EUROPEAN COMMISSION
Community Research

SEVENTH FRAMEWORK PROGRAMME

AMELI
Advanced Methodology for European Laeken Indicators

Deliverable 3.2

Variance Estimation for Indicators of Poverty and Social Exclusion

Version: 2011

Ralf Münnich and Stefan Zins

The project **FP7–SSH–2007–217322 AMELI** is supported by European Commission funding from the Seventh Framework Programme for Research.

http://ameli.surveystatistics.net/

Contributors to Deliverable 3.2

Chapter 1: Ralf Münnich and Stefan Zins, University of Trier.

Chapter 2: Ralf Münnich and Stefan Zins, University of Trier.

Chapter 3: Ralf Münnich and Stefan Zins, University of Trier.

Main responsibility

Ralf Münnich and Stefan Zins, University of Trier

Evaluators

Internal expert: Monique Graf, Swiss Federal Statistical Office.

Aim and Objectives of Deliverable 3.2

The main focus of this report is put on linearization methods to derive simple but still reliable variance estimates for indicators of poverty and inequality. Linearization methods enable estimating the variance of these statistics using standard variance estimation software developed for linear estimators. The report covers two areas of variance estimation. The first is the linearization of non-linear statistics, a general approach which allows also for the linearization of non-smooth statistics. The second one is the measurment of change in indicators values over time, which entails the need for variance estimation for measures of change. Variance estimation for change is then also addressed through an application of the already presented linearization methods to longitudinal data.

The results of a simulation study, where the linearization presented in this report have been employed, can be found in (see Alfons et al., 2011, chapter 9). Results from another, smaller simulation study, on the performance of variance estimation of change are given within this report.

Contents

Chapter 1

Introduction

In Bruch et al. (2011) the point estimator under consideration was the Horvitz-Thompson estimator of a population total. A special attention was put on different variance estimators and their expression for different sampling designs. However, in sample surveys it is required or a necessarty to employ non-linear estimators to obtain the statistics of interest. Unfortunatly, in most cases, the variance of non-linear estimators cannot be given in closed form. Furthermore, unbiased estimates of variances of non-linear estimators do not exist.

The two commonly used approaches to this problem are resampling and linearization methods. Resampling methods comprise the jackknife, balanced repeated replication and the bootstrap methods. The usage of resampling methods is more dependent on the sampling design at hand than on the estimator itself. Their procedures and modifications if applied in the presence of complex survey designs are subject of chapter 3 in Bruch et al. (2011), thus the focus in this report is put on linearization methods.

Linearization methods approximate the non-linear estimator by a linear function after which standard variance formulas for the given design can be applied to this linear approximation. This indirect approach estimates the asymptotic variance of an estimator which results in biased but typically consistent variance estimators (see Wolter, 2007, Chapter 6).

Suppose we have d different study variables and we want to estimate parameter θ of the finite population $\mathcal{U}$ of size N, which has the following form

$$\theta = f\left(\boldsymbol{\tau}\right) \ , \tag{1.1}$$

where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k, \ldots, \tau_d)$ and $\tau_k = \sum_{i \in \mathcal{U}} y_{ki}$, with y_{ki} as the observation of k-th study variable of the i-th element in $\mathcal{U}$. To estimate θ a sample s of size n is drawn from $\mathcal{U}$, according to a sampling design $p(s)$ (see Bruch et al., 2011, Chapter 1). Then we substitute the unknown $\boldsymbol{\tau}$ in (1.1) by $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \ldots, \hat{\tau}_k, \ldots, \hat{\tau}_d)$ which yields

$$\hat{\theta} = f\left(\hat{\boldsymbol{\tau}}\right) \ ,$$

with $\hat{\tau}_k = \sum_{i \in s} y_{ki} w_i$ as the estimated total of the k-th study variable and w_i is the survey weight of the i-th element in s. Further, it is assumed that $\hat{\tau}_k$ is a consistent esimator of τ_k.

First we examine the case of f being a linear function, because we will reduce the problem of estimating the variance of $\hat{\theta}$ if f is non-linear to that of the linear case. If f is a linear function we may write

$$\hat{\theta} = f\left(\hat{\boldsymbol{\tau}}\right) = a_0 + \sum_{k=1}^{d} a_k \hat{\tau}_k \ ,$$

where a_0 and a_k with k = 1, . . . , d are some constants. Then for the variance of $\hat{\theta}$ we have

$$\mathrm{V}(\hat{\theta}) = f\left(\boldsymbol{\tau}\right) = \sum_{k=1}^{d} a_k^2 \mathrm{V}(\hat{\tau}_k) + 2 \sum_{k=1}^{d} \sum_{\substack{l=1 \\ k<l}}^{d} a_k a_l \,\mathrm{Cov}\left(\hat{\tau}_k, \hat{\tau}_l\right) \ . \tag{1.2}$$

Expression (1.2) involves d variances and $d(d-1)2^{-1}$ covariances, that need to be estimated. If $w_i = \pi_i^{-1}$, i.e. the survey weight of the i-th element in s equals its inverse inclusion probability, then we can estimate $\mathrm{V}(\hat{\tau}_k)$ as described in Bruch et al. (2011, chapter 2) and the covariance $\mathrm{Cov}(\hat{\tau}_k, \hat{\tau}_l)$ by

$$\widehat{\mathrm{Cov}}(\hat{\tau}_k, \hat{\tau}_l) = \sum_{i \in s} \sum_{j \in s} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \frac{y_{ki}}{\pi_i} \frac{y_{lj}}{\pi_j} \ . \tag{1.3}$$

We now turn to the case of f being non-linear. Our aim is to use also in this case an expression for $V(\hat{\theta})$ that has the same form as in (1.2), which we know how to estimate. To achieve this the well-known Taylor series method can be employed, which give us a linear approximation of $\hat{\theta}$. The following theorm of the Taylor series can be found in Serfling (1980, p. 43f).

Let the function g defined on $\mathbb{R}^d$ possess continuous partial derivatives of order m at each point of an open set $\mathbb{S} \subset \mathbb{R}^d$. Let $\mathbf{x} \in \mathbb{S}$. For each point $\mathbf{y}$, $\mathbf{y} \neq \mathbf{x}$, such that the line segment $L(\mathbf{x}, \mathbf{y})$ joining $\mathbf{x}$ and $\mathbf{y}$ lies in $\mathbb{S}$, there exists a point $\mathbf{u}$ in the interior of $L(\mathbf{x}, \mathbf{y})$ such that

$$\begin{aligned} g(\mathbf{y}) =& g(\mathbf{x}) \\ &+ \sum_{k=1}^{m-1} \frac{1}{k!} \sum_{l_1=1}^{d} \cdots \sum_{l_k=1}^{d} \left[\frac{\partial^k g(p_1, \ldots, p_d)}{\partial p_{l_1} \ldots \partial p_{l_k}} \right]_{\mathbf{p}=\mathbf{x}} (y_{l_1} - x_{l_1}) \ldots (y_{l_k} - x_{l_k}) \\ &+ \frac{1}{m!} \sum_{l_1=1}^{d} \cdots \sum_{l_m=1}^{d} \left[\frac{\partial^m g(p_1, \ldots, p_d)}{\partial p_{l_1} \ldots \partial p_{l_m}} \right]_{\mathbf{p}=\mathbf{u}} (y_{l_1} - x_{l_1}) \ldots (y_{l_m} - x_{l_m}) \end{aligned} \tag{1.4}$$

where $\left[\frac{\partial^k g(p_1, \ldots, p_d)}{\partial p_1 \ldots \partial p_k} \right]_{\mathbf{p}=\mathbf{x}}$ is the k-th partial derivative of $g(\mathbf{p})$ with respect to $p_1, p_2, \ldots$, and p_k evaluated at $\mathbf{p} = \mathbf{x}$.

If the function f is continuously differentiable up to order two at each point in the open set $\mathbb{S}$ containing $\boldsymbol{\tau}$ and $\hat{\boldsymbol{\tau}}$, then we can use the expansion in (1.4) and write

$$\hat{\theta} - \theta = \sum_{k=1}^{d} \left[\frac{\partial f(p_1, \ldots, p_d)}{\partial \tau_k} \right]_{\mathbf{p}=\boldsymbol{\tau}} (\hat{\tau}_k - \tau_k) + R(\hat{\boldsymbol{\tau}}, \boldsymbol{\tau}) \ , \tag{1.5}$$

where

$$R(\hat{\boldsymbol{\tau}},\boldsymbol{\tau}) = \frac{1}{2!}\sum_{k=1}^{d}\sum_{l=1}^{d}\left[\frac{\partial^2 f(p_1,\ldots,p_d)}{\partial p_k \partial p_l}\right]_{\mathbf{p}=\ddot{\boldsymbol{\tau}}}(\hat{\tau}_k-\tau_k)(\hat{\tau}_l-\tau_l)$$

and $\ddot{\boldsymbol{\tau}}$ is in the interior of line segment $L(\boldsymbol{\tau},\hat{\boldsymbol{\tau}})$ joining $\boldsymbol{\tau}$ and $\hat{\boldsymbol{\tau}}$. For the remainder term R we have $R = O_p(r_n^2)$, where $r_n \to 0$ as $n \to \infty$. Further, we have $\hat{\theta}-\theta = O_p(r_n)$, thus, in most applications it is common practice to regard R as negligible in (1.5) for sample sizes large enough (see WOLTER, 2007, p. 227ff). This justifies the use of the following approximation

$$\hat{\theta}-\theta \approx \sum_{k=1}^{d}\left[\frac{\partial f(p_1,\ldots,p_d)}{\partial \tau_k}\right]_{\mathbf{p}=\boldsymbol{\tau}}(\hat{\tau}_k-\tau_k)\ , \tag{1.6}$$

note that in expression (1.6) only the linear part of the Taylor series is kept. Now we can use (1.6) to derive an approximation of the mean square error (MSE) of $\hat{\theta}$ which is given by

$$\begin{aligned}\mathrm{MSE}\left(\hat{\theta}\right) &\approx \mathrm{V}\left(\sum_{k=1}^{d}\left[\frac{\partial f(p_1,\ldots,p_d)}{\partial \tau_k}\right]_{\mathbf{p}=\boldsymbol{\tau}}\hat{\tau}_k\right) \\ &= \sum_{k=1}^{d} a_k^2 \mathrm{V}(\hat{\tau}_k) + 2\sum_{k=1}^{d}\sum_{\substack{l=1\\k<l}}^{d} a_k a_l \,\mathrm{Cov}\left(\hat{\tau}_k,\hat{\tau}_l\right)\ ,\end{aligned} \tag{1.7}$$

where $a_k = [\frac{\partial f(p_1,\ldots,p_d)}{\partial \tau_k}]_{\mathbf{p}=\boldsymbol{\tau}}$. Because, $\mathrm{MSE}(\hat{\theta}) = \mathrm{V}(\hat{\theta}) + \mathrm{Bias}(\hat{\theta})^2$, where $\mathrm{Bias}(\hat{\theta}) = \hat{\theta}-\theta$, we can approximate the variance of $\hat{\theta}$ by $\mathrm{MSE}(\hat{\theta})$, since $\mathrm{V}(\hat{\theta})$ is of higher order then $\mathrm{Bias}(\hat{\theta})^2$ for unbiased or at least consistent estimators (see WOLTER, 2007, p. 232). Thus, we can use (1.7) as an approximation to the design variance of $\hat{\theta}$.

To estimate (1.7) we could simply subsitute the variances and covariances with their corresponding estimates. This however might become unpractical if d the number of estimated totals in $\hat{\boldsymbol{\tau}}$ becomes large. To evade this problem WOODRUFF (1971) suggested the following transformation of y_{ki}

$$z_i = \sum_{k=1}^{d} a_k y_{ki}\ , \tag{1.8}$$

then we may write

$$\mathrm{MSE}\left(\hat{\theta}\right) \approx \mathrm{V}\left(\sum_{i\in s} w_i z_i\right)\ .$$

Transformation (1.8) is justified if the elementary estimates in $\boldsymbol{\tau}$ are linear, as it is the case for $\hat{\tau}_k$, then it is possible to change the order of summation in (1.7) (see ANDERSSON and NORDBERG, 1994). To estimate $\mathrm{MSE}(\hat{\theta})$ we need to replace the unknown a_k in (1.8) by an estimator $\hat{a}_k = [\frac{\partial f(p_1,\ldots,p_d)}{\partial \tau_k}]_{\mathbf{p}=\hat{\boldsymbol{\tau}}}$ which yields $\hat{z}_i$ the sample estimate of z_i. Finally we can estimate the asymptotic variance of $\hat{\theta}$ by $\hat{\mathrm{V}}(\sum_{i\in s} w_i \hat{z}_i)$, which has in case of $w_i = \pi_i^{-1}$ a well known solution for most survey designs used in practice (see BRUCH et al., 2011). Thus, it is possible to reduce the problem of estimating the (asymptotic) variance of a non-linear estimator to the one of estimating the variance of a single estimated total.

Chapter 2

Variance Estimation of Non-linear Statistics

2.1 Influence Functions

For variance estimation the Taylor method leads to proper results for statistics which can be expressed by functions which are continuously differentiable up to order two and are asymptotically normal. For statistics which do not meet these requirements, one approach however, is the concept of influence functions, invented by Hampel (1974), which up to now is widely used in the field of robust statistics. Heuristically spoken, an influence function gives a picture of the infinitesimal behaviour of the asymptotic value of a statistic. In other words, it measures the asymptotic bias caused by contamination in the observations on whose basis the statistic is estimated.

