Variance estimation of change in poverty rates: an application to the Turkish EU-SILC survey

Melike Oguz-Alper *[1] and Yves G. Berger[2]

[1]University of Southampton, United Kingdom, E-mail: M.OguzAlper@soton.ac.uk

[2]University of Southampton, United Kingdom, E-mail: Y.G.Berger@soton.ac.uk

Abstract

Interpreting changes between point estimates at different waves may be misleading, if we do not take the sampling variation into account. It is therefore necessary to estimate the standard error of these changes in order to judge whether or not the observed changes are statistically significant. This involves the estimation of temporal correlations between cross-sectional estimates, because correlations play an important role in estimating the variance of a change in the cross-sectional estimates. Standard estimator for correlations cannot be used, because of the rotation used in most panel surveys, such as the European Union Statistics on Income and Living Conditions (EU-SILC) surveys. Furthermore, as poverty indicators are complex functions of the data, they need a special treatment when estimating their variance. For example, poverty rates depend on poverty thresholds which are estimated from medians. We propose to use a multivariate linear regression approach to estimate correlations by taking into account of the variability of the poverty threshold. We apply the proposed approach to the Turkish EU-SILC survey data.

Keywords: Linearisation; multivariate regression; stratification; unequal inclusion probabilities.

*Funded by the Economic and Social Research Council (ESRC), United Kingdom

1 Introduction

In order to monitor the process towards agreed policy goals, particularly in the context of the Europe 2020 strategy, there is an interest in evaluating the evolution of social indicators. For the purpose of interpreting changes between indicators at different waves, it is important to estimate the standard error of these changes; so that we can judge whether or not observed changes are statistically significant. The poverty rate is an important policy indicator, especially within the context of the Europe 2020 Strategy. This rate is defined as the proportion of people with an equivalised total net income below 60% of the national median income (Eurostat 2003, p.2). This indicator is calculated from the European Union Statistics on Income and Living Conditions (EU-SILC) surveys (Eurostat 2012) which collect yearly information on income, poverty, social exclusion and living conditions from approximately 300 000 households across Europe. The poverty rate is a complex statistics unlike population totals or means; since, it is based on a poverty threshold computed from the median of the income distribution. Hence, there exist two sources of variability: one is due to the estimated threshold and the other one comes from the estimated proportion given the estimated threshold (e.g. Berger and Skinner 2003; Verma and Betti 2011).

Several methods to estimate the variance of the poverty rate like re-sampling and linearisation techniques have been discussed in the literature (e.g. Preston 1995; Deville 1999; Berger and Skinner 2003; Demnati and Rao 2004; Verma and Betti 2005; Osier 2009; Goedeme 2010; Verma and Betti 2011; Muennich and Zins 2011; Osier et al. 2013; Berger and Priam 2013). However, variance of change for the poverty rate has been studied in a limited number of papers (e.g. Betti and Gagliardi 2007; Muennich and Zins 2011; Osier et al. 2013; Berger and Priam 2013). Osier et al. (2013) proposed an estimator for the variance of change which takes into account the complexity of the sampling design such as stratification, unequal probabilities, clustering and rotation (see also Berger and Priam 2010, 2013). The proposed approach does not rely on neither the second order inclusion probabilities nor the re-sampling methods unlike its competitors (Betti and Gagliardi 2007; Wood 2008; Muennich and Zins 2011, p.20). It is based on a multivariate linear regression (general linear model) approach that can be easily implemented by any statistical software (Berger and Priam 2013). Berger et al. (2013) show how it can be implemented in SPSS.

The estimator proposed by Osier et al. (2013) ignores the sampling variability due to the poverty threshold, by treating the poverty rate as a ratio. In Sections 4 and 5, we show how this approach can be adjusted to take into account of the sampling variability of the poverty threshold. In Section 6, we compare the proposed approach with the more simple approach proposed by Osier et al. (2013) (see also Berger and Priam 2010, 2013) via a series of simulation. In Section 7, we apply the proposed approach to the Turkish EU-SILC survey data. The proposed variance estimator depends on a bandwidth used for the estimation of the density. We also show how sensitive the variance estimates are to the chosen bandwidth parameter by considering different bandwidth parameters.

2 Rotating sampling designs

With rotating panel surveys, it is common practice to select new units in order to replace old units that have been in the survey for a specified number of waves (e.g. Gambino and Silva 2009; Kalton 2009). The units sampled on both waves usually represent a large fraction of the first wave sample. This fraction is called the fraction of the common sample.

For example, for the EU-SILC surveys, this fraction is 75%. For the Canadian labour force survey and the British labour force survey, this fraction is 80%. For the Finish labour force survey, this fraction is 60%. We consider that the sample design is such that the common sample has a fixed number of units. Throughout this paper, we assume that the sampling fractions are negligible, that is, $(1 - \pi_{t;i}) \approx 1$, where $\pi_{t;i}$ denote the inclusion probabilities of unit i at wave t.

