$

où $AW(I,m,f)$ représente le poids d'agrégation pour l'échantillon complet et $IX(I,m,f,t)$ l'indice pour le mois t pour l'échantillon complet. Pour les différentes répliques, on procède de manière analogue.
Si les variances et covariances de $WI(I,M,t)$ et $WI(I,M,p)$ sont connues, alors la variance $IX(I,M,t)$ peut être estimée par une approximation à l'aide des séries de Taylor.

Les composantes de ces variances sont estimées comme exposé ci-dessous. Posons que:

$$WI(I,M,t) = \sum_{m \in M} WI(I,m,t)$$

où $WI(I,m,t)$ est l'indice pondéré pour l'aire m dans l'agrégat d'aires M.
Alors:

$$\begin{aligned} Var(WI(I,M,t)) &= \sigma^2_{WI}(I,M,t) \\ &= \sum_{m \in M} \sigma^2_{WI}(I,m,t) + \sum_{m \in M} \sum_{m' \neq M} Cov_A(I,m,m',t) \end{aligned}$$

où $\sigma^2_{WI}(I,m,t)$ est la variance de l'indice pondéré et $Cov_A(I,m,m',t)$ est la variance entre aires de l'indice au temps t.
La variance des indices pondérés est estimée par:

$$\begin{aligned} \sigma^2_{WI(I,m,t)} &= \frac{1}{2}((WI(I,m,impaire,t) - WI(I,m,f,t))^2 \\ &\quad +(WI(I,m,paire,t) - WI(I,m,f,t))^2) \end{aligned}$$

où $WI(I,m,impaire,t)$ et $WI(I,m,paire,t)$ sont les indices pondérés pour les répliques (r) paires et impaires.

$WI(I,m,impaire,t)$ et $WI(I,m,paire,t)$ sont définis par:

$$WI(I,m,impaire,t) = \frac{2}{NR(m)} \sum_{rimpaire} AW(I,m,r,t) \cdot IX(I,m,r,t)$$

et

$$WI(I,m,paire,t) = \frac{2}{NR(m)} \sum_{rpaire} AW(I,m,r,t) \cdot IX(I,m,r,t)$$

où $NR(m)$ dénotent le nombre de répliques dans l'aire m. Chaque aire appartient à l'une des quatre grandes régions. Chaque région est divisée en deux aires majeures, l'une est composée des indices des aires auto-représentatives (A) et l'autre est composée des indices des aires non-auto-représentatives ($non-A$). En tout, il y a donc huit aires majeures. Pour un agrégat plus large d'indices, des estimateurs des covariances entre indices pour chaque paire m et m' des différents indices dans la même aire majeure sont donnés par:

$$\begin{aligned} Cov_A(I,m,m',t) &= \frac{1}{2}((WI(I,m',impaire,t) - WI(I,m',f,t)) \\ &\quad \cdot(WI(I,m,impaire,t) - WI(I,m,f,t)) \\ &\quad +(WI(I,m',paire,t) - WI(I,m',f,t)) \\ &\quad \cdot(WI(I,m,paire,t) - WI(I,m,f,t))) \end{aligned}$$

Les covariances entre indices pour le mois pivot (p) sont estimées de façon identique. L'estimateur de la covariance entre $WI(I,M,t)$ et $WI(I,M,p)$ a deux composantes. La première comprend la somme, entre les indices dans l'agrégat d'aires, des covariances temporelles entre indices:

$$\begin{aligned} Cov_T(I,M,p,t) &= \sum_{m \in M} Cov_T(I,m,p,t) \\ &= \sum_{m \in M} \frac{1}{2}(((WI(I,m,impaire,p) - WI(I,m,t,p)) \\ &\quad \cdot(WI(I,m,impaire,t) - WI(I,m,f,t)) \\ &\quad +((WI(I,m,paire,p) - WI(I,m,f,p)) \\ &\quad \cdot(WI(I,m,paire,t) - WI(I,m,f,t))) \end{aligned}$$

Le second élément de la covariance comprend la somme, entre paire d'indices m et m' dans l'agrégat d'aires, des covariances mixtes entre indices pondérés pour l'indice m', au mois pivot p et l'indice m, au mois t:

$$\begin{aligned}
Cov_M(I,M,p,t) &= \sum_{m\in M}\sum_{m'\neq m} Cov_T(I,m',m,p,t) \\
&= \sum_{m\in M}\sum_{m'\neq m} \frac{1}{2}(((WI(I,m',impaire,p)-WI(I,m',f,p)) \\
&\quad \cdot(WI(I,m,impaire,t)-WI(I,m,f,t)) \\
&\quad +((WI(I,m',paire,p)-WI(I,m',f,p)) \\
&\quad \cdot(WI(I,m,paire,t)-WI(I,m,f,t)))
\end{aligned}$$

Alors la covariance totale entre $WI(I,M,t)$ et $WI(I,M,p)$ est donnée par:

$$Cov_{WI}(I,M,p,t) = Cov_T(I,M,p,t) + Cov_M(I,M,p,t)$$

Un estimateur de la variance non conditionnelle de $IX(I,M,t)$ est alors donné par:

$$\begin{aligned}
\sigma^2_{IX}(I,M,t) &= IX(I,M,p)^2\cdot(\frac{1}{WI(I,M,p)})^2\cdot\sigma^2_{WI}(I,M,t) \\
&\quad +(\frac{WI(I,M,t)}{WI(I,M,p)})^2\cdot\sigma^2_{WI}(I,M,p)-2\cdot(\frac{WI(I,M,t)}{WI(I,M,p)}) \\
&\quad \cdot Cov_{WI}(I,M,p,t)
\end{aligned}$$

La formule du biais estimé de l'indice est donnée par:

$$\begin{aligned}
Biais_{IX}(I,M,t) &= IX(I,M,p)\cdot(WI(I,M,t)\cdot\left(\frac{\sigma^2_{WI}(I,M,p)}{WI(I,M,p)^3}\right) \\
&\quad -\left(\frac{Cov_{WI}(I,M,p,t)}{WI(I,M,p)^2}\right))
\end{aligned}$$

L'écart quadratique moyen de l'indice est alors estimé par:

$$MSE_{IX}(I,M,t) = \sigma^2_{IX}(I,M,t) + (Biais_{IX}(I,M,t))^2$$

La variance du changement de prix
L'estimateur du changement de prix, $PC(I,M,t',t)$, entre le mois t' et le mois t pour l'agrégat d'articles I et l'agrégat d'aires M est calculé par:

$$PC(I,M,t',t) = \left(100 \cdot \left(\frac{WI(I,M,t)}{WI(I,M,t')}\right)\right) - 100$$

Alors, le changement de prix est un ratio d'indices pondérés entre deux périodes de temps.
La formule pour la variance estimée du changement de prix est donnée par:

$$\begin{aligned}\sigma^2_{PC}(I,M,t',t) &= \left(\frac{100}{WI(I,M,t')}\right)^2 \cdot (\sigma^2_{WI}(I,M,t) + \\ &\quad \left(\frac{WI(I,M,t)}{WI(I,M,t')}\right)^2 \cdot \sigma^2_{WI}(I,M,t') \\ &\quad -2 \cdot \left(\frac{WI(I,M,t)}{WI(I,M,t')}\right) \cdot Cov_{WI}(I,M,t',t))\end{aligned}$$

La formule du biais estimé pour le changement de prix est:

$$\begin{aligned}Biais_{PC}(I,M,t',t) &= 100 \cdot (WI(I,M,t) \cdot \left(\frac{\sigma^2_{WI}(I,M,t')}{(WI(I,M,t'))^3}\right) \\ &\quad -\frac{Cov_{WI}(I,M,t',t)}{(WI(I,M,t'))^2})\end{aligned}$$

Alors l'écart quadratique moyen du changement de prix est estimé par:

$$MSE_{PC}(I,M,t',t) = \sigma^2_{PC}(I,M,t',t) + (Biais_{PC}(I,M,t',t))^2.$$

Des résultats numériques ainsi que leur signification sont présentés dans la section 5.5.2 du cinquième chapitre.

Chapitre 4

Echantillons aléatoires pour l'ISPC

4.1 Introduction

Au cours du deuxième chapitre, nous avons présenté le plan d'échantillonnage de l'indice suisse des prix à la consommation. Nous avons vu que les échantillons des ménages et des loyers du logement sont aléatoires depuis mai 1993, mais que les échantillons des aires géographiques, des points de vente, des produits et des logements occupés par leur propriétaire ne le sont pas. Dans deux documents techniques, l'OFS (1990, 1993) [1] mentionne qu'il envisage, lors de révisions ultérieures de l'ISPC, d'introduire progressivement des échantillons aléatoires.

Note:

[1] *La révision de l'indice suisse des prix à la consommation. Concept de base.* (1990)
Révision de l'indice suisse des prix à la consommation. Conception du nouvel indice suisse des prix à la consommation. (1993)

Dans la première partie de ce chapitre, nous décrivons les recommandations internationales relatives à un plan d'échantillonnage aléatoire pour l'échantillon aréolaire et pour l'échantillon des points de vente et des produits. Dans la seconde partie du chapitre, nous présentons quelques propositions d'échantillons aléatoires pour le cas de l'indice suisse des prix à la consommation.

4.2 Les recommandations internationales

Les recommandations internationales qui sont présentées dans cette section ont pour source les documents de H. Picard (1985) et R. Turvey et al (1989). Nous verrons qu'elles sont les propositions faites au niveau international pour la sélection aléatoire d'échantillons. R. Turvey donne en exemple le plan d'échantillonnage de l'IPC américain (voir chapitre 3) qui est composé d'échantillons qui sont tous sélectionnés aléatoirement. Au cours du premier chapitre, nous avons vu que si l'on sélectionne des échantillons avec des méthodes aléatoires, on peut alors appliquer la théorie statistique afin de dériver les propriétés des estimateurs des échantillons. Les méthodes aléatoires font appel à la théorie des probabilités pour former l'échantillon. Dans un sondage aléatoire, chaque élément a une probabilité connue et non nulle d'inclusion dans l'échantillon. On notera que divers échantillons possibles peuvent être sélectionnés et qu'ils n'aboutissent pas tous au même résultat. La moyenne arithmétique de l'échantillon pour une variable d'intérêt est une variable aléatoire dont l'espérance mathématique est la moyenne de l'ensemble de la population; on parle alors d'estimateur sans biais. Cela signifie que la statistique calculée à partir de l'échantillon aléatoire est une variable aléatoire qui en moyenne vaut la variable calculée sur la population. Cependant, cette notion n'est pas suffisante pour juger de la qualité d'un échantillonnage. Il faut encore parler du risque lié à cet échantillonnage; cela signifie que l'on veut savoir comment l'ensemble de tous les réultats possibles se répartit autour de la moyenne; on jugera alors la dispersion des données en calculant la variance d'échantillonnage. En pratique, les populations étudiées sont grandes et les échantillons ont souvent plusieurs centaines d'individus; on peut alors utiliser le théorème central limite qui montre que la distribution de la variable aléatoire tend à se rapprocher de la loi normale. Cette

propriété asymptotique permet de construire des intervalles de confiance, c'est-à-dire de donner un intervalle aléatoire qui devrait couvrir le paramètre à estimer doit se situer avec une probabilité donnée. Fournir une estimation par intervalle permet de calculer une valeur précise, mais aussi un intervalle de valeurs. Avec ce dernier, on peut mesurer la probabilité que la vraie valeur de la population est hors de cet intervalle.
Au cours du chapitre 1, section 1.6, nous avons présenté le concept d'erreur quadratique moyenne (EQM) et nous avons vu que ce concept peut être appliqué à l'IPC (section I.6.1). Rappelons que l'EQM est composée de la variance, qui représente dans le cas de l'IPC l'erreur totale due aux erreurs d'échantillonnage, et du biais, qui fournit l'erreur de mesure qui est due à la mise sur pied des différents échantillons (erreur de couverture, omission de catégories de produits, inexactitude des données, erreurs d'homogénéité, non réponse). Relevons quelques réflexions émises R. Turvey au sujet de ce type d'erreurs. Les erreurs de couverture semblent relativement fréquentes pour les services; en effet, ces derniers sont parfois évincés par le fait qu'il est difficile de spécifier des produits particuliers pour certains d'entre eux, c'est notamment le cas des services bancaires et des assurances.
Les erreurs sur les coefficients de pondération semblent être moins graves que les erreurs dues à la collecte des prix. En effet, il faut remarquer qu'il s'avère plus grave de surestimer le changement de prix d'un article que de surestimer les dépenses pour ce même article qui servent à déterminer son coefficient budgétaire dans l'IPC. Les erreurs dues à la sélection et à l'observation des prix sont importantes et R. Turvey souligne que si des méthodes d'échantillonnage aléatoires sont utilisées, alors la qualité du sondage et la variance d'échantillonnage peuvent et doivent être calculés; pour ce faire, il préconise les techniques de rééchantillonnage. Notons qu'au chapitre suivant, nous proposons une méthode d'estimation de la variance de l'ISPC basée sur le rééchantillonnage. Si les articles et les points de vente sont sélectionnés avec une technique d'échantillonnage par choix raisonné, le contrôle ne peut être fait qu'au niveau d'une analyse critique des choix opérés; le risque principal réside alors dans le fait que les choix ne sont pas revus assez fréquemment afin de les tenir à jour.

Avant d'examiner les différents échantillons, nous présentons quelques remarques générales faites par H. Picard (1985). Ce dernier note qu'il faut distinguer entre un indice du lieu de résidence et un indice du lieu d'achat, car cela va déterminer le plan d'échantillonnage de l'IPC. L'indice du lieu de résidence est défini comme l'indice qui tend à mesurer les mouvements des prix payés par les ménages qui habitent dans une même zone géographique; il faut tenir compte de tous les achats effectués par les ménages, qu'ils soient faits ou non dans la zone géographique. L'indice du lieu d'achat est défini comme l'indice mesurant les variations des prix payés par tous les ménages dans les points de vente d'une même zone géographique. Le panier de référence doit alors être obtenu pour l'ensemble des ménages qui effectuent leurs achats dans une même zone géographique, indépendamment de leur lieu de résidence. Notons qu'un indice des prix à la consommation national estompe les écarts existants entre l'indice du lieu de résidence et l'indice du lieu d'achat. En pratique, la principale source d'information relative aux dépenses des ménages est l'enquête sur les dépenses de consommation des ménages, et cette dernière fournit les dépenses en fonction du lieu de résidence. Il en découle qu'en pratique, c'est principalement un indice du lieu de résidence que l'on calcule.

Une deuxième remarque concerne les sources d'information sur les prix; en principe, les ménages peuvent fournir les prix d'achat. Mais pratiquement, on substitue au prix d'achat le prix d'offre, car les prix sont observés dans les points de vente. Notons que les loyers représentent une exception, car ils sont observés chez les ménages et représentent le prix d'achat. Lorsque les prix sont fixés par l'Etat, par exemple les tarifs PTT, les prix peuvent être relevés directement sur les listes de prix officiels. D'autres documents administratifs peuvent aussi servir de base pour le relevé des prix, tels que les feuilles maladies pour les prix des prestations médicales, les dossiers des assurances (réparation automobile).

La dernière remarque concerne les méthodes de sélection des échantillons. H. Picard distingue la sélection avec probabilités proportionnelles aux consommations, avec probabilités proportionnelles aux quantités, avec probabilités proportionnelles aux populations et enfin la sélection équiprobable. Le tirage avec probabilités proportionnelles aux consommations est celui qui semble le plus approprié, à condition que l'on connaisse les consommations effectives des ménages pour chacun des éléments sélectionnés. Le tirage avec probabilités proportionnelles aux quantités n'est possible que si les observations du groupe sont homogènes et commensurables, la difficulté étant de rendre les unités de mesure cohérentes. Le tirage avec probabilités proportionnelles aux populations peut être utilisé pour sélectionner des zones ou sous-zones géographiques. Le tirage équiprobable peut être envisagé dans toutes les circonstances; cependant, il n'est à utiliser que lorsqu'il n'y a pas d'autre alternative possible.

Après ces quelques remarques préliminaires, passons à la description des échantillons.

4.2.1 Le plan d'échantillonnage aréolaire

Chez H. Picard, la notion de zone géographique recouvre un regroupement de communes, une commune urbaine ou une commune rurale. La stratification des zones géographiques est préconisée afin d'obtenir un indice national plus représentatif. La sélection des zones géographiques, à l'intérieur des strates, peut être effectuée avec une probabilité égale ou proportionnelle à la consommation ou à la population. Il s'agit de distinguer deux cas: premièrement, le cas où toutes les zones géographiques d'une strate sont retenues pour faire partie de l'échantillon et deuxièmement, le cas où l'on sélectionne des zones à l'intérieur d'une strate.

Examinons tout d'abord le cas où toutes les zones d'une strate sont retenues d'office. Ce cas est fréquent pour les grandes métropoles où une strate ne contient alors qu'une zone géographique. Si plusieurs zones géographiques sont dans la même strate, alors chacune d'elle est considérée comme une sous-strate ne contenant qu'une seule zone. Les indices des strates sont calculés comme des moyennes pondérées des indices des zones géographiques, où les coefficients de pondération sont estimés proportionnellement aux consommations de la strate. Pour estimer les coefficents de pondération, H. Picard distingue trois alternatives dépendant de l'enquête sur les dépenses de consommation des ménages. Si une enquête consommation (EC) est prévue sur toutes les zones géographiques de la strate, alors une estimation directe des consommations par zone pourra être réalisée. Si l'EC n'est réalisée que sur un échantillon de zones de la strate, il faut rechercher une relation entre la consommation des ménages et certaines caractéristiques de la zone géographique; cette relation servira à estimer les consommations des ménages de toutes les zones. La seconde alternative est celle où une EC ancienne est disponible; soit l'EC a porté sur toutes les zones et alors les consommations actuelles sont estimées en tenant compte des mouvements de la population, soit l'EC n'a porté que sur une partie des zones de la strate et alors il faut déterminer une relation entre la consommation des ménages et des caractéristiques de la zone géographique. A nouveau, cette relation servira à estimer les consommations des ménages de toutes les zones qu'il faudra encore actualiser en fonction des mouvements de la population. Si aucune EC n'existe, ou qu'elle est trop ancienne pour être utilisée, il faut procéder à une estimation élémentaire de la consommation par ménage.

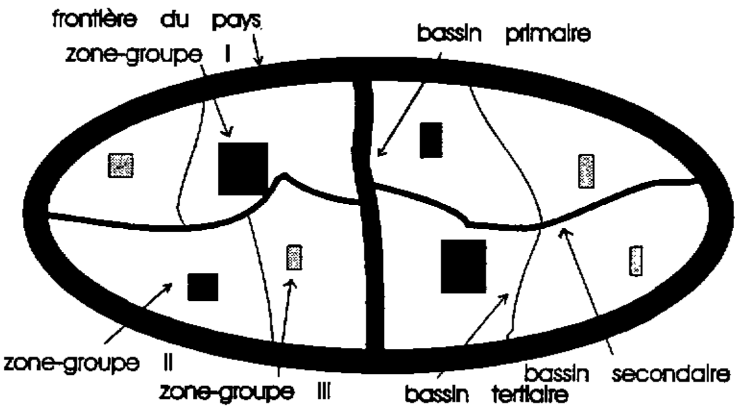
Considérons maintenant le second cas, où toutes les zones d'une strate ne sont pas retenues d'office. Il y aura alors sélection d'un certain nombre de zones géographiques parmi toutes celles qui sont contenues dans la strate. Dans ce cas aussi il faut considérer l'enquête sur la consommation. Si une EC est prévue, il faut alors l'étendre à toutes les zones tirées. Si l'EC donne des résultats pour chaque zone, la sélection des zones dans la strate est faite avec équiprobabilité. L'indice de la strate est calculé comme une moyenne pondérée des indice des zones choisies, la pondération étant égale à la part de la consommation des zones sélectionnées. Si l'EC donne des résultats uniquement au niveau de la strate, il faut alors supposer que les consommations par ménage sont égales dans toutes les zones de la strate. Enfin, si la sélection des zones est proportionnelle à la population de chaque zone, alors l'indice est estimé comme une moyenne des indices des zones tirées.

Une alternative à une sélection aléatoire de l'échantillon aréolaire est celle développée par Picard et qui fait appel à l'idée de bassins commerciaux. Un bassin commercial peut être défini comme une partie du territoire à l'intérieur duquel la population résidente effectue la majorité de ses achats. Les produits et services peuvent être classés en trois groupes:

1. Le premier groupe recouvre les produits peu courants. Ces derniers ne sont pas consommés fréquemment et leur achat nécessite un déplacement assez important. Par exemple: l'ameublement, l'automobile.
2. Le deuxième groupe de produits est composé des produits dits intermédiaires. Il s'agit d'articles achetés relativement fréquemment. Par exemple certains produits alimentaires (conserves), certains produits manufacturés (radio), ou certains services (dentiste).
3. Le troisième groupe de produits est composé des produits courants. C'est-à-dire d'articles achetés fréquemment et que l'on peut obtenir très facilement. Par exemple: les produits alimentaires de base (pain, légumes, etc), des produits industriels tels que les produits de nettoyage et certains services (coiffeur).

Les zones géographiques peuvent être classées en quatre groupes:
1. Le premier groupe est composé des zones qui ont une infrastructure complète où l'on trouve tous les biens et services.
2. Le deuxième groupe est composé des zones à infrastructure commerciale intermédiaire; on y trouve les produits dits intermédiaires et courants.
3. Le troisième groupe comprend les zones à infrastructure commerciale réduite; on y trouve les produits courants.
4. Le quatrième groupe est composé de zones à infrastructure commerciale inexistante; on n'y trouve que quelques produits de base.

Le schéma suivant permet de visualiser la notion de bassins commerciaux et de zones géographiques.



Pour sélectionner les zones géographiques en utilisant la notion de bassins commerciaux, il faut passer par plusieurs étapes. La première consiste à déterminer les bassins commerciaux primaires; le pays doit alors être dans un premier temps découpé en fonction de la consommation des produits peu courants. Pour procéder à ce découpage, on utilise comme point de départ la zone géographique la plus importante du groupe des zones géographiques à infrastructure commerciale complète. On détermine alors la portion de territoire national où la population se déplace principalement vers cette zone pour effectuer ses achats de produits peu courants. On répète le même procédé pour toutes les autres zones classées dans le groupe à infrastructure commerciale complète et on obtient ainsi un découpage du pays en bassins commerciaux primaires. La deuxième étape consiste à former les bassins commerciaux dits secondaires. Chaque bassin commercial primaire, défini au cours de la première étape, est découpé en bassins secondaires en fonction des achats de la population en produits intermédiaires. Puis on procède de manière analogue pour les bassins commerciaux tertiaires, qui eux sont découpés en fonction des achats des produits courants. L'étape suivante consiste à sélectionner tous les bassins secondaires contenant des zones géographiques du premier groupe; il s'agit des grandes métropoles nationales ou régionales et de leur zone d'influence immédiate. Puis on classifie les bassins secondaires à l'aide de critères tels que la taille de la population, la situation géographique, le caractère rural ou urbain du bassin et la nature des activités économiques du bassin. Dans chaque strate ainsi constituée, il faudra sélectionner un ou plusieurs bassins secondaires. Les bassins tertiaires contenant des zones du premier et du deuxième groupe sont retenus d'office. Puis on sélectionne encore au moins un autre bassin tertiaire dans chaque bassin secondaire. Cette sélection peut se faire en stratifiant les bassins tertiaires en fonction de critères identiques à ceux des bassins secondaires. Dans les bassins tertiaires contenant une zone du premier groupe, il faut observer toutes les catégories de produits. Dans les bassins tertiaires avec une zone du deuxième groupe on n'observe que les produits intermédiaires et les produits courants. Enfin dans les bassins tertiaires avec des zones du troisième groupe, on n'observe que les produits courants.

4.2.2 Le plan d'échantillonnage des points de vente et des produits

La méthode de référence, présentée ci-dessous, est exposée plus en détail chez R. Turvey et al. (1989) et chez H. Picard (1985). Elle a été inspirée par le plan d'échantillonnage aléatoire développé par le BLS (1988) pour l'IPC américain. On peut se rapporter au chapitre 3 de ce travail qui expose le plan d'échantillonnage aléatoire de l'IPC américain.

La méthode dite de référence propose une procédure d'échantillonnage aléatoire à deux niveaux. Le premier niveau consiste en la sélection d'un échantillon aléatoire de points de vente et le second d'un échantillon aléatoire de produits.

La première étape, qui consiste à sélectionner des points de vente, est basée sur une enquête sur les points d'achat des ménages. Pour chaque article acheté, on veut savoir quel montant a été dépensé et obtenir le nom et l'adresse exacte du lieu d'achat de l'article. Cela fournit une liste de points de vente avec le total des ventes pour chacun des points de vente et chaque article ou groupes d'articles pour l'échantillon de ménages. L'échantillon des points de vente est alors sélectionné à partir de la liste ainsi établie avec une probabilité de sélection proportionnelle au total des ventes. La seconde étape concerne le choix des produits dans les points de vente choisis. Dans chaque point de vente, et pour chaque catégorie de dépense, des produits précis doivent être sélectionnés afin de relever leur prix. Si le point de vente a été sélectionné plusieurs fois, alors un nombre correspondant d'articles doit être sélectionné.