In the context of this approach, first, the introduction of the concept of *statistical functionals* is advisable. Most statistics of interest are parameters of an underlying cumulative distribution function (cdf) F and most characteristics of F can be written as $T(F)$, where T is a functional from $\mathcal{F}$ to $\mathbb{R}^d$ with $\mathcal{F}$ as the collection of all cdf's on $\mathbb{R}^d$. Thus, many commonly used estimators can be defined as $T(F_n)$, where F_n is the empirical distribution function based on a sample vector $(y_1, \ldots, y_n)$, i.e.

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}[y_i \leq y] \; .$$

Then estimator $T(F_n)$ is called a *statistical functional*. For example, a sample moment is related to functional $T(F) = \int f(y) dF(y)$ with f as an integrable function, and the corresponding estimator would be $T(F_n) = \int f(y) dF_n(y) = n^{-1} \sum_{i=1}^{n} f(y_i)$ (see Shao, 2003, p. 338).

For the sample vector $(y_1, \ldots, y_n)$, whose elements are independent and identically distributed (iid), with distribution function F, the influence function of a statistic $T = T(F_n)$ at a point y is

$$IF(T, F, y) = \lim_{\epsilon \to 0} \frac{T\left((1-\epsilon)F + \epsilon \delta_y^{\bullet}\right) - T(F)}{\epsilon}, \tag{2.1}$$

where $\delta_y^\bullet = \mathbb{1}_{[y,\infty)}$ is a cumulative distribution function degenerated at point y (see SHAO (2003, p. 339 and 19)).

The derivation of influence functions requires differentials of T. There exist several versions of differentials, mainly the Gâteaux-differential, the d-Hadamard-differential and the d-Fréchet-differential. They are defined in the following (for the following definitions see SHAO, 2003, p. 338f).

For all three definitions, let $T : \mathcal{F}_0 \to \mathbb{R}$ be a real functional on a collection of all absolutely continuous cumulative distribution functions $\mathcal{F}_0$ on $\mathbb{R}^d$ and let $\mathcal{D} := \{c(F - G) : F, G \in \mathcal{F}_0, c \in \mathbb{R}\}$.

$T : \mathcal{F}_0 \to \mathbb{R}$ is Gâteaux-differentiable at $F \in \mathcal{F}_0$ if there is a linear functional $L_F : \mathcal{D} \to \mathbb{R}$ (i.e. $L_F(c_1\Delta_1 + c_2\Delta_2) = c_1 L_F(\Delta_1) + c_2 L_F(\Delta_2)\ \forall\ \Delta_j \in \mathcal{D}, c_j \in \mathbb{R}$) such that for all $\Delta \in \mathcal{D}$ and $F + t\Delta \in \mathcal{F}_0$:

$$\lim_{t\to 0}\left(\frac{T(F + t\Delta) - T(F)}{t} - L_F(\Delta)\right) = 0.$$

Let ϱ be a distance on $\mathcal{F}_0$ that is induced by a norm $\|\cdot\|$ on $\mathcal{D}$ (i.e. $d(F, G) = \|F - G\|$). $T : \mathcal{F}_0 \to \mathbb{R}$ is ϱ-Hadamard-differentiable at $F \in \mathcal{F}_0$ if there is a linear functional $L_F : \mathcal{D} \to \mathbb{R}$ such that for any sequence $t_j \to 0$ $(j \to 0)$ and $\Delta_j, \Delta \in \mathcal{D}$ with $\|\Delta_j - \Delta\| \to 0$, $F + t_j\Delta_j \in \mathcal{F}_0$:

$$\lim_{j\to\infty}\left(\frac{T(F + t_j\Delta_j) - T(F)}{t_j} - L_F(\Delta_j)\right) = 0.$$

Let ϱ be a distance on $\mathcal{F}_0$. $T : \mathcal{F}_0 \to \mathbb{R}$ is ϱ-Fréchet-differentiable in $F \in \mathcal{F}_0$ if there exists a linear functional $L_F : \mathcal{D} \to \mathbb{R}$ such that for all $\{F_j\} \subset \mathcal{F}_0$ with $d(F_j, F) \to 0$ $(j \to \infty)$:

$$\lim_{j\to\infty}\left(\frac{T(F_j) - T(F) - L_F(F_j - F)}{d(F_j, F)}\right) = 0.$$

For all three definitions, the functional L_F is called differential of T at F. Now if we define a function $h : \mathbb{R} \to \mathbb{R} = T(F + t\Delta)$, then Gâteaux-differentiability is equivalent to differentiability of the function h at $t = 0$, hence $L_F(\Delta) = h'(0)$ (see SHAO, 2003, p. 339).

The influence function of $T(F)$ at y is given by $L_F(\delta_y^\bullet - F) = IF(T, F, y)$. Now it is possible to describe the asymptotic behaviour of $T(F_n)$. If T is Gâteaux-differentiable in F then for $t = \frac{1}{\sqrt{n}}$, and $\Delta = \sqrt{n}\,(F_n - F)$

$$\sqrt{n}(T(F_n) - T(F)) = L_F(\sqrt{n}(F_n - F)) + R_n\ , \tag{2.2}$$

with R_n as some stochastic remainder. For $n \to \infty$ by the central limit theorem

$$L_F(\sqrt{n}(F_n - F)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} IF(T, F, y_i) \xrightarrow{d} N(0, \sigma_F^2)\ ,$$

if $E\,(IF(T, F, y_i)) = 0$, and $\sigma_F^2 = E\,(IF(T, F, y_i)^2) < \infty$. Thus, $T(F_n)$ is asymptotically normal if $R_n = o_p(1)$, i.e. $R_n \xrightarrow{p} 0$, which is unfortunately not ensured by Gâteaux-differentiability. Therefore, the ϱ-Hadamard-differentiability or ϱ-Fréchet-differentiability

is required. This means that first, L_F can be found by differentiating $h(t) = T(F + t\Delta)$ at $t = 0$ and then, it is checked whether T is ϱ-Hadamard' or ϱ-Fréchet-differentiable with a given distance ϱ on $\mathcal{F}_0$ (see Shao, 2003, p.340).

Note that in case of sample moments, e.g. for $T(G) = \int f(y)dF(G)$ with $G \in \mathcal{F}$, functional T is linear and therefore ϱ-Fréchet-differentiable of any ϱ. Further, it can be shown that if F is one-dimensional and $F(y)' > 0$ for all $y \in \mathbb{R}$, then the quantile functional $T(G) = G^{-1}$ is ϱ_∞-Hadamard-differentiable at F, where for $F \in \mathcal{F}$

$$\varrho_\infty = \|F_1 - F_2\|_\infty = \sup_{t \in \mathbb{R}^d} |F_1(t) - F_2(t)| ,$$

the distance induced by the sup-norm (see Shao, 2003, p. 321 and p. 341 and also Serfling, 1980, p. 216). Finally, via the total of the influence values $z_i = IF(T, F, y_i)$ it is possible to approximate the variance of an estimator $T(F_n)$.

2.2 Estimating Equations

Another approach to linearization is the usage of estimating equations. Estimating equations is a technique which can be used to derive both point estimates and their corresponding linearized values z_i used for variance estimation (see Binder and Patak, 1994). In particular, Kovacevic and Binder (1997) and Binder and Kovacevic (1993) used the method for estimating disparity and inequality measures and their corresponding sample errors. In the following, a brief overview of the framework of estimating equations will given firstly.

If we consider an infinite population and our variable of interest y has a continuous distribution function $\widetilde{F}(y, \theta)$ and a differentiable density function $\tilde{f}(y, \theta)$ for all $y \in \mathbb{R}$, it is assumed that the parameter of interest θ is the solution, θ_0 to

$$U(\theta) = \int u(y, \theta) d\widetilde{F}(y) = 0 , \tag{2.3}$$

(see Binder and Patak, 1994). If we define $u(y, \theta) = g'(y, \theta)$, and $g(y, \theta) = \log \widetilde{f}(y, \theta)$ then we have

$$u(y, \theta) = \frac{\partial \log \widetilde{f}(y, \theta)}{\partial \theta} ,$$

where u is called estimation function. For instance, for the population mean $\theta = \int y d\widetilde{F}(x)$ we have

$$u(y, \theta) = y - \theta ,$$

(see Binder and Patak, 1994). Whereas for a finite population, where θ can be expressed as function of the N distinct population units, Binder and Kovacevic (1993) describe θ as some solution, θ_0 to the equation

$$U(\theta) = \sum_{i=1}^{N} u(y_i, \theta) = 0 . \tag{2.4}$$

An estimator $\hat{\theta}_0$ of θ_0 is then the solution to the equation

$$\hat{U}(\theta) = \sum_{i \in s} w_i u(y_i, \theta) = 0 \ ,$$

where w_i is a weight for $i \in s$ and zero elsewhere. If we use Horvitz-Thompson type weights then $w_i = \pi_i^{-1}$, with π_i as the probability including the i-th element into the sample s (see BRUCH et al., 2011). The variance of $\hat{\theta}_0$ is estimated in the following way. Let $\hat{U}(\hat{\theta}_0) = \sum_{i=1}^{n} u(y_i, \hat{\theta}_0) = 0$, which can be rewritten as

$$\begin{aligned} \hat{U}(\hat{\theta}_0) = & \sum_{i=1}^{N} \Big(u(y_i, \hat{\theta}_0) - u(y_i, \theta_0) \Big) \\ & + \sum_{i=1}^{N} w_i u(y_i, \theta_0) \\ & + \sum_{i=1}^{N} \Big(u(y_i, \hat{\theta}_0) - u(y_i, \theta_0) \Big) (w_i - 1) \ . \end{aligned} \tag{2.5}$$

As $\hat{\theta}_0$ is a consistent estimator for θ_0, i.e. $\hat{\theta}_0 \xrightarrow{p} \theta_0$ for $n \to \infty$, the last term in (2.5) becomes negligible for large samples (see KOVACEVIC and BINDER, 1997). Furthermore, by using Taylor expansions we have

$$u(y_i, \hat{\theta}_0) - u(y_i, \theta_0) = \sum_{k=1}^{m} \left(\frac{\partial^k u(y_i, \theta)}{\partial^k \theta} \right)_{\theta = \theta_0} \Big(\hat{\theta}_0 - \theta_0 \Big)^m + o(|\hat{\theta}_0 \ - \theta_0|^m) \ ,$$

(see SERFLING, 1980, p. 45). Thus we can approximate the first term in (2.5) by

$$\Big(\hat{\theta}_0 - \theta_0 \Big) \sum_{i=1}^{N} \left(\frac{\partial u(y_i, \theta)}{\partial \theta} \right)_{\theta = \theta_0} + o(|\hat{\theta}_0 \ - \theta_0|) \ ,$$

and we have the following approximation for $\hat{\theta}_0 - \theta_0$

$$\hat{\theta}_0 - \theta_0 \approx \sum_{i=1}^{N} w_i z_i \ , \tag{2.6}$$

where

$$z_i = - \left[\left(\frac{\partial u(y_i, \theta)}{\partial \theta} \right)_{\theta = \theta_0} \right]^{-1} u(y_i, \theta_0) \ .$$

Finally, once the z_i have been obtained, we can approximate the variance of $\hat{\theta}_0$ by the variance of the estimated population total of the z_i, since

$$\mathrm{V}(\hat{\theta}_0 - \theta_0) = \mathrm{V}(\hat{\theta}_0) \approx \mathrm{V} \left(\sum_{i=1}^{N} w_i z_i \right) \ . \tag{2.7}$$

Note that in the following we will refer to z_i also as the linearized value, indicating that they allow for the usage of variance estimators of linear statistics, when estimating the variance of non-linear statistics.

2.3 Linearization of Poverty and Inequality Measures

In this section the influence or linearized values for the different poverty and inequality measures are derived. Therefore we follow first the approach of Deville (1999), where he used influence functions which differ slightly from the one described in equation (2.1). They are based on a measure M of the total mass N, the size of population, and not on the distribution function F (see Deville, 1999, p. 692). Thus, we have

$$IT\left(M, y\right) = \lim_{\epsilon \to \infty} \frac{T\left(M + \epsilon \delta_y^{\circ}\right) - T(M)}{\epsilon} \,, \tag{2.8}$$

where δ_y° is the unit mass at point $y \in \mathbb{R}$ and $M = \sum_{i \in \mathcal{U}} \delta_{y_i}^{\circ}$. For instance, an estimator for the total τ of variable y can be defined through an estimator $\hat{M}$ of M, since $\tau = \sum_{i \in \mathcal{U}} y_i = \int y dM$ is a functional of M (see Goga et al., 2009). A natural estimator of M would be $\hat{M} = \sum_{i=1}^{N} w_i \delta_y^{\circ}$, where $\hat{M}$ is a measure allocating a weight w_i to y_i for all $i \in s$ and zero elsewhere (see Deville, 1999).

For example, the influence function of $R = \frac{\int y dM}{\int x dM} = \frac{\tau_y}{\tau_x}$, the ratio between the total of variables y and x, with $y, x \in \mathbb{R}$, can be obtained by (see *rule 2* in Deville, 1999):

$$\begin{aligned} IT_{\mathrm{R}}\left(M, (y_i, x_i)\right) &= \frac{1}{\tau_x} IT(\tau_y, y_i) + \tau_y IT(\frac{1}{\tau_x}, x_i) \\ &= \frac{y_i}{\tau_x} - R\frac{x_i}{\tau_x} \,. \end{aligned}$$

In practice it is often not necessary to compose an influence function by applying (2.8) directly, which possibly requires rather complex limit calculations. This is due to the fact that there exist rules for deriving influence functions of different statistics which are essentially those for computing the derivative of a differentiable function (see Deville, 1999, p. 197).

