

Variance estimation by linearization for indicators of poverty and social exclusion in a person and household survey context Illustration with swiss SILC 2009

Eric Graf

Statistical Institute
University of Neuchâtel, Switzerland
www.unine.ch/statistics

NTTS, 7th of March 2013

Content

Introduction

Content

Introduction

Indicators for poverty and social exclusion

Empirical calculations

Parametric calculations

GB2 distribution and inequality measures

Content

Introduction

Indicators for poverty and social exclusion

- Empirical calculations

- Parametric calculations

- GB2 distribution and inequality measures

Linearization techniques

- Generalized linearization

- Influence functions for poverty and inequality measure

 - Estimation of income density function

 - Results from simulations

- Variance estimation of indicators under GB2

- Calibrated survey weights and linearization

Content

Introduction

Indicators for poverty and social exclusion

- Empirical calculations

- Parametric calculations

- GB2 distribution and inequality measures

Linearization techniques

- Generalized linearization

- Influence functions for poverty and inequality measure

 - Estimation of income density function

 - Results from simulations

- Variance estimation of indicators under GB2

- Calibrated survey weights and linearization

Illustrative results

Content

Introduction

Indicators for poverty and social exclusion

- Empirical calculations

- Parametric calculations

- GB2 distribution and inequality measures

Linearization techniques

- Generalized linearization

- Influence functions for poverty and inequality measure

 - Estimation of income density function

 - Results from simulations

- Variance estimation of indicators under GB2

- Calibrated survey weights and linearization

Illustrative results

Conclusions

Introduction : variance estimation

European projects

- ▶ **DACSEIS** (2001-2004, FP5) *Data Quality in Complex Surveys within the New European Information Society*
- ▶ **KEI** (2004-2008, FP6) *Knowledge Economy Indicators*
- ▶ **SAMPLE** (2008-2011, FP7) *Small Area Methods for Poverty and Living Condition Estimates* ↔ EU-SILC
- ▶ **AMELI** (2008-2011, FP7) *Advanced Methodology for European Laeken Indicators* : estimation, variance estimation, robustness,... ↔ EU-SILC
- ▶ *Handbook on Precision Requirements and Variance Estimation for ESS Household Surveys* : recommendations, precision requirements, best practices on variance estimation, computing standard errors for national and european statistics, references, ...

Introduction : variance estimation

European projects

- ▶ **DACSEIS** (2001-2004, FP5) *Data Quality in Complex Surveys within the New European Information Society*
- ▶ **KEI** (2004-2008, FP6) *Knowledge Economy Indicators*
- ▶ **SAMPLE** (2008-2011, FP7) *Small Area Methods for Poverty and Living Condition Estimates* ↔ EU-SILC
- ▶ **AMELI** (2008-2011, FP7) *Advanced Methodology for European Laeken Indicators* : estimation, variance estimation, robustness,... ↔ EU-SILC
- ▶ *Handbook on Precision Requirements and Variance Estimation for ESS Household Surveys* : recommendations, precision requirements, best practices on variance estimation, computing standard errors for national and european statistics, references, ...

Introduction : variance estimation

European projects

- ▶ **DACSEIS** (2001-2004, FP5) *Data Quality in Complex Surveys within the New European Information Society*
- ▶ **KEI** (2004-2008, FP6) *Knowledge Economy Indicators*
- ▶ **SAMPLE** (2008-2011, FP7) *Small Area Methods for Poverty and Living Condition Estimates* ↔ EU-SILC
- ▶ **AMELI** (2008-2011, FP7) *Advanced Methodology for European Laeken Indicators* : estimation, variance estimation, robustness,... ↔ EU-SILC
- ▶ *Handbook on Precision Requirements and Variance Estimation for ESS Household Surveys* : recommendations, precision requirements, best practices on variance estimation, computing standard errors for national and european statistics, references, ...

Introduction : variance estimation

European projects

- ▶ **DACSEIS** (2001-2004, FP5) *Data Quality in Complex Surveys within the New European Information Society*
- ▶ **KEI** (2004-2008, FP6) *Knowledge Economy Indicators*
- ▶ **SAMPLE** (2008-2011, FP7) *Small Area Methods for Poverty and Living Condition Estimates* ↔ EU-SILC
- ▶ **AMELI** (2008-2011, FP7) *Advanced Methodology for European Laeken Indicators* : estimation, variance estimation, robustness,... ↔ EU-SILC
- ▶ *Handbook on Precision Requirements and Variance Estimation for ESS Household Surveys* : recommendations, precision requirements, best practices on variance estimation, computing standard errors for national and european statistics, references, ...

Introduction : variance estimation

European projects

- ▶ **DACSEIS** (2001-2004, FP5) *Data Quality in Complex Surveys within the New European Information Society*
- ▶ **KEI** (2004-2008, FP6) *Knowledge Economy Indicators*
- ▶ **SAMPLE** (2008-2011, FP7) *Small Area Methods for Poverty and Living Condition Estimates* ↔ EU-SILC
- ▶ **AMELI** (2008-2011, FP7) *Advanced Methodology for European Laeken Indicators* : estimation, variance estimation, robustness,... ↔ EU-SILC
- ▶ *Handbook on Precision Requirements and Variance Estimation for ESS Household Surveys* : recommendations, precision requirements, best practices on variance estimation, computing standard errors for national and european statistics, references, ...

Goal

Through a household survey (example : SILC), we collect, among other measures, some income variable, call it y .

We intend to do inference to the population in publishing estimated values with confidence intervals for several indicators of poverty and social exclusion.



Press release

in this country, the Gini index is of $29\% \pm 1.5\%$.

Need of procedure to estimate the variance.

Sources of variability of a survey based estimator

- ▶ Sampling variance : the survey providing the salaries y originates from a probabilistic sample \Rightarrow **need to describe the sample-to-sample variations**
- ▶ Non-response variance, imputation variance, over-coverage variance, response-variance, ...
- ▶ Mostly the weights provided with the collected sample reflect some of the variability sources : sampling design, correction for unit non response, there may have been weight sharing, calibration, etc. **The weights are random too !**

Sources of variability of a survey based estimator

- ▶ Sampling variance : the survey providing the salaries y originates from a probabilistic sample \Rightarrow **need to describe the sample-to-sample variations**
- ▶ Non-response variance, imputation variance, over-coverage variance, response-variance, ...
- ▶ Mostly the weights provided with the collected sample reflect some of the variability sources : sampling design, correction for unit non response, there may have been weight sharing, calibration, etc. **The weights are random too !**

Sources of variability of a survey based estimator

- ▶ Sampling variance : the survey providing the salaries y originates from a probabilistic sample \Rightarrow **need to describe the sample-to-sample variations**
- ▶ Non-response variance, imputation variance, over-coverage variance, response-variance, ...
- ▶ Mostly the weights provided with the collected sample reflect some of the variability sources : sampling design, correction for unit non response, there may have been weight sharing, calibration, etc. **The weights are random too !**

Additional problems for Laeken indicators (ex. GINI)

- ▶ The **distribution of income is hard to model**, far away from a (log-)Normal distribution
- ▶ Most of the **indicators** for poverty and social exclusion **are non linear functions of y** .

How to estimate correctly their variance?!?

In most of the cases : many estimation methods lead to adequate estimator of variance under the commonly used sampling design, no unique methodology!

ATTENTION : **Bias** may also be a non negligible problem!

Additional problems for Laeken indicators (ex. GINI)

- ▶ The **distribution of income is hard to model**, far away from a (log-)Normal distribution
- ▶ Most of the **indicators** for poverty and social exclusion **are non linear functions of y** .