Un certain nombre de problèmes pratiques existent quant à la mise sur pied de la méthode de référence. Cette démarche est coûteuse. Il faut notamment effectuer une enquête auprès des ménages sur les points d'achat. Il serait possible d'envisager l'intégration de cette enquête à celle sur les dépenses de consommation. Cependant, cela rendrait cette dernière encore plus complexe. Le relevé de prix dans les points de vente est plus coûteux que les relevés centralisés. Le coût de cette méthode est aussi dû au fait qu'il faut respecter le principe de la collecte sur le lieu d'achat effectif. On peut être amené à faire des relevés qui sont dispersés, d'où des dépenses supplémentaires. La hiérarchie de classification des articles doit être détaillée, exhaustive et continuellement tenue à jour. Les collecteurs de prix doivent appliquer rigoureusement la procédure d'échantillonnage; il faut donc des agents collecteurs spécialisés et formés aux techniques statistiques. Enfin, une très bonne collaboration avec le personnel des points de vente choisis est nécessaire car les entretiens avec le personnel de vente peuvent prendre beaucoup de temps.

On a vu qu'il existe, face à la méthode de référence, un certain nombre d'inconvénients; R. Turvey et al (1989) et H. Picard (1985) proposent donc deux alternatives simplifiées concernant la sélection des points de vente. La première solution consiste en une enquête auprès des ménages comme dans la méthode de référence. Mais au lieu d'obtenir une liste détaillée des achats par produit et par point de vente, on se contente d'obtenir les achats par type de points de vente. Par exemple pour les produits laitiers, on obtient une estimation pour les supermarchés, les épiceries, les laiteries, etc. On peut interroger le ménage sur le point de vente exact où l'achat a été effectué mais au moment du codage, on ne s'intéressera qu'au type de point de vente. La sélection sera donc faite par type de point de vente. Le manque d'information sur le chiffre d'affaires de chacun des points de vente individuel entraîne l'hypothèse qu'ils ont tous la même valeur de vente pour un groupe donné d'articles. L'échantillon des points de vente est alors choisi avec une probabilité de sélection égale pour chacun d'entre eux. On suppose que les chiffres d'affaires relatifs aux produits du poste sont égaux dans tous les points de vente d'un type donné. L'indice par type de point de vente est alors une moyenne des indices des points de vente de l'échantillon. L'indice pour la zone géographique est une moyenne pondérée des indices par type de point de vente, les coefficients de pondération étant proportionnels aux montants des achats des ménages dans chaque type.

La seconde alternative est appliquée lorsque la décomposition des dépenses n'est connue qu'à un niveau supérieur, par exemple au niveau national ou à un niveau régional. On pose alors l'hypothèse que ce que l'on sait pour le niveau supérieur est aussi vrai pour les niveaux inférieurs, par exemple pour les villes. La sélection des points de vente se fait alors en appliquant la même méthode que celle décrite dans la première alternative.

4.2.3 Le plan d'échantillonnage des logements occupés par leur propriétaire

Dans le cas d'un indice des prix à la consommation, il faut mesurer l'évolution du prix de l'utilisation du logement; le problème consiste donc à trouver une mesure de l'évolution de ce prix. Les recommandations faites au niveau international, notamment chez R. Turvey (1989), distinguent trois types de solutions. Le premier type fait référence aux acquisitions nettes, c'est-à-dire au coût d'acquisition au prix du marché de logements neufs destinés à être habités par leur propriétaire. Cette première optique essaie de mesurer le changement à travers le temps des coûts des acquisitions nettes pour un logement occupé par son propriétaire (LOP). La deuxième approche fait référence à la notion d'équivalence-locative. Ici, l'évolution des prix de l'utilisation du logement occupé par le propriétaire est mesurée au moyen de loyers de logements comparables. Cette approche mesure le changement à travers le temps du coût d'utilisation pour un LOP. La troisième approche est celle du paiement. On cherche à mesurer les débours monétaires pour les logements occupés par leur propriétaire. Les débours monétaires incluent le paiement des intérêts hypothécaires et les autres frais d'utilisation, tels que les frais d'entretien, de réparations, les taxes, etc. Cette troisième méthode essaie de calculer le changement à travers le temps des paiements pour un LOP.

L'approche qu'il faut utiliser dépend du but de l'indice des prix à la consommation.

Dans une analyse macro-économique, l'IPC est utilisé comme indicateur général de l'inflation, et ceci par la mesure de la variation du niveau des prix. Comme les prix et les taux d'intérêt peuvent varier de façon indépendante, il faut utiliser une approche qui ne reflète pas les taux d'intérêt. L'approche par le paiement semble donc inappropriée pour une analyse économique, car les variations des prix passés et des taux d'intérêt passés ou présents peuvent affecter les dépenses au comptant des propriétaires. Cette remarque prévaut aussi pour l'approche par le coût lorsque des mouvements des taux d'intérêt l'affectent. Un indice limité aux coûts de réparation et d'entretien des LOP, ou encore un indice incluant les prix des nouvelles habitations est préférable pour mesurer l'inflation. Par contre, pour l'évaluation des revenus monétaires et, plus encore pour l'indexation, l'approche par le paiement semble meilleure; dans l'approche par le concept de la comptabilité nationale, c'est le cas de l'équivalence-locative qui doit être appliqué. Nous pouvons resumer ces trois approches comme suit:

1. *Acquisitions nettes*
   1.1 Valeur totale du marché des nouvelles habitations.

2. *Coût d'utilisation*
   2.1 Intérêts hypothécaires et coût de remplacement de la dépréciation.
   2.2 Coût d'opportunité d'une habitation au lieu de la vendre.
   2.3 Coût d'opportunité d'occuper un LOP au lieu de le louer.

3. *Paiement*
   3.1 Paiements comptants, achats, remboursements hypothécaires et intérêts hypothécaires.
   3.2 Remboursements hypothécaires.
   3.3 Intérêts hypothécaires.

4.2.4 L'optimisation du plan d'échantillonnage

Optimiser le plan d'échantillonnage signifie que pour un budget donné, on cherche à définir la procédure de collecte qui fournit la précision la meilleure pour les indices calculés. On cherche à savoir combien de zones géographiques doivent être sélectionnées et combien d'observations par poste de relevé et par zone géographique doivent être faites. On veut encore savoir s'il faut concentrer les observations dans certaines zones et quelle doit être la cadence des relevés. Il faut gérer deux objectifs opposés, à savoir la précision et le coût. En effet, si le nombre d'observations est grand et si elles sont dispersées on améliore la précision de l'indice. Cependant, une augmentation du nombre des observations et de leur dispersion entraîne une augmentation du coût de la collecte.

Le nombre d'observations

Dans un point de vente, on ne collecte généralement pas qu'un seul article. Il faut donc affecter à chaque relevé une partie des coûts communs. Si on analyse les constituants du coût des relevés, on y trouve les éléments suivants. Premièrement, le coût du relevé d'un article dépend de la facilité du relevé; il faut de plus s'assurer qu'il corresponde bien au produit décrit. Deuxièmement, dans un point de vente donné, on passe du temps à se présenter au responsable, à passer d'un rayon à l'autre (ce qui n'est pas négligeable dans des grandes surfaces), à obtenir le prix du produit (étiquetage, information par oral, etc). Un troisième élément est le déplacement d'un point de vente à l'autre; il faut alors compter les frais de transport et le temps passé pour le déplacement. Pour diminuer les coûts des relevés; on peut donc concentrer les observations dans un nombre restreint de points de vente ou de zones géographiques. H. Picard énumère trois principes quant au nombre d'observations à relever. Premièrement; le nombre d'observations doit être d'autant plus grand que la variabilité des produits d'un poste est grande. Deuxièmement; le nombre d'observations doit être d'autant plus grand que la pondération du poste est grande. Troisièmement; le nombre d'observations doit être d'autant plus petit que le coût unitaire du relevé est grand.

La répartition temporelle des relevés

Le calcul d'un indice moyen du mois entraîne un problème de répartition afin que l'on puisse garantir l'objectif de comparaison des relevés, par exemple avec l'année de base, avec le mois précédent ou avec le même mois pour l'année précédente. Le principe de base repose sur le fait que le relevé du prix d'un produit dans un point de vente déterminé doit être effectué à la même date dans un mois donné. Si on veut un indice moyen pour un mois, il faut alors répartir les observations uniformément sur l'ensemble du mois. Lorsque les produits ont des prix imposés par l'Etat et s'il y a modification du prix au cours du mois, H.Picard recommande alors d'observer le prix avant et après sa modification et de calculer une moyenne pondérée par le nombre de jour où le prix fixé était en vigueur. Pour les produits dont les variations de prix sont très rapides, Picard recommande de réaliser plusieurs relevés par mois dans chacun des points de vente pour les produits dont les changements sont importants. Où il y a difficultés de substitution (habillement, ameublement, etc), il est préférable, selon H. Picard, de diminuer le nombre d'observations mais de consacrer tout le temps nécessaire pour résoudre le problème de substitution. Pour ce type de produit, le prix varie en général relativement peu; on peut donc envisager de répartir les observations sur plusieurs mois.

4.3 La cas de la Suisse

Au cours du chapitre 2, nous avons présenté une description des échantillons sélectionnés pour l'indice suisse des prix à la consommation. Nous avons vu que l'échantillon de ménages est sélectionné aléatoirement depuis l'enquête sur la consommation des ménages de 1990. Les résultats de cette enquête ont permis de calculer les coefficients budgétaires qui sont utilisés pour le nouvel indice (base 100 en mai 1993). Dès mai 1993, l'échantillon des logements choisis pour le calcul de l'indice du loyer du logement est également aléatoire; rappelons que les logements occupés par leur propriétaire (LOP) n'ont pas été explicitement inclus dans le nouvel indice.
Au cours de cette section, nous ne reviendrons donc pas sur ces deux échantillons aléatoires et nous nous focaliserons sur l'échantillon aréolaire, sur l'échantillon des points de vente et des articles, et sur l'échantillon des logements occupés par leur propriétaire. Dans le rapport technique *"Indice suisse des prix à la consommation: Quelqucs recommandations relatives à son échantillonnage à venir"* par Y.Dodge, S. König, S. Pfaff et F. Mehran (mars 1992), nous avons émis des propositions quant aux échantillons aléatoires de l'ISPC; ces dernières sont présentées dans les sections suivantes.

4.3.1 Le plan d'échantillonnage aréolaire

La sélection aléatoire des communes nécessite en premier lieu la constitution d'une base de sondage. Pour sélectionner les communes où les relevés de prix sont effectués, il serait judicieux de connaître les dépenses de consommation pour chaque commune, afin de les sélectionner avec une probabilité proportionnelle à ces dépenses. Cela signifie qu'il est nécessaire de connaître à la fois les lieux de vente et l'importance des dépenses des consommateurs, ceci pouvant se mesurer au travers d'une enquête sur les points de vente où s'approvisionnent les consommateurs. L'importance des dépenses peut alors servir de probabilité de sélection

de la commune.
Une alternative est d'utiliser, comme base de sondage, la liste des communes. La population de chaque commune pouvant servir de probabilité de sélection lors d'un tirage aléatoire de l'échantillon.
A titre d'exemple, on peut envisager, parmi les différentes alternatives, les deux propositions suivantes pour la construction d'un plan de sondage aléatoire.

Première proposition:

On peut proposer un plan d'échantillonnage stratifié à un degré, avec une probabilité de sélection proportionnelle à la taille de la commune, et pour les cinq plus grandes communes, une probabilité de sélection égale à 1. Le premier degré de stratification serait les cantons, puis une commune serait sélectionnée aléatoirement, à l'intérieur de chaque canton, avec une probabilité de sélection proportionnelle à sa taille. De plus, les cinq communes (Bâle, Berne, Genève, Lausanne et Zurich) de plus de 100'000 habitants auraient une probabilité de sélection égale à 1. On obtiendrait ainsi un total de 31 communes participantes. L'un des avantages de cette proposition réside au niveau politique par le fait que tous les cantons sont représentés, et cet argument n'est pas négligeable lorsque l'on recherche un consensus.

Seconde proposition:

Une autre alternative peut résider dans un plan d'échantillonnage stratifié à deux degrés, le premier degré étant stratifié par groupe de communes en fonction de leur nombre d'habitants. Les groupes suivants pourraient être formés:

- Communes de plus de 100'000 habitants

- Communes de 50'000 à 100'000 habitants

- Communes de 10'000 à 50'000 habitants

- Communes de moins de 10'000 habitants

Le deuxième degré de stratification correspondrait aux régions linguistiques, chacun des groupes contenant les trois régions linguistiques. Puis à l'intérieur de chaque groupe et de chaque région, on pourrait sélectionner un échantillon aléatoire simple avec probabilité égale pour chaque commune. Notons que les communes du premier groupe auraient une probabilité de sélection égale à 1. Le deuxime groupe serait composé de quatre communes (Bienne, Lucerne, Saint-Gall, Winterthour), le troisième groupe compterait quelque 63 communes en Suisse alémanique, 21 en Suisse romande et trois en Suisse italienne. L'échantillon comprendrait les cinq communes du premier groupe, les quatres du deuxième groupe, une quizaine pour le troisième et six communes pour le quatrième groupe; ainsi, la taille totale de l'échantillon serait d'environ une trentaine de communes.

4.3.2 Le plan d'échantillonnage des points de vente et des articles

L'échantillon des points de vente

A l'heure actuelle en Suisse, on peut obtenir différentes informations sur les points de vente auprès desquels les ménages font leurs achats. En particulier, il existe des statistiques sur le chiffre d'affaires, le commerce de détail, etc.
A notre connaissance, l'ISPC n'utilise pas toutes les informations disponibles sur les points de vente. Selon un document technique (1990) publié par l'OFS, on remarque que la pondération des canaux de distribution, en vigueur jusqu'en mai 1993, a été déterminée d'une manière rudimentaire. L'OFS a souhaité par ailleurs affiner ces poids pour le nouvel indice (base 100 en mai 1993), car ils n'existaient que pour le groupe de l'alimentation et ils ne correspondaient plus au marché actuel.
Il semble nécessaire que l'OFS puisse disposer, pour l'ISPC, d'informations sur les points de vente répondant spécifiquement à ses besoins. Avant que l'OFS n'envisage de réaliser une enquête régulière sur les lieux d'achat des ménages, nous proposons de considérer la mise sur pied d'une enquête pilote. Les objectifs de cette enquête seraient de déterminer les points de vente où les ménages font leurs achats, de déterminer des types de canaux de distribution et de calculer des poids

de pondération pour chacun d'entre eux.

La sélection aléatoire des points de vente repose donc sur la mise sur pied d'une enquête pilote sur les lieux d'achat des ménages. Le plan de sondage de l'enquête pilote peut être envisagé de deux manières différentes.
Premièrement, l'enquête pilote pourrait être associée à l'enquête sur les dépenses de consommation. Par exemple, un dixième des ménages interrogés participerait en sus à l'enquête pilote. Cette alternative a le désavantage d'alourdir l'enquête de base.
La seconde alternative serait de réaliser une enquête parallèle, avec un plan de sondage identique, à celle sur les dépenses de consommation. Pour chaque dixième ménage choisi, on pourrait sélectionner un ménage supplémentaire qui, lui, ne participerait qu'à l'enquête pilote. Cette dernière devrait permettre d'obtenir des renseignements également sur le coût d'une enquête généralisée et sur la périodicité avec laquelle il faudrait la réaliser. Les ménages sélectionnés seraient questionnés sur l'ensemble de leurs dépenses, afin d'essayer d'obtenir des données sur un ensemble exhaustif de produits; qu'il s'agisse des articles où les observations de prix sont directes (alimentation), centralisées (PTT, CFF) ou relevées sur des listes (voitures, restaurants). Il faudrait interroger les ménages à partir de la nomenclature de produits, et leur demander dans quels points de vente ils achètent les produits correspondant aux rubriques; le nom exact et l'emplacement du point de vente devraient être spécifiés ainsi que le montant des dépenses. Un autre renseignement à obtenir serait de savoir quel est l'intervalle de temps entre deux achats consécutifs pour chaque produit de la nomenclature. On pourrait alors déterminer une subdivision par type de canal de distribution et estimer leurs poids de pondération, ces derniers serviraient alors de probabilité de sélection par type de canal de distribution. On procéderait ainsi, dans chaque commune participante, à une sélection aléatoire des points de vente. Avec cette procédure, il ne serait plus nécessaire, lors du calcul de l'ISPC, de pondérer les indices par canaux de distribution.
Une fois les points de vente sélectionnés, il faut choisir à l'intérieur de ces derniers les articles dont le prix est à relever.

L'échantillon des articles

La méthode de référence, utilisée pour le calcul de l'indice américain des prix à la consommation, pourrait être appliquée au cas de la Suisse. Le statisticien préconise naturellement la mise en place d'une telle méthode. Cependant en suivant les recommandations internationales, nous constatons qu'il peut y avoir quelques alternatives simplificatrices.

Une première alternative consiste en une sélection a priori d'articles représentatifs. Cette sélection a priori se fait d'une manière centralisée. Chaque groupe d'articles peut être divisé en sous-groupes, un choix raisonné d'articles étant fait à l'intérieur de chacun des sous-groupes. Dans le cas où les groupes d'articles sont homogènes, un ou plusieurs articles représentatifs peuvent être directement choisis. Les personnes chargées de collecter les prix dans les points de vente doivent alors choisir, pour chaque article, des variétés dont ils releveront les prix. Les variétés retenues doivent être typiques de l'article, elles doivent aussi être disponibles en stock et présentes sur le marché durant une certaine période de temps et leur qualité doit permettre de trouver des substituts si elles disparaissent. Ce choix représentatif doit être fait avec beaucoup de soin et être continuellement contrôlé. La seconde alternative concerne la sélection d'articles dans les points de vente. Les personnes chargées de relever les prix dans les points de vente sont également chargées de choisir les articles pour les groupes d'articles. Dans la première alternative, nous avons vu que ce choix est fait d'une manière centralisée et la méthode de décomposition des groupes d'articles nécessite une bonne connaissance du marché de la part du responsable des relevés. La solution la plus courante consiste à choisir la variété la plus vendue parmi les articles les plus vendus d'un groupe d'articles. Néanmoins, il faut être attentif au fait que si les ventes d'un article sont ventilées entre plusieurs variétés et qu'aucune ne prédomine, ce qui importe ce n'est pas la variété qui se vend le plus, mais aussi celle qui représente le mieux la variation du prix des articles qu'elle représente.

Le plan d'échantillonnage des logements occupés par leur propriétaire

Rappelons qu'à l'heure actuelle, l'OFS calcule un indice des habitations locatives (chapitre 2) et qu'aucune mesure n'est faite explicitement pour les logements occupés par leur propriétaire.

Nous avons vu que trois approches peuvent être envisagées pour calculer un indice des logements occupés par leur propriétaire. Il s'agit des acquisitions nettes, de l'équivalence-locative et de l'approche par les paiements.

Pour la Suisse, on peut envisager d'utiliser la notion d'équivalence-locative pour estimer l'indice des logements occupés par leur propriétaire. Pour cela, nous proposons que le recensement des habitations serve de base de sondage pour la sélection de l'échantillon de logements locatifs et de l'échantillon de logements occupés par leur propriétaire. Le plan d'échantillonnage des LOP serait identique à celui des habitations louées; rappelons que les logements sont aujourd'hui répartis en 20 catégories, premièrement par taille (habitations de 1 à 5 pièces) et deuxièmement par classe d'âge (de 0-5 ans, de 6-10 ans, de 11-20 ans et plus de 20 ans). Puis il s'agirait de calculer l'équivalence-locative en estimant le loyer implicite des propriétaires pour chaque unité de l'échantillon. Pour ce faire, l'enquêteur devrait poser au propriétaire la question suivante: "Si ce logement était loué, quelle en serait la location en tenant compte des frais d'entretien mais en excluant les charges (mazout, électricité, etc)?". S'il s'avérait que le propriétaire ne soit pas en mesure de donner une réponse satisfaisante, alors l'enquêteur devrait estimer le loyer par comparaison avec d'autres habitations du même type.

En conclusion, nous pouvons relever qu'au cours de ce chapitre, nous avons présenté la méthode de référence servant à mettre sur pied un plan d'échantillonnage aléatoire. La mise en place d'un tel plan pour la Suisse est importante car cela permettrait d'estimer la précision de l'IPC. C'est pour atteindre ce but qu'au chapitre suivant nous présentons une méthode d'estimation de la variance de l'ISPC qui pourrait être appliquée si les échantillons étaient aléatoires.

Chapitre 5

Estimation de la variance de l'ISPC: une proposition

5.1 Introduction

Après le calcul de toute statistique, on souhaite en estimer la précision. Cette dernière peut se mesurer au travers de l'estimation de la variance.

Pour calculer l'indice des prix à la consommation en Suisse, l'Office fédéral de la statistique a besoin de données issues de plusieurs échantillons. On a vu précédemment que ces échantillons sont les suivants: un échantillon de ménages, un échantillon de communes, un échantillon de points de vente, un échantillon d'articles et un échantillon de logements.
L'indice des prix à la consommation est une statistique ayant besoin, pour sa construction, d'un plan d'échantillonnage à plusieurs niveaux, chacun d'entre eux ayant un plan d'échantillonnage complexe (échantillon stratifié, à plusieurs degrés, etc).

Pour pouvoir calculer la variance d'une statistique, il faut que l'échantillon servant à sa construction soit sélectionné aléatoirement (par opposition à un choix raisonné). Comme l'ISPC n'est pas construit aujourd'hui sur des échantillons probabilistes, nous avons posé l'hypothèse que les échantillons sont sélectionnés aléatoirement afin de pouvoir établir une méthodologie permettant d'en estimer la variance.

Au vu de la complexité de l'IPC, on n'est plus en mesure d'appliquer les formules développées par la théorie classique de l'échantillonnage. En effet, nous sommes en présence d'une statistique construite sur cinq niveaux d'échantillonnage, et les formules analytiques sont développées pour un seul niveau d'échantillonnage (se référer notamment à W.G. Cochran (1963) et à M.H. Hansen, W.N. Hurwitz et W.G. Madow (1953)).

Pour estimer la variance, malgré cette complexité, on peut utiliser des techniques empiriques d'estimation de la variance basées sur le concept de rééchantillonnage ou réplication. Parmi ces méthodes, citons le Jackknife, le Bootstrap, les demi-échantillons équilibrés et la méthode des groupes aléatoires.

C'est la méthode des groupes aléatoires que nous avons retenue dans notre proposition. L'utilisation de la méthode des groupes aléatoires fournit un moyen de mesurer la variation totale de l'IPC à partir des variations calculées par sous-échantillonnage de l'échantillon. Le principe de base de cette méthode consiste à sélectionner deux ou plusieurs sous-échantillons en utilisant le même plan d'échantillonnage pour chacun d'eux. Puis on construit séparément un estimateur du paramètre d'intérêt pour chaque sous-échantillon ainsi qu'un estimateur à partir de tous les sous-échantillons. Enfin, on calcule la variance échantillonnale parmi toutes les estimations.

La complexité des calculs s'accroît avec le fait que la variation de l'IPC constitue un estimateur non linéaire. Une méthode fréquemment utilisée pour estimer la variance d'un estimateur non linéaire est d'approximer ce dernier par une fonction linéaire des observations. Cette approximation se calcule par les séries de Taylor. Nous combinerons donc la méthode des groupes aléatoires et les séries de Taylor pour estimer la variance du changement de niveau de l'ISPC.

Avant d'examiner en détail notre proposition pour le cas suisse, exposons les principes de base des techniques de rééchantillonnage.

5.2 Les techniques de rééchantillonnage

Dans cette section, on décrit les techniques de base des méthodes de rééchantillonnage. Celles-ci sont développées notamment chez B. Efron (1982) et K. Wolter (1985).

5.2.1 Les demi-échantillons équilibrés

Les demi-échantillons équilibrés, plus connu sous le terme anglais de Balanced Half-Samples (BHS), font référence à une méthode développée par McCarthy.

Techniques de base

Supposons que l'on veuille estimer la moyenne de la population $\bar{Y}$ d'un plan stratifié avec deux unités par strates et où les unités de chaque strate ont été sélectionnées par échantillonnage aléatoire simple avec remise.

Notons:

L :	le nombre de strates
N_h :	le nombre d'unités à l'intérieur de la strate h
$N = \sum_{h=1}^{L} N_h$:	la taille de la population
y_{h1}, y_{h2} :	les observations de la strate h $(h = 1 \cdots L)$

Un estimateur non biaisé de $\bar{Y}$ est:

$$\bar{y}_{st} = \sum_{h=1}^{L} \frac{N_h}{N} \bar{y}_h$$

où:

$$\bar{y}_h = (y_{h1} + y_{h2})/2.$$

L'estimateur de la $Var(\bar{y}_{st})$ est donné par:

$$\begin{aligned} v(\bar{y}_{st}) &= \sum_{h=1}^{L} (\frac{N_h}{N})^2 s_h^2/2 \\ &= \sum_{h=1}^{L} (\frac{N_h}{N})^2 d_h^2/4 \end{aligned}$$

où:

$$d_h = y_{h1} - y_{h2}.$$

Dans notre cas, on dispose de deux groupes aléatoires indépendants appelés aussi demi-échantillons; il s'agit de $(y_{11}, y_{21}, \ldots, y_{L2})$.