3 Estimation of change of a poverty rate

Let s_1 and s_2 be the samples selected at wave 1 and wave 2 respectively. Suppose, we wish to estimate the absolute net change $\Delta = \theta_2 - \theta_1$ between two population poverty rates θ_1 and θ_2, from wave 1 and wave 2 respectively. Suppose that Δ is estimated by $\widehat{\Delta} = \widehat{\theta}_2 - \widehat{\theta}_1$; where $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are the cross-sectional estimators of poverty rates defined by

$$\widehat{\theta}_1 = \frac{\widehat{\tau}_1}{\widehat{\tau}_2} = \frac{\sum_{i \in s_1} \delta\{y_{1;i} \leq 0.6\widehat{Y}_{1;0.5}\}\pi_{1;i}^{-1}}{\sum_{i \in s_1} \pi_{1;i}^{-1}} \quad \text{and} \quad \widehat{\theta}_2 = \frac{\widehat{\tau}_3}{\widehat{\tau}_4} = \frac{\sum_{i \in s_2} \delta\{y_{2;i} \leq 0.6\widehat{Y}_{2;0.5}\}\pi_{2;i}^{-1}}{\sum_{i \in s_2} \pi_{2;i}^{-1}},$$

where $y_{t;i}$ is the '*net equivalised income*' (see Eurostat 2003, p.2) of the individual i at wave t and $\widehat{Y}_{t;0.5}$ is the estimate of the median of the population income distribution at wave t $(t = 1, 2)$. The function $\delta\{A\} = 1$, when A is true, and $\delta\{A\} = 0$ otherwise.

The design-based variance of the estimator of change $\widehat{\Delta}$ is given by

$$\text{var}(\widehat{\Delta}) = \text{var}(\widehat{\theta}_1) + \text{var}(\widehat{\theta}_2) - 2\text{corr}(\widehat{\theta}_1, \widehat{\theta}_2)\sqrt{\text{var}(\widehat{\theta}_1)\text{var}(\widehat{\theta}_2)} \cdot \qquad (1)$$

Standard design-based estimators can be used to estimate the cross-sectional variances $\text{var}(\widehat{\theta}_1)$ and $\text{var}(\widehat{\theta}_2)$ (e.g. Deville 1999). The correlation $\text{corr}(\widehat{\theta}_1, \widehat{\theta}_2)$ is the most difficult part to estimate as $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are estimated from different samples because of the rotation.

Estimation of the covariance term has been discussed in several papers (Kish 1965, p.457-458; Tam 1984; Laniel 1987; Nordberg 2000; Holmes and Skinner 2000; Berger 2004; Qualité and Tillé 2008; Wood 2008; Muennich and Zins 2011).

Berger and Priam (2010, 2013) proposed a multivariate approach to estimate the correlation between functions of totals by incorporating the information related to the whole sample, $s = s_1 \cup s_2$. This approach can be used to estimate the variance of change between poverty rates when we ignore the sampling variability due to the estimated poverty threshold $0.6\widehat{Y}_{t;0.5}$, that is, when we treat the poverty rates as simple ratios.

When we treat the threshold as fixed, the change becomes a smooth function of four totals, that is, $\widehat{\Delta} = \mathrm{g}(\widehat{\boldsymbol{\tau}})$, where $\widehat{\boldsymbol{\tau}} = (\widehat{\tau}_1, \widehat{\tau}_2, \widehat{\tau}_3, \widehat{\tau}_4)^\top$ is a vector of four estimated totals. Berger and Priam (2010, 2013) showed that using the first-order Taylor approximation, the design-based variance of $\widehat{\Delta}$ can be estimated by

$$\widehat{\mathrm{var}}(\widehat{\Delta}) = \widehat{\mathbf{grad}}(\mathrm{g})^\top \widehat{\mathbf{var}}(\widehat{\boldsymbol{\tau}}) \widehat{\mathbf{grad}}(\mathrm{g}), \tag{2}$$

where $\widehat{\mathbf{grad}}(\mathrm{g})$ is the gradient of $\mathrm{g}(\widehat{\boldsymbol{\tau}})$ evaluated at $\widehat{\boldsymbol{\tau}}$, that is,

$$\widehat{\mathbf{grad}}(\widehat{\boldsymbol{\tau}}) = \frac{\partial \mathrm{g}(\widehat{\boldsymbol{\tau}})}{\partial \widehat{\boldsymbol{\tau}}} = \left(-\frac{1}{\widehat{\tau}_2}, -\frac{\widehat{\tau}_1}{\widehat{\tau}_2^2}, \frac{1}{\widehat{\tau}_3}, -\frac{\widehat{\tau}_3}{\widehat{\tau}_4^2} \right)^\top,$$

and $\widehat{\mathbf{var}}(\widehat{\boldsymbol{\tau}})$ is given by

$$\widehat{\mathbf{var}}(\widehat{\boldsymbol{\tau}}) = \widehat{\boldsymbol{D}}^\top \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{D}},$$