How to estimate correctly their variance?!?

In most of the cases : many estimation methods lead to adequate estimator of variance under the commonly used sampling design, no unique methodology!

ATTENTION : **Bias** may also be a non negligible problem!

Additional problems for Laeken indicators (ex. GINI)

- ▶ The **distribution of income is hard to model**, far away from a (log-)Normal distribution
- ▶ Most of the **indicators** for poverty and social exclusion **are non linear functions of y** .

How to estimate correctly their variance?!?

In most of the cases : many estimation methods lead to adequate estimator of variance under the commonly used sampling design, no unique methodology!

ATTENTION : **Bias** may also be a non negligible problem!

Additional problems for Laeken indicators (ex. GINI)

- ▶ The **distribution of income is hard to model**, far away from a (log-)Normal distribution
- ▶ Most of the **indicators** for poverty and social exclusion **are non linear functions of y** .

How to estimate correctly their variance?!?

In most of the cases : many estimation methods lead to adequate estimator of variance under the commonly used sampling design, no unique methodology!

ATTENTION : **Bias** may also be a non negligible problem!

Additional problems for Laeken indicators (ex. GINI)

- ▶ The **distribution of income is hard to model**, far away from a (log-)Normal distribution
- ▶ Most of the **indicators** for poverty and social exclusion **are non linear functions of y** .

How to estimate correctly their variance?!?

In most of the cases : many estimation methods lead to adequate estimator of variance under the commonly used sampling design, no unique methodology!

ATTENTION : **Bias** may also be a non negligible problem!

How to estimate the variance of complex statistics ?

- ▶ There are **no closed formulae** to compute the variance of inequality measures
- ▶ Two main philosophies :
 - ▶ **replication methods** (Jackknife, Bootstrap, Balanced repeated replications, Balanced Half-Samples, Random Group Methods)
 - ▶ **linearization methods** (Taylor Linearization, Linearization based on estimating equations, **Linearization based on influence functions**, Jackknife Linearization)
- ▶ Osier (2009), opting for linearization, based on Deville (1999)'s work cleared the way for variance estimation of some main poverty and inequality measures, Graf & Tillé (2013) revisited the results and brought some improvements

How to estimate the variance of complex statistics ?

- ▶ There are **no closed formulae** to compute the variance of inequality measures
- ▶ Two main philosophies :
 - ▶ **replication methods** (Jackknife, Bootstrap, Balanced repeated replications, Balanced Half-Samples, Random Group Methods)
 - ▶ **linearization methods** (Taylor Linearization, Linearization based on estimating equations, **Linearization based on influence functions**, Jackknife Linearization)
- ▶ Osier (2009), opting for linearization, based on Deville (1999)'s work cleared the way for variance estimation of some main poverty and inequality measures, Graf & Tillé (2013) revisited the results and brought some improvements

How to estimate the variance of complex statistics ?

- ▶ There are **no closed formulae to compute the variance** of inequality measures
- ▶ Two main philosophies :
 - ▶ **replication methods** (Jackknife, Bootstrap, Balanced repeated replications, Balanced Half-Samples, Random Group Methods)
 - ▶ **linearization methods** (Taylor Linearization, Linearization based on estimating equations, **Linearization based on influence functions**, Jackknife Linearization)
- ▶ Osier (2009), opting for linearization, based on Deville (1999)'s work cleared the way for variance estimation of some main poverty and inequality measures, Graf & Tillé (2013) revisited the results and brought some improvements

Indicators for poverty and social exclusion

Let's focus on five of them, formerly called *Laeken indicators* (Eurostat, 2005)

- ▶ At risk of poverty threshold ARPT
- ▶ At risk of poverty rate ARPR
- ▶ Relative median at-risk-of poverty gap RMPG
- ▶ Quintile share ratio (S_{80}/S_{20}) QSR
- ▶ Gini index GINI

Laeken indicators : empirical calculations

At risk of poverty threshold Let m be the median income,

$$\text{ARPT} = 0.6m$$

ARPT is an absolute measure (CHF, EUR, \$,...).

At risk of poverty rate ARPR is the share of persons with an income \leq ARPT. Let F be the distribution function of y ,

$$\text{ARPR} = P(Y < \text{ARPT}) = F^{-1}(\text{ARPT})$$

ARPR is a relative measure, it's scale free, $\text{ARPR} \in [0, 1]$.

Laeken indicators : empirical calculations

At risk of poverty threshold Let m be the median income,

$$\text{ARPT} = 0.6m$$

ARPT is an absolute measure (CHF, EUR, \$,...).

At risk of poverty rate ARPR is the share of persons with an income \leq ARPT. Let F be the distribution function of y ,

$$\text{ARPR} = P(Y < \text{ARPT}) = F^{-1}(\text{ARPT})$$

ARPR is a relative measure, it's scale free, $\text{ARPR} \in [0, 1]$.

Laeken indicators : empirical calculations

Relative median at-risk-of poverty gap The RMPG is the relative diff. between the ARPT and the median income of the poor (m_p) :

$$\text{RMPG} = \frac{\text{ARPT} - m_p}{\text{ARPT}} \in [0, 1]$$

Quintile share ratio (S_{80}/S_{20}) or degree of income inequality in the population. QSR is the ration of the total income of the persons in the top income quintile ($> q_{80}$) over the total income of the persons in the bottom quintile ($< q_{20}$),

$$\text{QSR} = \frac{\sum_{y_k > q_{80}} y_k}{\sum_{y_k < q_{20}} y_k}$$

Laeken indicators : empirical calculations

Relative median at-risk-of poverty gap The RMPG is the relative diff. between the ARPT and the median income of the poor (m_p) :

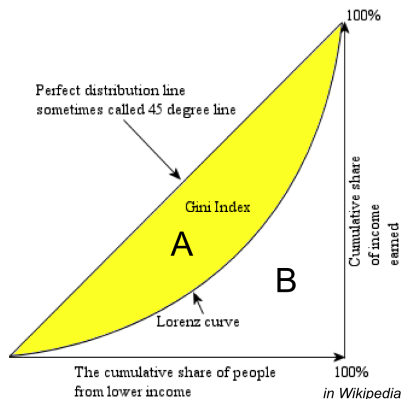
$$\text{RMPG} = \frac{\text{ARPT} - m_p}{\text{ARPT}} \in [0, 1]$$

Quintile share ratio (S_{80}/S_{20}) or degree of income inequality in the population. QSR is the ration of the total income of the persons in the top income quintile ($> q_{80}$) over the total income of the persons in the bottom quintile ($< q_{20}$),

$$\text{QSR} = \frac{\sum_{y_k > q_{80}} y_k}{\sum_{y_k < q_{20}} y_k}$$

Laeken indicators : empirical calculations

Gini index



$$\text{GINI} = 2A \in [0, 1]$$

$$A + B = 1/2$$

$$G = 2 \int_0^1 [\alpha - L(\alpha)] d\alpha.$$

If the y_k are n discrete realization of a random variable \mathcal{Y} , sorted by their ranks, an estimation of Gini is given by :

$$\hat{G} = \frac{2 \sum_{k \in S} k y_k}{n \sum_{k \in S} y_k} - \frac{n+1}{n}$$

Income distribution, GB2

- ▶ Knowledge about the **income distribution** of the population is of vital interest for every study of economic markets in order to govern economic and social decisions
- ▶ The study of the distribution of income is in the heart of inequality measures and more generally evaluations of social welfare
- ▶ The European project AMELI (2011), relying on the EU-SILC data, showed that **generalized beta distribution of the second kind GB2** can be fitted to the income data collected in a good way.