L'estimateur de $Var(\bar{y}_{st})$ est donné, par les groupes aléatoires, par (où α représente le nombre de demi-échantillons):

$$\begin{aligned} v_{ga}(\bar{y}_{st}) &= [2(2-1)]^{-1} \sum_{\alpha=1}^{2} (\bar{y}_{st,\alpha} - \bar{y}_{st})^2 \\ &= (\bar{y}_{st,1} - \bar{y}_{st,2})^2/4 \end{aligned}$$

où:

$$\bar{y}_{st,1} = \sum_{h=1}^{L} \frac{N_h}{N} \cdot y_{h1}$$

$$\bar{y}_{st,2} = \sum_{h=1}^{L} \frac{N_h}{N} \cdot y_{h2}$$

$$\bar{y}_{st} = (\bar{y}_{st,1} + \bar{y}_{st,2})/2.$$

$v_{ga}(\bar{y}_{st})$ est simple à calculer mais peu stable par rapport à $v(\bar{y}_{st})$, car $v_{ga}(\bar{y}_{st})$ n'a qu'un degré de liberté. Comment combiner alors la simplicité de $v_{ga}(\bar{y}_{st})$ et l'efficacité de $v(\bar{y}_{st})$?
Une méthode consiste à considérer les demi-échantillons comprenant une unité de chacune des strates. On obtient alors des demi-échantillons différents contenant des unités en commun et des unités différentes et ce d'une manière systématique. Comme il y a recouvrement, les demi-échantillons sont corrélés entre eux.

Supposons qu'un demi-échantillon soit formé en sélectionnant une unité de chaque strate.
Il y a 2^L demi-échantillons pour un échantillon donné. L'estimateur de $\bar{Y}$ calculé à partir des α demi-échantillons est donné par:

$$\bar{y}_{st,\alpha} = \sum_{h=1}^{L} \frac{N_h}{N}(\delta_{h1\alpha}y_{h1} + \delta_{h2\alpha}y_{h2})$$

$$\delta_{h1\alpha} = \left\{ \begin{array}{ll} 1 & \text{si l'observation } (h,1) \text{ fait} \\ & \text{partie du demi-échantillon} \\ & \alpha. \\ 0 & \text{sinon.} \end{array} \right\}$$

et

$$\delta_{h2\alpha} = 1 - \delta_{hi\alpha}$$

Remarque:

La moyenne des 2^L estimateurs de $\bar{y}_{st,\alpha}$ est égale à l'estimateur de $\bar{y}_{st}$ issu de l'échantillon donné. Cela provient du fait que chaque observation appartient à la moitié des 2^L demi-échantillons ($2^{L/2} = 2^{L-1}$).

$$\begin{aligned} \sum_{\alpha=1}^{2^L} \bar{y}_{st,\alpha}/2^L &= \sum_{h=1}^{L} \frac{N_h}{N}(y_{h1} + y_{h2})(2^{L-1}/2^L) \\ &= \bar{y}_{st} \end{aligned}$$

On peut construire un estimateur de la variance en termes de $\bar{y}_{st,\alpha}$. Si on définit:

$$\begin{aligned} \delta_h^{(\alpha)} &= 2\delta_{h1\alpha} - 1 \\ &= \left\{ \begin{array}{ll} 1 & \text{si l'observation } (h,1) \text{ fait} \\ & \text{partie du demi-échantillon} \\ & \alpha. \\ -1 & \text{si l'observation } (h,2) \\ & \text{fait partie du demi-échantillon} \\ & \alpha. \end{array} \right\} \end{aligned}$$

$$\bar{y}_{st,\alpha} - \bar{y}_{st} = \sum_{h=1}^{L} \frac{N_h}{N}\delta_h^{(\alpha)} d_h/2$$

et

$$(\bar{y}_{st,\alpha} - \bar{y}_{st})^2 = \sum_{h=1}^{2^L} (\frac{N_h}{N})^2 d_h^2/4 + \sum\sum \delta_h^{(\alpha)} \delta_{h'}^{(\alpha)} \frac{N_h}{N} \frac{N_{h'}}{N} d_h d_{h'}/2$$

Estimateur non biaisé de $Var\bar{y}_{st}$:

$$v(\bar{y}_{st}) = \sum_{\alpha=1}^{2^L} (\bar{y}_{st,\alpha} - \bar{y}_{st})^2/2^L$$

Si L est grand, le calcul s'avère difficile. Il faut donc calculer la moyenne seulement sur un échantillon des L. Cela simplifie les calculs, mais il faut choisir judicieusement les demi-échantillons.
Si k demi-échantillons sont choisis, on peut trouver l'estimateur de la variance suivant:

$$v_k(\bar{y}_{st}) = \sum_{\alpha=1}^{k} (\bar{y}_{st,\alpha} - \bar{y}_{st})^2/k.$$

Mais comment faut-il choisir les k demi-échantillons afin que $v_k(\bar{y}_{st})$ soit égal à $v(\bar{y}_{st})$?
Il faut que l'égalité suivante soit satisfaite:

$$\sum_{\alpha=1}^{k} \delta_h^{(\alpha)} \delta_{h'}^{(\alpha)} = 0$$

et cela pour tout $h < h' = 1, \cdots, L$.

Plackett et Burman (1946) ont présenté des méthodes qui permettent la construction de matrices orthogonales $k \cdot k$ (k est multiple de 4) et dont les colonnes satisfont l'égalité ci-dessus. Ces matrices sont appelées matrices d'Hadamard. Les k demi-échantillons contiennent alors toute l'information contenue dans les 2^L demi-échantillons. C'est pourquoi on parle de demi-échantillons équilibrés.

5.2.2 Le Jackknife

Le concept de base du Jackknife fut introduit par Quenouille en 1949. Le terme de Jackknife, introduit par Tuckey en 1956, vient du nom anglais du couteau scout à plusieurs lames prêt à être utilisé dans un grand nombre de cas. Cela pour montrer que c'est une méthode très flexible et qui s'adapte à un grand nombre de situations.

Techniques de base

Si $Y_1, \cdots, Y_n$ sont des variables aléatoires identiquement distribuées et indépendantes suivant une loi quelquonque, un estimateur $(\hat{\theta})$ du paramètre d'intérêt de la population (θ) peut être calculé à partir de l'échantillon complet.

L'estimateur de Quenouille:

Partitionons l'échantillon en k groupes de m observations chacun, avec l'hypothèse que n, m, k sont des entiers et que $n = m \cdot k$.
Posons que $\hat{\theta}_{(\alpha)}$ est l'estimateur de même forme que $\hat{\theta}$ et qu'il est calculé sur l'échantillon de taille $m(k-1)$ obtenu en omettant le groupe α et définissons:

$$\hat{\theta}_\alpha = k\hat{\theta} - (k-1)\hat{\theta}_{(\alpha)}$$

Les $\hat{\theta}_\alpha$ sont appelés pseudovaleurs. L'estimateur de Quenouille est la moyenne de $\hat{\theta}_\alpha$.

On peut supposer que les pseudovaleurs sont approximativement identiquement distribuées et indépendantes.
Posons que $\hat{\theta}_{(\cdot)}$ est la moyenne de k valeurs de $\hat{\theta}_{(\alpha)}$. L'estimateur du jackknife de la variance est donné par:

$$\begin{aligned} v_1(\hat{\hat{\theta}}) &= \frac{1}{k(k-1)} \sum_{\alpha=1}^{k} (\hat{\theta}_\alpha - \hat{\hat{\theta}})^2 \\ &= \frac{(k-1)}{k} \sum_{\alpha=1}^{k} (\hat{\theta}_\alpha - \hat{\theta}_{(\cdot)})^2 \end{aligned}$$

En pratique $v_1(\hat{\hat{\theta}})$ a été utilisé pour estimer la variance non seulement pour $\hat{\hat{\theta}}$ mais aussi pour $\hat{\theta}$.
Un autres estimateur est donné par:

$$v_2(\hat{\hat{\theta}}) = \frac{1}{k(k-1)} \sum_{\alpha=1}^{k} (\hat{\theta}_\alpha - \hat{\theta})^2$$

5.2.3 Le Bootstrap

L'idée du Bootstrap fut introduite en 1979 par B. Efron. Le principe de base est de rééchantillonner de façon indépendante B échantillons et de calculer pour chacun d'eux un estimateur du paramètre d'intérêt. Cette technique nécessite de longs calculs.

Techniques de base:
Supposons que l'on ait un échantillon de $y_1, \cdots, y_n$ variables aléatoires identiquement distribuées et indépendantes et de moyenne $\bar{y}$. Les étapes qu'il faut suivre pour appliquer la technique du Bootstrap sont les suivantes.

Algorithme du Bootstrap:

Premièrement:
Tirer un échantillon aléatoire simple avec remise de taille y_i^* à partir des valeurs observées $y_1, \cdots, y_n$ et calculer:

$$\hat{\theta}^* = g(\bar{y}^*)$$

où:

$$\bar{y}^* = \frac{\sum_{i=1}^{n} y_i^*}{n}.$$

Deuxièmement:
Refaire l'étape numéro un, d'une façon indépendante, un grand nombre de fois (B fois), et calculer les estimateurs pour chaque sous-échantillons tirés: $\hat{\theta}^{*1}, \cdots, \hat{\theta}^{*B}$.
Troisièmement:
L'estimateur de la variance de $\hat{\theta}$ par le Bootstrap est donné par:

$$v_B(a) = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^{*b} - \hat{\theta}_a^*)^2$$

où:

$$\hat{\theta}_a^* = \frac{\sum_{b=1}^{B} \hat{\theta}^{*b}}{B}.$$

5.2.4 Les groupes aléatoires

C'est la méthode des groupes aléatoires que nous avons retenue comme technique d'estimation de la variance dans le cadre notre proposition faite pour l'indice suisse des prix à la consommation. Historiquement, c'est l'une des premières techniques qui fut développée pour simplifier l'estimation de la variance lorsque l'on est en présence d'un plan d'échantillonnage complexe.

Techniques de base
Si on a un échantillon aléatoire simple de n observations $x_1, \ldots, x_n$, alors l'estimation usuelle de la variance S^2 est:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}.$$

Avec la méthode des groupes aléatoires, le principe de base pour estimer S^2 est le suivant. Posons que $\hat{\theta}$ est un estimateur du paramètre d'intérêt de la population θ. Premièrement, il faut diviser aléatoirement l'échantillon initial en k groupes. Les estimateurs du paramètre d'intérêt pour chaque groupe, $\hat{\theta}_\alpha$, sont calculés. Ces groupes sont appelés aussi des répliques. Puis on estime la variance en utilisant une expression de la forme:

$$v(\hat{\bar{\theta}}) = \sum_{\alpha=1}^{k} \frac{(\hat{\theta}_\alpha - \hat{\bar{\theta}})^2}{k(k-1)}$$

où $\hat{\theta}_\alpha$ sont des variables aléatoires non-corrélées et $\hat{\bar{\theta}}$ est défini par:

$$\hat{\bar{\theta}} = \sum_{\alpha=1}^{k} \frac{\hat{\theta}_\alpha}{k}.$$

L'avantage d'utiliser les répliques repose sur le fait qu'il y a possibilité de réduction de travail et de coût, ce qui peut être appréciable quand la variance d'un nombre considérable de statistiques doit être estimée. Dans les applications pratiques de sondage, les échantillons sont souvent tirés sans remise plutôt qu'avec remise. Néanmoins, dans ces cas-là, l'estimateur de la variance par la méthode des groupes aléatoires est largement utilisé, et surtout dans le cas des sondages fait sur une grande échelle.

Les estimateurs des répliques $\hat{\theta}_\alpha$ sont corrélés entre eux puisque l'échantillonnage est effectué sans remise. Or, l'estimateur calculé avec les répliques tend à estimer la variance comme si l'échantillon était sélectionné avec remise. Le prix que l'on doit alors payer est un biais dans l'estimateur de la variance.

La formation des répliques

Pour s'assurer que l'estimateur d'une réplique possède des propriétés statistiques acceptables, les répliques ne doivent pas être formées d'une façon arbitraire. Elles doivent être formées de manière à ce qu'elles aient essentiellement le même plan d'échantillonnage que l'échantillon initial. Pour cela, il faut suivre les règles suivantes qui dépendent du plan d'échantillonnage.

1. Si un échantillon à plan simple de taille n est tiré par échantillonnage aléatoire simple sans remise ou par échantillonnage à probabilités proportionnelles sans remise, alors les répliques devraient être formées en divisant l'échantillon initial aléatoirement. Ce qui signifie que la première réplique est obtenue en tirant un échantillon aléatoire simple sans remise de taille $m = \left[\frac{n}{k}\right]$ de l'échantillon initial. La seconde réplique est obtenue en tirant un échantillon aléatoire simple sans remise de taille m des $n - m$ unités restantes de l'échantillon initial et ainsi de suite. Si n/k n'est pas un entier, i.e. $n = km + q$ et $0 < q < k$, alors l'excédent q peut être omis des k répliques. Une autre manière de considérer l'excédent est d'ajouter une par une les unités à chacune des q premières répliques.

2. Si un échantillon systématique de taille n est tiré avec des probabilités égales ou inégales, alors les répliques doivent être formées en divisant l'échantillon initial d'une manière systématique. Cela peut se faire en générant un nombre entier aléatoirement entre 1 et k, disons α^*. La première unité de l'échantillon principal est alors assignée à la réplique α^*, la deuxième à la réplique $\alpha^* + 1$ et ainsi de suite modulo k.

3. Dans un échantillonnage à degrés multiples, les répliques doivent être construites en divisant les ultimes grappes ("ultimate clusters") i.e. les agrégats de toutes les unités élémentaires tirées

d'une même PSU en k groupes. Ainsi tous les degrés successifs tirés d'une PSU doivent être traités comme une seule unité au moment de former les répliques. La division des ultimes grappes en répliques est effectuée selon la règle 1. ou 2. Cela dépend de la nature du plan d'échantillonnage du 1er degré. Si le plan est aléatoire simple sans remise ou à probabilités proportionnelles sans remise, on utilisera la règle 1. Si le plan d'échantillonnage est systématique, alors on utilise la règle 2.

4. Pour un échantillonnage stratifié, nous avons deux *possibilités*. Premièrement, si on veut estimer la variance à l'intérieur d'une strate, on utilise une des trois règles précédentes selon la nature du plan d'échantillonnage de la strate. Par exemple, on utilisera la règle 3 si on a un plan à plusieurs degrés à l'intérieur de la strate. Deuxièmement, si on veut calculer la variance totale parmi toutes les strates, alors chaque réplique doit être un échantillon stratifié comprenant des unités de chaque strate. Dans ce cas, la première réplique est obtenue en tirant un échantillon aléatoire simple sans remise de taille $m_h = \frac{n_h}{k}$ de l'échantillon initial n_h de la strate h, pour $h = 1, \ldots, K$. La deuxième réplique est obtenue de la même manière en la tirant des $n_h - m_h$ unités restantes de la strate h. Les répliques restantes sont calculées de la même façon. S'il y a un excédent, i.e. $n_h = km_h + q_h$, elles peuvent être omises des k répliques ou alors ajoutées une par une aux q premières répliques. Si l'échantillon initial est tiré systématiquement à l'intérieur des strates, alors les répliques doivent aussi être formées de façon systématique. En d'autres termes, chaque réplique doit comprendre un sous-échantillon systématique de l'échantillon principal dans chaque strate.

5.3 Les séries de Taylor

Dans les applications pratiques de sondage, on se trouve souvent face à des estimateurs non-linéaires comme des ratios ou des différences de ratios (par exemple la variation de l'ISPC entre deux mois consécutifs est un ratio). Ceci rend le calcul de l'estimation de la variance du paramètre en question difficile. L'expression analytique exacte de la

variance échantillonnale d'un estimateur non-linéaire n'est souvent pas disponible, et moins encore l'estimateur non-biaisé de la variance.
Ce problème peut se résoudre en utilisant une procédure qui se résume en deux étapes. La première étape consiste à approximer l'estimateur par une fonction linéaire, et au cours la seconde étape on applique à cette approximation une formule d'estimation de la variance appropriée au plan d'échantillonnage. L'une des techniques de linéarisation parmi les plus connues est la méthode des séries de Taylor.

En suivant Wolter, K.M. (1985, pa. 236), posons que l'on a un ratio de deux totaux de population $\mathrm{R} = Y/X$, que l'on estime par $\hat{\mathrm{R}} = \hat{Y}/\hat{X}$ où $\hat{Y}$ et $\hat{X}$ représentent les estimateurs de Y et de X. En utilisant les séries de Taylor l'estimateur de la variance de $\hat{\mathrm{R}}$ est donné par:

$$\mathrm{v}(\hat{R}) \;=\; \hat{R}^2\left(\frac{v(\hat{Y})}{\hat{Y}^2}+\frac{v(\hat{X})}{\hat{X}^2}-2\frac{c(\hat{Y},\hat{X})}{\hat{X}\hat{Y}}\right)$$

où $v(\hat{Y})$, $v(\hat{X})$ et $c(\hat{Y},\hat{X})$ représentent les estimateurs de $Var(\hat{Y})$, $Var(\hat{X})$ et $Cov(\hat{Y},\hat{X})$.

Ainsi, nous avons adapté la théorie des groupes aléatoires pour calculer la variance de l'ISPC. Quant à l'estimation de la variance de la variation de l'indice entre deux mois consécutifs, elle a nécessité la combinaison entre la méthode des groupes aléatoires (estimateurs des variances et de la covariance) d'une part, et des séries de Taylor (estimateur de la variance du ratio) d'autre part. Les formules adaptées au cas de l'ISPC sont présentées dans la section suivante.

5.4 Proposition pour L'ISPC

5.4.1 Introduction

Dans cette section, nous proposons une méthodologie qui pourrait être appliquée, sous l'hypothèse d'échantillons aléatoires, pour estimer la variance de l'ISPC et la variance du changement de niveau de l'ISPC entre deux périodes.
Cette méthode a été testée sur des données réelles du groupe de l'alimentation pour la commune de Genève, grâce à la précieuse et étroite collaboration du service cantonal de statistique de Genève.

Notons que cette proposition a été élaborée avant l'entrée en vigueur du nouvel indice suisse des prix à la consommation en mai 1993. Nous avons donc utilisé pour notre étude la méthode de construction et de calcul de l'ISPC en vigueur jusqu'en mai 1993. Cependant, le concept méthodologique que nous proposons reste fondamentalement le même.

5.4.2 La méthodologie

Estimer la précision d'une statistique peut se faire au travers du calcul de la variance si on a utilisé des plans d'échantillonnage probabilistes pour sa construction. Cependant, comme cela a déjà été mentionné, à l'heure actuelle les échantillons servant à la construction de l'ISPC ne sont pas tous aléatoires. L'Office fédéral de la statistique (1990, 1993) envisage, lors de futures révisions, une possible mise sur pied d'échantillons probabilistes.

Cependant, la seule présence d'échantillons aléatoires n'est pas suffisante si l'on souhaite estimer la variance de l'ISPC. En effet, le plan d'échantillonnage étant à cinq niveaux, il n'est plus possible d'estimer la variance en ayant recours à la théorie classique de l'échantillonnage. Une méthode d'estimation de la variance basée sur le concept de rééchantillonnage est donc proposée.

Pour cela, on suggère d'utiliser la méthode des groupes aléatoires. Cette méthode est, d'un point de vue pratique, relativement aisée à appliquer. Elle est moins coûteuse (elle nécessite notamment moins de calculs) que les autres méthodes de rééchantillonnage tel que le Jackknife ou le Bootstrap.

La méthode des groupes aléatoires nécessite la présence de sous-échantillons que l'on appelle répliques. Une réplique peut être considérée comme un sous-échantillon de l'échantillon complet qui a le même plan d'échantillonnage que ce dernier.

On propose de construire les répliques à partir de l'échantillon aréolaire, c'est-à-dire à partir des aires géographiques, dans le cas de la Suisse les communes.

Notons tout d'abord qu'il s'agit de distinguer deux types de communes. Le premier type est constitué par les communes dont on supposera qu'elles ne représentent qu'elles-mêmes. On appelera alors une telle commune une unité primaire d'échantillonnage auto-représentative (UPE-AR).

Ce groupe de communes serait composé des communes de Bâle, Berne, Genève et Zurich. Nous avons choisi ces quatre communes car chacune d'entre elles a un poids de pondération qui lui est propre dans le calcul de l'ISPC. En règle générale, on retiendra dans ce type de communes celles qui sont retenues d'office dans un échantillon aléatoire. C'est-à-dire que cela touche les grandes villes ou métropoles d'un pays.

Le deuxième type englobe les communes qui représentent non seulement elles-mêmes mais aussi les autres communes. On parlera alors d'unité primaire d'échantillonnage non-auto-représentative (UPE-NAR). Ce groupe de communes serait composé des 44 communes restantes.

Pour construire les répliques pour les communes, il nous faut procéder de deux manières différentes selon qu'il s'agit des UPE-AR ou des UPE-NAR.

Dans le cas des UPE-AR, il faut disposer d'au moins deux répliques par UPE-AR. On peut alors procéder à au moins deux sélections de points de vente et de produits à l'intérieur de la commune. Chaque sélection a un plan d'échantillonnage identique et est appelée réplique. On obtient ainsi au moins deux échantillons de relevés de prix pour chaque unité primaire d'échantillonnage auto-représentative. A partir de chaque réplique on peut calculer un IPC. Puis, en appliquant la méthode des groupes aléatoires, on estime la variabilité à l'intérieur de la commune.

Dans le cas des UPE-NAR, on considère que chaque commune est une réplique des autres communes. A partir du calcul de l'IPC par commune, on peut alors calculer la variabilité de l'IPC pour les UPE-NAR. Il est naturellement possible d'envisager plusieurs groupes d'UPE-NAR, comme un groupe réunissant les communes de taille moyenne et un autre celles de petite taille.

A partir des 48 communes participantes, nous constituons un ensemble de répliques. En résumé on obtient, pour la Suisse, le tableau suivant:

	Communes	Nombre d'UPE	Nombres de répliques
UPE-AR	Bâle	1	2
	Berne	1	2
	Genève	1	2
	Zurich	1	4
UPE-NAR	Autres communes	44	44
Total	48 communes	48	54

En plus de la précision de l'indice lui-même, l'une des mesures parmi les plus intéressantes est la variance de la variation de l'indice entre des intervalles de temps (par exemple entre deux mois consécutifs ou entre deux mêmes mois à un an d'intervalle). On veut alors savoir si la variation du niveau de l'indice calculé est significative ou si elle est due aux erreurs d'échantillonnage.

Les relevés de prix de Genève, pour le groupe principal de l'alimentation, pour chaque poste de relevés (ou rubrique), ont été divisés aléatoirement en deux groupes ou répliques. Notons que cela ne correspond pas tout à fait à la proposition ci-dessus puisque on y suggère de diviser aléatoirement les points de vente en répliques. Cependant, pour des raisons pratiques (confidentialité des données), cela n'a pas été possible.
Un indice a été calculé pour chacune des deux répliques, ainsi qu'un indice regroupant tous les relevés de prix. Remarquons que ce dernier correspond à l'indice publié pour Genève.
Puis on a pu estimer, pour chaque rubrique ou poste de relevés, la variance par la méthode des groupes aléatoires.

La variance des rubriques au temps t est estimée par:

$$v(I_h^t) = \frac{\sum_{\alpha=1}^{2}(I_{h,\alpha}^t - I_h^t)^2}{2}$$

où I_h^t est l'indice de la rubrique h au temps t, $I_{h,\alpha}^t$ est l'indice de la rubrique h au temps t pour la réplique α (il y a deux répliques, $\alpha = 1$

et 2). Par exemple I_h^t est l'indice au mois d'août 1991 du *Lait entier, en emballage.*

On procède de façon similaire pour estimer la variance de l'indice de la rubrique h au temps $t-1$.

Les variances des indices au niveau supérieur d'agrégation (par exemple l'indice du *lait*) sont calculées en agrégant les variances pondérées de chaque rubrique appartenant au niveau inférieur.
Cette variance s'exprime par:

$$v(I_H^t) = \sum_{h=1}^{H} \left(\frac{w_h}{w_H}\right)^2 \cdot v(I_h^t)$$

où I_H^t designe l'indice au temps t d'une strate d'articles, par exemple le *Lait.*
w_h designe le coefficient de pondération de la rubrique h appartenant à la strate d'articles H. Par exemple, w_h représente le coefficient de pondération du *Lait entier en emballage.* w_H représente le coefficient budgétaire au niveau de la strate d'articles H, dans notre cas le poids de pondération pour le *Lait.*

Les variances calculées de cette façon incluent toutes les sources d'erreurs d'échantillonnage pour l'ensemble des niveaux du plan d'échantillonnage.

On s'intéresse encore à calculer la variance du changement de prix; voyons tout d'abord comment ce dernier s'exprime.
Le changement de prix entre deux périodes $t-1$ et t se calcule comme suit:

$$\Delta I^t = \frac{I^t - I^{t-1}}{I^{t-1}} = \frac{I^t}{I^{t-1}} - 1$$

On remarque que le changement de prix est un ratio de deux variables aléatoires (I^t et I^{t-1}).
Il faut alors estimer la variance du changement de prix en utilisant une méthode d'approximation linéaire; dans ce cas, on choisit d'utiliser les séries de Taylor.
En suivant Armknecht, P.A. (1991) et König, S. et Pfaff, S. (1991) *Indice des prix à la consommation des USA et calcul de sa variance*, on peut exprimer alors la variance du changement de prix par:

$$
\begin{aligned}
Var(\Delta I^t) &= Var\left(\frac{I^t}{I^{t-1}}\right) \\
&= \frac{1}{(I^{t-1})^2}\left(v(I^t) + \left(\frac{I^t}{I^{t-1}}\right)^2 \cdot v(I^{t-1}) - 2\frac{I^t}{I^{t-1}}Cov(I^t, I^{t-1})\right)
\end{aligned}
$$

où $v(I^t)$ et $v(I^{t-1})$ sont estimés par la méthode des groupes aléatoires, comme exposé ci-dessus. Et $Cov(I^t, I^{t-1})$ représente la covariance entre l'indice au temps t et l'indice au temps $t-1$. Cette covariance se calcule de la façon suivante, ce qui correspond à la méthode des groupes aléatoire pour la covariance.

$$
\begin{aligned}
Cov(I^t, I^{t-1}) &= ((I_1^{t-1} - I^{t-1}) \cdot (I_1^t - I^t) \\
&\quad +(I_2^{t-1} - I^{t-1}) \cdot (I_2^t - I^t))/2
\end{aligned}
$$

où I_1^t et I_1^{t-1} sont les indices au temps (t) et $(t-1)$ pour la réplique 1. I_2^t et I_2^{t-1} sont les indices au temps (t) et $(t-1)$ pour la réplique 2.