2.3.1 The At-risk-of-poverty Rate

This subsection presents the linearization of the at-risk-of-poverty rate (ARPR). The following exposure harkens back to Osier (2009), where the influence functions for the different statistics are derived from the rule of derivation stated by Deville (1999). Since the indicator ARPR is directly connected with the at-risk-of-poverty threshold (ARPT), we firstly derive the influence function of the ARPT and then the influence function of the ARPR. The ARPT is defined as 60% of the median income

$$\mathrm{ARPT} = 0.6\,\mathrm{MED}(M) = T(M) \,,$$

where $\mathrm{MED}(M)$ is the median income and F the cumulative income distribution, $F(M, y) = \frac{1}{N} \sum_{i \in \mathcal{U}} \mathbb{1}(y_i \leq y)$, with y_i as the income of the i-th element in $\mathcal{U}$. By definition $F(M, \mathrm{MED}(M)) = 0.5$, thus the respective influence function $IT_F(M, \mathrm{MED}(M))$ is equal to 0 for all $i \in \mathcal{U}$.

In general, for a functional of the form $F(M, \mathrm{MED}(M))$ one can apply *rule 7* given by Deville (1999) to obtain its influence function. Because of the significance of this rule

in the context of our applications, we will state it here explicitly. If $S(M)$ is a functional in $\mathbb{R}^d$ and $T(M, \lambda(M))$ is a family of functionals with $\lambda(M) \in \mathbb{R}^d$ the influence function of $T_S = T(M, S(M))$ is given by

$$IT_{T_S} = IT_{T_S}(M, S(M)|S(M)fixed) + \left(\frac{\partial T(M)}{\partial \lambda}\right)_{\lambda=S(M)} IT_S(M, y_i) \ . \tag{2.9}$$

Now using (2.9) we may write

$$\begin{aligned} 0 =& IT_F(M, \text{MED}(M)|\text{MED}(M)fixed) \\ &+ \left[\left(\frac{\partial F(M,y)}{\partial y}\right)_{y=\text{MED}(M)}\right] IT_{\text{MED}}(M, y_i) \ . \end{aligned} \tag{2.10}$$

The first term in (2.10) is the influence function of $F(M, MED(M))$ with respect to M holding $MED(M)$ constant. The second term accounts for the influence of MED(M). Thus, the first term in (2.10) is given with

$$IT_F(M, \text{MED}(M)|\text{MED}(M)fixed) = \frac{1}{N}\left(\mathbb{1}[y_i \leq \text{MED}(M)] - 0.5\right) \ .$$

Hence, the influence function of the median is

$$IT_{\text{MED}}\left(M, y_i\right) = -\frac{1}{NF'[\text{MED}(M)]}\left(\mathbb{1}[y_i \leq \text{MED}(M)] - 0.5\right) \ , \tag{2.11}$$

where F' is the derivative of F. It follows from this result that the influence function of the ARPT is given by, $IT_{\text{ARPT}} = 0.6\, IT_{\text{MED}}\left(MED(M), y_i\right)$, which can be written as

$$IT_{\text{ARPT}}\left(M, y_i\right) = -\frac{0.6}{NF'(\text{MED}(M))}\left(\mathbb{1}[y_i \leq \text{MED}(M)] - 0.5\right) \ . \tag{2.12}$$

After these preparations for the ARPT, we can go on to the ARPR. The ARPR is defined as the share of persons with an income below the at-risk-of-poverty threshold ARPT(M),

$$\text{ARPR} = F\left(M, \text{ARPT}(M)\right) \ .$$

Its influence function can be obtained by applying the rule in (2.9), hence

$$\begin{aligned} IT_{\text{ARPR}}\left(M, y_i\right) = \ & IT_F\left(M, \text{ARPT}(M)|\text{ARPT(M)}fixed\right) \\ & + \left[\frac{dF(M,y)}{dy}|_{y=\text{ARPR}(M)}\right] IT_{\text{ARPR}}(M). \end{aligned}$$

Furthermore, the influence function of F with respect to M holding ARPT(M) constant is given by

$$IT_F\left(M, \text{ARPT}(M)|ARPT(M)fixed\right) = \frac{\mathbb{1}[y_i \leq \text{ARPT}(M)] - \text{ARPR}(M)}{N}.$$

In analogy to the influence function of the ARPT, this formula can be expanded to

$$IT_{\text{ARPT}}\left(M, y_i\right) = -\frac{0.6}{F'[\text{MED}(M)]}\left(\frac{\mathbb{1}[y_i \leq \text{MED}(M)] - 0.5}{N}\right).$$

With $F'(y)$ strictly non-negative for all y, we have

$$\begin{aligned} IT_{\text{ARPR}}(M, y_i) =& \frac{1}{N}\left(\mathbb{1}[y_i \leq \text{ARPT}(M)] - \text{ARPR}(M)\right) \\ & - \frac{0.6F'[\text{ARPT}(M)]}{F'[MED(M)]}\left(\frac{\mathbb{1}[y_i \leq \text{MED}(M)] - 0.5}{N}\right) \ . \end{aligned} \tag{2.13}$$

2.3.2 The Relative Median Poverty Gap

The relative median poverty gap (RMPG) is defined as the difference between the ARPT and the median income MED_p of the persons whose income is lower than the ARPT, the difference being expressed relatively to the ARPT:

$$\text{RMPG} = \frac{\text{ARPT} - \text{MED}_p}{\text{ARPT}} = 1 - \frac{\text{MED}_p}{\text{ARPT}} \, .$$

Because the indicators ARPR and RMPG are connected with each other via the ARPT the following explanations also rely on the ones made by Osier (2009).

The influence function of the RMPG can be written as

$$IT_{\text{RMPG}}\,(M, y_i) = -\frac{IT_{\text{MED}_p}(M, y_i)}{\text{ARPT}(M)} - \frac{\text{MED}_p(M)}{\text{ARPT}(M)^2} IT_{\text{ARPT}}(M, y_i). \tag{2.14}$$

Since the influence function of the ARPT has already been derived, the last remaining issue is the derivation of $IT_{\text{MED}_p}(M)$, the influence function of the median income for persons who are below 60% of the median income. By definition $F(M, \text{MED}_p(M)) = \frac{1}{2}F[M, \text{ARPT}(M)]$, thus we can conclude that the corresponding influence functions for all $i \in \mathcal{U}$ is

$$IT_F\,(M, \text{MED}_p(M)) = \frac{1}{2} IT_{\text{ARPR}}(M, y_i) \, ,$$

with $IT_{\text{ARPR}}(M, y_i)$ given in (2.13). Rearranging the above expression yields

$$\begin{aligned} IT_{\text{MED}_p, y_i}(M) =& \frac{1}{2}\left(IT_{\text{ARPR}}(M, y_i)\right) \\ & - \frac{\mathbb{1}[y_i \leq \text{MED}_p(M)] - F(M, \text{MED}_p(M))}{NF'[\text{MED}_p(M)]} \, . \end{aligned} \tag{2.15}$$

Finally, the influence function for the RMPG can be obtained by substituting (2.15) and (2.12) into (2.14).

2.3.3 The Quintile Share Ratio

The quintile share ratio (QSR) is a simple inequality measure. It denotes the ratio between the total income of the richest 20% and the poorest 20% of a population. A formal expression of the QSR would be

$$\begin{aligned} \text{QSR} &= \frac{\sum_{i \in \mathcal{U}} \left(y_i - y_i \cdot \mathbb{1}[y_i \leq q_{0.8}]\right)}{\sum_{i \in \mathcal{U}} y_i \mathbb{1}[y_i \leq q_{0.2}]} \\ &= \frac{\int y dM - \int y \mathbb{1}[y \leq q_{0.8}] dM}{\int y \mathbb{1}[y \leq q_{0.2}] dM} \, , \end{aligned}$$

where $q_\alpha = F^{-1}(\alpha)$ and $\alpha \in [0, 1]$. An alternative is to define the QSR as the ratio between the mean income of richest 20%, μ_{R} and the the poorest 20%, μ_{P}, where

$$\mu_{\text{r}} = \sum_{i \in \mathcal{U}} \left(y_i - y_i \cdot \mathbb{1}[y_i \leq q_{0.8}]\right) \Big/ \sum_{i \in \mathcal{U}} (1 - 0.8) \;\; \text{and}$$

$$\mu_{\mathrm{p}} = \sum_{i\in\mathcal{U}} y_i \mathbb{1}[y_i \leq q_{0.2}] \Big/ \sum_{i\in\mathcal{U}} 0.2 \ ,$$

(see HULLIGER and MÜNNICH, 2006). This may in turn be defined as a function of four totals

$$\mathrm{QSR} = \frac{\mu_{\mathrm{r}}}{\mu_{\mathrm{p}}} = \frac{\tau_1}{\tau_2} \Big/ \frac{\tau_3}{\tau_4} \ , \tag{2.16}$$

where τ_1 and τ_3 correspond to the nominator and τ_2 and τ_4 to the denominator in μ_{r} and μ_{p}, respectively. The influence function for the QSR can now be obtained by first deriving the influence functions of each of the four totals in (2.16). For τ_2 and τ_4 this is straightforward if we assume $\tau_1 = \tau_2 = \int 0.2 dM$, and accordingly $IT_{\tau_k}(M, y_i) = 0.2$, $\forall i \in \mathcal{U}$ and $k \in \{2, 3\}$. Total τ_1 is given by

$$\tau_1 = \int y dM - \int y \mathbb{1}[y \leq q_{0.8}] dM \ .$$

For $\tau = \int y dM$ we have $IT_{\tau}(M, y_i) = y_i$ and by defining $Y(M, y) = \int y_i \mathbb{1}[y_i \leq y] dM$ and using inference rule (2.9) we have

$$\begin{aligned} IT_{Y(q_\alpha)}(M, y_i) =& IT_{Y(q_\alpha)}\left(M, q_\alpha(M) | q_\alpha(M) fixed\right) \\ &+ \left[\frac{dY(M, y)}{dy}|_{y=q_\alpha(M)}\right] IT_{q_\alpha}(M, y_i) \ . \end{aligned} \tag{2.17}$$

From (2.11) it is clear that

$$IT_{q_\alpha}(M, y_i) = -\frac{1}{NF'[q_\alpha(M)]} \left(\mathbb{1}[y_i \leq q_\alpha(M)] - \alpha\right) \tag{2.18}$$

and for the first term in (2.17) OSIER (2009, p. 184f) shows that

$$IT_{Y(q_\alpha)}\left(M, q_\alpha(M) | q_\alpha(M) fixed\right) = y_i \mathbb{1}[y_i \leq q_\alpha(M)] \ . \tag{2.19}$$

Inserting (2.18) and (2.19) into (2.17) yields the influence function of $Y(q_\alpha)$. However, this approach requires the derivatives of two discontinuous step functions F and Y, which can be avoided as LANGEL and TILLÉ (2011) show. They propose the following influence function

$$IT_{Y(q_\alpha)}(M, y_i) = (y_i - q_\alpha)\mathbb{1}[y_i \leq q_\alpha] + \alpha q_\alpha \ , \tag{2.20}$$

(see also HULLIGER and MÜNNICH, 2006). Now we are able to give the influence functions of the four totals in (2.16)

$$\begin{aligned} IT_{\tau_1}(M, y_i) &= y_i - \left((y_i - q_{0.8})\mathbb{1}[y_i \leq q_{0.8}] + 0.8 q_{0.8}\right) \ , \\ IT_{\tau_2}(M, y_i) &= 0.2 \ , \\ IT_{\tau_3}(M, y_i) &= (y_i - q_{0.2})\mathbb{1}[y_i \leq q_{0.2}] + 0.2 q_{0.2} \ , \\ IT_{\tau_4}(M, y_i) &= 0.2 \ , \end{aligned}$$

which are used to compose the influence functions of μ_{r} and μ_{p}

$$\begin{aligned} IT_{\mu_{\mathrm{r}}}(M, y_i) &= \left(IT_{\tau_1}(M, y_i) - \mu_{\mathrm{r}} 0.2\right) \frac{1}{\tau_2} \ , \\ IT_{\mu_{\mathrm{p}}}(M, y_i) &= \left(IT_{\tau_3}(M, y_i) - \mu_{\mathrm{p}} 0.2\right) \frac{1}{\tau_4} \ . \end{aligned}$$

Finally, we have

$$IT_{\mathrm{QSR}} = \left(IT_{\mu_{\mathrm{r}}}(M, y_i) - QSR \cdot IT_{\mu_{\mathrm{p}}}(M, y_i)\right) \frac{\tau_4}{\tau_3} \ .$$

2.3.4 The Gini Coefficient

The last indicator which we consider is the Gini coefficient. However, to linearize the Gini coefficient we use the estimating equations approach, in contrast to the influence function which we used until now. The Gini coefficient (GINI) can be defined as

$$\text{GINI} = \frac{1}{\mu} \int_0^\infty (2F(y) - 1)\, y dF(y) \; . \tag{2.21}$$