Income distribution, GB2

- ▶ Knowledge about the **income distribution** of the population is of vital interest for every study of economic markets in order to govern economic and social decisions
- ▶ The study of the distribution of income is in the heart of inequality measures and more generally evaluations of social welfare
- ▶ The European project AMELI (2011), relying on the EU-SILC data, showed that **generalized beta distribution of the second kind GB2** can be fitted to the income data collected in a good way.

Income distribution, GB2

- ▶ Knowledge about the **income distribution** of the population is of vital interest for every study of economic markets in order to govern economic and social decisions
- ▶ The study of the distribution of income is in the heart of inequality measures and more generally evaluations of social welfare
- ▶ The European project AMELI (2011), relying on the EU-SILC data, showed that **generalized beta distribution of the second kind GB2** can be fitted to the income data collected in a good way.

Generalized beta distribution of the second kind (GB2)

- ▶ **four parameters** distribution : $GB2(a, b, p, q)$.
- ▶ has been developed by McDonald (1984).

$$f_{GB2}(y; a, b, p, q) = \frac{a}{b \cdot B(p, q)} \frac{(y/b)^{ap-1}}{(1+(y/b)^a)^{p+q}}$$

- ▶ density :
where $B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt$ is the beta function.

Empirical studies on income - see for ex. Jenkins (2007); Dastrup et al. (2007); Kleiber et Kotz (2003); Sepanski et Kong (2007) - show that **the GB2 fits well with such data** and it is often more suitable than other four-parameter distribution.
Results from AMELI (2011) **confirm for EU-SILC**.

Generalized beta distribution of the second kind (GB2)

- ▶ **four parameters** distribution : $GB2(a, b, p, q)$.
- ▶ has been developed by McDonald (1984).

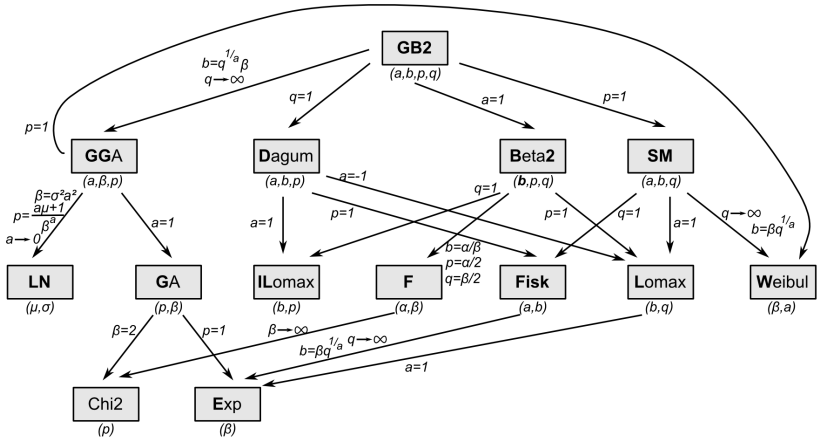
$$f_{GB2}(y; a, b, p, q) = \frac{a}{b \cdot B(p, q)} \frac{(y/b)^{ap-1}}{(1+(y/b)^a)^{p+q}}$$

- ▶ density :
where $B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt$ is the beta function.

Empirical studies on income - see for ex. Jenkins (2007); Dastrup et al. (2007); Kleiber et Kotz (2003); Sepanski et Kong (2007) - show that **the GB2 fits well with such data** and it is often more suitable than other four-parameter distribution.
Results from AMELI (2011) **confirm for EU-SILC**.

GB2 Law I

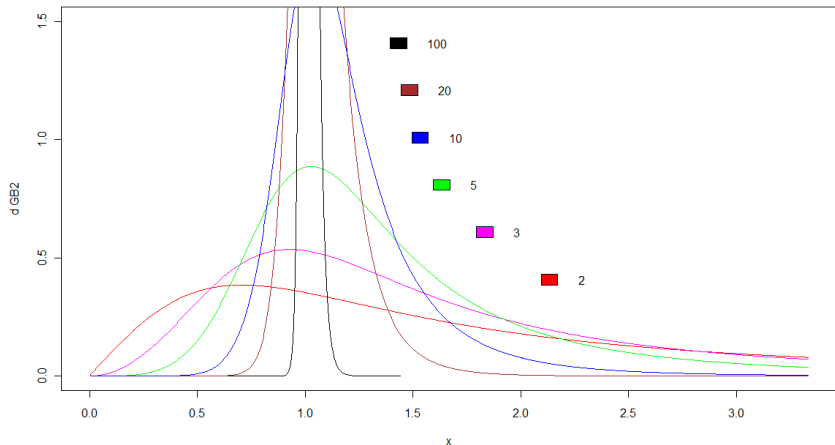
Many other probability distributions can be seen as special cases of the GB2



GB2 Law II

GB2 densities, a varies, $b = 1$, $p = 1$, $q = 0.5$

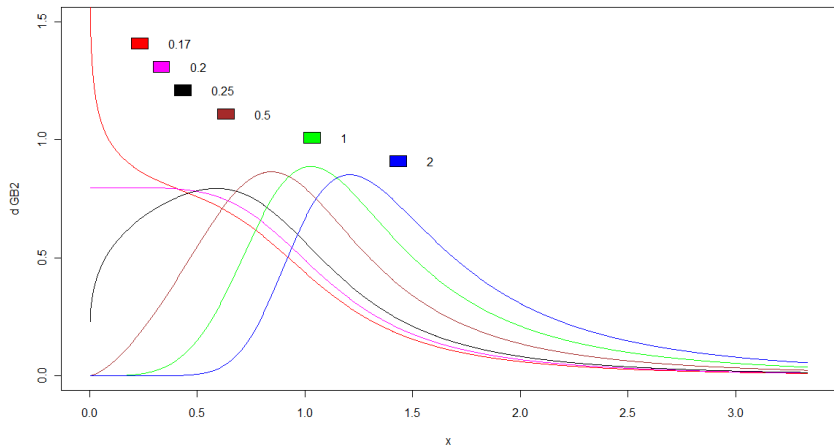
→ \pm sharp distribution



GB2 Law III

GB2 densities, $a = 5$, $b = 1$, p varies, $q = 0.5$

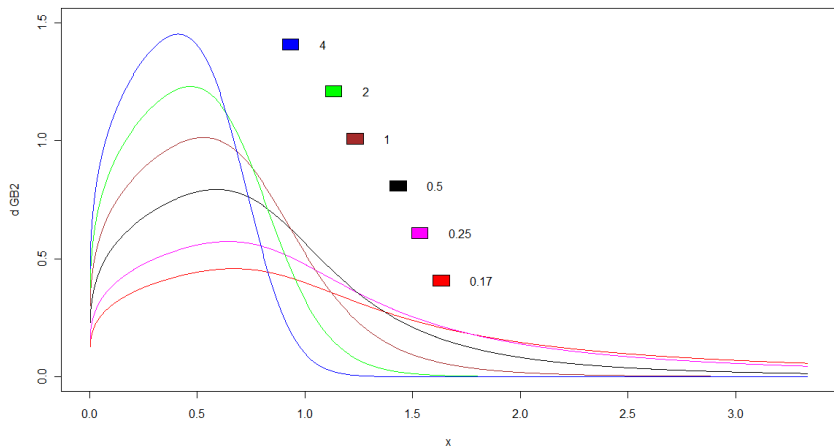
→ playing on the shape of the left tail



GB2 Law IV

GB2 densities, $a = 5$, $b = 1$, $p = 0.25$, q varies

→ playing on the shape of the right tail



GB2 distribution and inequality measures

Advantage of a parametric estimation for a distribution of income : there are **explicit formulae for the inequality measures as functions of the four parameters $\theta = (a, b, p, q)$** of the GB2 adjusted to the data. McDonald (1984), Graf, M. (2009), Ameli (2011)

- ▶ At risk of poverty threshold ARPT(a, b, p, q)
- ▶ At risk of poverty rate ARPR(a, b, p, q)
- ▶ Relative median at-risk-of poverty gap RMPG(a, b, p, q)
- ▶ Quintile share ratio (S_{80}/S_{20}) QSR(a, b, p, q)
- ▶ Gini index GINI(a, b, p, q)

All programmed and available in statistical software R, package ("GB2").