L'agrégation au niveau géographique

En disposant de données pour l'ensemble des communes, il serait possible de calculer la variance du changement de prix au niveau national. Le principe de base resterait le même que celui qui vient d'être exposé. La variance de l'indice ou du groupe principal, au temps t et $t-1$, s'exprimerait par:

$$v(I^t) = \sum_{j=1}^{5} g_j \cdot v(I_j^t)$$

où g_j est le coefficient de pondération des communes. I_j^t est l'indice au temps t pour la commune j. Rappelons que ces coefficients sont de 0.2 pour Zurich, 0.1 pour Bâle, 0.1 pour Berne, 0.1 pour Genève et 0.5 pour l'ensemble des autres communes de la Suisse.

5.5 Résultats numériques

En appliquant cette méthodologie, nous avons procédé à deux types de calculs. Premièrement, nous avons calculé la variance de l'IPC du groupe de l'alimentation de Genève, pour les mois de juin à octobre 1991. Deuxièmement, on a estimé la variance de la variation de prix, pour le groupe de l'alimentation de Genève, entre plusieurs mois consécutifs de l'année 1991. Les programmes informatiques et les données détaillées ayant servi aux calculs ci-dessous sont exposés en annexe A.

Remarque: On a choisi de présenter l'écart-type au niveau du groupe de l'alimentation, car au niveau des rubriques, les échantillons étant parfois de très petite taille (quelques observations), on ne peut alors pas s'attendre à une grande stabilité des estimateurs de la variance.

Dans le tableau suivant, on trouve l'IPC de l'alimentation pour Genève pour cinq mois de 1991, l'écart-type de l'IPC calculé en utilisant la méthode des groupes aléatoires, ainsi que l'intervalle de confiance à 95%.

Mois	Indice 1982 = 100	écart-type estimé (σ)	intervalle de confiance à 95%
Juin 91	133.37	0.41	$132.56 \leq \mu \leq 134.17$
Juillet 91	131.51	0.17	$131.18 \leq \mu \leq 131.84$
Août 91	131.42	0.24	$130.95 \leq \mu \leq 131.89$
Septembre 91	131.53	0.17	$131.20 \leq \mu \leq 131.86$
Octobre 91	131.96	0.18	$131.61 \leq \mu \leq 132.31$

En posant l'hypothèse que l'indice suit une loi normale, on peut calculer un intervalle de confiance à 95%. Par exemple pour le mois de juin 1991:

$$133.37 - 1.96 \cdot \sigma \leq \mu \leq 133.37 + 1.96 \cdot \sigma$$

Cela signifie que l'intervalle de confiance contient, avec une probabilité de 95%, la vraie valeur de l'indice. Les autres intervalles de confiance, se trouvant dans la quatrième colonne du tableau, s'interprètent de manière identique.

Dans le tableau suivant, nous donnons les résultats de l'estimation de la variance de la variation de prix, pour le groupe de l'alimentation de Genève, entre plusieurs mois consécutifs de l'année 1991. On y trouve la variation (en %) de l'IPC entre deux mois consécutifs et l'écart-type estimé (en %) pour l'IPC.

	variation en % de l'IPC	écart-type estimé en % σ
juin-juillet 91	-1.39	0.46
juillet-août 91	-0.06	0.26
août-septembre 91	0.08	0.20
septembre-octobre 91	0.31	0.24

Par la théorie statistique, nous savons que le ratio de deux lois normales est une loi de Cauchy. Cependant, nous avons construit un intervalle de confiance en posant l'hypothèse que la variation de prix du groupe de l'alimentation pour Genève suit une loi normale. Notons qu'un test d'adéquation du χ^2 nous a permis de vérifier l'hypothèse de normalité du changement de prix. Celui-ci est présenté chez König, S. et Pfaff, S. (1992) *"Indice des prix à la consommation en Suisse et calcul de sa variance"*.
En construisant un intervalle de confiance, avec un niveau de confiance de 95%, on veut vérifier si la variation en pourcent de l'indice est significativement différente de zéro. Si la variation en pourcent de l'indice est comprise dans l'intervalle $(0 - 1.96 \cdot \sigma \leq \mu \leq 0 + 1.96 \cdot \sigma)$, alors on peut conclure que la variation de l'IPC n'est pas significative et qu'elle n'est due qu'à l'aléa des échantillons. Au contraire, si la variation de l'indice n'est pas comprise dans cet intervalle de confiance, alors on peut dire qu'elle est significative et qu'il y a eu une variation de prix significative entre les deux périodes considérées.
Calculons donc l'intervalle de confiance avec un niveau de confiance de 95% pour chacune des variations de l'indice.

La variation de juin-juillet 1991:
Intervalle de confiance: $-0.9016 \leq \mu \leq +0.9016$.
La variation en pourcent de l'indice se trouve en dehors de l'intervalle de confiance. On peut alors penser que la variation est significative. Cela signifie que la variation des prix entre juin et juillet 1991 n'est pas due qu'aux aléas de l'échantillonnage, mais qu'il y a eu une variation (diminution) significative du niveau des prix pour l'IPC de l'alimentation.
La variation de juillet-août 1991:
Intervalle de confiance: $-0.5096 \leq \mu \leq +0.5096$.

La variation en pourcent de l'indice se trouve dans l'intervalle de confiance. Cela indique qu'il n'y a pas de variation significative de l'indice de l'alimentation. Donc il semble que la variation de l'indice entre juillet et août ne soit due qu'à l'aléa de l'échantillonnage.
La variation de août-septembre 1991:
Intervalle de confiance: $-0.392 \leq \mu \leq +0.392$.
La variation en pourcent de l'indice se situe dans l'intervalle de confiance. Cela semble indiquer, comme dans l'exemple précédent, qu'il n'y a pas de variation significative de l'indice de l'alimentation.
La variation de septembre-octobre 1991:
Intervalle de confiance: $-0.4704 \leq \mu \leq +0.4704$.
La variation en pourcent de l'indice se trouve dans l'intervalle de confiance. Cela signifie, comme nous l'avons vu ci-dessus, qu'il n'y a pas de variation significative de l'indice de l'alimentation.

De tels tests pourraient naturellement être effectués pour la variation annuelle d'un indice. On pourrait ainsi estimer si la variation annuelle est significative ou si elle n'est due qu'aux aléas de l'échantillonnage.

5.5.1 Comparaison internationale

Dans cette section, nous présentons quelques résultats numériques relatifs à des études menées aux Etats-Unis, en France, aux Pays-Bas, en Suède et en Italie. Chacun de ces pays a utilisé sa propre méthode pour procéder à des estimations de variances. Nous devons remarquer que ce n'est pas toujours la variance totale de l'IPC qui a été mesurée, mais qu'il s'agit souvent de l'estimation de la variance d'une des composantes de l'IPC, comme la variance due à l'échantillon des ménages.

Les Etats-Unis
Dès 1978, le Bureau of Labor Statistics (BLS) a estimé la variance de l'IPC américain. L'écart-type, calculé sur la période 1978-1986, est conditionnel. En effet, il comprend la variation d'échantillonnage provenant de l'échantillon de prix et de loyers, mais il n'inclut pas l'écart-type dû aux coefficients des coûts initiaux. Une description détaillée de la méthode est présenté au cours du chapitre 3, ainsi que chez BLS (1988), P.A. Armknecht (1991) et S. König et S. Pfaff (1991).

Le tableau 1 indique la variation moyenne en pourcentage sur deux

mois consécutifs, l'erreur type moyenne et la fourchette des erreurs types pour l'IPC des Etats-Unis, par groupes majeurs entre janvier 1978 et décembre 1986.
Le tableau 2 indique la variation moyenne en pourcentage sur douze mois, l'erreur type moyenne et la fourchette des erreurs types pour l'IPC des Etats-Unis par groupes majeurs entre janvier 1978 et décembre 1986.

Tableau 1: USA - Calcul des variations et des erreurs types sur deux mois consécutifs de janvier 1978 à décembre 1986

Catégorie de de dépenses	Variation moyenne en %	Erreur type moyenne en %	Erreur type minimum	Erreur type maximum
Tous articles	0.5	0.05	0.04	0.09
Alimentation et boissons	0.5	0.08	0.05	0.18
Habitation	0.5	0.11	0.07	0.23
Logement	0.6	0.17	0.09	0.27
Combustibles et services publics divers	0.5	0.18	0.09	0.85
Equipement et tenue du ménage	0.4	0.17	0.07	0.36
Habillement et entretien	0.3	0.27	0.14	0.68
Transports	0.5	0.06	0.03	0.12
Soins médicaux	0.7	0.12	0.05	0.73
Loisirs	0.4	0.18	0.10	0.76
Autres biens et services	0.6	0.10	0.04	0.55

Source: P.A. Armknecht (1991) Replication and Regression: modern solutions to age old problems in price index construction.

Tableau 2: USA - Calcul des variations et erreurs types sur douze mois de janvier 1978 à décembre 1986

Catégorie de de dépenses	Variation moyenne en %	Erreur type moyenne en %	Erreur type minimum	Erreur type maximum
Tous articles	6.6	0.12	0.09	0.15
Alimentation et boissons	5.4	0.18	0.12	0.32
Habitation	7.0	0.22	0.15	0.31
Logement	7.8	0.31	0.22	0.43
Combustibles et services				
publics divers	7.7	0.48	0.19	1.00
Equipement et tenue du ménage	4.5	0.34	0.22	0.77
Habillement et entretien	3.5	0.73	0.40	1.18
Transports	6.8	0.17	0.11	0.22
Soins médicaux	9.0	0.43	0.25	0.89
Loisirs	5.7	0.52	0.37	0.96
Autres biens et services	8.2	0.35	0.19	0.60

Source: P.A. Armknecht (1991) Replication and Regression: modern solutions to age old problems in price index construction.

La France

Dans le tableau 3, concernant la France, on trouve les écarts type dus aux choix des variétés et des agglomérations, des indices par secteurs et de l'indice général en 1987 et 1988. Ces chiffres proviennent d'un article écrit par P. Ardilly et F. Guglielmetti (1991).

Tableau 3: France - Ecarts type dus aux choix des variétés et des agglomérations en 1987 et 1988

	1987 Ecart-type	1988 Ecart-type
Alimentation	0.061	0.061
Habillement	0.140	0.140
Autres manufacturés	0.073	0.070
Services	0.096	0.082
Indice général	0.045	0.042

Source: P. Ardilly et F. Guglielmetti (1991) Précision de l'indice des prix français et optimisation des échantillons.

Les Pays-Bas

Le tableau 4 est issu d'un article du Bureau central de statistique des Pays-Bas écrit par B. Balk (1991). Dans ce tableau, on trouve la variance conditionnelle due à l'échantillonnage des ménages pour l'indice des prix à la consommation de 1990. L'année de base est 1985.

Tableau 4: Les Pays-Bas - Variance conditionnelle due à l'échantillonnage des ménages pour l'IPC de 1990

	Indice	Erreur type
Consommation totale		
Ensemble des ménages	104.3	0.04
Ménage de salariés à faible revenu	103.7	0.06
Ménage de salariés à revenu élevé	104.5	0.07
Alimentation, boissons, tabac		
(ensemble des ménages)	101.3	0.04
Mobilier, etc. (ensemble des ménages)	106.4	0.07

Source: B. Balk (1991) Estimating the precision of a consumer price index; some experiences from the Netherlands.

La Suède

La sourcc des chiffres mentionnés dans le tableau 5 est un article de J. Dalén (1991) écrit pour Statistics Sweden. On y trouve la variance de l'indice des prix à la consommation de décembre 1990, sur la base de décembre 1989. La variance calculée comprend la variabilité due aux prix des articles, la variabilité due aux différences de prix selon les points de vente et l'interaction entre les différences de prix selon les articles et selon les points de vente.

Tableau 5: La Suède - Variance de l'IPC de décembre 1990, sur la base de décembre 1989

	Enquête sur les prix locaux	Enquête sur les prix de catalogue
Variance article	0.1328	0.0466
Variance point de vente	0.0798	0.0662
Variance interaction	0.0566	0.0162
Variance totale	0.2692	0.1290

Source: J. Dalén (1991) The work with an error model and error calculations in the Swedish CPI.

L'Italie

En Italie, une étude menée par L. Biggieri et A. Giommi (1987) a permis d'estimer la variabilité de l'IPC due à l'échantillonnage de l'enquête sur la consommation des familles. Le calcul a été effectué en utilisant la méthode des groupes aléatoire (RG), la méthode du Balanced Half Sample (BHS) et la méthode du Jackknife (J). Les résultats sont donnés dans le tableau 6.

Tableau 6: L'Italie - La variabilité de l'IPC due à l'échantillonnage de l'enquête sur la consommation des familles

	Décembre 1984		Novembre 1985		Décembre 1985	
	IPC	Erreur type	IPC	Erreur type	IPC	Erreur type
RG	106.68	0.024	114.51	0.061	115.04	0.075
BHS	106.68	0.031	114.51	0.048	115.04	0.047
J	106.68	0.031	114.51	0.048	115.04	0.049

Source: L. Biggieri et A. Giommi (1987) On the accuracy and precision of the consumer price index.

5.5.2 Conclusion

Nous remarquons que comparer les différentes études internationales s'avère difficile. En effet, chaque pays a utilisé des méthodes différentes, et la variance qui a été estimée ne représente pas toujours la variance totale de l'IPC comme dans le cas de Genève. Les Etats-Unis ont estimé la variance du changement de prix de l'IPC. L'Italie a utilisé des méthodes basées sur le concept de rééchantillonnage pour estimer la variabilité due à l'échantillonnage des ménages. La méthode utilisée par la Suède permet de savoir quelle composante entre l'échantillon d'articles ou l'échantillon de points de vente a la variance la plus grande. Cette information indique alors quel est l'échantillon le plus sensible. Cela permet de savoir sur quel échantillon il faut porter une attention toute particulière lors de sa sélection.
On constate que l'estimation de la variance de l'IPC n'en est encore qu'à ses premiers pas. Mais il ne fait aucun doute que ce sujet deviendra important à l'avenir, compte tenu de l'importance de l'IPC dans la vie économique. L'Office fédéral de la statistique (1993) et la Commission de statistique conjoncturelle et sociale (1992) reconnaissent par ailleurs que c'est un thème sur lequel il faudra se pencher lors des futures révisions de l'indice suisse des prix à la consommation.

Chapitre 6

L'IPC et l'usage de la médiane

6.1 Introduction

Ce chapitre a pour but de s'interroger sur une possible utilisation de la médiane dans le calcul de l'indice des prix à la consommation. Nous commencerons par définir la médiane, donner quelques exemples de calcul et citer ses avantages et inconvénients. Puis nous examinerons dans quel contexte et pour quel type d'analyse la médiane a été utilisée dans le cadre de l'indice des prix à la consommation; pour ce faire, nous étudierons un article de B. Barberi. Enfin, nous émettons quelques suggestions relatives à l'utilisation de la médiane dans le calcul de l'indice des prix à la consommation; dans cette dernière section, nous suggérons l'utilisation de la médiane lors du calcul de l'indice à l'intérieur d'un canal de distribution et pour un poste de relevé.

6.2 La médiane

La médiane, qui est une mesure de tendance centrale, peut être définie comme étant la valeur centrale des variables, lorsque ces dernières sont rangées par ordre croissant ou décroissant.
Notons que lorsque le nombre d'observations est impair, la médiane

correspond à la $((n+1)/2)$ observation. Si le nombre d'observations est pair, alors la médiane correspond à la $(0.5 \cdot ((n/2)+(n/2+1)))$ observation.
Dans le cas d'une courbe de fréquences, la médiane peut être définie comme étant la variable qui divise l'aire de la courbe en deux parties égales.
Illustrons par les exemples suivants le calcul de la médiane:

Exemple 6.1: Cas d'une variable discrète.
Voici les notes obtenue lors d'un examen par 9 étudiants:

$$4 \quad 5 \quad 3 \quad 4 \quad 6 \quad 5 \quad 5 \quad 4 \quad 2$$

Dans ce cas, la médiane se calcule comme suit. Premièrement, il faut classer les observations par ordre croissant (ou décroissant):

$$2 \quad 3 \quad 4 \quad 4 \quad 4 \quad 5 \quad 5 \quad 5 \quad 6$$

La médiane correspond alors à la $[(n+1)/2]$ observation, c'est-à-dire, dans notre exemple, à la cinquième observation:

$$Me = 4$$

Pour illustrer le cas où le nombre d'observations est pair, considérons l'exemple suivant. Dans une classe de 12 élèves, les notes obtenues lors d'un examen sont les suivantes:

$$2 \quad 4 \quad 6 \quad 5 \quad 5 \quad 3 \quad 4 \quad 5 \quad 4 \quad 3 \quad 6 \quad 5$$

Comme précédemment, la première étape consiste à les ranger par ordre:

$$2 \quad 3 \quad 3 \quad 4 \quad 4 \quad 4 \quad 5 \quad 5 \quad 5 \quad 5 \quad 6 \quad 6$$

La médiane est alors calculée ainsi:

$$\begin{aligned} Me &= \frac{1}{2}((n/2)obs + (n/2+1)obs) \\ &= \frac{1}{2}(4+5) \\ &= 4,5 \end{aligned}$$

Exemple 6.2: Cas d'une variable continue.
En raison du groupement des observations par classe, on situe dans un premier temps la médiane dans une classe, que l'on appelle la classe médiane. La classe i est la classe médiane si:

$$F_{i-1} < \frac{1}{2} < F_i$$

Dans l'exemple ci-dessous, on a un tableau représentant la distribution de 150 individus selon leur salaire:

No. de classe	classes	effectifs n_i	effectifs cumulés N_i	fréquences cumulées F_i
1	1000-1500	25	25	0.166
2	1500-2000	40	65	0.433
3	2000-2500	75	140	0.933
4	2500-3000	10	150	1.000

La classe médiane correspond à la classe numéro 3 ($n/2 = 150/2 = 75$). On suppose que les 75 individus de la troisième classe sont répartis uniformément dans la classe. L'individu qui occupe la position médiane a la 75 ème place, mais la dixième place dans la classe. Le salaire médian est alors calculé comme suit:

$$Me = 2000 + \frac{(2500 - 2000) \cdot (75 - 65)}{(140 - 65)} = 2066,66$$

6.2.1 Les avantages et inconvénients de la médiane

Yule et Kendall posent six conditions pour qu'une mesure de tendance centrale (moyenne, médiane) soit considérée comme bonne. Ces conditions sont énumérées ci-dessous:

1. La mesure doit être définie précisément, afin qu'il n'y ait pas de libre appréciation.

2. Elle doit être basée sur toutes les observations, sinon ce n'est pas une caractéristique de la distribution entière.

3. Les propriétés doivent être simples et compréhensibles.
4. Les calculs doivent être rapides et aisés.
5. Cette mesure doit être affectée le moins possible par les fluctuations d'échantillonnage.
6. Le traitement algébrique doit être facile. Si deux ou plusieurs séries d'observations sont données, la mesure de tendance centrale des séries combinées doit être possible.

Les trois premières conditions sont remplies tant par la moyenne arithmétique que par la médiane. Cette dernière se calcule plus facilement que la moyenne, mais celle-ci est moins sensible aux fluctuations d'échantillonnage. S'il y a plusieurs séries d'observations que l'on doit combiner en une seule série, la moyenne résume ces séries par la moyenne de leurs composantes. La médiane ne permet pas en général de résumer un ensemble de séries en une seule série (la valeur de la médiane résultante dépend de la forme des distributions des séries et pas seulement de leurs médianes). La médiane n'est pas influencée par des valeurs extrêmes ou aberrantes, comme c'est le cas pour la moyenne. Cette propriété est l'un des avantages principaux de la médiane. Lorsque la loi de distribution suivie par les observations n'est pas connue, il n'est pas toujours aisé de savoir quel estimateur est le meilleur. Il faut alors utiliser un estimateur qui soit efficace dans n'importe quelle situation. Ce type d'estimateur est dit robuste. Pour des distributions qui ne suivent pas une loi normale, la médiane est plus efficace que la moyenne arithmétique. L'une des propriétés de la médiane est que la somme des écarts (déviations en valeur absolue) entre chacune des valeurs d'un ensemble et une valeur x_o fixée est minimale lorsque la valeur x_o est la médiane:

$$\sum_i \mid x_i - Me \mid \leq \sum_i \mid x_i - x_o \mid si x_o \neq Me$$

6.3 L'emploi de la médiane dans l'analyse statistique des prix

Le titre de cette section reprend volontairement celui d'un article publié par B. Barberi (1965) dans le Bulletin de l'Institut International de Statistique (35ème session) et cette section est consacrée à l'analyse qu'il a effectuée. Barberi ne s'est pas préoccupé de la méthode de calcul utilisée pour construire un indice des prix à la consommation. En effet, il a travaillé sur des indices déjà existants. Son étude a consisté à analyser les causes des variations des indices de prix dans le temps. Barberi a donc appliqué l'idée d'Edgeworth concernant l'utilisation de la médiane comme outil statistique pour analyser les changements de prix. En effet, ces derniers préoccupent tous les économistes et statisticiens qui recherchent une explication scientifique aux comportements des prix.

Dans un premier temps nous décrivons ci-dessous l'analyse de Barberi, puis sa démarche, sur des données réelles pour la ville de Zurich, et l'on constatera que l'on arrive aux mêmes conclusions que Barberi.

6.4 L'analyse de Barberi

Exposons les considérations méthodologiques faites par Barberi qui touchent à la nature et la valeur scientifique des indices des prix. Barberi remarque que les discussions portent premièrement sur la nature et la signification des indices de prix en tant que mesure de grandeur économique, tels que niveau général des prix, pouvoir d'achat de la monnaie; et deuxièmement sur des questions d'ordre technique, comme: l'application des méthodes d'échantillonnage pour le choix des articles dont le prix est à relever et des points de vente.
Barberi souligne que la théorie des indices peut être organisée autour de trois concepts. Premièrement, le concept comptable qui se rattache aux constructions d'indices composés, obtenus comme rapport entre agrégats ou comme moyenne arithmétique pondérée des prix relatifs. Deuxièmement, le concept monétaire qui voit l'étude des prix comme l'étude de la parité de la monnaie. Troisièmement, le concept physique où l'on considère les variations de prix comme étant dues aux diverses parties de l'économie.
La médiane joue un rôle très important, selon Barberi, si on admet les trois concepts mentionnés ci-dessus. Barberi fait remarquer que F.Y. Edgeworth soutient l'utilisation de la médiane et préconise son usage en lieu et place de la moyenne lorsque l'on dispose d'un échantillon de prix. Il fait aussi remarquer que les prix (ou leur indice) de divers

produits ne forment pas un ensemble homogène, mais sont constitués de divers ensembles. La courbe de distribution de l'ensemble des prix des divers produits est à ensivager comme un ensemble de différentes familles de courbes symétriques.
Dans son analyse, Barberi s'intéresse au comportement des prix individuels des divers produits et il préconise l'utilisation de la médiane des observations. Il ne tient pas compte de l'importance économique des produits, car il considère chaque produit comme un élément distinct de la collectivité à laquelle il appartient. Pour Barberi, l'emploi de la médiane permet de suivre le trend général de l'évolution des prix et de voir les caractéristiques dynamiques et certains aspects de la dispersion des prix dans le temps.

Pour son analyse, Barberi a distingué deux cas, le premier cas étant constitué par une analyse des prix des produits alimentaires de l'Italie pour quatre années consécutives (1961, 1962, 1963 et 1964). Le choix de Barberi s'est porté sur 76 articles individuels. Il a exclu les articles dit composés, dont les indices sont obtenus comme moyenne de plusieurs indices élémentaires et qui sont généralement indiqués par l'expression "produits divers" ou "autres produits". L'analyse de Barberi est composée de trois étapes. La première étape consiste en la formation de la distribution des indices. La deuxième étape consiste en la détermination des valeurs caractéristiques des distributions des prix et la troisième étape est consacrée à la détermination des distributions de prix par quartile. Pour le second cas, concernant des produits non alimentaires et des services, Barberi a choisi 86 produits manufacturés et 54 services et produits divers. L'analyse fut la même que pour celle du premier cas. Barberi arrive à la conclusion que la hausse des prix à la consommation est le résultat des hausses de produits dans les secteurs influencés, non pas par la demande des consommateurs, mais par d'autres facteurs de type institutionnel.

6.5 L'analyse de Barberi appliquée au cas de Zurich

A partir de l'article de Barberi, nous avons appliqué son analyse sur des données de la ville de Zurich, en choisissant deux mois consécutifs (août

1991 et septembre 1991), pour les rubriques du groupe de l'alimentation. Commençons tout d'abord par examiner la distribution des indices des prix à la consommation des 127 produits alimentaires.