The estimation function for the GINI is given by

$$u(y, \text{GINI}) = (2F(y) - 1)\, y - \text{GINI}\, y \; .$$

Let J be a function with J(p) = 2p-1, then we can write the estimation equation

$$\begin{aligned}
\hat{U}(\widehat{\text{GINI}}) &= \int \Big(J[\hat{F}(y)]y - Gy \Big)\, d\hat{F} \\
&\approx \int \Big(J[\hat{F}(y)] - J[F(y)] \Big)\, y dF(y) - (\widehat{\text{GINI}}) \int y dF(y) \\
&\quad + \int \Big(J[\hat{F}(y)]y - \text{GINI} y \Big)\, d\hat{F}(y) \; ,
\end{aligned}$$

where $\hat{F}(y) = \sum_{i=1}^n w_i \frac{\mathbb{1}[y_i \leq y]}{\sum_{i=1}^n w_i}$. Using the approximation

$$\begin{aligned}
\Big(J[\hat{F}(y)] - J[F(y)] \Big)\, y dF(y) &\approx \int \Big(\hat{F}(y) - F(y) \Big)\, J'[F(y)] y dF(y) \\
&= \int \hat{F}(y) J'[F(y)] y dF(y) + \text{E}\left(F(y) J'[F(y)] y \right) \; ,
\end{aligned}$$

and letting

$$\begin{aligned}
\int \hat{F}(y) J'[F(y)] y dF(y) &= \int \int_0^y J'[F(y)] y d\hat{F}(x) dF(y) \\
&= \int \left(\int_y^\infty J'[F(x)] x dF(x) \right) dF(y) \; ,
\end{aligned}$$

we see that

$$\widehat{\text{GINI}} - \text{GINI} \approx \int z_i d\hat{F}(y) \; , \tag{2.22}$$

with

$$z_i = \frac{1}{\int y dF(y)} \left(\int_{F(y)}^1 J'[p] F^{-1} dp + J[F(y)]y - \text{GINI} y - \text{E}\left(F(y) J'[F(y)] y \right) \right) \; .$$

Using the sum notation, z_i equals

$$z_i = \frac{1}{\overline{y}N}\left(\sum_{i=1}^{N} 2\frac{\mathbb{1}(y_i \geq y)}{N} + (2F(y_i) - 1 - \text{GINI})y_i - \frac{2F(y_i)y_i}{N}\right) ,$$

with $\text{GINI} = \frac{1}{\overline{y}N}\sum_{i=1}^{N}(2F(y_i)-1)\,y_i$ and $\overline{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$, thus we can write $\frac{2F(y_i)y_i}{N} = \overline{y}(\text{GINI}+1)$, which leads to

$$z_i = \frac{2}{\overline{y}N}\left(\sum_{i=1}^{N}\frac{\mathbb{1}(y_i \geq y)}{N} + \left(F(y_i) - \frac{\text{GINI}+1}{2}\right)y_i - \frac{\overline{y}}{2}(\text{GINI}+1)\right) .$$

The same result can be found in KOVACEVIC and BINDER (1997) but with the use of a more general and flexible framework.

2.4 Variance Estimators for Poverty and Inequality Measures

After deriving in section 2.3 the influence function or linearized values z_i for the indicators ARPR, RMPG, QSR, and GINI, this section gives an oveview on how to estimate the variance for this indicators. We consider variance estimator of the following kind

$$\hat{\text{V}}\left(\sum_{i\in s} w_i z_i\right) , \tag{2.23}$$

where the exact form of $\hat{\text{V}}(.)$ depends on the sampling design but it has a well known solution for most sampling designs used in practice. The case of $w_i = \pi_i^{-1}$ is treated in BRUCH et al. (2011, chapter 2). If the survey weights w_i are calibrated on some auxiliary information,

$$\hat{\text{V}}\left(\sum_{i\in s} w_i e_i\right) , \tag{2.24}$$

where $e_i = z_i - \mathbf{x}_i^T\beta$, further $\mathbf{x}_i$ is a vector of auxiliary variables observed for the i-th sampling element and β is the vector of regression coefficients from the regression of z on $\mathbf{x}$ (on the use of calibration weights see DEVILLE and SÄRNDAL, 1992; SÄRNDAL, 2007). Variance estimator (2.24), which can also be derived through linearization, is well known for estimating the variance for the general regression or calibration estimators. But DEVILLE (1999) noted that if a non-linear statistic θ, that allows for the derivation of a linearized value z_i, is estimated using calibrated weights w_i, then e_i is the linearized value for $\hat{\theta}$.

However, before it is possible to compute either (2.23) or (2.24), the z_i's or e_i's have to be replaced by sample estimates $\hat{z}_i$ or $\hat{e}_i$, because they involve some unknown quantities. Table 2.1 gives an overview on the point estimators and their corresponding $\hat{z}_i$ values. To

obtain the sample residuals $\hat{e}_i = \hat{z}_i - \mathbf{x}_i^T \hat{\beta}$ an estimator $\hat{\beta}$ for β has to be calculated, which can be done by a weighted least square estimator (see Deville and Särndal, 1992).

Further, as table 2.1 shows, for some of the indicators we need $\hat{F}'$ an estimator of the derivative of the estimated cumulative distribution function $\hat{F}$, which is unfortunately in this case always 0 or not defined. This problem can be circumvented by using $\widetilde{\hat{F}}'$ instead, which is the derivative of $\widetilde{\hat{F}}$, a smooth function of $\hat{F}'$. We suggest to use kernel density estimators, with a Gaussian kernel, then we have for all $y \in \mathbb{R}$:

$$\widetilde{\hat{F}}' = \frac{1}{\hat{N} h \sqrt{2\pi}} \sum_{i \in s} w_i \exp\left[-\frac{(y - y_i)^2}{2h^2} \right] , \tag{2.25}$$

which is strictly non-negative. The choice of bandwith parameter h in (2.25) is crucial for the accuracy of $\widetilde{\hat{F}}'(y)$ in terms of its mean integrated squared error

$$\mathrm{E} \int (\widetilde{\hat{F}}' - F'(y))^2 \, dy \, .$$

Traditionally the choice of smoothing parameter, including the bandwidth h, remains with the applicant who can examine the outcome with graphical tools. However, various ways of data-based choices of h have been proposed in the literature. See e.g. Jones et al. (1996) for an overview of these methods. There are some data-based bandwidth selectors for gaussian kernels implemented in `R` (cf. R Development Core Team, 2010), see for instance the function `bw.nrd`.

Table 2.1: Summary of Point Estimators and Linearized Values for Variance Estimation

Indicator θ	Point Estimator $\hat{\theta}$	Estimated linearized value ${}_{\hat{\theta}}\hat{z}_i$
ARPT	$0.6\,\hat{F}^{-1}(0.5)$	$-\frac{0.6}{\hat{N}\hat{F}'(\hat{F}^{-1}(0.5))}\left(\mathbb{1}[y_i \leq \hat{F}^{-1}(0.5)] - 0.5\right)$
ARPR	$\hat{F}(\widehat{\text{ARPT}})$	$\frac{1}{N}\left(\mathbb{1}[y_i \leq \widehat{\text{ARPT}}] - \widehat{\text{ARPT}}\right) - \frac{0.6\hat{F}'[\widehat{\text{ARPR}}]}{\hat{F}'(\hat{F}^{-1}(0.5))}\left(\frac{\mathbb{1}[y_i \leq \hat{F}^{-1}(0.5)] - 0.5}{\hat{N}}\right)$
MED_p	$\hat{F}^{-1}(0.5\,\widehat{\text{ARPR}})$	$\frac{1}{2}\left[{}_{\widehat{\text{ARPR}}}\hat{z}_i\right] - \frac{\mathbb{1}[y_i \leq \widehat{\text{MED}_p}] - F(\widehat{\text{MED}_p})}{\hat{N}F'(\widehat{\text{MED}_p})}$
RMPG	$1 - \frac{\widehat{\text{MED}_p}}{\widehat{\text{ARPT}}}$	$-\frac{\widehat{\text{ARPT}}\,{}_{\widehat{\text{MED}_p}}\hat{z}_i - \widehat{\text{MED}_p}\,{}_{\widehat{\text{ARPT}}}\hat{z}_i}{\widehat{\text{ARPT}}^2}$
QSR	$\frac{\hat{\mu}_\text{r}}{\hat{\mu}_\text{p}}$	$\frac{y_i - \left[(y_i - \hat{F}^{-1}(0.8))\mathbb{1}[y_i \leq \hat{F}^{-1}(0.8)] + 0.8\hat{F}^{-1}(0.8)\right]}{\sum_{i\in s} w_i y_i \mathbb{1}[y_i \leq \hat{F}^{-1}(0.2)]} - \frac{\widehat{\text{QSR}}\left[(y_i - \hat{F}^{-1}(0.2))\mathbb{1}[y_i \leq \hat{F}^{-1}(0.2)] + 0.2\hat{F}^{-1}(0.2)\right]}{\sum_{i\in s} w_i y_i \mathbb{1}[y_i \leq \hat{F}^{-1}(0.2)]}$
GINI	$\frac{1}{\hat{N}\hat{\mu}}\sum_{i\in s} w_i(2\,\hat{F}(y_i) - 1)y_i$	$\frac{2\left[\sum_{i\in s}\frac{w_i\mathbb{1}(y_i \geq y)}{\hat{N}} + \left(\hat{F}(y_i) - \frac{\widehat{\text{GINI}}+1}{2}\right)y_i\right]}{\overline{y}N} - \frac{\widehat{\text{GINI}}+1}{\hat{N}}$

$$\hat{F}(y) = \sum_{i\in s} w_i \mathbb{1}(y_i \leq y)(\sum_{i\in s} w_i)^{-1} \qquad \hat{\mu}_\text{p} = \sum_{i\in s} w_i y_i \mathbb{1}[y_i \leq \hat{F}^{-1}(0.2)]\frac{1}{\sum_{i\in s} w_i 0.2}$$

$$\hat{F}^{-1}(p) = \inf\left\{y \in \mathbb{R} : p \leq \hat{F}(y)\right\} \qquad \hat{\mu}_\text{r} = \sum_{i\in s} w_i\left(y_i - y_i \cdot \mathbb{1}[y_i \leq \hat{F}^{-1}(0.8)]\right)\frac{1}{\sum_{i\in s} w_i(1-0.8)}$$

$$\hat{N} = \sum_{i\in s} w_i$$

Chapter 3

Measures of a Change in Indicators of Poverty and Social Exclusion

3.1 Introduction

This chapter deals with the topic of estimating a change in time between measurements of indicators for poverty and social cohesion. The consideration of this topic is linked to one of the main purposes of the EU-SILC as a tool to monitor the process towards agreed policy goals. Indicator values, which are annually published, are usually estimated from the cross-sectional data sets of the EU-SILC survey. However, the EU-SILC survey also has a longitudinal aspect, as it is set up as a rotational sampling scheme. This permits the analysis of individual level changes over time, but it also introduces a dependency between estimated indicator values.

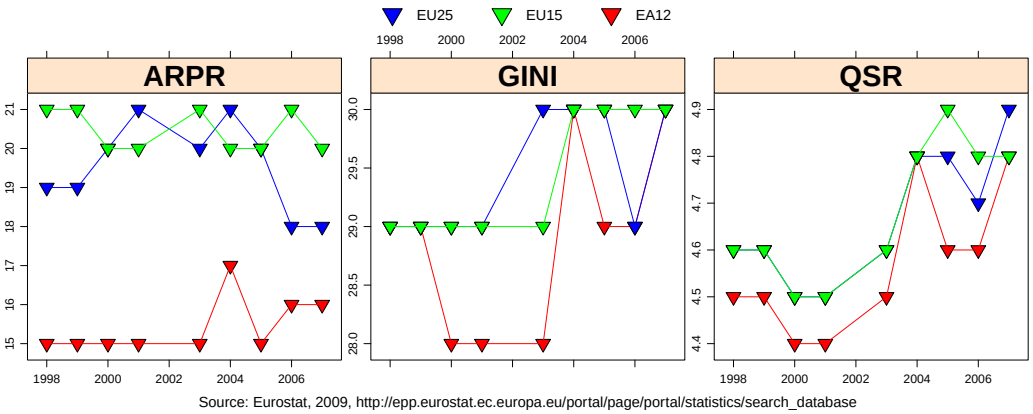

Figure 3.1: Values of Indicators ARPR, GINI and QSR from 1998 to 2007

Figure 3.1 shows the development of the poverty indicator ARPR and the income inequality measures GINI and QSR for the groups of the EU25, EU15, and EA12 (EU25: EU member states in 2004, EU15: EU member states in 1995, EA12: Euro area from 2001 - 2006). Reading the point estimator naively may lead to an over-interpretation of the data. The question is to what extent observed changes over time are due to actual differences

in the indicator values and not sampling errors. Thus, rather than only observing the absolute change in the measurement of an indicator from one time point to another it is of interest to test whether the observed changes are significant in the statistical sense. For this reason, variance estimation for measures of change is required. This would allow to conduct a statistical test on which basis a null hypothesis, which for instance claims that no change has occurred at all, may be rejected or not on a certain confidence level.

There are mainly two problems which arise when trying to estimate the variance for measures of change in two indicator estimates.

1. The statistics in question are highly non-linear. Thus, as it has been outlined in chapter 2, basic variance estimation formulas cannot be applied directly.

2. The EU-SILC survey used to estimate the indicator values has a longitudinal aspect. Hence, correlation through time between indicators has to be taken into account.