GB2 distribution and inequality measures

Advantage of a parametric estimation for a distribution of income : there are **explicit formulae for the inequality measures as functions of the four parameters $\theta = (a, b, p, q)$** of the GB2 adjusted to the data. McDonald (1984), Graf, M. (2009), Ameli (2011)

- ▶ **At risk of poverty threshold** ARPT(a, b, p, q)
- ▶ **At risk of poverty rate** ARPR(a, b, p, q)
- ▶ **Relative median at-risk-of poverty gap** RMPG(a, b, p, q)
- ▶ **Quintile share ratio (S_{80}/S_{20})** QSR(a, b, p, q)
- ▶ **Gini index** GINI(a, b, p, q)

All programmed and available in statistical software R, package ("GB2").

GB2 distribution and inequality measures

Advantage of a parametric estimation for a distribution of income : there are **explicit formulae for the inequality measures as functions of the four parameters $\theta = (a, b, p, q)$** of the GB2 adjusted to the data. McDonald (1984), Graf, M. (2009), Ameli (2011)

- ▶ **At risk of poverty threshold** ARPT(a, b, p, q)
- ▶ **At risk of poverty rate** ARPR(a, b, p, q)
- ▶ **Relative median at-risk-of poverty gap** RMPG(a, b, p, q)
- ▶ **Quintile share ratio (S_{80}/S_{20})** QSR(a, b, p, q)
- ▶ **Gini index** GINI(a, b, p, q)

All programmed and available in statistical software **R**, package ("GB2").

Generalized linearization technique

- ▶ Usable for **estimating the variance of sample-based non linear statistics** (Deville, 1999)
- ▶ Relies on the concept of **influence functions** proposed in robust statistic (Hampel, 1974)
- ▶ Applied by Osier (2009) for several EU-SILC indicators

Idea : what is the **influence** of a unit k on the parameter of interest $A = T(M)$ viewed as a functional of a measure M ?

M is a measure allowing some mass to each unit :

- ▶ $M(k) = M_k = 1$ for $k \in U$
- ▶ $\hat{M}(k) = \hat{M}_k = w_k$ for $k \in U$

Generalized linearization technique

- ▶ Usable for **estimating the variance of sample-based non linear statistics** (Deville, 1999)
- ▶ Relies on the concept of **influence functions** proposed in robust statistic (Hampel, 1974)
- ▶ Applied by Osier (2009) for several EU-SILC indicators

Idea : what is the **influence** of a unit k on the parameter of interest $A = T(M)$ viewed as a functional of a measure M ?

M is a measure allowing some mass to each unit :

- ▶ $M(k) = M_k = 1$ for $k \in U$
- ▶ $\hat{M}(k) = \hat{M}_k = w_k$ for $k \in U$

Influence function

One identifies the influence of the unit k on A at the population or sample level by an infinitesimal variation of the weight of the unit.

Influence function of A

$$I[T(M)]_k = z_k = \lim_{t \rightarrow 0} \frac{T(M + t\delta_k) - T(M)}{t}, \text{ for all } k \in U$$

Empirical influence function of A

$$I[T(\hat{M})]_k = \hat{z}_k = \lim_{t \rightarrow 0} \frac{T(\hat{M} + t\delta_k) - T(\hat{M})}{t}, \text{ for all } k \in s$$

Example : for the population total $Y = \sum_{k \in U} y_k$, $Y = T(M)$,
 $\hat{Y} = T(\hat{M}) = \sum_{k \in s} w_k y_k$, $z_k = y_k$ $k \in U$, $\hat{z}_k = w_k y_k$ $k \in s$

Influence function

One identifies the influence of the unit k on A at the population or sample level by an infinitesimal variation of the weight of the unit.

Influence function of A

$$I[T(M)]_k = z_k = \lim_{t \rightarrow 0} \frac{T(M + t\delta_k) - T(M)}{t}, \text{ for all } k \in U$$

Empirical influence function of A

$$I[T(\hat{M})]_k = \hat{z}_k = \lim_{t \rightarrow 0} \frac{T(\hat{M} + t\delta_k) - T(\hat{M})}{t}, \text{ for all } k \in s$$

Example : for the population total $Y = \sum_{k \in U} y_k$, $Y = T(M)$,
 $\hat{Y} = T(\hat{M}) = \sum_{k \in s} w_k y_k$, $z_k = y_k$ $k \in U$, $\hat{z}_k = w_k y_k$ $k \in s$

Generalized linearization technique : main result

Under asymptotic conditions described in Deville (1999), that are in principle satisfied if the sample is "big enough", **the variance of the estimated total of \hat{z}_k is an approximation of the variance of the (complex) statistic \hat{A}** :

$$\text{var} \left[\sum_{k \in s} \hat{z}_k w_k(s) \right] \approx \text{var}(\hat{A})$$

Influence functions for poverty and inequality measure

See Osier (2009), Langel & Tillé (2011, 2012), Graf & Tillé (2013)

Suppose y_k ordered, $A = \text{GINI}$

$$\hat{z}_k^{\text{GINI}} = \frac{1}{\hat{N}\hat{Y}} \left[2\hat{N}_k(y_k - \hat{Y}_k) + \hat{Y} - \hat{N}y_k - \hat{G}(\hat{Y} + y_k\hat{N}) \right]$$

where

- ▶ $\hat{N}_k = \sum_{l \in S} w_l \mathbf{1}_{[y_l \leq y_k]}$ is the cumulated sum of the w_k up to k
- ▶ $\hat{N} = \sum_{k \in S} w_k$ the estimated total size of the population
- ▶ $\hat{Y}_k = \sum_{l \in S} w_l y_l \mathbf{1}_{[y_l \leq y_k]} / \hat{N}_k$ the weighted mean partial sum

Influence functions for poverty and inequality measure

See Osier (2009), Langel & Tillé (2011, 2012), Graf & Tillé (2013)

Suppose y_k ordered, $A = \text{GINI}$

$$\hat{z}_k^{\text{GINI}} = \frac{1}{\widehat{N}\widehat{Y}} \left[2\widehat{N}_k(y_k - \widehat{Y}_k) + \widehat{Y} - \widehat{N}y_k - \widehat{G}(\widehat{Y} + y_k\widehat{N}) \right]$$

where

- ▶ $\widehat{N}_k = \sum_{\ell \in S} w_\ell \mathbf{1}_{[y_\ell \leq y_k]}$ is the cumulated sum of the w_k up to k
- ▶ $\widehat{N} = \sum_{k \in S} w_k$ the estimated total size of the population
- ▶ $\widehat{Y}_k = \sum_{\ell \in S} w_\ell y_\ell \mathbf{1}_{[y_\ell \leq y_k]} / \widehat{N}_k$ the weighted mean partial sum