	Sept. 91		Août 91	
Indices	f	F	f	F
Total		127		127
80- 90	1	126	1	126
90-100	6	120	6	120
100-110	10	110	11	109
110-120	23	87	23	86
120-130	28	59	26	60
130-140	26	33	24	36
140-150	11	22	13	23
150-160	10	12	8	15
160-170	4	8	5	10
170-180	1	7	2	8
180-190	3	4	3	5
190-200	1	3	-	-
200-210	1	2	2	3
210-220	-	-	1	2
220-230	1	1	-	-
230-240	-	-	1	1
240-250	1	0	1	0

La seconde étape consiste à calculer les valeurs caractéristiques des distributions des prix des 127 denrées alimentaires. Avant de présenter le tableau des différentes valeurs, commençons par exposer les différentes caractéristiques qui sont calculées.
Barberi donne les valeurs du premier quartile (Q_1), de la médiane (M) et du troisième quartile (Q_3), et il calcule quatres mesures de "skewness" ou mesures de symétrie.
La relation entre les quartiles Q_1 , Q_3 et la médiane M est définie par:

$$k = \frac{(Q_1 + Q_3) - 2M}{Q_3 - Q_1}$$

k donne une mesure de la symétrie de la distribution; on définit k également comme suit:

$$k = \frac{2(\bar{Q} - M)}{(M - Q_1) + (Q_3 - M)}$$

avec moyenne des quartiles $\bar{Q} = (Q_1 + Q_3)/2$.
En posant:

$$S_{\bar{q}} = \bar{Q} - M$$

$$\bar{S}_q = \frac{(M - Q_1) + (Q_3 - M)}{2}$$

on obtient finalement:

$$k = \frac{S_{\bar{q}}}{\bar{S}_q}$$

où le numérateur représente l'écart de la moyenne des quartiles à la médiane et le dénominateur représente la moyenne des écarts absolus des quartiles à la médiane.
Pour Barberi le rapport k peut s'interpréter comme un écart standardisé qui donne, en même temps, par la valeur numérique une mesure de dispersion relative, et par le signe le type d'asymétrie de la distribution.
On peut maintenant présenter le tableau qui regroupe les valeurs caractéristiques de distributions:

Mois	Q_1	M	Q_3	$\bar{Q}$	$S_{\bar{q}}$	$\bar{S}_q$	k
Août	115.78	128.23	142.95	129.36	+1.13	13.58	+0.0832
Septembre	115.58	128.70	141.36	128.47	-0.23	12.89	-0.0179

La troisième étape permet d'obtenir une vue d'ensemble de l'évolution des prix. Pour cela, on construit un tableau d'entrées-sorties dont chaque ligne indique dans quel quartile sont allés se placer, pour le mois de septembre, les prix qui au mois d'août occupaient les quartiles indiqués dans la première colonne.

	Sep.			
août	Q_1	Q_2	Q_3	Q_4
Q_1	28	3	-	-
Q_2	2	25	3	-
Q_3	2	5	24	3
Q_4	1	-	4	27
Total	33	33	31	30

On remarque qu'environ 82% des prix restent dans le même quartile que le mois précédent. Cette conclusion est la même que celle de l'analyse de Barberi. En effet, ce dernier a conclu qu'il y a une règle de permanence des prix à rester dans le même quartile. Cette analyse permet de constater que l'on arrive aux mêmes types de résultats que Barberi, pour des époques fort éloignées l'une de l'autre. De plus, dans le cas de Barberi, c'est l'analyse des mouvements de prix d'un pays, en l'occurence l'Italie, qui est faite, et dans le cas de Zurich, ce sont les variations de prix pour une ville qui sont étudiés.

6.6 La médiane et le calcul de l'IPC

Nous avons vu précédemment, notamment au chapitre 2, que l'indice suisse des prix à la consommation est calculé en utilisant l'indice de Laspeyres des prix. Nous avons montré que les indices élémentaires servant à déterminer l'indice d'une rubrique sont calculés avec une moyenne arithmétique. Dans cette section, nous proposons de calculer l'indice d'une rubrique non pas à partir de la moyenne des indices élémentaires, mais à partir de la médiane des indices élémentaires. Pour ce faire, nous allons illustrer cette proposition par un exemple numérique. Puis nous comparerons les deux méthodes.

Commençons par rappeler la formule théorique de l'indice de Laspeyres des prix:

$$IL^t = \frac{\sum_{j=1}^{n} q_j^0 \cdot p_j^t}{\sum_{j=1}^{n} q_j^0 \cdot p_j^0}$$

où p_j^t est le prix de l'article j au temps t, p_j^0 est le prix de l'article j au temps 0 et q_j^0 est le poids de pondération de l'article j au temps 0.

Examinons maintenant l'exemple numérique suivant. Nous avons calculé un indice en utilisant premièrement la méthode qu'on applique pour calculer l'ISPC (voir chapitre 2), et deuxièmement, nous construisons en parallèle un indice dont les étapes de calcul sont identiques, sauf en ce qui concerne la deuxième étape de la méthode, où nous utilisons la médiane. Les données fictives suivantes servent de base aux exemples numériques.

Imaginons que l'on veuille calculer l'IPC de la rubrique du lait, entre mars 1992 ($t = 0$) et avril 1992 ($t = 1$), pour une commune. Dans le tableau ci-dessous, on trouve les données par canal de distribution et pour deux périodes consécutives ainsi que les poids de pondération des postes de relevés.

Lait ($w = 1.245$)	Migros		Coop		Autres	
	t_0	t_1	t_0	t_1	t_0	t_1
Lait entier en	1.50	1.55	1.45	1.40	1.60	1.55
emballage	1.60	1.55	1.55	1.60	1.45	1.70
($w = 0.374$)	1.70	1.90	1.70	1.80	1.60	1.35
	1.45	1.60			1.55	1.55
	1.50	1.75				
Lait drink ($w = 0.374$)	1.60	1.65	1.55	1.55	1.60	1.50
	1.80	1.80	1.65	1.75	1.70	1.75
	1.80	1.70	1.55	1.80	1.50	1.55
	1.80	1.90	1.75	1.75		
			1.80	1.85		
Autre Lait en	1.70	1.80	1.75	1.75	1.90	2.00
emballage	1.90	2.00	1.80	1.85	1.85	1.85
($w = 0.062$)	1.85	1.80	2.00	2.10	1.80	1.75
Lait entier en vrac	1.70	1.70	1.70	1.75	1.80	1.85
($w = 0.435$)	1.75	1.85			1.70	1.65
					1.70	1.70
					1.75	1.80
					1.70	1.70

A partir de ces données, calculons maintenant les indices. Les poids de pondération des canaux de distribution sont les suivants: Migros $a_1 = 0.20$, Coop $a_2 = 0.15$ et autres canaux $a_3 = 0.65$. On calcule tout d'abord les indices élémentaires des observations individuelles de prix (I^t_{hije}), et l'on obtient le tableau suivant:

	Migros	Coop	Autres
Lait entier en emball.	103.34	96.55	96.87
	96.87	103.22	117.24
	111.76	105.88	84.37
	110.34		100.00
	116.67		
Lait drink	103.12	100.00	93.75
	100.00	106.06	102.94
	94.44	116.13	103.34
	105.55	100.00	
		102.78	
Autre lait en emball.	105.88	100.00	105.26
	105.26	102.78	100.00
	97.30	105.00	97.22
Lait entier en vrac	100.00	102.94	102.77
	105.71		97.06
			100.00
			102.86
			100.00

La deuxième étape consiste à calculer l'indice à l'intérieur d'un canal de distribution et pour un poste de relevé. C'est au cours de cette étape que l'on va utiliser dans un cas la moyenne et dans l'autre cas la médiane pour déterminer l'indice.

Lait entier en emballage:

$$I^t_{moy,migros} = \frac{103.34 + 96.87 + 111.76 + 110.34 + 116.67}{5} = 107.89$$

$$I^t_{med,migros} = 110.34$$

$$I^t_{moy,coop} = \frac{96.55 + 103.22 + 105.88}{3} = 101.88$$

$$I^t_{med,coop} = 103.22$$

$$I^t_{moy,autre} = \frac{96.87 + 117.24 + 84.37 + 100.00}{4} = 99.62$$

$$I^t_{med,autre} = \frac{96.87 + 100.00}{2} = 98.43$$

Lait drink:

$$I^t_{moy,migros} = \frac{103.12 + 100.00 + 94.44 + 105.55}{4} = 100.78$$

$$I^t_{med,migros} = \frac{100.00 + 103.12}{2} = 101.56$$

$$I^t_{moy,coop} = \frac{100.00 + 106.06 + 116.13 + 100.00 + 102.78}{5} = 104.99$$

$$I^t_{med,coop} = 102.78$$

$$I^t_{moy,autre} = \frac{93.75 + 102.94 + 103.34}{3} = 100.01$$

$$I^t_{med,autre} = 102.94$$

Autre lait en emballage:

$$I^t_{moy,migros} = \frac{105.88 + 105.26 + 97.30}{3} = 102.81$$

$$I^t_{med,migros} = 105.26$$

$$I^t_{moy,coop} = \frac{100.00 + 102.78 + 105.00}{3} = 102.59$$

$$I^t_{med,coop} = 102.78$$

$$I^t_{moy,autre} = \frac{105.26 + 100.00 + 97.22}{3} = 100.82$$

$$I^t_{med,autre} = 100$$

Lait entier en vrac:

$$I^t_{moy,migros} = \frac{100.00 + 105.71}{2} = 102.85$$

$$I^t_{med,migros} = 102.85$$

$$I^t_{moy,coop} = 102.94$$

$$I^t_{med,coop} = 102.94$$

$$I^t_{moy,autre} = \frac{102.77 + 97.06 + 100.00 + 102.86 + 100.00}{5} = 100.54$$

$$I^t_{med,autre} = 100.00$$

Dans la troisième étape, on calcule l'indice par poste de relevé, une fois pour l'indice calculé avec la moyenne et une fois pour l'indice calculé avec la médiane.

Lait entier en emballage:

$$I^t_{moy} = (107.89 \cdot 0.2) + (101.88 \cdot 0.15) + (99.62 \cdot 0.65) = 101.613$$

$$I^t_{med} = (110.34 \cdot 0.2) + (103.22 \cdot 0.15) + (98.43 \cdot 0.65) = 101.53$$

Lait drink:

$$I^t_{moy} = (100.78 \cdot 0.2) + (104.99 \cdot 0.15) + (100.01 \cdot 0.65) = 100.911$$

$$I^t_{med} = (101.56 \cdot 0.2) + (102.78 \cdot 0.15) + (102.94 \cdot 0.65) = 102.64$$

Autre lait en emballage:

$$I^t_{moy} = (102.81 \cdot 0.2) + (102.59 \cdot 0.15) + (100.82 \cdot 0.65) = 101.483$$

$$I^t_{med} = (105.26 \cdot 0.2) + (102.78 \cdot 0.15) + (100.00 \cdot 0.65) = 101.469$$

Lait entier en vrac:

$$\begin{aligned} I^t_{moy} &= (102.85 \cdot 0.2) + (102.94 \cdot 0.15) + (100.54 \cdot 0.65) \\ &= 101.362 \end{aligned}$$

$$\begin{aligned} I^t_{med} &= (102.85 \cdot 0.2) + (102.94 \cdot 0.15) + (100.00 \cdot 0.65) \\ &= 101.011 \end{aligned}$$

L'étape suivante consiste à calculer l'indice de la rubrique du Lait:

$$\begin{aligned} I^t_{moy,Lait} &= ((101.61 \cdot 0.374) + (100.91 \cdot 0.374) + (101.48 \cdot 0.062) \\ &\quad +(101.36 \cdot 0.435))/1.245 \\ &= 101.30 \end{aligned}$$

$$
\begin{aligned}
I_{med,lait} &= ((101.53 \cdot 0.374) + (102.64 \cdot 0.374) \\
&\quad +(101.47 \cdot 0.062) + (101.01 \cdot 0.435))/1.245 \\
&= 101.68
\end{aligned}
$$

Ainsi dans notre exemple numérique, on obtient l'indice 101.30 pour la rubrique du Lait lorsque le calcul se fait avec la méthode traditionnelle, et l'indice s'élève à 101.68 pour la méthode de calcul basée sur la médiane. Il y a donc dans notre cas un écart de 0.38 en chiffres absolus et une variation de 36%. En conclusion, nous constatons que l'utilisation de la médiane, à la place de la moyenne dans la deuxième étape du calcul, permet de ne pas subir l'influence de valeurs extrêmes. L'utilisation de la médiane pourrait donc s'avérer utile lorsque l'on est en présence de produits dont les prix sont très dispersés, cela afin qu'une observation extrême n'influence pas trop l'ensemble de l'indice.

Chapitre 7

Conclusion

L'indice des prix à la consommation est l'un des indicateurs économiques parmi les plus importants, puisqu'il sert à mesurer l'inflation, à déflater d'autres statistiques économiques et à indexer les salaires et les rentes.

Au vu de son importance économique, il apparaît donc comme essentiel que l'IPC soit statistiquement fiable. Si l'on suit les recommandations internationales, les échantillons doivent être sélectionnés aléatoirement. Ce type de sélection permet alors de mesurer la précision de l'estimateur calculé. Dans le cas de l'IPC, cela signifie que l'on peut

1. calculer la variance de l'IPC. On peut alors déterminer, pour une probabilité donnée, l'intervalle de confiance à l'intérieur duquel la vraie valeur de l'IPC se situe;

2. calculer la variance du changement de prix de l'indice entre deux périodes (comme deux mois consécutifs ou deux mêmes mois à un an d'intervalle). Cela permet de déterminer si la variation de l'IPC est significative ou non. C'est-à-dire si elle est due à une variation réelle des prix ou si elle n'est attribuable qu'aux aléas de l'échantillonnage.

Nous avons vu qu'en pratique, les échantillons de l'ISPC ne sont pas tous sélectionnés selon les techniques d'échantillonnage aléatoire et qu'il n'est donc pas possible, à l'heure actuelle, de mesurer ces variances. En Suisse, la première étape devrait donc consister à mettre sur pied un plan d'échantillonnage aléatoire. Une fois cela réalisé, il serait alors possible de mesurer les erreurs dues à l'échantillonnage. C'est dans cette optique que s'orientent nos propositions et notre méthode développées aux chapitres 4 et 5.

Soulignons que l'estimation de la variance de l'IPC n'en est encore qu'à ses premiers pas au niveau international; citons notamment le cas de la Suède, des Pays-Bas ou de l'Italie qui s'intéressent à cette problématique. Le cas des Etat-Unis est unique à l'heure actuelle. En effet, c'est le seul pays à disposer d'un plan d'échantillonnage aléatoire et à calculer la variance de l'IPC. Nous avons vu du reste que la méthode américaine sert de référence au niveau international.

Les résultats numériques que nous avons obtenus pour Genève sont de deux types. Cependant, avant de les présenter, nous rappelons que nous avons dû travailler en posant l'hypothèse que les échantillons étaient aléatoires, ce qui n'est pas le cas en pratique. L'interprétation des résultats doit donc être faite en gardant à l'esprit cette hypothèse de base.

Nous avons tout d'abord calculé la variance de l'IPC de Genève pour le groupe de l'alimentation, pour les mois de juin à octobre 1991. Nous résumons ces résultats dans le tableau ci-dessous:

Mois	Indice 1982 = 100	Ecart-type estimé (σ)	Intervalle de confiance à 95%
Juin 91	133.37	0.41	$132.56 \leq \mu \leq 134.17$
Juillet 91	131.51	0.17	$131.18 \leq \mu \leq 131.84$
Août 91	131.42	0.24	$130.95 \leq \mu \leq 131.89$
Septembre 91	131.53	0.17	$131.20 \leq \mu \leq 131.86$
Octobre 91	131.96	0.18	$131.61 \leq \mu \leq 132.31$

Les intervalles de confiance que nous avons déterminés permettent de conclure, avec une probabilité de 95%, que la vraie valeur de l'indice est comprise dans cet intervalle. Par exemple, pour le mois d'août 1991, nous avons trouvé un indice de 131.42, et l'intervalle de confiance (à 95%) qui s'y rattache se situe entre 130.95 et 131.89. Cela signifie que si l'on avait observé les prix de toutes les transactions économiques pour l'ensemble des ménages, alors la valeur de l'indice (que nous appelons vraie valeur) se situerait, avec 95% de chances, entre ces deux bornes. Nous constatons donc que plus la variance calculée est petite (faible dispersion), plus l'intervalle de confiance est étroit et plus la valeur calculée de l'indice est précise et est proche de la vraie valeur de l'IPC.

Dans un deuxième temps, nous nous sommes intéressés au calcul de la variance du changement de prix du groupe de l'alimentation pour Genève entre deux mois successifs. Grâce à la construction d'un intervalle $(0 - 1.96 \cdot \sigma \leq \mu \leq 0 + 1.96 \cdot \sigma)$, avec un niveau de confiance de 95%, on peut vérifier si la variation en pourcent de l'indice est significativement différente de zéro. C'est-à-dire que si la variation en pourcent de l'indice est comprise dans cet intervalle, alors on peut conclure que la variation de l'IPC n'est pas significative par rapport à zéro et qu'elle n'est due qu'aux aléas des échantillons. Au contraire, si la variation de l'indice n'est pas comprise dans cet intervalle de confiance, alors on peut dire qu'elle est significative et qu'il y a eu une variation réelle des prix dans la période considérée. Dans le cas de Genève, nous avons étudié la variation du changement de prix du groupe de l'alimentation pour quatre périodes successives (juin-juillet 1991, juillet-août 1991, août-septembre 1991 et septembre-octobre 1991). Ces résultats sont présentés dans le tableau suivant:

	Variation en % de l'IPC	Ecart-type estimé en % σ
Juin-juillet 91	-1.39	0.46
Juillet-août 91	-0.06	0.26
Août-septembre 91	0.08	0.20
Septembre-octobre 91	0.31	0.24

Après l'estimation de la variance de la variation de l'IPC du groupe de l'alimentation de Genève et de l'intervalle de confiance, on conclut que seule la variation de juin-juillet 1991 est significative et que la variation des prix est réelle. Pour toutes les autres périodes, nous avons conclu que la variation n'est pas significative. Cela veut dire que la variation de l'indice ne serait due qu'aux aléas de l'échantillonnage.

Aussi, nous constatons que lorsque l'on publie uniquement l'indice des prix à la consommation, il y a tout un ensemble d'informations statistiques d'une extrême importance, autant pour l'analyse économique que pour les négociations entre partenaires sociaux d'un pays, qui est occulté. A l'heure actuelle, lorsque l'indice varie, on conclut uniquement à une augmentation ou diminution du niveau des prix, donc de l'inflation.
Penchons-nous sur l'indexation des rentes AVS, des salaires et de toutes les négociations qui y sont liées dans les différentes branches de l'économie suisse. Que devons-nous penser des résultats de ces négociations si ni la variance de l'ISPC, ni la variance du changement de prix ne peuvent être calculées? Imaginons un instant que la variation annuelle de l'ISPC s'établisse à 2%. Aujourd'hui les discussions sur l'indexation des salaires utilisent comme base de négociation ce résultat en admettant que l'inflation a été de 2% durant l'année écoulée. Si nous pouvions calculer la variance du changement de prix de l'IPC et que l'on aboutisse à la conclusion que pour cette année-là, il n'y a pas de variation significative de prix, qu'adviendrait-il alors de l'indexation? Dans un cas comme celui-ci, l'indexation représenterait une augmentation réelle du salaire, avec toutes les conséquences économiques que cela signifie, notamment au niveau de la compétitivité, pour l'économie suisse. Nous sommes bien conscients que la mise en place d'une telle méthode s'avérerait très coûteuse et qu'il serait nécessaire d'estimer le coût d'une telle opération, ce qui n'était pas l'un des objectifs de ce travail. Cependant, il faudrait comparer ce coût à l'ensemble des avantages qu'engendrerait une plus grande précision de l'ISPC pour l'économie suisse.

L'établissement d'un plan d'échantillonnage aléatoire ainsi que la mesure des erreurs d'échantillonnage sont reconnus comme des thèmes importants et qu'il faudra étudier lors des prochaines révisions de l'ISPC. Pour illustrer ce propos, citons notamment l'Office fédéral de la statistique (OFS) et la Commission de statistique conjoncturelle et sociale (CSCS):

- OFS: " *Au plan international, des tentatives d'établir en termes statistiques la précision des résultats de l'indice, ou plus exactement la marge d'erreur (erreurs aléatoires, intervalles de confiance) sont en cours. Certains pays (Les Etats-Unis, Les Pays-Bas) en publient les résultats. Les recherches nécessaires doivent être faites, éventuellement avec le concours de chercheurs intéressés.*"

  Source: Révision de l'indice suisse des prix à la consommation. Conception du nouvel indice suisse des prix à la consommation. Page 71. OFS (1993).

- CSCS: "*Par ailleurs, elle prend acte du fait que certains problèmes méthodologiques importants n'ont pas trouvé de solution au cours de la révision, à savoir:*
  - le traitement des services financiers et des intérêts;
  - le traitement des logements et des maisons individuelles occupés par leur propriétaire;
  - l'exécution de relevés de prix aléatoires, l'estimation de l'erreur d'échantillonnage;
  - la mesure des différences aux changements de qualité.
  Elle soutient l'OFS dans son intention d'aborder l'étude de ces questions au terme de la révision en cours, pour assurer un encadrement scientifique continu et pour préparer à temps les révisions à venir."
  Source: Prise de position de la Commission de statistique conjoncturelle et sociale (CSCS) sur la conception présentée par l'Office fédéral de la statistique (OFS). Annexe 4. Révision de l'indice suisse des prix à la consommation. Conception du nouvel indice suisse des prix à la consommation. Page 80. OFS (1993).

Nous concluons ce travail en espérant qu'il puisse, peut-être, apporter quelques éléments de réflexion aux organes chargés du calcul et de la construction de l'indice suisse des prix à la consommation.

Annexe A

Programmes informatiques et données

Dans cette annexe, on trouve les programmes et données relatives au calcul de l'estimation de la variance de l'indice des prix à la consommation du groupe de l'alimentation de Genève. Les données sont relatives au mois de juin, juillet, août, septembre et octobre 1991. Rappelons que les données ont été fournies par le service cantonal de statistique du canton de Genève.