This means that the covariance between two non-linear, and non-smooth statistics, has to be estimated. This chapter is structured according to these tasks. Section 3.2 introduces a framework for two-dimensional samples. Section 3.2.1 presents a methodology for estimating the covariance between cross-sectional estimates obtained from overlapping samples which focuses more on design aspects. Section 3.2.2 gives a brief description on how the method of linearization can also be used to approximate the variance of measures of change between poverty and inequality indicators. Finally, section 3.2.3 presents some results from a simulation study based on the EU-SILC longitudinal data set.

3.2 Variance Estimation in Overlapping Samples

In our framework change is estimated by differences in cross-sectional estimates, hence there is a need to estimate the covariance between estimators which are based on samples with a partial overlap. Therefore, we first extend our framework of sampling on one occasion to a two-dimensional sampling design. Let y_0 and y_1 be variables of interest measured on two different occasions by samples s_0 and s_1, respectively. Samples s_0 and s_1 are selected by sampling designs $p_0(.)$ and $p_1(.)$ from population $\mathcal{U}$, which for now is assumed to be composed of the same elements $\{1, \ldots, i, \ldots, N\}$ at both occasions. Both designs $p_0(.)$ and $p_1(.)$ are of fixed size n_0 and n_1, respectively, and without replacement (cf. BRUCH et al., 2011, Section 2.1). Further let the matched sample $s_{01} = s_0 \cap s_1$ be non-empty and of size n_{01}. In additon, we denote $s_{0\backslash 1} = s_0 \backslash s_1$ and $s_{1\backslash 0} = s_1 \backslash s_0$. Thus, we have three disjoint samples $s_{0\backslash 1}$, s_{01}, and $s_{1\backslash 0}$ which are of size $n_{0\backslash 1}$, n_{01} and $n_{1\backslash 0}$, respectively (cf. GOGA et al., 2009, p. 694).

A two-dimensional sampling design can then be defined as a probability distribution function $p(.)$ on selecting a sample $\mathbf{s} = (s_0, s_1) \in \{\mathcal{P}(\mathcal{U})^2, s_0 \subset \mathcal{P}(\mathcal{U}), s_1 \subset \mathcal{P}(U)\}$ (cf. GOGA et al., 2009, p. 696), satisfying the following properties:

$$\sum_{\mathbf{s} \in \{\mathcal{P}(\mathcal{U})^2\}} p(\mathbf{s}) = 1 \qquad \text{and} \qquad p(\mathbf{s}) \geq 0 \quad \forall\, \mathbf{s} \in \mathcal{P}(\mathcal{U})^2 \qquad .$$

For a more general approach to multidimensional sampling see Goga (2003), Chapter 2. Now we can distinguish between cross-sectional inclusion probabilities, which give the probabilities of an element being included in one of the samples s_0 or s_1 and longitudinal inclusion probabilities, which give the probabilities of elements of being jointly in s_0 and s_1.

Let π_{0i} and π_{1j} be the probabilities of including the i-th and j-th element in sample s_0 and s_1, respectively. Further, let π_{0ij} and π_{1ij} denote the probabilities of including element i and j jointly in sample s_0, and elements i and j jointly in sample s_1, respectively. For the longitudinal inclusion probabilities, let π_{01i} denote the probability for including element i both in sample s_0 and s_1 and π_{01ij} denote the joint probability of including the i-th element in sample s_0 and the j-th element in s_1 (cf. Tam, 1984).

A simple example of a two-dimensional sampling design can be made by assuming that samples $s_{0\setminus 1}$, s_{01}, $s_{1\setminus 0}$ are selected by a simple random sample (SRS) without replacement (WoR). Such a design could be implemented by first selecting sample s_0 by SRS WoR from $\mathcal{U}$ and sample s_{01} from s_0 also by SRS WoR. Then $s_{1\setminus 0}$ is selecting from $\mathcal{U}\setminus s_0$ through SRS WoR (see sampling plan A, Tam, 1984). The inclusion probabilities for such a design are given by

$$\pi_{tij} = \begin{cases} \frac{n_0}{N} & i = j \wedge i, j \leq N, t = 0 \\ \frac{n_0 \cdot (n_0 - 1)}{N \cdot (N-1)} & i \neq j \wedge i, j \leq N, t = 0 \\ \\ \frac{n_1}{N} & i = j \wedge i, j \leq N, t = 1 \\ \frac{n_1 \cdot (n_1 - 1)}{N \cdot (N-1)} & i \neq j \wedge i, j \leq N, t = 1 \end{cases}, \tag{3.1}$$

and

$$\pi_{01ij} = \begin{cases} \frac{n_{01}}{N} & i = j \wedge i, j \leq N \\ \frac{n_0 n_1 - n_{01}}{N \cdot (N-1)} & i \neq j \wedge i, j \leq N, \end{cases}, \tag{3.2}$$

(cf. Tam, 1984).

3.2.1 Variance Estimation of Measures of a Change

A straightforward measure of change from time $t = 0$ to $t = 1$ can be a difference or the ratio between two estimators θ_0 and θ_1, where θ_0 corresponds to the characteristic of interest at base time $t = 0$ and θ_1 the value of the same characteristic but at time $t = 1$. If we wish to estimate the absolute change

$$\theta_\Delta = \theta_1 - \theta_0 \tag{3.3}$$

a natural estimator would be

$$\hat{\theta}_\Delta = \hat{\theta}_1 - \hat{\theta}_0 \, . \tag{3.4}$$

The variance of $\hat{\theta}_\Delta$ is given by

$$\mathrm{V}(\theta_\Delta) = \mathrm{V}(\hat{\theta}_1) + \mathrm{V}(\hat{\theta}_0) - 2\mathrm{Cov}(\hat{\theta}_0\,, \hat{\theta}_1)\;. \tag{3.5}$$

If we aim to estimate the difference between totals of study variable y_0 and y_1, i.e. $\theta_1 = \tau_1 = \sum_{i\in\mathcal{U}} y_{1i}$ and $\theta_0 = \tau_0 = \sum_{i\in\mathcal{U}} y_{0i}$, estimator (3.4) would be calculated through cross-sectional estimators

$$\hat{\theta}_1 = \hat{\tau}_1 = \sum_{i\in s_1} \frac{y_{1i}}{\pi_{1i}} \quad \text{and} \quad \hat{\theta}_0 = \hat{\tau}_0 = \sum_{i\in s_0} \frac{y_{0i}}{\pi_{0i}}\;, \tag{3.6}$$

thus we have

$$\hat{\tau}_\Delta = \hat{\tau}_1 - \hat{\tau}_0\;. \tag{3.7}$$

The variance of estimator (3.7) is given by

$$\begin{aligned}
\mathrm{V}(\hat{\tau}_\Delta) =& \mathrm{V}(\hat{\tau}_1) + \mathrm{V}(\hat{\tau}_0) - 2\mathrm{Cov}(\hat{\tau}_0\,, \hat{\tau}_1) && \text{(3.8a)}\\
=& \sum_{i\in\mathcal{U}} \pi_{1i}(1-\pi_{1i}) \left(\frac{y_i}{\pi_{1i}}\right)^2 + \sum_{i\in\mathcal{U}} \sum_{\substack{j\in\mathcal{U}\\ j\neq i}} \left(\frac{\pi_{1ij}}{\pi_{1i}\pi_{1j}} - 1\right) y_{1i}y_{1j} && \text{(3.8b)}\\
&+ \sum_{i\in\mathcal{U}} \pi_{0i}(1-\pi_{0i}) \left(\frac{y_i}{\pi_{0i}}\right)^2 + \sum_{i\in\mathcal{U}} \sum_{\substack{j\in\mathcal{U}\\ j\neq i}} \left(\frac{\pi_{0ij}}{\pi_{1i}\pi_{1j}} - 1\right) y_{0i}y_{0j} && \text{(3.8c)}\\
&-2\left[\sum_{i\in\mathcal{U}} \left(\frac{\pi_{01i}}{\pi_{0i}\pi_{1i}} - 1\right) y_{0i}y_{1i} - \sum_{i\in\mathcal{U}} \sum_{\substack{j\in\mathcal{U}\\ j\neq i}} \left(\frac{\pi_{01ij}}{\pi_{0i}\pi_{1j}} - 1\right) y_{0i}y_{1j}\right]\;. && \text{(3.8d)}
\end{aligned}$$

Terms (3.8b) and (3.8c) correspond to the variance of the Horwitz-Thompson estimators $\hat{\tau}_1$, and $\hat{\tau}_0$, respectively (see SÄRNDAL et al., 1992, p. 44). However, term (3.8d) equals $2\mathrm{Cov}(\hat{\tau}_0, \hat{\tau}_1) = 2\mathrm{E}\left(\hat{\tau}_0 - \tau_0\right)\left(\hat{\tau}_1 - \tau_1\right)$, two times the covariance between estimators $\hat{\tau}_0$ and $\hat{\tau}_1$. Term (3.8d) can be directly derived form TAM (1984), where he gives the covariance between two sample means.

For designs, where samples s_0, s_{01} and s_1 have been selected by SRS WoR ,QUALITÉ and TILLÉ (2008) gave, conditional on n_{01}, an unbiased variance estimator for (3.8a) of the following form:

$$\begin{aligned}
\hat{\mathrm{V}}(\hat{\tau}_\Delta) =& N^2\left(1 - \frac{n_1}{N}\right)\frac{\hat{\sigma}^2_{y1}}{n_1}\\
&+N^2\left(1 - \frac{n_0}{N}\right)\frac{\hat{\sigma}^2_{y0}}{n_0}\\
&-2N^2\left(1 - \frac{n_0 n_1}{N\mathrm{E}(n_{01})}\right)\frac{\hat{\sigma}^2_{y01}\mathrm{E}(n_{01})}{n_0 n_1}\;,
\end{aligned} \tag{3.9}$$

where

$$\hat{\sigma}^2_{y1} = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(y_{1i} - \frac{1}{n_1}\sum_{j=1}^{n_1} y_{1j})^2$$

$$\hat{\sigma}^2_{y0} = \frac{1}{n_0 - 1}\sum_{i=1}^{n_0}(y_{0i} - \frac{1}{n_0}\sum_{j=1}^{n_0} y_{0j})^2$$
$$\hat{\sigma}^2_{y01} = \frac{1}{n_{01} - 1}\sum_{i=1}^{n_{01}}(y_{0i} - \frac{1}{n_{01}}\sum_{j=1}^{n_{01}} y_{0j})(y_{1i} - \frac{1}{n_{01}}\cdot\sum_{j=1}^{n_{01}} y_{1j})$$

Variance estimator (3.9) can be applied to designs where $\mathrm{E}(n_{01})$ depends only on n_1, n_0 and N or if n_{01} is not a random number (e.g. sampling plan A in Tam, 1984). It may be interesting to note that if $s_{01} = \emptyset$, i.e there is no overlap between samples s_0 and s_1, $\mathrm{Cov}(\hat{\tau}_0\,, \hat{\tau}_1) = -N\sigma_{y01}$, with $\sigma_{y01} = \frac{1}{N-1}\sum_{i=1}^{N}(y_{0i} - \frac{1}{N}\sum_{j=1}^{N} y_{1j})(y_{1i} - \frac{1}{N}\sum_{j=1}^{N} y_{1j})$. Thus, in this case a positive correlation between τ_1 and τ_0 will inflate the variance of $\hat{\tau}_\Delta$.

Another approach was given by Berger (2004), which allows for a more general form of estimator (3.4) by assuming that θ_1 and θ_0 are functions of a finite number of totals, that is $\theta_0 = f\,(\tau_{01},\ \ldots,\ \tau_{0P_0})$ and $\theta_1 = \mathrm{f}\,(\tau_{11},\ \ldots,\ \tau_{1P_1})$ (see also Nordberg, 2000). The variance of estimator (3.4) can be approximated through its first order Taylor series:

$$\hat{\theta}_\Delta - \theta_\Delta \approx \sum_{k=1}^{P_0+P_1} \frac{\partial \mathrm{f}(\tau)}{\partial \tau_k}\,(\hat{\tau}_k - \tau_k)\ , \tag{3.10}$$

where $\boldsymbol{\tau} = (\tau_{01},\ \ldots,\ \tau_{0P_0},\ \tau_{11},\ \ldots,\ \tau_{1P_1})$ (see Andersson and Nordberg, 1994). Then the variance of (3.4) is approximated by

$$\mathrm{V}(\hat{\theta}_\Delta) \approx \sum_{k=1}^{P_0+P_1}\left(\frac{\partial f(\boldsymbol{\tau})}{\partial \tau_k}\right)^2 \mathrm{V}(\hat{\tau}_k) + \sum_{k=1}^{P_0+P_1}\sum_{\substack{l=1\\k\neq l}}^{P_0+P_1} \frac{\partial f(\boldsymbol{\tau})}{\partial \tau_j}\frac{\partial f(\boldsymbol{\tau})}{\partial \tau_l}\,\mathrm{Cov}(\hat{\tau}_k\,, \hat{\tau}_l)\ , \tag{3.11}$$

which can be estimated by

$$\widehat{\mathrm{V}}(\hat{\theta}_\Delta) \approx \sum_{k=1}^{P_0+P_1}\left(\frac{\partial f(\hat{\boldsymbol{\tau}})}{\partial \hat{\tau}_k}\right)^2 \widehat{\mathrm{V}}(\hat{\tau}_k) + \sum_{k=1}^{P_0+P_1}\sum_{\substack{l=1\\k\neq l}}^{P_0+P_1} \frac{\partial f(\hat{\tau})}{\partial \hat{\tau}_j}\frac{\partial f(\hat{\tau})}{\partial \hat{\tau}_l}\widehat{\mathrm{Cov}}(\hat{\tau}_k\,, \hat{\tau}_l)\ , \tag{3.12}$$

with $\hat{\boldsymbol{\tau}} = (\hat{\tau}_{01},\ \ldots,\ \hat{\tau}_{0P_0},\ \hat{\tau}_{11},\ \ldots,\ \hat{\tau}_{1P_1})$.