$A = \text{QSR}$ Quintile share ratio

$$\hat{z}_k^{\text{QSR}} = \frac{y_k - \left\{ y_k H \left(\frac{0.8\hat{N} - \hat{N}_{k-1}}{w_k} \right) + \hat{Q}_{0.8} [0.8 - \mathbf{1}_{[y_k < \hat{Q}_{0.8}]}] \right\}}{\hat{Y}_{0.2}} \\ - \frac{(\hat{Y} - \hat{Y}_{0.8}) \left\{ y_k H \left(\frac{0.2\hat{N} - \hat{N}_{k-1}}{w_k} \right) + \hat{Q}_{0.2} [0.2 - \mathbf{1}_{[y_k < \hat{Q}_{0.2}]}] \right\}}{\hat{Y}_{0.2}^2}$$

where

$$\blacktriangleright H(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1. \end{cases}$$

$A = \text{ARPT}$ At risk of poverty threshold

$$\hat{z}_k^{\text{ARPT}} = -\frac{0.6}{f(\hat{m})} \frac{1}{\hat{N}} [\mathbf{1}_{[y_k \leq \hat{m}]} - 0.5]$$

$A = \text{ARPR}$ At risk of poverty rate

$$\hat{z}_k^{\text{ARPR}} = \frac{1}{\hat{N}} \left(\mathbf{1}_{[y_k \leq \widehat{\text{ARPT}}]} + \widehat{\text{ARPR}} \right) - f(\widehat{\text{ARPT}}) \hat{z}_k^{\text{ARPT}}$$

$A = \text{ARPT}$ At risk of poverty threshold

$$\hat{z}_k^{\text{ARPT}} = -\frac{0.6}{f(\hat{m})} \frac{1}{\hat{N}} [\mathbf{1}_{[y_k \leq \hat{m}]} - 0.5]$$

$A = \text{ARPR}$ At risk of poverty rate

$$\hat{z}_k^{\text{ARPR}} = \frac{1}{\hat{N}} \left(\mathbf{1}_{[y_k \leq \widehat{\text{ARPT}}]} + \widehat{\text{ARPR}} \right) - f(\widehat{\text{ARPT}}) \hat{z}_k^{\text{ARPT}}$$

$A = m_p$ Median of poor

$$\hat{z}_k^{m_p} = \frac{1}{f(\hat{m}_p)} \frac{\hat{z}_k^{\text{ARPR}}}{2} - \frac{1}{\hat{N}} (\mathbf{1}_{[y_k \leq \hat{m}_p]} - F(\hat{m}_p))$$

$A = \text{RMPG}$ Relative median at-risk-of poverty gap

$$\hat{z}_k^{\text{RMPG}} = \frac{\hat{m}_p \hat{z}_k^{\text{ARPT}} - \widehat{\text{ARPT}} \hat{z}_k^{m_p}}{\widehat{\text{ARPT}}^2}$$

$A = m_p$ Median of poor

$$\hat{z}_k^{m_p} = \frac{1}{f(\hat{m}_p)} \frac{\hat{z}_k^{\text{ARPR}}}{2} - \frac{1}{\hat{N}} (\mathbf{1}_{[y_k \leq \hat{m}_p]} - F(\hat{m}_p))$$

$A = \text{RMPG}$ Relative median at-risk-of poverty gap

$$\hat{z}_k^{\text{RMPG}} = \frac{\hat{m}_p \hat{z}_k^{\text{ARPT}} - \widehat{\text{ARPT}} \hat{z}_k^{m_p}}{\widehat{\text{ARPT}}^2}$$

Estimation of income density function f

The **influence functions** or **linearized variables** of the ARPT, ARPR, m_p and RMPG **require to estimate the value of f in some points.**

Deville (1999) and Osier (2009) propose to proceed through **Gaussian kernel density estimation**

$$\textcircled{1} \quad \hat{f}_1(x) = \frac{1}{h\sqrt{2\pi}} \frac{1}{\hat{N}} \sum_{k \in S} w_k \exp \left[-\frac{(x-y_k)^2}{2h^2} \right]$$

- ▶ h is the *bandwidth* that Osier estimates by $\hat{h} = \hat{\sigma} \hat{N}^{-0.2}$
- ▶ $\hat{\sigma}$ the empirical standard deviation of the income variable \mathcal{Y} .
- ▶ The **estimation of σ is non robust** (sensitive to extreme values)
- ▶ In surveys, one encounter often some **concentration of observations around some values** what may cause trouble with a fixed bandwidth.

Estimation of income density function f

The **influence functions** or **linearized variables** of the ARPT, ARPR, m_p and RMPG **require to estimate the value of f in some points.** Deville (1999) and Osier (2009) propose to proceed through **Gaussian kernel density estimation**

$$\textcircled{1} \quad \hat{f}_1(x) = \frac{1}{h\sqrt{2\pi}} \frac{1}{\hat{N}} \sum_{k \in S} w_k \exp \left[-\frac{(x-y_k)^2}{2h^2} \right]$$

- ▶ h is the *bandwidth* that Osier estimates by $\hat{h} = \hat{\sigma} \hat{N}^{-0.2}$
- ▶ $\hat{\sigma}$ the empirical standard deviation of the income variable \mathcal{Y} .
- ▶ The **estimation of σ is non robust** (sensitive to extreme values)
- ▶ In surveys, one encounter often some **concentration of observations around some values** what may cause trouble with a fixed bandwidth.

Two proposed alternatives for density estimation

② Estimate through the logarithm

$$\hat{f}_2(x) = \frac{\hat{f}_1(\log(x))}{x}$$

Diminishes problems due to extreme values in the GK density estimation

③ Nearest neighbours with minimal bandwidth $\hat{f}_{NNMB}(x) = \frac{p(x)}{nh(x)}$

At each point of the distribution, use at least p neighbours and a minimal bandwidth $h(x) \geq h_{opt}$ where h_{opt} is the *rule of thumb* of Silverman (1986) to determine the bandwidth.

This solution is more robust and avoids problems arising from the concentration of observations around some values.

$$\hat{f}_3(x) = \frac{\hat{f}_{NNMB}(\log(x))}{x}$$

Two proposed alternatives for density estimation

② Estimate through the logarithm

$$\hat{f}_2(x) = \frac{\hat{f}_1(\log(x))}{x}$$

Diminishes problems due to extreme values in the GK density estimation

③ Nearest neighbours with minimal bandwidth $\hat{f}_{NNMB}(x) = \frac{p(x)}{nh(x)}$

At each point of the distribution, use at least p neighbours and a minimal bandwidth $h(x) \geq h_{opt}$ where h_{opt} is the *rule of thumb* of Silverman (1986) to determine the bandwidth.