Calcul des variances pour l'alimentation juin et juillet 1991

```
PROGRAM ALIMENT(input,output,juillet1,gares1);
CONST
    nligne=128;
    ncolonne=10;
TYPE
    indlig=1..nligne;
    indcol=1..ncolonne;
    matrice=array[indlig,indcol] of real;
VAR
    mat1:matrice;
    a,i,j:integer;
    juillet1,gares1:text;
    varil:array[1..128,1..2] of real;
    svarip1,svarip2,covt,vartay:real;
PROCEDURE LIREMATRICE;
VAR i,j:integer;
BEGIN
    for i:=1 to nligne do
    begin
       for j:=1 to ncolonne do
       begin
          read(juillet1,mat1[i,j]);
       end;
    end;
    close(juillet1);
END;
PROCEDURE VARIANCEREP;
BEGIN
for i:=1 to 128 do
begin
    varil[i,1]:=(((mat1[i,5]-mat1[i,7])*(mat1[i,5]-mat1[i,7])) +
      ((mat1[i,6]-mat1[i,7])*(mat1[i,6]-mat1[i,7])))/2;
end;
for i:=1 to 128 do
```

```
begin
    VariI[i,2]:=(((matI[i,1]-matI[i,3])*(matI[i,1]-matI[i,3])) +
      ((matI[i,2]-matI[i,3])*(matI[i,2]-matI[i,3])))/2;
end;
END;
PROCEDURE SOMMEVARP(a,b:integer);
BEGIN
svarip1:=0;
svarip2:=0;
for i:=a to (b+I) do
begin
    svarip1:=svarip1+((matI[i,8]/matI[(a-1),8])*(matI[i,8]/
      matI[(a-1),8]))*variI[i,1];
    svarip2:=svarip2 + ((matI[i,4]/matI[(a-1),4])*(matI[i,4]/
      matI[(a-1),4]))*variI[i,2];
end;
write(garesI,svarip1:6:4);
writeln(garesI,' ',svarip2:6:4);
END;
PROCEDURE SOMCOV(i:integer);
BEGIN
    covt:=((matI[i,I]-matI[i,3])*(matI[i,5]-matI[i,7])+
      (matI[i,2]-matI[i,3])*(matI[i,6]-matI[i,7]))/2;
END;
PROCEDURE TAYLOR(a:integer);
BEGIN
vartay:=1/(matI[a,3]*matI[a,3])*(svarip1+(matI[a,7]/matI[a,3])*
(matI[a,7]/matI[a,3])*svarip2-2*(matI[a,7]/matI[a,3])*covt);
write(gares1,vartay:10:8);
writeln(gares1);
END;
BEGIN(*PROGRAMME PRINCIPAL*);
reset(juilletI);
rewrite(garesI);
writeln(gares1,'Variance Variance Variance du cbangement');
writeln(gares1,'juillet(t) juin(t-1) de prix entre t et t-1');
writeln(garesI);
```

```
liramatrice;
    variancerep;
    sommevarp(2,127);
    somcov(1);
    taylor(1);
    writeln(gares1);
close(gares1);
END.
```

Calcul des variances pour l'alimentation juillet et août 1991

```
PROGRAM ALIMENT(input,output,août1,gares1);
CONST
    nligne=128;
    ncolonne=10;
TYPE
    indlig=1..nligne;
    indcol=1..ncolonne;
    matrice=array[indlig,indcol] of real;
VAR
    mat1:matrice;
    a,i,j:integer;
    août1,gares1:text;
    vari1:array[1..128,1..2] of real;
    svarip1,svarip2,covt,vartay:real;
PROCEDURE LIREMATRICE;
VAR i,j:integer;
BEGIN
    for i:=1 to nligne do
    begin
        for j:=1 to ncolonne do
        begin
            read(août1,mat1[i,j]);
        end;
    end;
```

```
    close(août1);
END;
PROCEDURE VARIANCEREP;
BEGIN
for i:=1 to 128 do
begin
    variI[i,I]:=(((matI[i,5]-matI[i,7])*(matI[i,5]-matI[i,7])) +
      ((matI[i,6]-matI[i,7])*(matI[i,6]-matI[i,7])))/2;
end;
for i:=1 to 128 do
begin
    VariI[i,2]:=(((matI[i,1]-matI[i,3])*(matI[i,1]-matI[i,3])) +
      ((matI[i,2]-matI[i,3])*(matI[i,2]-matI[i,3])))/2;
end;
END;
PROCEDURE SOMMEVARP(a,b:integer);
BEGIN
svarip1:=0;
svarip2:=0;
for i:=a to (b+1) do
    begin
       svarip1:=svarip1+((matI[i,8]/matI[(a-1),8])*(matI[i,8]/
          matI[(a-1),8]))*variI[i,I];
       svarip2:=svarip2 + ((matI[i,4]/matI[(a-1),4])*(matI[i,4]/
          matI[(a-1),4]))*variI[i,2];
end;
write(gares1,svarip1:6:4);
writeln(gares1,' ',svarip2:6:4);
END;
PROCEDURE SOMCOV(i:integer);
BEGIN
    covt:=((matI[i,I]-matI[i,3])*(matI[i,5]-matI[i,7])+
      (matI[i,2]-matI[i,3])*(matI[i,6]-matI[i,7]))/2;
END;
PROCEDURE TAYLOR(a:integer);
BEGIN
vartay:=1/(matI[a,3]*matI[a,3])*(svarip1+(matI[a,7]/matI[a,3])*
```

```
(matI[a,7]/matI[a,3])*svarip2-2*(matI[a,7]/matI[a,3])*covt);
write(gares1,vartay:10:8);
writeln(gares1);
END;
BEGIN(*PROGRAMME PRINCIPAL*);
reset(août1);
rewrite(gares1);
writeln(gares1,'Variance Variance Variance du changement');
writeln(gares1,'août(t) juillet(t-1) de prix entre t et t-1');
writeln(gares1);
lirematrice;
    variancerep;
    sommevarp(2,127);
    somcov(1);
    taylor(1);
    writeln(gares1);
close(gares1);
END.
```

Calcul des variances pour l'alimentation août et septembre 1991

```
PROGRAM ALIMENT(input,output,septembre1,gares1);
CONST
    nligne=128;
    ncolonne=10;
TYPE
    indlig=1..nligne;
    indcol=1..ncolonne;
    matrice=array[indlig,indcol] of real;
VAR
    matI:matrice;
    a,i,j:integer;
    septembre1,gares1:text;
    variI:array[1..128,1..2] of real;
    svarip1,svarip2,covt,vartay:real;
PROCEDURE LIREMATRICE;
```

```
VAR i,j:integer;
BEGIN
    for i:=1 to nligne do
    begin
      for j:=1 to ncolonne do
      begin
        read(septembre1,matI[i,j]);
      end;
    end;
    close(septembre1);
END;
PROCEDURE VARIANCEREP;
for i:=1 to 128 do
begin
    variI[i,1]:=(((matI[i,5]-matI[i,7])*(matI[i,5]-matI[i,7])) +
      ((matI[i,6]-matI[i,7])*(matI[i,6]-matI[i,7])))/2;
end;
for i:=1 to 128 do
begin
    VariI[i,2]:=(((matI[i,I]-matI[i,3])*(matI[i,I]-matI[i,3])) +
      ((matI[i,2]-matI[i,3])*(matI[i,2]-matI[i,3])))/2;
end;
END;
PROCEDURE SOMMEVARP(a,b:integer);
BEGIN
svarip1:=0;
svarip2:=0;
for i:=a to (b+I) do
    begin
      svarip1:=svarip1+((matI[i,8]/matI[(a-1),8])*(matI[i,8]/
        matI[(a-1),8]))*variI[i,I];
      svarip2:=svarip2 + ((matI[i,4]/matI[(a-I),4])*(matI[i,4]/
        matI[(a-I),4]))*variI[i,2];
    end;
write(gares1,svarip1:6:4);
writeln(garesI,' ',svarip2:6:4);
END;
```

```
PROCEDURE SOMCOV(i:integer);
BEGIN
    covt:=((matI[i,1]-matI[i,3])*(matI[i,5]-matI[i,7])+
      (matI[i,2]-matI[i,3])*(matI[i,6]-matI[i,7]))/2;
END;
PROCEDURE TAYLOR(a:integer);
BEGIN
vartay:=1/(matI[a,3]*matI[a,3])*(svarip1+(matI[a,7]/matI[a,3])*
(matI[a,7]/matI[a,3])*svarip2-2*(matI[a,7]/matI[a,3])*covt);
write(gares1,vartay:10:8);
writeln(gares1);
END;
BEGIN(*PROGRAMME PRINCIPAL*);
reset(septembre1);
rewrite(gares1);
writeln(gares1,'Variance Variance Variance du changement');
writeln(gares1,'septembre(t) août (t-1) de prix entre t et t-1');
writeln(gares1);
lirematrice;
    variancerep;
    sommevarp(2,127);
    somcov(1);
    taylor(1);
    writeln(gares1);
close(gares1);
END.
```

Calcul des variances pour l'alimentation septembre et octobre 1991

```
PROGRAM ALIMENT(input,output,octobre1,gares1);
CONST
    nligne=128;
    ncolonne=10;
TYPE
    indlig=1..nligne;
```

```
    indcol=1..ncolonne;
    matrice=array[indlig,indcol] of real;
VAR
    matI:matrice;
    a,i,j:integer;
    octobre1,garesI:text;
    variI:array[1..128,1..2] of real;
    svarip1,svarip2,covt,vartay:real;
PROCEDURE LIREMATRICE;
VAR i,j:integer;
BEGIN
    for i:=1 to nligne do
    begin
       for j:=1 to ncolonne do
       begin
          read(octobre1,matI[i,j]);
       end;
    end;
    close(octobre1);
END;
PROCEDURE VARIANCEREP;
BEGIN
for i:=1 to 128 do
begin
    variI[i,1]:=(((matI[i,5]-matI[i,7])*(matI[i,5]-matI[i,7])) +
      ((matI[i,6]-matI[i,7])*(matI[i,6]-matI[i,7])))/2;
end;
for i:=1 to 128 do
begin
    VariI[i,2]:=(((matI[i,1]-matI[i,3])*(matI[i,1]-matI[i,3])) +
      ((matI[i,2]-matI[i,3])*(matI[i,2]-matI[i,3])))/2;
end;
END;
PROCEDURE SOMMEVARP(a,b:integer);
BEGIN
svarip1:=0;
svarip2:=0;
```

```
for i:=a to (b+1) do
    begin
      svarip1:=svarip1+((matI[i,8]/matI[(a-1),8])*(matI[i,8]/
         matI[(a-1),8]))*variI[i,1];
      svarip2:=svarip2 + ((matI[i,4]/matI[(a-1),4])*(matI[i,4]/
         matI[(a-1),4]))*variI[i,2];
end;
write(gares1,svaripI:6:4);
writeln(gares1,' ',svarip2:6:4);
END;
PROCEDURE SOMCOV(i:integer);
BEGIN
covt:=((matI[i,1]-matI[i,3])*(matI[i,5]-matI[i,7])+
(matI[i,2]-matI[i,3])*(matI[i,6]-matI[i,7]))/2;
END;
PROCEDURE TAYLOR(a:integer);
BEGIN
vartay:=1/(matI[a,3]*matI[a,3])*(svarip1+(matI[a,7]/matI[a,3])*
(matI[a,7]/matI[a,3])*svarip2-2*(matI[a,7]/matI[a,3])*covt);
write(gares1,vartay:10:8);
writeln(gares1);
END;
BEGIN(*PROGRAMME PRINCIPAL*);
reset(octobre1);
rewrite(gares1);
writeln(gares1,'Variance Variance Variance du changement');
writeln(gares1,'octobre(t) septembre (t-1) de prix entre t et t-1');
writeln(garesI);
lirematrice;
    variancerep;
    sommevarp(2,127);
    somcov(I);
    taylor(1);
    writeln(gares1);
close(gares1);
END.
```

Données de Genève pour le mois de juin 1991

Dans les tableaux suivants, on donne les indices au niveau des rubriques, c'est-à-dire au dernier niveau où l'on a un poids de pondération.

Indice pour la réplique 1 = A
Indice pour la réplique 2 = B
Indice pour le total = T
Poids de pondération = wi
Numéro de la ligne = i

Rubriques	A	B	T	wi	i
Alimentation	133.75	133.01	133.37	21.000	1
Lait et produits laitiers	124.51	124.61	124.53	3.924	
Lait	124.71	124.71	124.71	1.245	
Lait entier en emballage	121.92	121.92	121.92	0.374	2
Lait "drink"	126.39	126.39	126.39	0.374	3
Autre lait en emballage	166.33	166.33	166.33	0.062	4
Lait entier en vrac	119.73	119.73	119.73	0.435	5
Beurre	114.12	113.67	113.84	0.494	
Beurre de table	101.89	101.57	101.73	0.193	6
Beurre de cuisine	121.97	121.43	121.61	0.301	7
Fromage	128.38	128.30	128.34	1.319	
Fromage à pâte dure	126.95	126.83	126.88	0.593	8
Fromage à pâte mi-dure	131.04	130.78	130.93	0.396	9
From.pâte molle, from. frais	127.57	127.68	127.63	0.198	10
Fromage fondu	128.07	128.43	128.22	0.132	11
Produits laitiers frais	127.76	129.00	128.15	0.449	
Yogourt	129.25	130.88	129.78	0.359	12
Séré	116.78	116.78	116.78	0.045	13
Autres prod. lait. frais	126.83	126.23	126.45	0.045	14
Crème	116.77	117.80	117.25	0.299	
Crème entière	114.77	116.49	115.57	0.179	15
Crème à café	119.75	119.75	119.75	0.120	16
Autres produits laitiers	130.00	128.77	129.45	0.118	17
Oeufs	155.40	157.06	156.13	0.361	
Oeufs du pays	137.52	138.11	137.82	0.199	18
Oeufs étrangers	177.36	180.34	178.62	0.162	19

Viande, charcut., saucisses	123.54	123.23	123.26	4.317	
Viande de boeuf (sans os)	121.67	120.95	121.32	0.972	
Entrecôte (boeuf)	126.35	126.35	126.35	0.126	20
Beefsteak dans la cuisse	126.87	126.79	126.86	0.215	21
Rôti de boeuf	124.17	124.21	124.19	0.165	22
Bouilli de boeuf	117.72	117.68	117.70	0.175	23
Ragoût de boeuf	116.73	117.06	117.31	0.126	24
Viande de boeuf hachée	116.79	112.38	114.28	0.165	25
Viande de veau 1er. chx. (s/os)	134.75	135.18	134.83	0.284	
Tranche de veau	136.89	138.36	137.80	0.102	26
Rôti de veau	132.19	132.20	132.23	0.054	27
Ragoût de veau	139.19	138.88	138.39	0.085	28
Emincé de veau	124.08	124.06	124.01	0.043	29
Viande de porc (sans os)	122.13	122.31	121.91	1.480	
Tranche (jambon, cou)	113.13	113.04	112.98	0.266	30
Tranche (filet, fil.mignon)	120.65	122.00	120.70	0.133	31
Rôti de porc	115.44	114.44	114.69	0.311	32
Cotelettes de porc	111.36	113.20	111.47	0.237	33
Ragoût de porc	121.05	120.70	120.53	0.252	34
Jambon de derrière	129.99	130.37	130.13	0.163	35
Lard maigre	174.78	174.88	175.03	0.118	36
Viande d'agneau	123.91	124.36	124.04	0.074	37
Abats	115.53	115.50	115.52	0.077	38
Charcuterie et saucisses	125.42	125.52	125.48	1.096	39
Volaille	122.78	118.75	120.48	0.288	40
Conserves de viande	112.10	112.10	112.10	0.046	41
Poissons	130.79	134.95	132.77	0.306	
Poissons frais	145.75	156.00	150.63	0.124	42
Poissons congelés	135.33	135.33	135.33	0.089	43
Poissons en boîte	106.52	106.52	106.52	0.093	44
Huiles, graisses comestibles	111.17	111.02	111.09	0.269	
Huiles comestibles	110.31	110.24	110.29	0.149	45
Graisses comestibles	116.87	117.75	117.32	0.028	46
Margarine	110.84	110.22	110.49	0.092	47
Produits à base de céréales	141.27	141.32	141.31	2.491	
Pain	146.14	146.16	146.15	0.787	
Pain bis	149.42	149.42	149.42	0.315	48
Pain mi-blanc	143.87	143.87	143.87	0.315	49
Pains spéciaux	144.10	144.21	144.15	0.157	50

Autres art. de boulangerie	147.64	147.49	147.57	1.265	
Petite boulangerie	168.68	168.68	168.68	0.316	51
Boulangerie fine	156.71	156.64	156.66	0.316	52
Biscottes et biscuits	124.45	124.45	124.45	0.253	53
Pâtisserie, confiserie	138.05	137.62	137.84	0.380	54
Pâte à gâteau	125.51	124.78	125.41	0.139	55
Farine	125.34	128.17	126.50	0.078	56
Riz	93.22	94.04	93.81	0.056	57
Pâtes alimentaires	105.52	106.39	105.95	0.139	58
Céréales pour petit déj.	111.77	112.33	112.05	0.027	59
Pommes de terre	140.41	142.72	142.73	0.223	
Pommes de terre	137.33	140.78	140.81	0.149	60
Pommes de terre à encaver	146.62	146.62	146.62	0.074	61
Légumes	141.47	144.38	142.12	1.309	
Aubergines	142.08	142.08	142.08	0.000	62
Haricots nains	199.91	198.38	198.45	0.092	63
Champignons de Paris	111.63	110.23	111.11	0.026	64
Fenouil	176.15	177.67	178.23	0.039	65
Carottes	138.38	126.44	134.42	0.092	66
Ail	106.78	104.47	105.09	0.013	67
Choux-fleurs	120.06	125.11	120.47	0.079	68
Choux de Bruxelles	183.96	183.96	183.96	0.000	69
Choux rouges	154.33	154.33	154.33	0.000	70
Choux blancs	164.21	164.21	164.21	0.000	71
Choux frisés	217.59	177.61	215.34	0.013	72
Poireaux verts	274.11	300.50	304.97	0.013	73
Poivrons verts	161.59	167.80	160.57	0.092	74
Endives	112.70	112.70	112.70	0.000	75
Chicorée de Trévise	181.29	200.21	180.97	0.013	76
Chicorée scarole	220.01	220.01	220.01	0.000	77
Concombres	105.44	120.25	121.37	0.105	78
Salade pommée	144.08	150.28	140.63	0.170	79
Salade mélangée	135.68	135.68	135.68	0.000	80
Mâche	203.98	203.98	203.98	0.000	81
Betteraves rouges	134.06	134.06	134.06	0.000	82
Céleri pomme	133.72	133.72	133.72	0.000	83
Asperges blanches	131.00	118.93	124.70	0.079	84
Tomates	136.25	143.27	137.73	0.365	85
Courgettes	159.49	154.90	156.61	0.026	86
Oignons	108.84	107.93	108.15	0.092	87

Fruits	195.26	180.63	188.82	1.449	
Pommes	198.29	199.28	199.20	0.362	
Golden delicious 1	209.33	207.15	209.95	0.130	88
Granny smith I	121.44	124.17	122.06	0.087	89
Autres pommes I	234.49	237.30	235.84	0.145	90
Pommes II	180.70	180.70	180.70	0.000	91
Autres fruits	194.25	174.42	185.36	1.087	
Bananes	143.81	146.39	147.17	0.188	92
Poires	312.25	307.94	309.47	0.043	93
Fraises	230.43	176.34	201.31	0.293	94
Grapefruits blancs	94.66	97.85	97.26	0.043	95
Kaki	137.46	137.46	137.46	0.000	96
Pastèques	180.50	164.40	161.00	0.015	97
Melons	252.92	236.89	250.34	0.043	98
Noix	113.79	113.79	113.79	0.000	99
Arachides grillées	114.88	114.88	114.88	0.000	100
Oranges blondes	182.07	167.24	175.73	0.043	101
Clémentines	153.08	153.08	153.08	0.000	102
Mandarines	137.12	137.12	137.12	0.000	103
Abricots	170.06	154.74	170.47	0.188	104
Cerises	163.53	169.67	166.55	0.043	105
Nectarines	129.10	129.10	129.10	0.000	106
Pêches	233.17	220.00	224.29	0.145	107
Quetsches "pruneaux"	162.67	162.67	162.67	0.000	108
Raisin blanc	148.07	148.07	148.07	0.000	109
Citrons	113.46	112.20	113.42	0.043	110
Conserves légumes et fruits	110.91	110.91	110.91	0.404	
Conserves légumes	111.71	111.71	111.71	0.273	111
Conserves fruits	103.52	103.52	103.52	0.081	112
Légumes et fruits congelés	118.52	118.52	118.52	0.050	113
Confiture, miel, mélasse	107.44	107.44	107.44	0.165	114
Plats cuisinés	115.95	115.95	115.95	0.385	115
Potages en sachets	132.22	132.22	132.22	0.084	116
Epices et sauces	121.53	121.53	121.53	0.349	117
Sucre	114.76	116.79	115.68	0.116	118
Aliments fortifiants	112.19	112.19	112.19	0.110	119
Chocolat	106.47	106.50	106.48	0.484	
Chocolat en plaque	104.70	104.94	104.82	0.387	120
Autre chocolat	113.53	112.69	113.08	0.097	121

Café	95.49	95.61	95.55	0.420	
Café en grains	87.38	87.56	87.47	0.288	122
Café soluble en poudre	113.16	113.16	113.16	0.132	123
Thé	125.40	125.39	125.40	0.063	124
Repas au restaurant	138.85	138.85	138.85	3.771	
Menus du jour	137.24	137.24	137.24	1.131	125
Plats chauds	139.29	139.29	139.29	1.697	126
Plats froids	140.55	140.55	140.55	0.754	127
Desserts	137.77	137.77	137.77	0.189	128

Données de Genève pour le mois de juillet 1991

Dans les tableaux suivants, on donne les indices au niveau des rubriques, c'est-à-dire au dernier niveau où l'on a un poids de pondération.

Indice pour la réplique 1 = A
Indice pour la réplique 2 = B
Indice pour le total = T
Poids de pondération = wi
Numéro de la ligne = i

Rubriques	A	B	T	wi	i
Alimentation	131.35	131.89	131.51	21.000	1
Lait et produits laitiers	124.54	124.87	124.71	3.924	
Lait	124.71	124.71	124.71	1.245	
Lait entier en emballage	121.92	121.92	121.92	0.374	2
Lait "drink"	126.39	126.39	126.39	0.374	3
Autre lait en emballage	166.33	166.33	166.33	0.062	4
Lait entier en vrac	119.73	119.73	119.73	0.435	5
Beurre	114.50	114.84	114.67	0.494	
Beurre de table	103.41	104.28	103.85	0.193	6
Beurre de cuisine	121.61	121.61	121.61	0.301	7
Fromage	128.30	128.50	128.39	1.319	
Fromage à pâte dure	126.88	126.88	126.88	0.593	8
Fromage à pâte mi-dure	130.85	131.06	130.96	0.396	9
From.pâte molle, from. frais	128.09	128.03	128.06	0.198	10
Fromage fondu	127.36	128.83	127.94	0.132	11
Produits laitiers frais	129.25	130.15	129.74	0.449	
Yogourt	131.18	131.48	131.38	0.359	12
Séré	116.72	118.94	117.73	0.045	13
Autres prod. lait. frais	126.45	130.77	128.70	0.045	14
Crème	114.41	115.88	115.27	0.299	
Crème entière	112.35	113.29	112.82	0.179	15
Crème à café	117.49	119.75	118.94	0.120	16
Autres produits laitiers	130.69	130.59	130.56	0.118	17
Oeufs	156.16	156.16	156.16	0.361	
Oeufs du pays	137.88	137.87	137.87	0.199	18
Oeufs étrangers	178.62	178.62	178.62	0.162	19

Viande, charcut., saucisses	123.31	123.74	123.45	4.317	
Viande de boeuf (sans os)	121.12	121.46	121.00	0.972	
Entrecôte (boeuf)	126.23	125.07	125.08	0.126	20
Beefsteak dans la cuisse	127.89	128.20	127.78	0.215	21
Rôti de boeuf	123.60	123.53	123.57	0.165	22
Bouilli de boeuf	117.33	117.36	117.35	0.175	23
Ragoût de boeuf	115.45	118.74	115.83	0.126	24
Viande de boeuf hachée	114.28	114.28	114.28	0.165	25
Viande de veau 1er. chx. (s/os)	134.71	134.71	134.75	0.284	
Tranche de veau	138.01	138.01	138.01	0.102	26
Rôti de veau	131.83	131.79	131.88	0.054	27
Ragoût de veau	137.90	137.94	138.01	0.085	28
Emincé de veau	124.22	124.21	124.18	0.043	29
Viande de porc (sans os)	122.33	122.26	122.09	1.480	
Tranche (jambon, cou)	114.09	113.29	113.56	0.266	30
Tranche (filet, fil.mignon)	122.11	122.51	122.12	0.133	31
Rôti de porc	115.59	115.24	115.34	0.311	32
Cotelettes de porc	113.87	114.97	113.82	0.237	33
Ragoût de porc	120.53	119.79	119.79	0.252	34
Jambon de derrière	130.54	130.55	130.48	0.163	35
Lard maigre	168.43	169.16	169.03	0.118	36
Viande d'agneau	124.95	124.96	124.95	0.074	37
Abats	114.89	115.57	115.24	0.077	38
Charcuterie et saucisses	124.80	126.55	125.89	1.096	39
Volaille	122.47	121.28	121.84	0.288	40
Conserves de viande	112.10	112.10	112.10	0.046	41
Poissons	130.16	131.95	131.12	0.306	
Poissons frais	144.19	148.61	146.55	0.124	42
Poissons congelés	135.33	135.33	135.33	0.089	43
Poissons en boîte	106.52	106.52	106.52	0.093	44
Huiles, graisses comestibles	111.03	111.48	111.25	0.269	
Huiles comestibles	110.90	111.34	111.11	0.149	45
Graisses comestibles	116.77	118.01	117.28	0.028	46
Margarine	109.49	109.72	109.63	0.092	47
Produits à base de céréales	141.43	141.52	141.47	2.491	
Pain	146.59	146.69	146.63	0.787	
Pain bis	149.42	149.42	149.42	0.315	48
Pain mi-blanc	144.86	145.11	144.97	0.315	49
Pains spéciaux	144.38	144.36	144.37	0.157	50

Autres art. de boulangerie	147.69	147.53	147.61	1.265	
Petite boulangerie	168.68	168.68	168.68	0.316	51
Boulangerie fine	156.66	156.66	156.66	0.316	52
Biscottes et biscuits	124.45	124.45	124.45	0.253	53
Pâtisserie, confiserie	138.24	137.71	137.99	0.380	54
Pâte à gâteau	125.11	126.00	125.55	0.139	55
Farine	126.22	126.28	126.21	0.078	56
Riz	93.04	92.54	92.79	0.056	57
Pâtes alimentaires	105.55	107.04	106.20	0.139	58
Céréales pour petit déj.	111.34	112.48	111.90	0.027	59
Pommes de terre	136.76	137.98	137.79	0.223	
Pommes de terre	131.86	133.70	133.41	0.149	60
Pommes de terre à encaver	146.62	146.62	146.62	0.074	61
Légumes	129.47	130.72	130.19	1.309	
Aubergines	149.65	149.93	156.68	0.013	62
Haricots nains	195.10	191.34	194.48	0.092	63
Champignons de Paris	112.78	112.84	112.50	0.026	64
Fenouil	134.76	136.36	131.38	0.039	65
Carottes	135.08	134.02	131.12	0.092	66
Ail	100.71	102.44	101.15	0.013	67
Choux-fleurs	86.48	88.00	89.94	0.065	68
Choux de Bruxelles	183.96	183.96	183.96	0.000	69
Choux rouges	154.33	154.33	154.33	0.000	70
Choux blancs	164.21	164.21	164.21	0.000	71
Choux frisés	194.98	192.79	183.03	0.013	72
Poireaux verts	171.77	177.54	170.54	0.013	73
Poivrons verts	135.92	129.88	135.26	0.092	74
Endives	112.70	112.70	112.70	0.000	75
Chicorée de Trévise	180.97	180.97	180.97	0.000	7 6
Chicorée scarole	220.01	220.01	220.01	0.000	77
Concombres	100.24	103.84	98.61	0.105	78
Salade pommée	94.38	99.48	98.60	0.157	79
Salade mélangée	135.68	135.68	135.68	0.000	80
Mâche	203.98	203.98	203.98	0.000	81
Betteraves rouges	134.06	134.06	134.06	0.000	82
Céleri pomme	133.72	133.72	133.72	0.000	83
Asperges blanches	124.70	124.70	124.70	0.000	84
Tomates	137.63	138.84	138.90	0.484	85
Courgettes	120.85	131.51	122.24	0.026	86
Oignons	122.27	126.80	125.74	0.079	87

Fruits	172.00	176.29	172.57	1.449	
Pommes	192.48	192.78	192.87	0.246	
Golden delicious I	212.11	216.41	211.89	0.072	88
Granny smith I	122.82	128.33	127.28	0.087	89
Autres pommes I	245.90	237.67	242.73	0.087	90
Pommes II	180.70	180.70	180.70	0.000	91
Autres fruits	167.81	172.92	168.42	1.203	
Bananes	135.32	146.64	138.97	0.145	92
Poires	319.50	315.46	317.63	0.072	93
Fraises	200.64	223.36	210.33	0.058	94
Grapefruits blancs	107.28	95.58	104.55	0.029	95
Kaki	137.46	137.46	137.46	0.000	96
Pastèques	139.23	159.02	147.96	0.043	97
Melons	177.75	202.37	177.21	0.116	98
Noix	113.79	113.79	113.79	0.000	99
Arachides grillées	114.88	114.88	114.88	0.000	100
Oranges blondes	175.22	185.37	180.89	0.029	101
Clémentines	153.00	153.08	153.08	0.000	102
Mandarines	137.12	137.12	137.12	0.000	103
Abricots	148.34	139.19	138.57	0.218	104
Cerises	175.39	164.61	170.19	0.087	105
Nectarines	168.06	182.06	174.50	0.145	106
Pêches	165.68	168.82	170.16	0.218	107
Quetsches "pruneaux"	162.67	162.67	162.67	0.000	108
Raisin blanc	148.07	148.07	148.07	0.000	109
Citrons	109.90	111.00	111.10	0.043	110
Conserves légumes et fruits	110.91	110.91	110.91	0.404	
Conserves légumes	111.71	111.71	111.71	0.273	111
Conserves fruits	103.52	103.52	103.52	0.081	112
Légumes et fruits congelés	118.52	118.52	118.52	0.050	113
Confiture, miel, mélasse	107.44	107.44	107.44	0.165	114
Plats cuisinés	115.95	115.95	115.95	0.385	115
Potages en sachets	132.22	132.22	132.22	0.084	116
Epices et sauces	121.53	121.53	121.53	0.349	117
Sucre	115.90	115.90	115.90	0.116	118
Aliments fortifiants	112.19	112.19	112.19	0.110	119
Chocolat	106.84	106.63	106.76	0.484	
Chocolat en plaque	105.27	105.29	105.30	0.387	120
Autre chocolat	113.08	111.99	112.16	0.097	121

Café	95.17	93.77	94.38	0.420	
Café en grains	86.92	84.88	85.78	0.288	122
Café soluble en poudre	113.16	113.16	113.16	0.132	123
Thé	126.15	125.55	125.88	0.063	124
Repas au restaurant	138.85	138.85	138.85	3.771	
Menus du jour	137.24	137.24	137.24	1.131	125
Plats chauds	139.29	139.29	139.29	1.697	126
Plats froids	140.55	140.55	140.55	0.754	127
Desserts	137.77	137.77	137.77	0.189	128

Données de Genève pour le mois d'août 1991

Dans les tableaux suivants, on donne les indices au niveau des rubriques, c'est-à-dire au dernier niveau où l'on a un poids de pondération.