The following sampling procedure was considered by Berger (2004): First, select sample s_0 by unequal probability from $\mathcal{U}$ and s_{01} by SRS WoR from s_0, and sample $s_{1\backslash 0}$ of size $n_{1\backslash 0}$ is drawn from $\mathcal{U}\backslash s_0$ by SRS WoR. Then, we have $\pi_{0i} = \sum_{s_0 \ni i} p_0(s_0)$ and the probability of selecting the i-th element into s_1 given s_0 is $\pi_{1|0i} = \frac{n_{01}}{n_0}\mathcal{I}_{0,i} + \frac{n_{1\backslash 0}}{N-n_0}(1 - \mathcal{I}_{0,i})$, where $\mathcal{I}_{0,i}$ $= 1$ if $i \in s_0$ and $\mathcal{I}_{0,i} = 0$ else. Finally, $\pi_{1i} = E(\pi_{1|0i}) = \frac{n_{01}}{n_0}\frac{n_{1\backslash 0}}{N-n_0}\pi_{0i} + \frac{n_{1\backslash 0}}{N-n_0}(1 - \pi_{0i})$.

Term (3.12) involves the estimation of $P_0 + P_1$ variances and $\frac{(P_0+P_1)(P_0+P_1-1)}{2}$ covariances. Berger (2004) does this by extending his approach of estimating variances for maximum entropy sampling schemes to two-dimensional samples, which makes it possible to incorporate unequal probability sampling into the design. The actual two-dimensional sampling design is approximated by a design, where $p_0(s_0)$, $p_{1\backslash 0}(s_{1\backslash 0})$ and $p_{01}(s_{01})$ are of fixed size and $p_0(s_0)$ and $p_1(s_1|s_0)$ are conditional Poisson sampling schemes, given n_0, n_1 and n_{01}, respectively (cf. Berger, 2004, p. 454f). It is then assumed that, under the Poisson sampling scheme, vector $\mathbf{t} = (\boldsymbol{\tau}, n_0, n_1, n_{01})^T \sim N(E(\mathbf{t}), \Sigma_{\mathbf{t}})$, i.e. $\mathbf{t}$ is multivariate

normally distributed with mean $E(\mathbf{t})$ and covariance matrix $\Sigma_{\mathbf{t}}$ (see also BRUCH et al., 2011, section 2.3). Thus, the covariance matrix of vector $\boldsymbol{\tau}$ under the Poisson sampling scheme is given by

$$\Sigma_{\boldsymbol{\tau}|\mathbf{n}} = \Sigma_{\boldsymbol{\tau}} - \Sigma_{\boldsymbol{\tau}\mathbf{n}}\Sigma_{\mathbf{n}}{}^{-1}\Sigma_{\boldsymbol{\tau}\mathbf{n}} \,, \tag{3.13}$$

where $\mathbf{n} = (n_0, n_1, n_{01})^T$, $\Sigma_{\boldsymbol{\tau}\mathbf{n}}$ denotes the covariance matrix between $\boldsymbol{\tau}$ and $\mathbf{n}$, and $\Sigma_{\mathbf{n}}$ the covariance matrix of $\mathbf{n}$. Estimates for the terms in (3.13) can be obtained by estimating matrix $\Sigma_{\mathbf{t}}$, which is composed of four submatrices, with

$$\Sigma_{\mathbf{t}} = \begin{bmatrix} \Sigma_{\boldsymbol{\tau}} & \Sigma_{\boldsymbol{\tau}\mathbf{n}} \\ \Sigma_{\mathbf{n}\boldsymbol{\tau}} & \Sigma_{\mathbf{n}} \end{bmatrix} \tag{3.14}$$

$$= \mathbf{A}^T \mathrm{V}(\mathcal{I})\, \mathbf{A} \,, \tag{3.15}$$

where $\mathcal{I} = (\mathcal{I}_0^T, \mathcal{I}_1^T, \mathcal{I}_{01}^T)^T$ with $\mathcal{I}_0 = (\mathcal{I}_{0,1}, \ \ldots, \mathcal{I}_{0,N})$, $\mathcal{I}_1 = (\mathcal{I}_{1,1}, \ \ldots, \mathcal{I}_{1,N})$ and $\mathcal{I}_{01} = (\mathcal{I}_{0,1}\mathcal{I}_{1,1}, \ \ldots, \mathcal{I}_{0,N}\mathcal{I}_{1,N})$, and $\mathcal{I}_{1,i} = 1$ if $i \in s_1$ and $\mathcal{I}_{1,i} = 0$ otherwise. So $\mathcal{I}$ is a matrix of dimension $N \times 3$ with the sample indicators of s_0, s_1 and s_{01} as column vectors. Under Poisson sampling the covariance matrix $\mathrm{V}(\mathcal{I})$ can be shown to be composed of six different $N \times N$ diagonal matrices (see BERGER, 2004, p. 464f.). This is due to the property of Poisson sampling that elements within and between samples s_0, s_1 and s_{01} are drawn independently from each other. This allows to write $\Sigma_{\mathbf{t}}$ as a matrix of population totals which can then be estimated by their corresponding Horvitz-Thompson estimators (see BERGER, 2004, p. 465).

If we consider the special case of $P_0 = P_1 = 1$, then matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} \mathbf{y}_0^* & \mathbf{0} & \mathbf{1}_N & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{y}_1^* & \mathbf{0} & \mathbf{1}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1}_N \end{bmatrix} \,,$$

with $\mathbf{y}_0^* = (\frac{y_{01}}{\pi_{01}}, \ \ldots, \frac{y_{0i}}{\pi_{0i}}, \ \ldots, \frac{y_{0N}}{\pi_{0N}})^T$, $\mathbf{y}_1^* = (\frac{y_{11}}{\pi_{11}}, \ \ldots, \frac{y_{1i}}{\pi_{1i}}, \ \ldots, \frac{y_{1N}}{\pi_{1N}})^T$, and $\mathbf{1}_N$ a vector of $N \times 1$ ones. Matrix $\Sigma_{\mathbf{t}}$ has then the dimension 5×5. Note that for $P_0 > 1$ and $P_1 > 1$ $\mathbf{y}_0^*$ and $\mathbf{y}_1^*$ would be replaced by accordant matrices of dimension $N \times P_0$ and $N \times P_1$, respectively.

An estimator of (3.14) would then be

$$\widehat{\Sigma}_{\mathbf{t}} = \begin{bmatrix} \widehat{\Sigma}_{\boldsymbol{\tau}} & \widehat{\Sigma}_{\boldsymbol{\tau}\mathbf{n}} \\ \widehat{\Sigma}_{\mathbf{n}\boldsymbol{\tau}} & \widehat{\Sigma}_{\mathbf{n}} \end{bmatrix} \,, \tag{3.16}$$

where

$$\widehat{\Sigma}_{\boldsymbol{\tau}} = \begin{pmatrix} \sum\limits_{i\in s_0} \frac{(1-\pi_0)y_{0i}^2}{\pi_0^2} & \sum\limits_{i\in s_{01}} \frac{\left(1-\frac{\pi_1}{g}\right)y_{0i}y_{1i}}{\pi_0\pi_1} \\ \ldots & \sum\limits_{i\in s_1} \frac{(1-\pi_1)(y_1)_i^2}{\pi_1^2} \end{pmatrix} \,,$$

$$\widehat{\Sigma}_{\boldsymbol{\tau}\mathbf{n}} = \begin{pmatrix} \sum\limits_{i\in s_0} \frac{(1-\pi_0)y_{0i}}{\pi_0} & \sum\limits_{i\in s_{01}} \frac{\left(1-\frac{\pi_1}{g}\right)y_{0i}}{\pi_0} & \sum\limits_{i\in s_{01}} \frac{(1-\pi_0)y_{0i}}{\pi_0} \\ \sum\limits_{i\in s_{01}} \frac{\left(1-\frac{\pi_1}{g}\right)y_{1i}}{\pi_1} & \sum\limits_{i\in s_1} \frac{(1-\pi_1)y_{1i}}{\pi_1} & \sum\limits_{i\in s_{01}} \frac{(1-\pi_1)y_{1i}}{\pi_1} \end{pmatrix} \,,$$

$$\widehat{\Sigma}_{\mathbf{n}} = \begin{pmatrix} n_0(1-\pi_0) & n_{01}(1-\frac{\pi_1}{g}) & n_{01}(1-\pi_0) \\ \dots & n_1(1-\pi_1) & n_{01}(1-\pi_1) \\ \dots & \dots & n_{01}(1-g\pi_0) \end{pmatrix} ,$$

and $g = \frac{n_{01}}{n_0}$. Finally, we can estimate $\Sigma_{\boldsymbol{\tau}|\mathbf{n}}$, the covariance matrix between estimates τ_{01} and τ_{11}, via

$$\widehat{\Sigma}_{\boldsymbol{\tau}|\mathbf{n}} = \widehat{\Sigma}_{\boldsymbol{\tau}} - \widehat{\Sigma}_{\boldsymbol{\tau}\mathbf{n}}\widehat{\Sigma}_{\mathbf{n}}^{-1}\widehat{\Sigma}_{\boldsymbol{\tau}\mathbf{n}} .$$

The explicit expression for the covariance between τ_{01} and τ_{11} is omitted here because it is lengthy and of little value for general applications.

3.2.2 Linearization of Bivariate Statistical Functionals

The aim of this subsection is to show how linearization is used to approximate the variance of estimators depending on two-dimensional samples. For this purpose the concept of influence functions, as presented in section 2.1, is extended to the case of statistical functionals depending on a multi-dimensional measure $\mathbf{M}$. Thus, we introduce a two-variate functional $\mathbf{T}(\mathbf{M})$ with $\mathbf{M} = (M_0, M_1)$, where in analogy to the one-sample case, $M_0 = \sum_{i=1}^{N} \delta^{\circ}_{y_{0i}}$ and $\delta^{\circ}_{y_{0i}}$ denotes a measure taking mass one for each y_{0i} with $i \in \mathcal{U}$ and zero elsewhere, with M_1 defined in a similar way (see Goga et al., 2009, p. 694). As for the one-dimensional case in section 2.3, an intuitive estimator of M_0 can be defined as $\hat{M}_0 = \sum_{i=1}^{N} w_{0i}\delta^{\circ}_{u_{0i}}$, where w_{0i} is the survey weight associated with the i-th element for $i \in s_0$ and zero elsewhere. To estimate M_1 an estimator $\hat{M}_1$ is defined in analogy to $\hat{M}_0$. Note that we do not consider the case of any composite estimator which requires knowledge of the matched sample s_{01}. In other words there is no interaction between M_0 and M_1, estimator $\mathbf{T}(\hat{\mathbf{M}})$ consists of cross-sectional $T(\hat{M}_0)$ and $T(\hat{M}_1)$, which are constants with respect to $\hat{M}_0$ and $\hat{M}_1$, respectively (on the variance of composite estimators see Goga et al., 2009). This limitation is justified by the fact that change in an indicator value is often simply measured by the difference or ratio between two cross-section estimates.

The variance of an estimator of the form $\mathbf{T}(\hat{\mathbf{M}}) = \mathbf{T}\left(\hat{M}_0, \hat{M}_1\right)$ can be approximated by its associated *partial influence functions* (see Pires and Branco, 2002), which equal the linearized variables and, as in the one-sample case, the variance of $\mathbf{T}(\hat{\mathbf{M}})$ equals the variance of their estimated total. Hence, the asymptotic variance of $\mathbf{T}(\hat{\mathbf{M}})$ equals

$$\mathrm{V}\left(\sum_{t=0}^{1}\sum_{i\in s_t} w_{ti}z_{ti}\right) , \tag{3.17}$$

where z_{ti} is the partial Gâteaux derivative of $\mathbf{T}$ with respect to M_t, i.e

$$z_{ti} = IT_t\left(\mathbf{M}, \mathbf{g}_{ti}\right) = \lim_{\epsilon\to\infty} \frac{\mathbf{T}\left(M_0 + \epsilon\delta_{u_0}, M_1\right) - \mathbf{T}(M_0,\, M_1)}{\epsilon} ,$$

with $\mathbf{g}_{ti} \in \mathbb{R}^d$ as the variable of interest measured at time t of the i-th unit in $\mathcal{U}$ (for the proof see Goga et al., 2009).