This solution is more robust and avoids problems arising from the concentration of observations around some values.

$$\hat{f}_3(x) = \frac{\hat{f}_{NNMB}(\log(x))}{x}$$

Results from simulations

Relative bias of the variance obtained on 1000 simple random samples without replacement on the Swiss SILC 2009 data ($y =$ equivalized household income, $N = 17,534$)

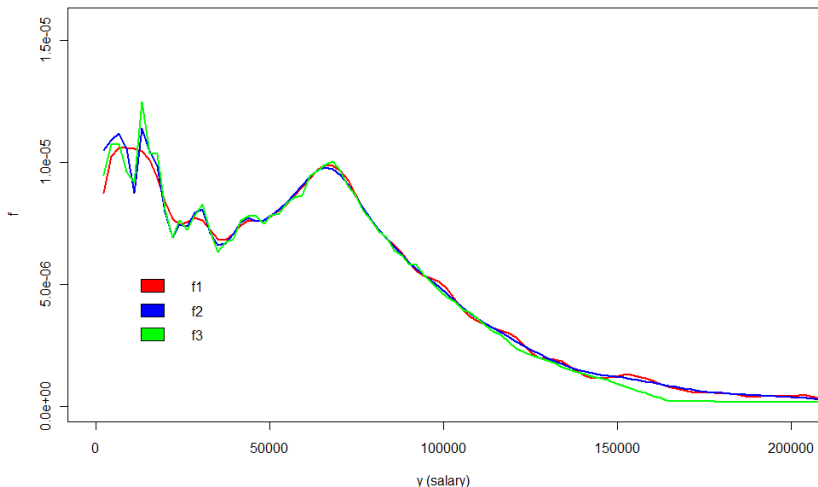
Indicator	Sample sizes (sampling rate)					
	$n = 500(2.9\%)$			$n = 1000(5.7\%)$		
	\hat{f}_1	\hat{f}_2	\hat{f}_3	\hat{f}_1	\hat{f}_2	\hat{f}_3
GINI	0.02			0.00		
QSR	-0.02			-0.02		
ARPT	-0.08	-0.05	0.00	-0.12	-0.09	-0.04
ARPR	-0.14	-0.10	-0.04	-0.10	-0.07	0.00
RMPG	-0.04	0.01	-0.07	-0.02	0.02	-0.05
m_p	-0.17	-0.08	-0.08	-0.15	-0.07	-0.07
m	-0.08	-0.05	0.00	-0.11	-0.08	-0.03

Results from simulations

Relative bias of the variance obtained on 1000 simple random samples without replacement on the Swiss SILC 2009 data (y = personal income for a subgroup of employed persons, $N = 7,922$)

Indicator	Sample sizes (sampling rate)					
	$n = 500(6.3\%)$			$n = 1000(12.6\%)$		
GINI	-0.05			-0.04		
QSR	-0.13			-0.05		
	\hat{f}_1	\hat{f}_2	\hat{f}_3	\hat{f}_1	\hat{f}_2	\hat{f}_3
ARPT	0.06	0.04	0.03	0.08	0.05	0.04
ARPR	-0.04	-0.03	0.00	-0.06	-0.05	-0.02
RMPG	0.53	0.05	-0.03	0.53	0.05	-0.02
m_p	0.71	0.14	-0.00	0.71	0.14	-0.01
m	0.06	0.03	0.02	0.07	0.04	0.03

Density estimations for personal income



Variance estim. of indicators under GB2 : fitting the GB2

Pseudo maximum likelihood estimation of the parameters of the GB2 - AMELI (2011), Graf & Nedyalkova (2013)

$$l_n(\theta) = \sum_{i=1}^n w_i \log f_{GB2}(y_i; \theta)$$

Partial derivatives of the pseudo-loglikelihood function are set to 0 and solve for θ :

$$l'_n(\theta) = \sum_{i=1}^n w_i \underbrace{\frac{\partial}{\partial \theta} \log f_{GB2}(y_i; \theta)}_{:=u_i(y_i; \theta)}.$$

The scores $u_i(y_i; \theta)$ are 4×1 vectors of partial derivatives of the GB2 density function in y_i . *They can be seen as the linearized parameters with respect to the GB2.*

Variance estim. of indicators under GB2 : fitting the GB2

Pseudo maximum likelihood estimation of the parameters of the GB2 - AMELI (2011), Graf & Nedyalkova (2013)

$$l_n(\theta) = \sum_{i=1}^n w_i \log f_{GB2}(y_i; \theta)$$

Partial derivatives of the pseudo-loglikelihood function are set to 0 and solve for θ :

$$l'_n(\theta) = \sum_{i=1}^n w_i \underbrace{\frac{\partial}{\partial \theta} \log f_{GB2}(y_i; \theta)}_{:=u_i(y_i; \theta)}.$$

The scores $u_i(y_i; \theta)$ are 4×1 vectors of partial derivatives of the GB2 density function in y_i . *They can be seen as the linearized parameters with respect to the GB2.*

Variance estim. of indicators under GB2 : fitting the GB2

Pseudo maximum likelihood estimation of the parameters of the GB2 - AMELI (2011), Graf & Nedyalkova (2013)

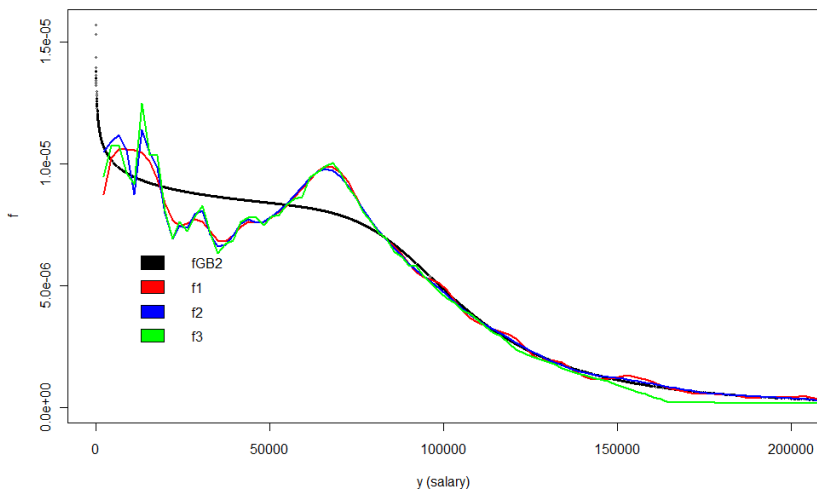
$$l_n(\theta) = \sum_{i=1}^n w_i \log f_{GB2}(y_i; \theta)$$

Partial derivatives of the pseudo-loglikelihood function are set to 0 and solve for θ :

$$l'_n(\theta) = \sum_{i=1}^n w_i \underbrace{\frac{\partial}{\partial \theta} \log f_{GB2}(y_i; \theta)}_{:= u_i(y_i; \theta)}.$$

The scores $u_i(y_i; \theta)$ are 4×1 vectors of partial derivatives of the GB2 density function in y_i . *They can be seen as the linearized parameters with respect to the GB2.*

Density estimations for personal income



Variance estimation of the parameters of the GB2

Approximate $l'_n(\hat{\theta}_n)$ by the **first two terms** of a Taylor series around θ (**linearization!**):

$$\begin{aligned}l'_n(\hat{\theta}_n) &\approx l'_n(\theta) + l''_n(\theta)(\hat{\theta}_n - \theta) \stackrel{!}{=} 0 \\ \Rightarrow \hat{\theta}_n - \theta &\approx [-l''_n(\theta)]^{-1} l'_n(\theta)\end{aligned}$$

Get the approximated variance by the so-called **Huber sandwich estimator** (Huber (1967), White (1980, 1982), Freedman(2006), Pfeffermann and Sverchkov (2003)):

$$\widehat{\text{var}}(\hat{\theta}_n) \approx [l''_n(\hat{\theta}_n)]^{-1} \widehat{\text{var}}[l'_n(\hat{\theta})] [l''_n(\hat{\theta}_n)]^{-1}$$

with $\widehat{\text{var}}[l'_n(\hat{\theta}_n)] = \widehat{\text{var}}[\sum_{i=1}^n w_i u_i(y_i; \hat{\theta}_n)]$ the **estimated design variance** of the estimated total from the u_j .