Indice pour la réplique 1 = A
Indice pour la réplique 2 = B
Indice pour le total = T
Poids de pondération = wi
Numéro de la ligne = i

Rubriques	A	B	T	wi	i
Alimentation	131.63	131.40	131.42	21.000	1
Lait et produits laitier	124.62	124.45	124.54	3.924	
Lait	124.71	124.71	124.71	1.245	
Lait entier en emballage	121.92	121.92	121.92	0.374	2
Lait "drink"	126.39	126.39	126.39	0.374	3
Autre lait en emballage	166.33	166.33	166.33	0.062	4
Lait entier en vrac	119.73	119.73	119.73	0.435	5
Beurre	114.45	114.12	114.28	0.494	
Beurre de table	103.28	102.46	102.86	0.193	6
Beurre de cuisine	121.61	121.61	121.61	0.301	7
Fromage	128.73	128.48	128.64	1.319	
Fromage à pâte dure	126.88	126.88	126.88	0.593	8
Fromage à pâte mi-dure	131.33	131.23	131.28	0.396	9
From.pâte molle, from. frais	127.85	128.06	128.00	0.198	10
Fromage fondu	130.62	128.13	129.59	0.132	11
Produits laitier frais	129.74	129.47	129.61	0.449	
Yogourt	131.38	130.99	131.19	0.359	12
Séré	117.73	116.08	116.86	0.045	13
Autres prod. lait. frais	128.70	130.81	129.77	0.045	14
Crème	113.28	113.05	113.17	0.299	
Crème entière	112.82	112.82	112.82	0.179	15
Crème à café	113.98	113.39	113.70	0.120	16
Autres produits laitiers	129.56	129.59	129.58	0.118	17
Oeufs	155.76	155.76	155.76	0.361	
Oeufs du pays	137.87	137.87	137.87	0.199	18
Oeufs étrangers	177.73	177.73	177.73	0.162	19

Viande, charcut., saucisses	123.21	122.74	122.89	4.317	
Viande de boeuf (sans os)	120.13	120.77	120.76	0.972	
Entrecôte (boeuf)	126.36	124.92	126.14	0.126	20
Beefsteak dans la cuisse	126.49	126.78	126.86	0.215	21
Rôti de boeuf	123.30	123.27	123.29	0.165	22
Bouilli de boeuf	115.68	115.82	115.75	0.175	23
Ragoût de boeuf	112.74	115.84	116.09	0.126	24
Viande de boeuf hachée	114.28	116.28	115.05	0.165	25
Viande de veau (1er ch./ s/os)	133.97	133.84	134.03	0.284	
Tranche de veau	137.25	137.25	137.25	0.102	26
Rôti de veau	131.24	131.42	131.44	0.054	27
Ragoût de veau	137.07	136.43	137.07	0.085	28
Emincé de veau	123.46	123.67	123.67	0.043	29
Viande de porc (sans os)	120.16	120.87	120.21	1.480	
Tranche (jambon, cou)	108.89	113.17	108.88	0.266	30
Tranche (filet, fil.mignon)	121.04	121.24	121.25	0.133	31
Rôti de porc	113.15	113.64	113.57	0.311	32
Cotelettes de porc	112.46	112.25	112.61	0.237	33
Ragoût de porc	116.71	117.11	117.43	0.252	34
Jambon de derrière	131.41	131.59	131.30	0.163	35
Lard maigre	170.35	167.44	167.94	0.118	36
Viande d'agneau	124.09	124.95	124.52	0.074	37
Abats	115.63	115.80	115.72	0.077	38
Charcuterie et saucisses	125.55	125.49	125.41	1.096	39
Volaille	133.07	120.54	126.30	0.288	40
Conserves de viande	113.44	112.10	113.44	0.046	41
Poissons	135.32	133.21	134.52	0.306	
Poissons frais	156.20	151.56	154.06	0.124	42
Poissons congelés	135.33	135.54	135.54	0.089	43
Poissons en boîte	107.48	106.52	107.48	0.093	44
Huiles, graisses comestibles	111.10	111.11	111.09	0.269	
Huiles comestibles	110.69	110.72	110.68	0.149	45
Graisses comestibles	117.15	117.14	117.14	0.028	46
Margarine	109.93	109.90	109.91	0.092	47
Produits à base de céréales	142.47	142.46	142.42	2.491	
Pain	146.75	147.30	147.00	0.787	
Pain bis	149.42	149.42	149.42	0.315	48
Pain mi-blanc	144.97	146.32	145.57	0.315	49
Pains spéciaux	144.96	145.04	145.00	0.157	50

Autres articles de boulang.	148.62	148.49	148.51	1.265	
Petite boulangerie	168.82	169.17	168.99	0.316	51
Boulangerie fine	158.49	159.23	158.71	0.316	52
Biscottes et biscuits	126.35	124.70	125.50	0.253	53
Pâtisserie, confiserie	138.44	138.20	138.32	0.380	54
Pâte à gâteau	130.40	127.79	128.77	0.139	55
Farine	126.03	126.00	126.02	0.078	56
Riz	93.27	93.11	93.19	0.056	57
Pâtes alimentaires	109.26	109.51	109.39	0.139	58
Céréales pour petit déj.	112.34	114.23	113.36	0.027	59
Pommes de terre	120.82	127.92	122.84	0.223	
Pommes de terre	108.01	118.63	111.03	0.149	60
Pommes de terre à encaver	146.62	146.62	146.62	0.074	61
Légumes	131.18	130.26	130.09	1.309	
Aubergines	143.55	149.27	149.18	0.013	62
Haricots nains	189.72	194.92	191.23	0.079	63
Champignons de Paris	104.96	109.25	108.63	0.026	64
Fenouil	111.75	116.85	111.88	0.026	65
Carottes	117.38	117.81	117.47	0.065	66
Ail	100.60	99.62	100.47	0.026	67
Choux-fleurs	101.49	97.82	98.87	0.065	68
Choux de Bruxelles	183.96	183.96	183.96	0.000	69
Choux rouges	165.31	159.00	160.93	0.013	70
Choux blancs	150.20	158.86	152.08	0.013	71
Choux frisés	166.70	169.32	170.73	0.013	72
Poireaux verts	143.54	148.20	144.43	0.013	73
Poivrons verts	146.52	130.61	132.58	0.092	74
Endives	112.70	112.70	112.70	0.000	75
Chicorée de Trévise	180.97	180.97	180.97	0.000	76
Chicorée scarole	224.68	228.37	224.97	0.013	77
Concombres	94.55	89.04	93.14	0.079	78
Salade pommée	115.71	112.11	113.84	0.210	79
Salade mélangée	134.17	135.68	134.40	0.026	80
Mâche	201.08	201.54	201.81	0.013	81
Betteraves rouges	134.06	134.06	134.06	0.000	82
Céleri pomme	133.72	133.72	133.72	0.000	83
Asperges blanches	124.70	124.70	124.70	0.000	84
Tomates	139.29	139.39	139.26	0.419	85
Courgettes	87.26	107.54	99.94	0.026	86
Oignons	118.23	120.51	116.67	0.079	87

Fruits	170.55	167.21	167.50	1.449	
Pommes	209.87	210.75	209.05	0.187	
Golden delicious I	227.57	229.73	225.44	0.072	88
Granny smith I	126.76	129.03	127.81	0.043	89
Autres pommes I	241.80	240.56	241.18	0.072	90
Pommes II	180.70	180.70	180.70	0.000	91
Autres fruits	164.72	160.76	161.34	1.262	
Bananes	127.24	125.90	125.92	0.145	92
Poires	319.57	332.92	326.40	0.101	93
Fraises	210.33	210.33	210.33	0.000	94
Grapefruits blancs	107.57	108.46	109.53	0.015	95
Kaki	137.46	137.46	137.46	0.000	96
Pastèques	102.54	112.97	111.41	0.015	97
Melons	134.09	139.12	134.40	0.087	98
Noix	113.79	113.79	113.79	0.000	99
Arachides grillées	114.88	114.88	114.88	0.000	100
Oranges blondes	191.44	183.99	188.65	0.029	101
Clémentines	153.08	153.08	153.08	0.000	102
Mandarines	137.12	137.12	137.12	0.000	103
Abricots	147.53	148.45	148.69	0.145	104
Cerises	170.19	170.19	170.19	0.000	105
Nectarine	144.14	142.62	134.38	0.159	106
Pêches	169.84	135.83	150.13	0.174	107
Quetsches "pruneaux"	176.55	182.63	186.95	0.072	108
Raisin blanc	163.67	160.44	160.75	0.277	109
Citrons	113.27	111.87	113.32	0.043	110
Conserves légumes et fruits	111.12	111.56	111.77	0.404	
Conserves légumes	111.71	112.65	112.65	0.273	111
Conserves fruits	104.57	103.52	104.57	0.081	112
Légumes et fruits congelés	118.52	118.67	118.67	0.050	113
Confiture, miel, mélasse	108.85	107.44	108.85	0.165	114
Plats cuisinés	115.95	116.88	116.88	0.385	115
Potages en sachets	137.92	132.22	137.92	0.084	116
Epices et sauces	121.53	123.71	123.71	0.349	117
Sucre	115.76	115.76	115.76	0.116	118
Aliments fortifiants	113.41	112.19	113.41	0.110	119
Chocolat	107.01	106.84	106.91	0.484	
Chocolat en plaque	105.61	105.39	105.48	0.387	120
Autre chocolat	112.61	112.61	112.61	0.097	121

Café	93.40	95.25	94.31	0.420	
Café en grains	84.34	87.23	85.85	0.288	122
Café soluble en poudre	113.16	112.77	112.77	0.132	123
Thé	125.84	125.82	125.83	0.063	124
Repas au restaurant	140.23	140.70	140.59	3.771	
Menus du jour	138.97	137.65	138.36	1.131	125
Plats chauds	141.53	140.69	141.15	1.697	126
Plats froids	138.90	145.28	142.56	0.754	127
Desserts	141.46	140.65	141.17	0.189	128

Données de Genève pour le mois de septembre 1991

Dans les tableaux suivants, on donne les indices au niveau des rubriques, c'est-à-dire au dernier niveau où l'on a un poids de pondération.

Indice pour la réplique 1 = A
Indice pour la réplique 2 = B
Indice pour le total = T
Poids de pondération = wi
Numéro de la ligne = i

Rubriques	A	B	T	wi	i
Alimentation	131.58	131.55	131.53	21.000	1
Lait et produits laitiers	124.45	124.92	124.91	3.924	
Lait	125.53	126.83	126.82	1.245	
Lait entier en emballage	123.59	123.06	123.40	0.374	2
Lait "drink"	127.44	128.22	127.85	0.374	3
Autre lait en emballage	166.33	166.33	166.33	0.062	4
Lait entier en vrac	119.73	123.25	123.25	0.435	5
Beurre	113.48	113.92	113.70	0.494	
Beurre de table	100.81	101.93	101.37	0.193	6
Beurre de cuisine	121.61	121.61	121.61	0.301	7
Fromage	128.53	128.49	128.45	1.319	
Fromage à pâte dure	126.11	127.16	126.72	0.593	8
Fromage à pâte mi-dure	131.79	132.06	131.90	0.396	9
From.pâte molle, from. frais	129.01	127.18	128.00	0.198	10
Fromage fondu	127.13	125.69	126.57	0.132	11
Produits laitiers frais	129.37	129.31	129.34	0.449	
Yogourt	130.57	130.56	130.57	0.359	12
Séré	119.32	118.92	119.11	0.045	13
Autres prod. lait. frais	129.77	129.77	129.77	0.045	14
Crème	111.25	110.13	110.68	0.299	
Crème entière	109.77	107.93	108.83	0.179	15
Crème à café	113.45	113.42	113.44	0.120	16
Autres produits laitiers	130.32	131.88	131.12	0.118	17
Oeufs	155.76	155.76	155.76	0.361	
Oeufs du pays	137.87	137.87	137.87	0.199	18
Oeufs étrangers	177.73	177.73	177.73	0.162	19

Viande, charcut., saucisses	123.21	123.29	123.31	4.317	
Viande de boeuf (sans os)	122.01	121.99	122.20	0.972	
Entrecôte (boeuf)	127.73	126.30	127.46	0.126	20
Beefsteak dans la cuisse	127.22	128.80	128.44	0.215	21
Rôti de boeuf	124.17	124.46	124.31	0.165	22
Bouilli de boeuf	117.35	117.41	117.37	0.175	23
Ragoût de boeuf	117.88	118.27	118.37	0.126	24
Viande de boeuf hachée	116.80	115.05	115.97	0.165	25
Viande de veau 1er. chx. (s/os)	134.01	135.13	135.30	0.284	
Tranche de veau	137.03	137.36	137.13	0.102	26
Rôti de veau	129.20	130.85	130.80	0.054	27
Ragoût de veau	138.43	137.56	138.43	0.085	28
Emincé de veau	124.15	130.40	130.40	0.043	29
Viande de porc (sans os)	120.81	121.02	120.81	1.480	
Tranche (jambon, cou)	108.88	108.88	108.88	0.266	30
Tranche (filet, fil.mignon)	121.43	121.39	121.39	0.133	31
Rôti de porc	115.21	115.07	115.01	0.311	32
Cotelettes de porc	113.04	112.94	112.92	0.237	33
Ragoût de porc	119.52	120.30	119.43	0.252	34
Jambon de derrière	130.74	130.84	130.89	0.163	35
Lard maigre	166.42	167.89	167.24	0.118	36
Viande d'agneau	123.75	122.43	122.82	0.074	37
Abats	115.72	115.72	115.72	0.077	38
Charcuterie et saucisses	125.14	124.46	124.82	1.096	39
Volaille	125.07	127.01	126.09	0.288	40
Conserves de viande	113.44	113.44	113.44	0.046	41
Poissons	137.02	135.59	136.41	0.306	
Poissons frais	160.23	156.72	158.73	0.124	42
Poissons congelés	135.54	135.54	135.54	0.089	43
Poissons en boîte	107.48	107.48	107.48	0.093	44
Huiles, graisses comestibles	112.35	112.70	112.50	0.269	
Huiles comestibles	112.29	112.30	112.28	0.149	45
Graisses comestibles	117.84	118.97	118.41	0.028	46
Margarine	110.79	111.42	111.07	0.092	47
Produits à base de céréales	142.15	142.26	142.22	2.491	
Pain	147.00	147.00	147.00	0.787	
Pain bis	149.42	149.42	149.42	0.315	48
Pain mi-blanc	145.57	145.57	145.57	0.315	49
Pains spéciaux	145.00	145.00	145.00	0.157	50

Autres art. de boulangerie	148.51	148.51	148.51	1.265	
Petite boulangerie	168.99	168.99	168.99	0.316	51
Boulangerie fine	158.71	158.71	158.71	0.316	52
Biscottes et biscuits	125.50	125.50	125.50	0.253	53
Pâtisserie, confiserie	138.32	138.32	138.32	0.380	54
Pâte à gâteau	123.05	124.57	124.05	0.139	55
Farine	126.86	127.12	126.99	0.078	56
Riz	92.94	92.89	92.91	0.056	57
Pâtes alimentaires	109.91	110.27	110.10	0.139	58
Céréales pour petit déj.	113.40	113.19	113.29	0.027	59
Pommes de terre	116.40	116.88	117.07	0.223	
Pommes de terre	101.40	102.11	102.40	0.149	60
Pommes de terre à encaver	146.62	146.62	146.62	0.074	61
Légumes	129.27	127.65	127.82	1.309	
Aubergines	125.48	123.23	123.94	0.013	62
Haricots nains	201.13	192.17	197.60	0.065	63
Champignons de Paris	112.54	110.77	110.92	0.039	64
Fenouil	104.08	101.06	105.63	0.026	65
Carottes	100.86	110.05	109.83	0.065	66
Ail	96.06	99.42	96.01	0.026	67
Choux-fleurs	103.65	102.07	102.74	0.079	68
Choux de Bruxelles	167.62	152.74	167.53	0.039	69
Choux rouges	144.65	147.15	148.69	0.013	70
Choux blancs	138.37	135.03	140.08	0.013	71
Choux frisés	148.07	145.71	145.39	0.013	72
Poireaux verts	133.23	126.11	133.06	0.026	73
Poivrons verts	112.22	108.41	111.42	0.092	74
Endives	112.70	112.70	112.70	0.000	75
Chicorée de Trévise	142.06	146.49	148.36	0.039	76
Chicorée scarole	223.86	223.29	223.88	0.052	77
Concombres	92.90	102.18	98.22	0.065	78
Salade pommée	127.95	136.28	130.79	0.158	79
Salade mélangée	134.40	134.40	134.40	0.065	80
Mâche	195.06	195.87	196.66	0.013	81
Betteraves rouges	145.46	136.26	144.47	0.013	82
Céleri pomme	148.30	152.16	155.58	0.013	83
Asperges blanches	124.70	124.70	124.70	0.000	84
Tomates	123.23	112.94	111.36	0.277	85
Courgettes	92.50	89.06	89.22	0.026	86
Oignons	116.17	118.42	117.89	0.079	87

Fruits	169.75	169.40	168.75	1.449	
Pommes	224.12	232.79	230.32	0.260	
Golden delicious I	224.76	239.65	237.49	0.101	88
Granny smith I	127.81	127.81	127.81	0.000	89
Autres pommes I	245.72	253.36	249.08	0.101	90
Pommes II	185.37	185.01	185.18	0.058	91
Autres fruits	157.86	155.53	155.29	1.189	
Bananes	125.42	127.57	126.78	0.159	92
Poires	324.14	323.14	322.53	0.116	93
Fraises	210.33	210.33	210.33	0.000	94
Grapefruits blancs	113.77	122.29	121.67	0.015	95
Kaki	137.46	137.46	137.46	0.000	96
Pastèques	113.38	123.27	123.03	0.015	97
Melons	131.36	117.72	129.21	0.058	98
Noix	113.79	113.79	113.79	0.000	99
Arachides grillées	114.88	114.88	114.88	0.000	100
Oranges blondes	177.17	189.89	181.32	0.029	101
Clémentines	153.08	153.08	153.08	0.000	102
Mandarines	137.12	137.12	137.12	0.000	103
Abricots	148.69	148.69	148.69	0.000	104
Cerises	170.19	170.19	170.19	0.000	105
Nectarines	123.54	136.31	139.47	0.043	106
Pêches	153.24	149.66	151.13	0.058	107
Quetsches "pruneaux"	213.51	200.05	197.03	0.072	108
Raisin blanc	137.02	133.27	132.34	0.581	109
Citrons	111.80	110.87	111.84	0.043	110
Conserves légumes et fruits	111.77	111.77	111.77	0.404	
Conserves légumes	112.65	112.65	112.65	0.273	111
Conserves fruits	104.57	104.57	104.57	0.081	112
Légumes et fruits congelés	118.67	118.67	118.67	0.050	113
Confiture, miel, mélasse	108.85	108.85	108.85	0.165	114
Plats cuisinés	116.88	116.88	116.88	0.385	115
Potages en sachets	137.92	137.92	137.92	0.084	116
Epices et sauces	123.71	123.71	123.71	0.349	117
Sucre	115.68	115.65	115.66	0.116	118
Aliments fortifiants	113.41	113.41	113.41	0.110	119
Chocolat	107.63	107.60	107.61	0.484	
Chocolat en plaque	106.28	106.34	106.31	0.387	120
Autre chocolat	113.01	112.61	112.81	0.097	121

Café	96.06	95.37	95.63	0.420	
Café en grains	88.41	87.39	87.78	0.288	122
Café soluble en poudre	112.77	112.77	112.77	0.132	123
Thé	126.43	126.48	126.46	0.063	124
Repas au restaurant	140.59	140.59	140.59	3.771	
Menus du jour	138.36	138.36	138.36	1.131	125
Plats chauds	141.15	141.15	141.15	1.697	126
Plats froids	142.56	142.56	142.56	0.754	127
Desserts	141.17	141.17	141.17	0.189	128

Données de Genève pour le mois d'octobre 1991

Dans les tableaux suivants, on donne les indices au niveau des rubriques, c'est-à-dire au dernier niveau où l'on a un poids de pondération.