To illustrate the approach, we consider the following example of a change in a ratio (see also GOGA et al., 2009, p. 694). The variable of interest $\mathbf{g}$ is bivariate with $\mathbf{g}_{ti} = (y_{ti}, x_{ti})$

$$T(\mathbf{M}) = R_\Delta = R_1 - R_0 = \frac{\tau_3}{\tau_4} - \frac{\tau_1}{\tau_2} = \frac{\int y_1 dM_1}{\int x_1 dM_1} - \frac{\int y_0 dM_0}{\int x_0 dM_0} \tag{3.18}$$

The values of partial influence functions of R_Δ are given by

$$IT_0(\mathbf{M}, (y_{0i}, x_{0i})) = z_{0i} = -\frac{1}{\tau_2}(y_{0i} - R_0 x_{0i})$$
$$IT_1(\mathbf{M}, (y_{1i}, x_{1i})) = z_{1i} = \frac{1}{\tau_4}(y_{1i} - R_1 x_{1i}) \ ,$$

the linearized variables of R_0 and R_1 (see also GOGA et al., 2009, p. 695). If we substitute M_0 and M_1 by their estimators $\hat{M}_0$ and $\hat{M}_1$ in equation (3.18), we obtain an estimator $\hat{R}_\Delta$ for R_Δ. Now, the variance of $\hat{R}_\Delta$ can be approximated by

$$\mathrm{V}\left(\sum_{i \in s_1} z_{1i} w_{1i} + \sum_{i \in s_0} z_{0i} w_{0i}\right). \tag{3.19}$$

3.2.3 Rotational Samples: An Application to the EU-SILC Survey

The Monte-Carlo study is a design-based simulation which aims to evaluate empirically the accuracy of the proposed variance estimators for a nonlinear functional $T(\hat{M}_0, \hat{M}_1)$. Estimator $T(\hat{M}_0, \hat{M}_1)$ is based on a two-dimensional sample $\mathbf{s} = (s_0, s_1)$. The study was set up following the rotational sampling scheme used in the EU-SILC survey. Figure 3.2 illustrates the rotational pattern used in the simulation. A more detailed description of the rotational scheme of EU-SILC can be found in VERMA et al., 2007. The survey has

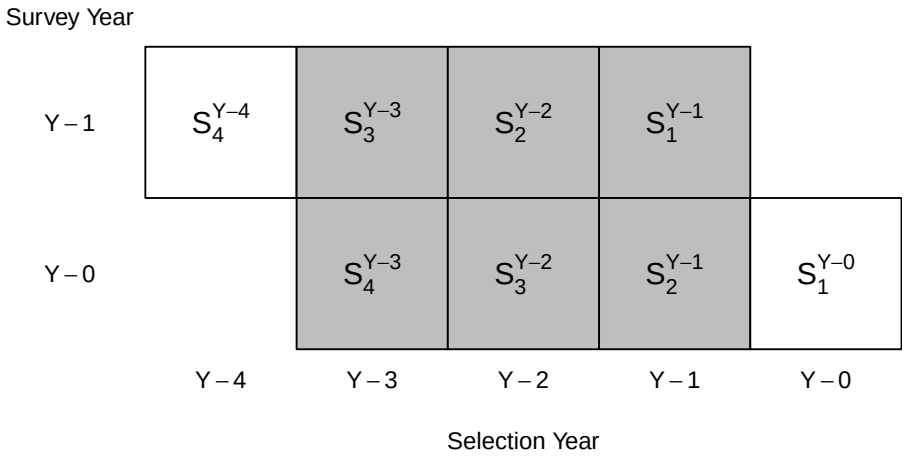

Figure 3.2: Rotational samples in EU-SILC

four rotational groups, all households in one group are surveyed over four successive years

which adds the longitudinal dimension to the dataset. In figure 3.2 s_3^{Y-3} represents a sample of households drawn in year $Y-3$ and surveyed for 3 successive years, where Y indicates the present year. The observation period includes years $Y-1$ and $Y-0$ and the two cross-sectional samples available for $Y-1$ and $Y-0$ are $s_0 = \{s_t^{Y-t}\}$ and $s_1 = \{s_t^{Y-t+1}\}$, respectively, with $t = \{1,2,3,4\}$. Elements in s_4^{Y-4} have stayed over four years in the survey and are replaced by the newly drawn sample s_1^{Y-0}. Hence, the rotational pattern produces an overlap between s_0 and s_1 of 75%.

For the simulation study EU-SILC longitudinal data was chosen as a finite population. The data includes observations from all EU member states excluding Bulgaria, Romania, and Malta and plus Iceland and Norway. The annual equivalized disposable household income was selected as the variable of interest y. The data was then preperated that the simulation incorporates following assumptions:

1. A stable population over time
2. No outliers in the population
3. The rotational groups are independent of each other.

The population is be stable over time, i.e. there are no death and births in the population, because only those households were included that have been present at all considered years. Further, only those households have been considered with $0 \leq y < 150,000$. The truncation of the income distribution was done for two reasons. First, the GINI, as given in formula (2.21), is not defined for negative incomes and second, the QSR should not have been subject to the effects of outliers, which are present in the original EU-SILC data set. In preparation of the sampling process the eligible households where randomly allocated into four rotational groups $\mathcal{U}^{Y-4}$, $\mathcal{U}^{Y-3}$, $\mathcal{U}^{Y-2}$, and $\mathcal{U}^{Y-1}$.

The actual sampling plan consists of two steps:

1. s_0 is drawn by selecting four stratified random samples of households $\{s_t^{Y-t}\}$ with equal size independently from groups U^{Y-t}
2. For the rotational part a stratified random sample of households $\{s_1^{Y-0}\}$ with the same size as s_4^{Y-4} is selected from $\mathcal{U}^{Y-4}$

Stratification is done after countries with proportional allocation of the samples sizes and households are selected without replacement. Note that s_4^{Y-4} and s_1^{Y-0} are bothe selected from $\mathcal{U}^{Y-4}$, i.e. the no overlapping parts of s_1 and s_0 are independent of each other. Because the population is stationary we have $s_3^{Y-3} = s_4^{Y-3}$, $s_2^{Y-2} = s_3^{Y-2}$, and $s_1^{Y-1} = s_2^{Y1}$, thus we may drop suffix t and denote the matched sample by $s_{01} = \{s^{Y-t}\}$ for $t = \{1,2,3\}$.

The boxplots in figure 3.3 show the distribution of the relative bias of 10000 variance estimates of measures of change. The data set used in the simulation was the EU-SILC 2006 longitudinal data and the change is estimated for the indicators QSR, GINI and ARPR for 2005 to 2006. The upper half refers to the results for a sample fraction of 1% and the lower for 10%. The left half of the figure shows the results for the difference and the right one for ratio as a measure of change. Further, HT over a panel denotes the use of

Horvitz-Thompson weights and CAL the use of calibration weights for point estimation. The calibration was done on person level using the age and sex variable.

The asymptotic variance for the difference, as given in 3.17, is estimated by its equivalent for stratified random samples. To simplify the notation, rotational groups are treated like strata. Set $\mathcal{H}$ denotes the cross-classification of rotational groups and strata (countries) and set $\mathcal{H}^{01}$ is the same but without rotational group $\mathcal{U}^{Y-4}$. Thus, we us the following variance estimator for $\hat{\theta}_\Delta$

$$\begin{aligned}\hat{\mathrm{V}}(\hat{\theta}_\Delta) = &\sum_{h\in\mathcal{H}} N_h^2 \left(1-\frac{n_h}{N_h}\right)\frac{\hat{\sigma}^2_{0_h}}{n_h} \\ &+ \sum_{h\in\mathcal{H}} N_h^2 \left(1-\frac{n_h}{N_h}\right)\frac{\hat{\sigma}^2_{1_h}}{n_h} \\ &+ \sum_{h\in\mathcal{H}^{01}} 2N_h^2 \left(1-\frac{n_{01_h}}{N_h}\right)\frac{\hat{\sigma}^2_{01_h}}{n_{01_h}} ,\end{aligned} \tag{3.20}$$

where

$$\begin{aligned}\hat{\sigma}^2_{0_h} &= \frac{1}{n_h}\sum_{i\in s_{0_h}} \left(z_{hi}-\overline{z}_{0_h}\right)^2 , \\ \hat{\sigma}^2_{1_h} &= \frac{1}{n_h}\sum_{i\in s_{1_h}} \left(z_{hi}-\overline{z}_{1_h}\right)^2 , \\ \hat{\sigma}^2_{01_h} &= \frac{1}{n_{01_h}}\sum_{i\in s_{01_h}} \left(z_{hi}-\overline{z}_{0_h}\right)\left(z_{hi}-\overline{z}_{1_h}\right) ,\end{aligned}$$

and

$$\begin{aligned}\overline{z}_{0_h} &= \frac{1}{n_{0_h}}\sum_{i\in s_{0_h}} z_{hi} , \\ \overline{z}_{1_h} &= \frac{1}{n_{1_h}}\sum_{i\in s_{1_h}} z_{hi} .\end{aligned}$$

The asymptotic variance of the ratio $\hat{\theta}_1\hat{\theta}_0^{-1}$ is estimated by

$$\frac{1}{\hat{\theta}_0^2}\hat{\mathrm{V}}_1 + \left(\frac{\hat{\theta}_1}{\hat{\theta}_0}\right)^2 \hat{\mathrm{V}}_0 - 2\frac{\hat{\theta}_1}{\hat{\theta}_0}\widehat{\mathrm{Cov}}_{01} ,$$

where $\hat{\mathrm{V}}_0$ equals the first term in (3.20), $\hat{\mathrm{V}}_1$ the second and $\widehat{\mathrm{Cov}}_{01}$ the third. In case of the calibration estimators for variance estimation not the linearized variable z was used but e, the residual of the regression z on $\mathbf{x}$, where $\mathbf{x}$ are is the auxiliary information used for calibration (see section 2.4).

Figure 3.3 reveals that the variance estimators work well. The black cross, which indicates the mean value of over the 10,000 estimates, is always well within the interval $[-0.1, 0.1]$. As one would expect, the variance estimates become more accurate with increasing sample

size. In general, there is no big difference in accuracy of variance estimators neither with respect to the statistic nor the usage of Horvitz-Thompson or calibration weights. The only exception are the variance estimates from the ARPR, which seem to suffer from a bias if the sampling fraction is high.

Figure 3.4 shows the confidence interval (CI) coverage rates for the 10,000 variance estimates, again for sampling fractions of 1% and 10%. For all indicators CIs have been created with 90% and 95% confidence level. The two red horizontal lines, at 0.9 and 0.95 are the benchmarks for the coverage rates and the black lines build a bandwidth of ± 0.1 around them. The figure shows that almost all variance estimators result in CI with coverage rates well within the bandwidth around the benchmark. Further there is no big difference between the estimators using Horvitz-Thompson and calibration weights.

Figure 3.5 shows so called funnel plots (see Hulliger and Pooda, 2010; Templ et al., 2011, section 2.9), which are a possible application of variance estimates of a change in indicator values. They visualise the test decision if the null hypothesis, that no change has occurred at all, i.e. $H_0 := \Delta = \Delta_0 = 0$, can be rejected or not. Each of the three plots show the time line, in solid black, that connects the estimates $\hat{\theta}_t$ of the ARPR, the GINI, or the QSR for four consecutive years. The funnel between two adjacent years can be interpreted in the following way. The vertical line of the three triangles or funnels in each plot correspond to a CI build around estimate $\hat{\theta}_1$ but using a variance estimate $\widehat{\mathrm{V}}(\widehat{\Delta})$. The dashed horizontal line corresponds to estimate $\hat{\theta}_0$, if it is outside the CI around $\hat{\theta}_1$, H_0 can be rejected and vice versa. In the example in Figure 3.5 it is the case that albeit all indicator estimates decrease over time none of the changes between one year and the next is regarded as significant.