Variance estimation of the parameters of the GB2

Approximate $l'_n(\hat{\theta}_n)$ by the **first two terms** of a Taylor series around θ (**linearization!**):

$$l'_n(\hat{\theta}_n) \approx l'_n(\theta) + l''_n(\theta)(\hat{\theta}_n - \theta) \stackrel{!}{=} 0$$

$$\Rightarrow \hat{\theta}_n - \theta \approx [-l''_n(\theta)]^{-1} l'_n(\theta)$$

Get the approximated variance by the so-called **Huber sandwich estimator** (Huber (1967), White (1980, 1982), Freedman(2006), Pfeffermann and Sverchkov (2003)):

$$\widehat{\text{var}}(\hat{\theta}_n) \approx [l''_n(\hat{\theta}_n)]^{-1} \widehat{\text{var}}[l'_n(\hat{\theta})] [l''_n(\hat{\theta}_n)]^{-1}$$

with $\widehat{\text{var}}[l'_n(\hat{\theta}_n)] = \widehat{\text{var}}[\sum_{i=1}^n w_i u_i(y_i; \hat{\theta}_n)]$ the **estimated design variance** of the estimated total from the u_j .

Variance of the indicators expressed as a function of the 4 GB2-parameters $\theta = (a, b, p, q)'$

Apply the **Delta method** (Davison (2003), AMELI (2011)).
 $\hat{A} = A(\hat{\theta}_n)$ (for example $A = \text{GINI}, \text{ARPR}, \text{QSR}, \dots$) then :

$$\widehat{\text{var}}(\hat{A}) = \frac{\partial \hat{A}'}{\partial \hat{\theta}_n} \widehat{\text{var}}(\hat{\theta}_n) \frac{\partial \hat{A}}{\partial \hat{\theta}_n}$$

The partial derivatives of the indicators with respect to the 4 parameters can be computed numerically.

Calibrated survey weights and linearization

Weighting scheme of the swiss SILC : main steps (Graf, 2008)

- ▶ stratified random cluster sample \Rightarrow sampling weights
- ▶ non response corrections on Grid, Household, Individual levels
 \Rightarrow NR-correction weights
- ▶ weight sharing (see Lavallée, 2002)
- ▶ panels combinations
- ▶ calibration on known totals

Massiani (2012) detailed the procedure to estimate the variance of a total taking all these steps into account, i.e. take the randomness of the weights into account.

Calibrated survey weights and linearization

Weighting scheme of the swiss SILC : main steps (Graf, 2008)

- ▶ stratified random cluster sample \Rightarrow sampling weights
- ▶ non response corrections on Grid, Household, Individual levels
 \Rightarrow NR-correction weights
- ▶ weight sharing (see Lavallée, 2002)
- ▶ panels combinations
- ▶ calibration on known totals

Massiani (2012) detailed the procedure to estimate the variance of a total taking all these steps into account, i.e. take the randomness of the weights into account.

Calibrated survey weights and linearization

Weighting scheme of the swiss SILC : main steps (Graf, 2008)

- ▶ stratified random cluster sample \Rightarrow sampling weights
- ▶ non response corrections on Grid, Household, Individual levels
 \Rightarrow NR-correction weights
- ▶ weight sharing (see Lavallée, 2002)
- ▶ panels combinations
- ▶ calibration on known totals

Massiani (2012) detailed the procedure to estimate the variance of a total taking all these steps into account, i.e. take the randomness of the weights into account.

Calibrated survey weights and linearization

Weighting scheme of the swiss SILC : main steps (Graf, 2008)

- ▶ stratified random cluster sample \Rightarrow sampling weights
- ▶ non response corrections on Grid, Household, Individual levels
 \Rightarrow NR-correction weights
- ▶ weight sharing (see Lavallée, 2002)
- ▶ panels combinations
- ▶ calibration on known totals

Massiani (2012) detailed the procedure to estimate the variance of a total taking all these steps into account, i.e. take the randomness of the weights into account.

Calibrated survey weights and linearization

Weighting scheme of the swiss SILC : main steps (Graf, 2008)

- ▶ stratified random cluster sample \Rightarrow sampling weights
- ▶ non response corrections on Grid, Household, Individual levels
 \Rightarrow NR-correction weights
- ▶ weight sharing (see Lavallée, 2002)
- ▶ panels combinations
- ▶ calibration on known totals

Massiani (2012) detailed the procedure to estimate the variance of a total taking all these steps into account, i.e. take the randomness of the weights into account.

Calibrated survey weights and linearization

Weighting scheme of the swiss SILC : main steps (Graf, 2008)

- ▶ stratified random cluster sample \Rightarrow sampling weights
- ▶ non response corrections on Grid, Household, Individual levels
 \Rightarrow NR-correction weights
- ▶ weight sharing (see Lavallée, 2002)
- ▶ panels combinations
- ▶ calibration on known totals

Massiani (2012) detailed the procedure to estimate the variance of a total taking all these steps into account, i.e. take the randomness of the weights into account.

Variance estimation for calibrated estimators of totals

- ▶ estimated total of y_k : $\sum_{k \in S} w_k y_k$
- ▶ estimated total of some linearized variable z_k : $\sum_{k \in S} w_k z_k$
- ▶ estimated total of GB2 scores u_k : $\sum_{k \in S} w_k u_k$

Deville (1999) shows that to estimate the variance of such a calibrated estimator (Deville & Särndal, 1992) is equivalent to estimate the design variance of the residuals from the regression of the variable of interest on the auxiliary variables used to calibrate.
 \Rightarrow the residuals of that regression are the linearized variable with respect to the calibration.

Variance estimation for calibrated estimators of totals

- ▶ estimated total of y_k : $\sum_{k \in S} w_k y_k$
- ▶ estimated total of some linearized variable z_k : $\sum_{k \in S} w_k z_k$
- ▶ estimated total of GB2 scores u_k : $\sum_{k \in S} w_k u_k$

Deville (1999) shows that **to estimate the variance of such a calibrated estimator** (Deville & Särndal, 1992) **is equivalent to estimate the design variance of the residuals from the regression of the variable of interest on the auxiliary variables used to calibrate.**
⇒ the residuals of that regression are the linearized variable with respect to the calibration.

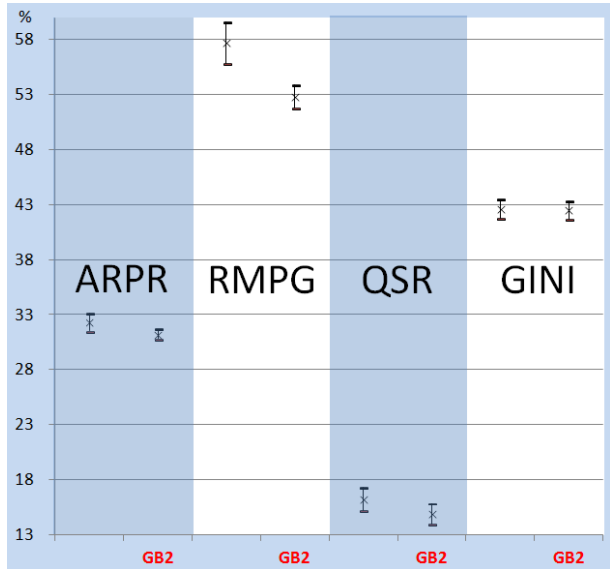
Variance estimation for calibrated estimators of totals

- ▶ estimated total of y_k : $\sum_{k \in S} w_k y_k$
- ▶ estimated total of some linearized variable z_k : $\sum_{k \in S} w_k z_k$
- ▶ estimated total of GB2 scores u_k : $\sum_{k \in S} w_k u_k$

Deville (1999) shows that **to estimate the variance of such a calibrated estimator** (Deville & Särndal, 1992) **is equivalent to estimate the design variance of the residuals from the regression of the variable of interest on the auxiliary variables used to calibrate.**
 \Rightarrow the residuals of that regression are the linearized variable with respect to the calibration.