Indice pour la réplique 1 = A
Indice pour la réplique 2 = B
Indice pour le total = T
Poids de pondération = wi
Numéro de la ligne = i

Rubriques	A	B	T	wi	i
Alimentation	131.68	131.88	131.96	21.000	1
Lait et produits laitiers	124.88	124.86	124.83	3.924	
Lait	126.95	127.30	127.06	1.245	
Lait entier en emballage	123.51	124.93	124.00	0.374	2
Lait "drink"	128.82	128.79	128.80	0.374	3
Autre lait en emballage	162.50	160.98	161.80	0.062	4
Lait entier en vrac	123.25	123.25	123.25	0.435	5
Beurre	114.75	114.28	114.51	0.494	
Beurre de table	104.07	102.86	103.44	0.193	6
Beurre de cuisine	121.61	121.61	121.61	0.301	7
Fromage	128.46	127.81	128.11	1.319	
Fromage à pâte dure	127.57	125.47	126.51	0.593	8
Fromage à pâte mi-dure	132.50	132.15	132.39	0.396	9
From.pâte molle, from. frais	125.88	126.55	126.21	0.198	10
Fromage fondu	124.20	127.18	125.29	0.132	11
Produits laitiers frais	129.31	128.81	129.05	0.449	
Yogourt	130.56	130.56	130.56	0.359	12
Séré	118.92	117.08	117.99	0.045	13
Autres prod. lait. frais	129.77	126.61	128.07	0.045	14
Crème	108.56	110.64	109.51	0.299	
Crème entière	105.30	108.78	106.88	0.179	15
Crème à café	113.42	113.42	113.42	0.120	16
Autres produits laitiers	129.89	131.49	130.57	0.118	17
Oeufs	155.61	155.81	155.72	0.361	
Oeufs du pays	137.50	137.87	137.70	0.199	18
Oeufs étrangers	177.85	177.85	177.85	0.162	19

Viande, charcut., saucisses	123.58	123.07	123.49	4.317	
Viande de boeuf (sans os)	121.84	120.24	121.23	0.972	
Entrecôte (boeuf)	126.15	126.17	126.18	0.126	20
Beefsteak dans la cuisse	129.05	129.08	129.02	0.215	21
Rôti de boeuf	124.22	124.25	124.24	0.165	22
Bouilli de boeuf	116.97	116.93	116.95	0.175	23
Ragoût de boeuf	117.76	118.01	117.98	0.126	24
Viande de boeuf hachée	115.05	105.42	111.33	0.165	25
Viande de veau 1er. chx. (s/os)	135.35	138.22	136.54	0.284	
Tranche de veau	137.93	145.36	140.28	0.102	26
Rôti de veau	133.37	132.18	132.34	0.054	27
Ragoût de veau	136.33	136.09	136.48	0.085	28
Emincé de veau	129.76	133.06	133.06	0.043	29
Viande de porc (sans os)	121.88	120.24	121.39	1.480	
Tranche (jambon, cou)	108.88	108.88	108.88	0.266	30
Tranche (filet, fil.mignon)	121.13	121.18	121.18	0.133	31
Rôti de porc	117.46	115.07	116.34	0.311	32
Cotelettes de porc	111.55	111.88	111.94	0.237	33
Ragoût de porc	123.65	120.30	123.17	0.252	34
Jambon de derrière	131.49	131.13	131.21	0.163	35
Lard maigre	167.34	160.06	164.69	0.118	36
Viande d'agneau	122.43	122.43	122.43	0.074	37
Abats	115.72	115.72	115.72	0.077	38
Charcuterie et saucisses	125.44	125.04	125.29	1.096	39
Volaille	123.57	128.37	126.22	0.288	40
Conserves de viande	113.44	113.44	113.44	0.046	41
Poissons	135.68	135.54	135.61	0.306	
Poissons frais	156.93	156.58	156.76	0.124	42
Poissons congelés	135.54	135.54	135.54	0.089	43
Poissons en boîte	107.48	107.48	107.48	0.093	44
Huiles, graisses comestibles	110.26	110.43	110.66	0.269	
Huiles comestibles	111.96	111.79	111.89	0.149	45
Graisses comestibles	119.27	119.44	119.35	0.028	46
Margarine	104.77	105.49	106.04	0.092	47
Produits à base de céréales	142.34	142.43	142.39	2.491	
Pain	147.00	147.00	147.00	0.787	
Pain bis	149.42	149.42	149.42	0.315	48
Pain mi-blanc	145.57	145.57	145.57	0.315	49
Pains spéciaux	145.00	145.00	145.00	0.157	50

Autres art. de boulangerie	148.29	148.38	148.33	1.265	
Petite boulangerie	168.99	168.99	168.99	0.316	51
Boulangerie fine	158.71	158.71	158.71	0.316	52
Biscottes et biscuits	125.50	125.50	125.50	0.253	53
Pâtisserie, confiserie	137.59	137.87	137.71	0.380	54
Pâte à gâteau	127.08	127.07	127.07	0.139	55
Farine	127.05	126.83	126.94	0.078	56
Riz	92.99	93.25	93.12	0.056	57
Pâtes alimentaires	111.28	111.90	111.63	0.139	58
Céréales pour petit déj.	113.03	114.23	113.68	0.027	59
Pommes de terre	116.16	117.41	116.35	0.223	
Pommes de terre	101.04	102.90	101.32	0.149	60
Pommes de terre à encaver	146.62	146.62	146.62	0.074	61
Légumes	132.44	134.39	134.10	1.309	
Aubergines	150.81	137.19	138.18	0.013	62
Haricots nains	196.68	199.16	197.32	0.052	63
Champignons de Paris	111.71	108.95	108.99	0.039	64
Fenouil	104.16	107.33	104.68	0.026	65
Carottes	114.61	111.22	111.98	0.065	66
Ail	127.98	112.34	128.63	0.026	67
Choux-fleurs	104.58	112.00	109.51	0.079	68
Choux de Bruxelles	129.73	163.73	129.04	0.052	69
Choux rouges	147.80	145.73	146.71	0.026	70
Choux blancs	136.56	135.03	135.73	0.013	71
Choux frisés	134.37	135.88	137.25	0.013	72
Poireaux verts	115.00	116.37	114.76	0.026	73
Poivrons verts	121.33	126.54	123.71	0.052	74
Endives	144.58	141.54	145.01	0.105	75
Chicorée de Trévise	138.18	141.47	141.92	0.039	76
Chicorée scarole	181.77	165.14	166.67	0.052	77
Concombres	122.97	118.24	119.20	0.052	78
Salade pommée	129.72	130.26	131.09	0.118	79
Salade mélangée	128.24	131.26	133.27	0.065	80
Mâche	161.50	159.65	155.29	0.013	81
Betteraves rouges	134.35	141.27	137.16	0.039	82
Céleri pomme	137.02	138.94	140.64	0.026	83
Asperges blanches	124.70	124.70	124.70	0.000	84
Tomates	133.27	139.54	143.43	0.226	85
Courgettes	101.73	104.09	103.47	0.013	86
Oignons	117.22	116.44	116.45	0.079	87

Fruits	166.94	169.13	169.65	1.449	
Pommes	225.48	226.25	229.07	0.405	
Golden delicious I	219.10	211.14	222.30	0.159	88
Granny smith I	127.81	127.81	127.81	0.000	89
Autres pommes I	251.38	253.76	251.81	0.159	90
Pommes II	189.83	203.58	199.91	0.087	91
Autres fruits	144.23	146.97	146.60	1.044	
Bananes	127.42	128.18	124.69	0.174	92
Poires	315.89	313.59	317.04	0.087	93
Fraises	210.33	210.33	210.33	0.000	94
Grapefruits blancs	125.59	120.46	120.28	0.029	95
Kaki	164.58	183.70	166.77	0.029	96
Pastèques	123.27	123.27	123.27	0.000	97
Melons	117.72	117.72	117.72	0.000	98
Noix	110.08	114.64	113.43	0.014	99
Arachides grillées	116.72	115.51	115.96	0.029	100
Oranges blondes	188.57	191.39	189.78	0.043	101
Clémentines	153.08	153.08	153.08	0.000	102
Mandarines	137.12	137.12	137.12	0.130	103
Abricots	148.69	148.69	148.69	0.000	104
Cerises	170.19	170.19	170.19	0.000	105
Nectarines	136.31	136.31	136.31	0.000	106
Pêches	149.66	149.66	149.66	0.000	107
Quetsches "pruneaux"	200.05	200.05	200.05	0.000	108
Raisin blanc	122.03	127.09	128.16	0.466	109
Citrons	111.30	111.52	111.29	0.043	110
Conserves légumes et fruits	111.77	111.77	111.77	0.404	
Conserves légumes	112.65	112.65	112.65	0.273	111
Conserves fruits	104.57	104.57	104.57	0.081	112
Légumes et fruits congelés	118.67	118.67	118.67	0.050	113
Confiture, miel, mélasse	108.85	108.85	108.85	0.165	114
Plats cuisinés	116.88	116.88	116.88	0.385	115
Potages en sachets	137.92	137.92	137.92	0.084	116
Epices et sauces	123.71	123.71	123.71	0.349	117
Sucre	115.60	115.80	115.70	0.116	118
Aliments fortifiants	113.41	113.41	113.41	0.110	119
Chocolat	107.56	107.99	107.77	0.484	
Chocolat en plaque	105.85	105.90	105.88	0.387	120
Autre chocolat	114.37	116.31	115.27	0.097	121

Café	94.73	94.64	94.68	0.420	
Café en grains	86.47	86.33	86.39	0.288	122
Café soluble en poudre	112.77	112.77	112.77	0.132	123
Thé	125.27	125.07	125.18	0.063	124
Repas au restaurant	140.59	140.59	140.59	3.771	
Menus du jour	138.36	138.36	138.36	1.131	125
Plats chauds	141.15	141.15	141.15	1.697	126
Plats froids	142.56	142.56	142.56	0.754	127
Desserts	141.17	141.17	141.17	0.189	128

Bibliographie

[1] Adelman, I.V. (1958). *A new approach to the construction of index numbers.* Review of Economics and Statistics, 40, n.3, pa. 240-249.

[2] Allen, R.G.D. (1975). *Index numbers in theory and practice.* The MacMillan Press Ltd. London.

[3] Andersson, C., Forsman, G. et Wretman, J. (1987). *On the measurement of errors in the swedish consumer price index.* Bulletin de l'IIS, Vol. 3, pa. 155-171., (Conférence de Tokyo, 1987).

[4] Ardilly, P. et Guglielmetti, F. (1991). *Précision de l'indice des prix français et optimisation des échantillons.* Réunion commune CEE/OIT sur les indices des prix à la consommation. 18-21 novembre 1991. Organisation Internationale du Travail. Genève.

[5] Armknecht, P.A. (1991). *Replication and Regression: modern solutions to age old problems in price index construction.* BLS, U.S.A. Réunion commune CEE/OIT sur les indices des prix à la consommation. 18-21 novembre 1991. Organisation Internationale du Travail. Genève.

[6] Balk, B.M. (1980). *On characteristicity of price indices.* Statistical studies, n.26.

[7] Balk, B.M. (1991). *Estimating the precision of a consumer price index; some experiences from the Netherlands.* Netherlands Central Bureau of Statistics. Réunion commune CEE/OIT sur les indices des prix à la consommation. 18-21 novembre 1991. Organisation Internationale du Travail. Genève.

[8] Balk, B.M. and Kersten, H.M.P. (1986).*On the precision of consumer price indices caused by the sampling variability of budget surveys.* Journal of Economic and Social Measurement, 14, pa. 19-35.

[9] Banerjee, K.S. (1959). *Precision in the construction of cost of living index numbers.* Sankhya, 21, Parts 3 and 4, pa. 393-400.

[10] Banerjee, K.S. (1960). *Calculation of sampling errors for index numbers.* Sankhya, 22, Parts 1 and 2, pa. 119-130.

[11] Banerjee, K.S. (1956). *A note on the optimal allocation of consumption items in the construction of a cost of living index.* Econometrica, 24, No.3, pa. 294-295.

[12] Barberi, B. (1965). *L'emploi de la médiane dans l'analyse statistique des prix.* Bulletin de l'Institut International de Statistique. Vol. (41) 1. 35ème session. Beograd. 1965.

[13] Biggeri, L. et Giommi, A. (1987). *On the accuracy and precision of the consumer price indices. Methods and applications to evaluate the influence of the sampling of households.* Bulletin de l'IIS, Vol. 3, pa. 137-154. Conférence de Tokyo.

[14] BIT. (1986). *Indices des prix à la consommation.* Vol. 1. Sources et méthodes Statistiques. BIT. Genève.

[15] BIT. (1988). *Recommandations internationales en vigueur sur les statistiques du travail.* BIT. Genève.

[16] BLS. (1988). *BLS Handbook of Methods.* Bulletin 2285. Avril. Bureau of Labor Statistics, U.S. Department of Labor. U.S.A.

[17] Bowley, A.L. (1928). *Notes on index numbers.* Economic Journal, 38, pa. 216-237.

[18] Bowley, A.L. (1926). *The influence on the precision of index-numbers of correlation between the prices of commodities.* Journal of the Royal Statistical Society, 89, pa. 300-319.

[19] Bowley, A.L. (1911). *The measurement of the accuracy of an average.* Journal of the Royal Statistical Society, 75, pa. 77-88.

[20] Bowley, A.L. (1897). *Relations between the accuracy of an average and that of its constituent parts.* Journal of the Royal Statistical Society, 60(4), December, pa. 855-866.

[21] Carli, G.R. (1764). *Del valore etc.* Opere scelte di Carli, ed. Custodi, vol. 1, p. 299.

[22] Cochran, W.G. (1963). *Sampling techniques.* John Wiley. New York.

[23] CSCS. (1982). *Réexamen et refonte de la méthode de calcul de l'indice suisse des prix à la consommation.* Rapport du 7 septembre 1982 de la CSCS à l'intention du Conseil fédéral. Berne.

[24] CSCS. (1987). *Objectifs et principes de la révision de l'indice suisse des prix à la consommation.* La Vie économique. 6ème fascicule, juin. Pa. 378-380. Berne.

[25] Dalén, J. (1991). *The Impact of Sampling and Non-Sampling Errors on the Swedish Consumer Price Index.* Bulletin de l'Institut International de Statistique, communications libres, livraison 1, pa. 151-152. 48ème session, Le Caire. 9-17 septembre 1991.

[26] Dalén, J. (1991). *The work with an error model and error calculations in the Swedish CPI.* Statistics Sweden. Réunion commune CEE/OIT sur les indices des prix à la consommation. 18-21 novembre 1991. Organisation Internationale du Travail. Genève.

[27] Dippo, C.S. and Jacobs, C.A. (1983). *Area sample redesign for the consumer price index.* Proceedings of the American Statistical Association, section on survey research methods, Washington, pa. 118-123.

[28] Dodge, Y. (1993). *Statistique. Dictionnaire encyclopédique.* Dunod. Paris.

[29] Dodge, Y. (1989). *Mathématiques de base pour économistes.* Presses Académiques Neuchâtel. Neuchâtel.

[30] Dodge, Y., Mehran, F. et Rousson, M. (1990). *Statistique.* Presses Académiques Neuchâtel. Neuchâtel.

[31] Dodge, Y., Mehran, F., König, S. et Pfaff, S. (1991). *Can the Swiss Consumer Price Index Samples be divided into Replicates for Variance Estimation ?* Bulletin de l'Institut International de Statistique, communications libres, livraison 1, pa. 173-174. 48ème session, Le Caire. 9-17 septembre 1991.

[32] Dodge, Y., Mehran, F., König, S. et Pfaff, S. (1992). *Indice suisse des prix à la consommation: Quelques recommandations relatives à son échantillonnage à venir.* Rapport final-FNRS, Université de Neuchâtel.

[33] Dussaix, A.-M. et Grobras, J.-M. (1993). *Les sondages: principes et méthodes.* Que sais-je? Presses Universiatires de France. Paris.

[34] Craig, J. (1969). *On the elementary treatment of index numbers.* Journal of the Royal Statistical Society, C, 18, pa. 141-152.

[35] Drobisch, M.W. (1871). *Über Mittelgrössen und die Andwendbarkeit derselben auf die Berechnung des Steigens und Sinkens des Geldwerts.* Berichte über die Verhandlungen der König. Sachs. Ges. Wiss. Leipzig. Math-Phy. Klasse 23, 25.

[36] Dutot, C. (1738). *Réflexions politiques sur les finances et le commerce.* The Hague.

[37] Edgeworth, F.Y. (1925). *The plurality of index numbers.* Economic Journal, 35, pa. 279-388.

[38] Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* Society for Industrial and Applied Mathematics. Philadelphia.

[39] Evelyn, Sir George Shuckburgh. (1798). *An account of some endeavours to ascertain a standard of weight and measure.* Phil. Trans., 113.

[40] Fisher, I. (1922). *The making of index numbers.* Boston and New York, houfton Mifflin.

[41] Fleetwood, W. (1707). *Chronicon Preciosum.* London. (Second edition 1745).

[42] Forsyth, F.G. (1978). *The practical construction of a chain price index number.* Journal of the Royal Statistical Society, 141, Part 3, pa. 348-358.

[43] Généreux, P.A. (1983). *Impact of the choice of formulae on the canadian consumer price index.* Price level measurement: Proceedings from a conference sponsored by Statistics Canada, Diewert and Montmarquette, pa. 489-511.

[44] Gourieroux, Ch. (1981). *Théorie des sondages.* Economica. Paris.

[45] Grais, B. (1977). *Méthodes statistiques.* Dunod. Paris.

[46] Grais, B. (1979). *Statistique descriptive.* 2ème édition. Dunod. Paris.

[47] Granville, W.A., Smith, P.F. et Longley, W.R. (1970). *Eléments de calcul différentiel et intégral.* Vuibert. Paris.

[48] Grosbras, J.-M. (1987).*Méthodes statistiques des sondages.* Economica. Paris.

[49] Hagemann, R.P. (1982). *The variability of inflation rates across household types.* Journal of Money, Credit and Banking, 4, November, pa. 494-510.

[50] Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1960). *Sample survey methods and theory.* Vol. 1 et 11. John Wiley. New York.

[51] Hansen, B. and Lucas, E.F. (1984). *On the accuracy of index numbers.* The review of Income and Wealth, 30, pa. 25-38.

[52] Hofsten, E. von (1959). *Price Indexes and Sampling.* Sankhya, 21 pa. 401-403.

[53] Kalton, G. (1979). *Ultimate cluster sampling.* Journal of the Royal Statistical Society, A, 142, 2, pa. 211-221.

[54] Kendall, M.G. (1969). *The early history of index numbers.* Studies in the history of probability and statistics. Vol. 11. Pa. 51-62. Edited by M. Kendall and R.L. Plackett. Griffin, London. 1977.

[55] Kersten, H.M.P. (1985). *Nonresponse assessment of a consumer price index.* Journal of Business and Economic Statistics, 3, pa. 336-343.

[56] Kiaer, A. (1896). *Observations et expériences concernant des dénombrements représentatifs.* Bulletin de l'IIS, 9, liv. 2, pa. 176-183 (Congrès de Berne, 1895).

[57] Kiear, A. (1899). *Sur les méthodes représentatives ou typologiques appliquées à la statistique.* Bulletin de l'IIS, 11, liv. 1, pa. 180-185 (Congrès de Saint-Petersbourg, 1897).

[58] Kiear, A. (1903). *Sur les méthodes représentatives ou typologiques appliquées à la statistique.* Bulletin de l'IIS, 13, liv. 1, pa. 66-78, (Congrès de Budapest, 1901).

[59] Kiear, A. (1905). *Discours sans intitulé sur la méthode représentative.* Bulletin de l'IIS, 14, liv. 1, pa. 119-134, (Congrès de Berlin, 1903).

[60] Kish, L. (1959). *Variances for indexes from comples samples.* American Statistical Association, Proceedings of the Social Statistical Section, Washington D.C., pa. 190-199.

[61] König, S. (1991). *Estimation de la variance. Méthodes basées sur le concept de réplication.* Mémoire du postgrade en statistique. Université de Neuchâtel.

[62] König, S. et Pfaff, S. (1991). *Indice des prix à la consommation des USA et calcul de sa variance.* Rapport technique 1991-1. Université de Neuchâtel.

[63] König, S. et Pfaff, S. (1991). *Estimation de la variance: Cas des plans de sondage complexes.* Rapport technique 1991-2. Université de Neuchâtel.

[64] König, S. et Pfaff, S. (1991). *Estimation de la variance. Cas des plans de sondage complexes. Programmes.* Annexe au rapport technique 1991-2. Université de Neuchâtel.

[65] König, S. et Pfaff, S. (1992). *Indice des prix à la consommation en Suisse et calcul de sa variance.* Rapport technique 1992-1. Université de Neuchâtel.

[66] König, S. et Pfaff, S. (1992). *Indice des prix à la consommation en Suisse et calcul de sa variance. Programmes.* Annexe au rapport technique 1992-1. Université de Neuchâtel.

[67] Koop, J.C. (1986). *Estimating variance of a consumer price index and some comments on inference.* Journal of Official Statistics, 57, pa.74-76.

[68] Kott, Ph.S. (1984). *A superpopulation theory approach to the design of price index estimators with small sampling biases.* Journal of Business and Economic Statistics, 2, pa. 83-90.

[69] Laspeyres, E. (1864). *Hamburger Warenpreise 1850-1863 und die Kalifornisch-australischen Gold-entdeckung seit 1848.* Jahrbücher für Nationaloekonomie und Statistik,3. 81.

[70] Laspeyres, E. (1871). *Die Berechnung einer mittleren Warenpreissteigerung.* Jahrbücher für Nationaloekonomie und Statistik, 16, 296.

[71] Leaver, S.G., Weher, W.L., Cohen, M.P. and Archer, K.P. (1983). *Item-outlet sample redesign for the 1987 U.S. consumer price index revision.* Proceedings of the 46th Session of ISI, Tokyo, vol.52 (3), pa. 173-185.

[72] Lowe, J. (1822). *The present state of England etc.* London.

[73] Mairesse, J. (1988). *Estimation et sondages. Cinq contributions à l'histoire de la statistique.* Economica. Paris.

[74] Malaguerra, C. et Koch, D. (1991). *Révision de l'indice suisse des prix à la consommation.* Bulletin de l'Institut International de Statistique, communications libres, livraison 2, pa. 425-427. 48ème session, Le Caire. 9-17 septembre 1991.

[75] Neyman, J. (1934). *On the two different aspects of the representative method: the method of stratified sampling and the method*

of purposive selection. J. Roy. Statist. Assoc., 97, pa. 558-606, discussion pa. 607-625.

[76] Neyman, J. (1938). *Contribution ot the theory of sampling human populations*. J. Amer. Stat. Assoc., 33, pa. 101-116.

[77] Norwood, J.L. (1983). *Problems in the measurement of consumer prices.* Proceedings of the 44th Session of ISI, Madrid, vol.50 (1), pa. 148-169.

[78] OFIAMT. (1977). *Budgets de ménages de salarié en 1975.* La Vie économique. 2ème fascicule, février. pa. 71-81. Berne.

[79] OFIAMT. (1977). *Nouvelles bases et méthodes du calcul de l'indice suisse des prix à la consommation.* 11e fascicule, novembre. pa. 664-665. La Vie économique. Berne.

[80] OFIAMT. (1977). *Budgets des ménages de salariés en 1975.* 92e numéro spécial de La Vie économique. Berne.

[81] OFIAMT. (1977). *L'indice suisse des prix à la consommation.* 89e numéro spécial de La Vie économique. Berne.

[82] OFIAMT. (1978). *Budgets des ménages de salariés en 1977.* La Vie économique. 6ème fascicule, juin. pa. 347-364. Berne.

[83] OFIAMT. (1982). *Modification de la méthode de calcul de l'indice suisse des prix à la consommation.* La Vie économique. 5ème fascicule, mai. pa. 328. Berne.

[84] OFIAMT. (1982). *Budgets des ménages de salariés et de retraités en 1981.* 5e fascicule, mai. pa. 329-356. La Vie économique. Berne.

[85] OFIAMT. (1985). *Nouvelle méthode de calcul de l'indice suisse des prix à la consommation. Base décembre 1982=100.* 97e numéro spécial de La Vie économique. Berne.

[86] OFIAMT. (1985). *Indice suisse des prix à la consommation. Plausibilité du panier-type de marchandises et du schéma de pondération, des relevés et des mouvements de prix.* Tirage à part de La Vie économique, août, 8e fascicule. Berne.

[87] OFS. (1987). *Budgets des ménages de salariés et de retraités en 1986*. 6e fascicule, juin. Pa. 428-449. La Vie économique. Berne.

[88] OFS. (1990). *La révision de l'indice suisse des prix à la consommation. Concept de base*. Documents techniques, OFS. Berne.

[89] OFS. (1993). *Révision de l'indice suisse des prix à la consommation. Conception du nouvel indice suisse des prix à la consommation*. Documents techniques, OFS. Berne.

[90] OFS. (1993). *Le nouvel indice suisse des prix à la consommation: mai 1993 = 100*. Actualités OFS. Berne.

[91] OFS. (1993). *La statistique suisse des prix: amélioration des informations destinées au public, aux milieux politiques, économiques et scientifiques*. Communiqué de presse, mars 1993. Dept. fédéral de l'intérieur. Berne.

[92] Ohlsson, E. (1991). *The Sampling Variance of the Swedish Consumer Pirce Index*. Bulletin de l'Institut International de Statistique, communications libres, livraison 2, pa. 504-505. 48ème session, Le Caire. 9-17 septembre 1991.

[93] OIT. (1987). *Rapport II. Indices des prix à la consommation*. Quatorzième Conférence internationale des statisticiens du travail. Genève, 28 octobre-6 novembre 1987. Bureau International du Travail. Genève.

[94] OIT. (1987). *Rapport de la Conférence*. Quatorzième Conférence internationale des statisticiens du travail. Genève, 28 octobre-6 novembre 1987. Bureau International du Travail. Genève.

[95] Paasche, H. (1874). *Über die Preisentwicklung der letzten Jahre, nach den Hamburger Börsennotierungen*. Jahrbücher für Nationaloekonomie und Statistik, 23, 168.

[96] Picard, H. (1985). *L'échantillonnage dans les indices des prix à la consommation*. Bureau International du Travail. Genève.

[97] Rao, J.N.K. et Wu, C.F.J. (1984). *Bootstrap Inference for Sample Surveys*. Proceedings of the Section on Survey Research Methods ASA, Washington D.C., pa. 106-112.

[98] Sellwood, D.J. (1978). *Reduction of errors in a consumer price index*. Bulletin of Labor Statistics, n.1, International Labour Office, Geneva, pa. IX-XI.

[99] Singh, D. and Chaudhary, F.S. (1986). *Theory and analysis of sample survey designs*. John Wiley. New Delhi.

[100] Tchuprov, A. (1923). *On the mathematical expectation of the moments of frequency distributions in the case of correlated observations*. Metron, 2, pa.461-493 et pa. 646-683.

[101] Thakurta, B.K.G. (1967). *On a problem in conumers' price index number*. Applied Statistics, 16, pa. 42-45.

[102] Turvey, R. et al. (1989). *Consumer price indices. An ILO manual*. Bureau International du Travail. Genève.

[103] US Department of Labor, Bureau of Labor Statistics. (1988). *BLS Handbook of Methods*. Bulletin 2285. April 1988.

[104] Wolter, K.M. (1985). *Introduction to variance estimation*. Springer-Verlag. New York.

[105] Wilkerson, M. (1967). *Sampling error in the consumer price index*. Journal of the American Statistical Association, 62, n.319, pa. 899-914.

[106] Woodruf, R.S. and Causey, B.D. (1976). *Computerized method for approximating the variance of a complicated estimate*. Journal of the American Statistical Association, vol.71,n.354, pa. 315-321.

[107] Zahn, D.A. and Killion, R.A. (1975). *A direct calculation of the variance estimate for a complex ratio statistic with applications to the Florida price level index study*. Proceedings of the American Statistical Association, Business and Economic Statitics Section, pa. 635-639.