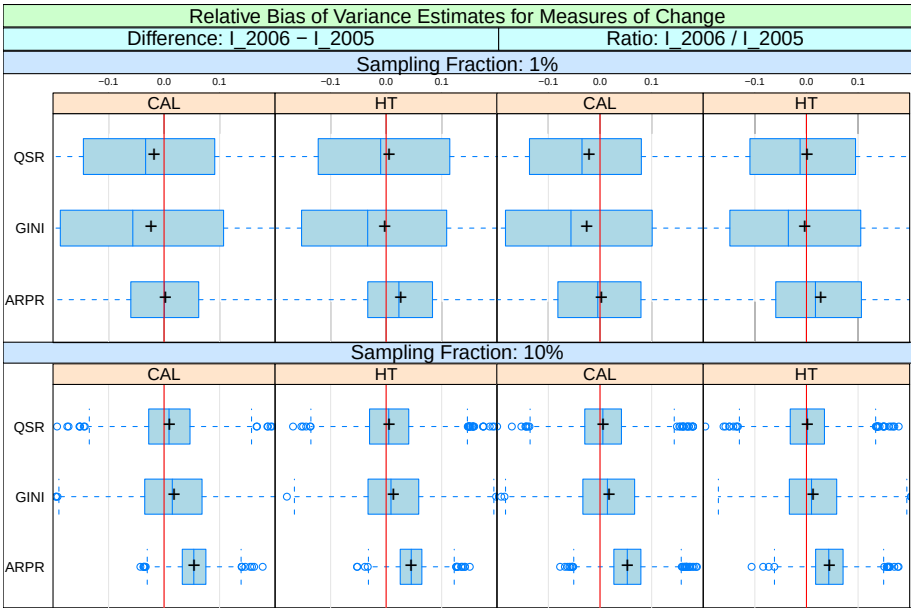

Figure 3.3: Relative Bias of Variance Estimates for Measures of Change

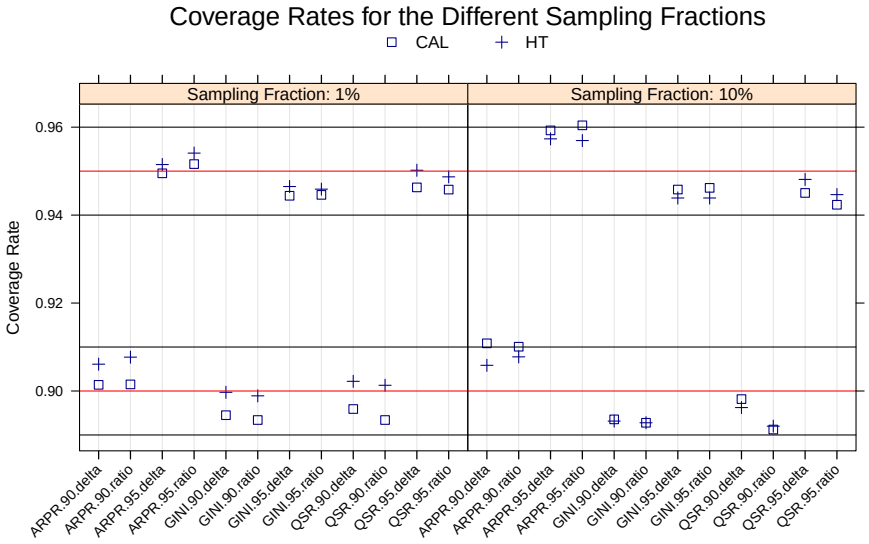

Figure 3.4: Coverage Rates for the Different Sampling Fractions

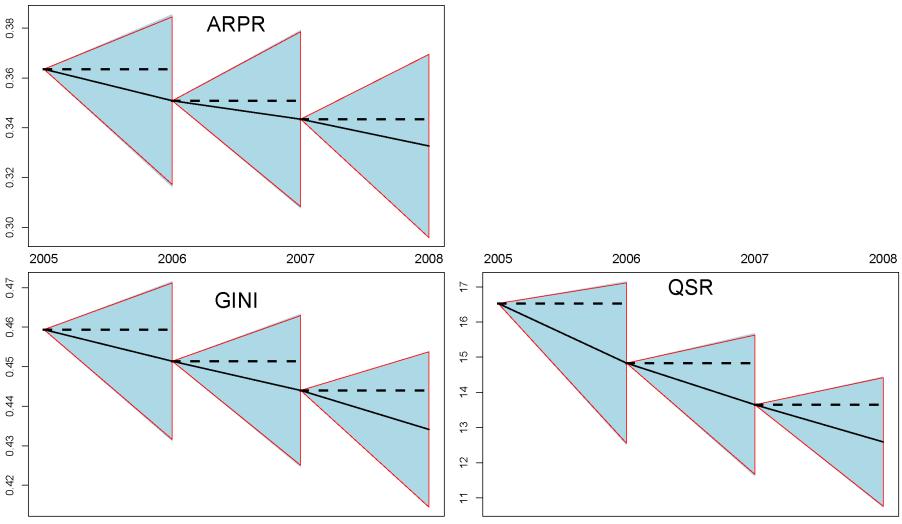

Figure 3.5: Funnel Plot; Test of Significant Change: $H_0 := \Delta = \Delta_0 = 0$

Chapter 4

Summary and Outlook

In the following we draw some conclusions from the application of the presented methodology, as well as giving an outlook and adding a few thoughts on how the presented methods can be extended to account for more problems encountered in the practice of estimating from samples surveys.

The availability of linear variance estimators for indicators of poverty and inequality can be of great value because its simplifies the computation of standard sample errors, because it allows for the use of variance estimators for linear statistics, which can be derived for sampling designs of nearly arbitrary complexity. Thus, they have the potential to facilitate the use of statistical inference not only when producing official statistics but also in the area of economic and social sciences, when analyses are based on sample data. But it should be mentioned that linearization can also be problematic. To apply the linearized values for variance estimation we need that the remainder term R in expansion 2.2 converges in probability to zero. This is ensured, as Demnati and Rao (2004) noted, if we have iid sample data, which would be reasonable to assume for simple random sampling or simple stratified sampling where we have iid observation within each strata. Serfling (1980, section 6.2.2) gave some advice on how to handle the remainder term R. But for more complex survey design, like multi-stage sampling involving cluster sampling at higher stages the iid assumption might not be sustainable. This would especially be true if there is a large cluster effect. Wolter (2007) also notes that one should be cautious if using approximations in the presence of highly skewed populations, although he does this in the context of the Taylor series. Experience with linearization have shown that in general for samples sizes large enough first-order approximations (as (1.5)) deliver variance estimates that allow for the construction of valued confidence intervals. However, even if the population itself is not highly skewed certain sampling designs might have a similar effect by creating highly skewed sample distributions of observations. This might render a (first-order) approximation, like the one in (1.6), inadequate.

For estimation of change in indicator values, linearization provides a convenient approach to estimate the variance of a change, especially if variance estimates for cross-sectional have already been obtained by linearization. However, in practice rotational sampling schemes are often somewhat more complex. For instance, in the simulation study presented in section 3.2.3 the rotational parts of the two-dimensional sample are constructed to be independent of each other. This is a simplification that might not be applicable

to most rotational schemes, as it conceivable that sample elements that dropped out of the survey are excluded from being selected into the replacement sample. TAM (1984) already considered sampling plans where the rotational parts are not selected independently of each other. By using the results from TAM (1984) variance estimators (3.20) can be readily adjusted to account for such sample schemes.

Another degree of complexity found in many rotational samples applied in practice is multi-stage sampling (e.g. the French Labour Force Survey, see PLACE, 2008). For instance, if PSUs are selected at an earlier stage and rotation is then carried out within PSU, then the rotational parts can be a source of a secondary correlation between estimates, assuming a certain degree of similarity between elements in a common PSU (see STEEL and MCLAREN, 2009, section 3). This raises the question of how to estimate covariances from a part of the sample that have no elements in common. A possible approach to this matter is to treat it as a problem of missing data, appearing in the non-overlapping part of the sample (see PLACE (2008)). In this case missing observations would be determined by the rotational scheme, thus, the unobserved values could be regarded as *missing at random* , given the duration how long a element stays in the sample.

Finally, there is the problem that the population under consideration may not be stable over time, as assumed in section 3.2.3. There are elements that depart from the population (deaths) or appear new (births). There are a number of authors that accommodated this fact in their analysis. LANIEL (1987) generalised the results of TAM (1984) by removing the assumption of a stable population, (on this matter see also NORDBERG (2000); WOOD (2008)). The problem can in part be described as estimation in a dual frame survey, i.e. the non-rotational part of sample is drawn from one frame and the rotational part from another. Both frames overlap and together they cover the population of interest, however, this does not consider the attrition within the non-rotational part.

Bibliography

Alfons, A., **Bruch, C.**, **Filzmoser, P.**, **Graf, M.**, **Hulliger, B.**, **Kolb, J.-P.**, **Lehtonen, R.**, **Lussmann, D.**, **Meraner, A.**, **Münnich, R.**, **Nedyalkova, D.**, **Schoch, T.**, **Templ, M.**, **Valaste, M.**, **Veijanen, A.** and **Zins, S.** (**2011**): *Report on the Simulation Results*. Research Project Report WP7 – D7.1, FP7-SSH-2007-217322 AMELI.
URL http://ameli.surveystatistics.net

Andersson, C. and **Nordberg, L.** (**1994**): *A Method for Variance Estimation of Non-Linear Functions of Totals in Surveys - Theory and Software Implementation*. Journal of Official Statistics, 10 (4), pp. 395 – 405.

Berger, Y. G. (**2004**): *Variance estimation for measures of change in probability sampling*. The Canadian Journal of Statistics. La Revue Canadienne de Statistique, 32 (4), pp. 451–467.

Binder, D. A. and **Kovacevic, M. S.** (**1993**): *Estimating Some Measures of Income Inequality from Survey Data: An Application of the Estimating Equation Approach*. Journal of the American Statistical Association, (550), pp. 550–555.

Binder, D. A. and **Patak, Z.** (**1994**): *Use of estimating functions for estimation from complex surveys*. Journal of the American Statistical Association, 89 (427), pp. 1035–1043.

Bruch, C., **Münnich, R.** and **Zins, S.** (**2011**): *Variance Estimation for Complex Surveys*. Research Project Report WP3 – D3.1, FP7-SSH-2007-217322 AMELI.
URL http://ameli.surveystatistics.net

Demnati, A. and **Rao, J.** (**2004**): *Linearization variance estimators for survey data*. Survey Methodology, 30 (1), pp. 17 – 26.

Deville, J.-C. (**1999**): *Variance estimation for complex statistics and estimators: Linearization and residual techniques*. Survey Methodology, 25 (2), pp. 193– 203.

Deville, J.-C. and **Särndal, C.-E.** (**1992**): *Calibration estimators in survey sampling*. Journal of the American Statistical Association, 87 (418), pp. 376–382.

Goga, C. (**2003**): *Estimation de la variance dans les sondages à plusieurs échantillons et prise en compte de l'information auxiliaire par des modèles nonparamétriques*, ph.D. Dissertation, Université de Rennes II, Haute Bretagne, France.

Goga, C., **Deville, J.-C.** and **Ruiz-Gazen, A.** (**2009**): *Use of functionals in linearization and composite estimation with application to two-sample survey data*. Biometrika, 96 (3), pp. 691–709.

Hampel, F. R. (**1974**): *The influence curve and its role in robust estimation*. Journal of the American Statistical Association, 69, pp. 383–393.

Hulliger, B. and **Münnich, R.** (**2006**): *Variance Estimation for Complex Surveys in the Presence of Outliers*. Proceedings of the American Statistical Association, Survey Research Methods Section, pp. 3153–3161.

Hulliger, B. and **Pooda, D. L.** (**2010**): *Assessment of Sustainable Development and Environmental Indicators. Report commissioned by the Federal Statistical Office, Section Environment, Sustainable Development, Agriculture.* Technical report, School of Business, University of Applied Sciences Northwestern Switzerland (FHNW).

Jones, M. C., **Marron, J. S.** and **Sheather, S. J.** (**1996**): *A brief survey of bandwidth selection for density estimation*. Journal of the American Statistical Association, 91 (433), pp. 401–407.

Kovacevic, M. S. and **Binder, D. A.** (**1997**): *Variance Estimation for Measures of Income Inequality and Polarization - The Estimating Equations Approach*. Journal of Official Statistics, 13 (1), pp. 41 – 58.

Langel, M. and **Tillé, Y.** (**2011**): *Statistical inference for the quintile share ratio*. Journal of Statistical Planning and Inference, 141 (8), pp. 2976–2985.

Laniel, N. (**1987**): *Variances for a Rotating Sample from a Changing Population*. Proceedings of the American Statistical Association, Survey Research Methods Section.

Nordberg, L. (**2000**): *On Variance Estimation for Measures of Change When samples are Coordinated by the Use of Permanent Random Numbers*. Journal of Official Statistics, 16 (4), pp. 363–378.

Osier, G. (**2009**): *Variance estimation for complex indicators of poverty and inequality using linearization techniques*. Survey Research Methods, 3 (3), pp. 167 – 195.

Pires, A. M. and **Branco, J. A.** (**2002**): *Partial influence functions*. Journal of Multivariate Analysis, 83 (2), pp. 451–468.

Place, D. (**2008**): *Longitudinal Variance Estimation in the French Labour Force Survey*. Proceedings of the European Conference on Quality in Offical Statistics.

Qualité, L. and **Tillé, Y.** (**2008**): *Variance Estimation of changes in repeated surveys and its application to the Swiss survey of value added*. Survey Methodology, 34 (2), pp. 173 – 181.

R Development Core Team (**2010**): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL http://www.R-project.org/

Särndal, C.-E. (**2007**): *The calibration approach in survey theory and practice*. Survey Methodology, 33 (2), pp. 99 – 119.

Serfling, R. J. (**1980**): Approximation theorems of mathematical statistics. New York: Wiley, wiley Series in Probability and Mathematical Statistics.

Shao, J. (**2003**): Mathematical Statistics. New York: Springer, second ed.

Särndal, C.-E., **Swensson, B.** and **Wretman, J.** (**1992**): Model Assisted Survey Sampling. New York: Springer-Verlag.

Steel, D. and **McLaren, C.** (**2009**): *Design and Analysis of Survey Repeated over Time*. **Rao, C.** (editor) Handbook of Statistics - Sample Surveys: Inference and Analysis, *Handbook of Statistics*, vol. 29, Part 2, pp. 289 – 249, Elsevier.

Tam, S. M. (**1984**): *On covariances from overlapping samples*. The American Statistician, 38 (4), pp. 288–289.

Templ, M., **Hulliger, B.**, **Alfons, A.**, **Lussmann, D.** and **Filzmoser, P.** (**2011**): *Visualisation Tools*. Research Project Report WP8 – D8.2, FP7-SSH-2007-217322 AMELI.
URL http://ameli.surveystatistics.net

Verma, V., **Betti, G.** and **Ghellini, G.** (**2007**): *Cross-sectional and Longitudinal Weighting in a Rotational Houshold Panel: Application to EU-SILC*. Statistics in Transition, 8 (1), pp. 5–50.

Wolter, K. M. (**2007**): Introduction to variance estimation. New York: Springer.

Wood, J. (**2008**): *On the Covariance Between Related Horvitz-Thompson Estimators*. Journal of Official Statistics, 24 (1), pp. 53–78.

Woodruff, R. S. (**1971**): *A Simple Method for Approximating the Variance of a Complicated Estimate*. Journal of the American Statistical Association, 66 (334), pp. 411–414.