Illustrative results

y = personal income
for a subgroup of
salaried individuals
 $n = 7,922$



Concluding remarks

- ▶ Taylor linearization and generalized linearization techniques are of many use in the field of survey methodology, these methods work very well with large samples (say > 1000 units)
- ▶ In small samples, asymptotic properties may fail and some linearized variables (ex. for the median) can become non robust to extreme values (Croux, 1998)!
- ▶ Generalized linearization permits to estimate variances for widely used complex indicators, **but one must be careful how the income density is estimated in such computations**. We've shown some good working methods which reduce the bias of the variance.
- ▶ For the poverty and inequality measures considered, empirical and GB2-parametrical calculations provide very similar results.
- ▶ Work in progress : take imputation into account in the variance estimation through generalized linearization.

Concluding remarks

- ▶ Taylor linearization and generalized linearization techniques are of many use in the field of survey methodology, these methods work very well with large samples (say > 1000 units)
- ▶ In small samples, asymptotic properties may fail and some linearized variables (ex. for the median) can become non robust to extreme values (Croux, 1998)!
- ▶ Generalized linearization permits to estimate variances for widely used complex indicators, **but one must be careful how the income density is estimated in such computations**. We've shown some good working methods which reduce the bias of the variance.
- ▶ For the poverty and inequality measures considered, empirical and GB2-parametrical calculations provide very similar results.
- ▶ Work in progress : take imputation into account in the variance estimation through generalized linearization.

Concluding remarks

- ▶ Taylor linearization and generalized linearization techniques are of many use in the field of survey methodology, these methods work very well with large samples (say > 1000 units)
- ▶ In small samples, asymptotic properties may fail and some linearized variables (ex. for the median) can become non robust to extreme values (Croux, 1998)!
- ▶ Generalized linearization permits to estimate variances for widely used complex indicators, **but one must be careful how the income density is estimated in such computations**. We've shown some good working methods which reduce the bias of the variance.
- ▶ For the poverty and inequality measures considered, empirical and GB2-parametrical calculations provide very similar results.
- ▶ Work in progress : take imputation into account in the variance estimation through generalized linearization.


Concluding remarks

- ▶ Taylor linearization and generalized linearization techniques are of many use in the field of survey methodology, these methods work very well with large samples (say > 1000 units)
- ▶ In small samples, asymptotic properties may fail and some linearized variables (ex. for the median) can become non robust to extreme values (Croux, 1998)!
- ▶ Generalized linearization permits to estimate variances for widely used complex indicators, **but one must be careful how the income density is estimated in such computations**. We've shown some good working methods which reduce the bias of the variance.
- ▶ For the poverty and inequality measures considered, empirical and GB2-parametrical calculations provide very similar results.
- ▶ Work in progress : take imputation into account in the variance estimation through generalized linearization.

Concluding remarks

- ▶ Taylor linearization and generalized linearization techniques are of many use in the field of survey methodology, these methods work very well with large samples (say > 1000 units)
- ▶ In small samples, asymptotic properties may fail and some linearized variables (ex. for the median) can become non robust to extreme values (Croux, 1998)!
- ▶ Generalized linearization permits to estimate variances for widely used complex indicators, **but one must be careful how the income density is estimated in such computations**. We've shown some good working methods which reduce the bias of the variance.
- ▶ For the poverty and inequality measures considered, empirical and GB2-parametrical calculations provide very similar results.
- ▶ Work in progress : take imputation into account in the variance estimation through generalized linearization.

Cited references

- ▶ AMELI (2011), Deliverable 2.1, Graf, M., Nedyalkova, D., Münnich, R., Seger, J., Zins, S.. *Parametric Estimation of Income Distributions, and Indicators of Poverty and Social Exclusion*.
- ▶ Croux, C. (1998). *Statistics & Probability Letters. Limit behaviour of the empirical influence function of the median*.
- ▶ Dastrup, S. R., Hartshorn, McDonald, J. B.. *Journal of Economic Inequality. The impact of taxes and transfer payments on the distribution of income : A parametric comparison*.
- ▶ Davison, A. (2003). Cambridge University Press. *Statistical Models*.
- ▶ Deville, J.-C., Särndal, K. E. (1992). *Journal of ASA. Calibration Estimators in Survey Sampling*.
- ▶ Deville, J.-C. (1999). *Survey Methodology. Variance estimation for complex statistics and estimators : linearization and residual techniques*.
- ▶ Eurostat (2005). Working papers and studies, Office for Official Publications of the European Communities, Luxembourg. *The continuity of indicators during the transition between ECHP and EU-SILC*.
- ▶ Freedman, D. A. (2006). *American Statistician. On the so-called "Hubert sandwich estimator" and "robust standard errors"*.
- ▶ Graf, E. (2008). Rapport de méthodes. Office Fédéral de la Statistique. *Pondérations du SILC pilote*.
- ▶ Graf, E., and Tillé, Y. (2013). Submitted paper. *Estimation de variance par linéarisation pour des indices de pauvreté et d'exclusion sociale*.
- ▶ Graf, M. (2009). *JSM Proceedings. An Efficient Algorithm for the Computation of the Gini Coefficient of the Generalized Beta Distribution of the Second Kind*.
- ▶ Graf, M. and Nedyalkova, D. (2013). In press. Review of Income and Wealth. *Modeling of income and indicators of poverty and social exclusion using the Generalized Beta Distribution of the Second Kind*.
- ▶ Hampel, F. R. (1974). *Journal of the American Statistical Association. The influence curve and its role in robust estimation*.
- ▶ Huber, P. J. (1967). In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. *The behavior of maximum likelihood estimates under nonstandard conditions*. 

- ▶ Jenkins, S. P. (2007). *Inequality and the GB2 Income Distribution*.
- ▶ Kleiber, C., Kotz, S. (2003). Wiley. *Statistical Size Distributions in Economics and Actuarial Sciences*.
- ▶ McDonald James B. (1984). *Econometrica. Some Generalized Functions for the Size Distribution of Income*.
- ▶ Langel, M. and Tillé, Y. (2011). *Journal of Statistical Planning and Inference. Statistical inference for the quintile share ratio*.
- ▶ Langel, M. and Tillé, Y. (2012). In Press in *Journal of the Royal Statistical Society - Series A. Variance estimation of the Gini index : Revisiting a result several times published*.
- ▶ Lavallée, P. (2002). *Le sondage indirect ou la méthode généralisée du partage des poids*.
- ▶ Massiani, A. (2012). *Survey Methodology. Estimation of the variance of cross-sectional indicators for the SILC survey in Switzerland*.
- ▶ Osier, G. (2009). *Survey Research Methods. Variance estimation for complex indicators of poverty and inequality using linearization techniques*.
- ▶ Pfefferman, D. and Sverchkov, M. Y. (2003). Wiley, New York. in Skinner, C. and Chambers, R. editors. *Analysis of Survey Data. Fitting generalized linear model under informative sampling*.
- ▶ Sepanski, J. H. and Kong, J. (2007). *Electronic Journal of Statistics. A Family of Generalized Beta Distributions for Income*.
- ▶ Silverman, B. W. (1986). Chapman and Hall, London. *Density Estimation for Statistics and Data Analysis*.
- ▶ White, H. (1980). *Econometrica. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity*.
- ▶ White, H. (1982). *Econometrica. Maximum likelihood estimation of misspecified models*.
- ▶ European projects : DACSEIS, KEI, SAMPLE, AMELI, see <http://www.uni-trier.de>