with

$$\widehat{\boldsymbol{D}} = \mathrm{diag}\left\{ \sqrt{\widehat{\mathrm{var}}(\widehat{\tau}_1)\widehat{\Sigma}_{11}^{-1}}, \sqrt{\widehat{\mathrm{var}}(\widehat{\tau}_2)\widehat{\Sigma}_{22}^{-1}}, \sqrt{\widehat{\mathrm{var}}(\widehat{\tau}_3)\widehat{\Sigma}_{33}^{-1}}, \sqrt{\widehat{\mathrm{var}}(\widehat{\tau}_4)\widehat{\Sigma}_{44}^{-1}} \right\}$$

where $\widehat{\boldsymbol{\Sigma}}$ is the Ordinary Least Square (OLS) estimator of the residual covariance matrix $\boldsymbol{\Sigma}$ of the multivariate linear regression model in (3) proposed by Berger and Priam (2010, 2013); $\widehat{\text{var}}(\widehat{\tau}_k)$ is the design-based variance estimator of the Horvitz and Thompson (1952) estimator of $\widehat{\tau}_k$, and $\widehat{\Sigma}_{kk}^{-1}$ is the k-th diagonal element of $\widehat{\boldsymbol{\Sigma}}$ ($k = 1, 2, 3, 4$). Berger and Priam (2010, 2013) showed that (2) gives an approximately unbiased estimator for the variance of change.

Let $\breve{p}_{t;i} = \delta\{y_{t;i} \leq 0.6\widehat{Y}_{t;0.5}\}\pi_{t;i}^{-1}$ and $w_{t;i} = \pi_{t;i}^{-1}$. The multivariate model is given as follows,

$$\begin{pmatrix} \breve{p}_{1;i} \\ w_{1;i} \\ \breve{p}_{2;i} \\ w_{2;i} \end{pmatrix} = \begin{pmatrix} \alpha_{1;1} z_{1;i} + \alpha_{1;2} z_{2;i} + \alpha_{1;3} z_{1;i} \times z_{2;i} \\ \beta_{1;1} z_{1;i} + \beta_{1;2} z_{2;i} + \beta_{1;3} z_{1;i} \times z_{2;i} \\ \alpha_{2;1} z_{1;i} + \alpha_{2;2} z_{2;i} + \alpha_{2;3} z_{1;i} \times z_{2;i} \\ \beta_{2;1} z_{1;i} + \beta_{2;2} z_{2;i} + \beta_{2;3} z_{1;i} \times z_{2;i} \end{pmatrix} + \boldsymbol{\epsilon}_i. \tag{3}$$

The vector of the residuals follow a multivariate distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}$. Rotation of the sampling design is incorporated into the model through the model covariates: $z_{t;i} = \delta\{i \in s_t\}$ and $z_{1;i} \times z_{2;i} = \delta\{i \in s_1, i \in s_2\}$. It should be noted that the correlations $\widehat{\text{corr}}(\widehat{\tau}_k, \widehat{\tau}_\ell)$, with $(k, \ell = 1, 2, 3, 4)$, are obtained from the estimated residual covariance matrix $\widehat{\boldsymbol{\Sigma}}$. The covariance terms on the non-diagonal part of the matrix $\widehat{\textbf{var}}(\widehat{\boldsymbol{\tau}})$ are based on those estimated correlations $\widehat{\text{corr}}(\widehat{\tau}_k, \widehat{\tau}_\ell)$ and the estimated cross-sectional variance terms $\widehat{\text{var}}(\widehat{\tau}_k)$. Note that this approach also accounts for a multi-stage sampling, using an 'ultimate cluster approach' (e.g. Osier et al. 2013; Di Meglio et al. 2013).

Berger and Priam (2010, 2013) showed that the multivariate approach gives estimates which are approximately equal to the Hansen and Hurwitz (1943) variance estimator (e.g. Holmes and Skinner 2000).

The proposed approach can be easily extended to a stratified sampling. In this case, we assume that the sample sizes within each stratum are fixed (non-random) quantities. The model covariates $z_{t;i}$ are replaced by the stratum wave indicators $z_{th;i} = \delta\{i \in s_{th}\}$, where s_{th} is the sample for the stratum h at wave t. As the rotation is done within each stratum, we consider the interactions $z_{th;i} \times z_{(t+1)h;i}$.

4 Allowing for the variability of the poverty threshold

Note that in (2), the variability of the poverty threshold is not taken into account because we treat $\widehat{\theta}_1$ and $\widehat{\theta}_2$ as ratios. Treating the poverty threshold as fixed might lead over-estimation of the variances (e.g. Preston 1995; Berger and Skinner 2003; Verma and Betti 2011). Verma and Betti (2011) compared the ratio variance estimator (i.e. when the poverty threshold is treated as fixed) with linearisation and Jackknife repeated replication. They found that the ratio variance estimator over-estimated the standard errors for all the poverty measures and several complex statistics. However, these findings are related to the cross-sectional estimators and do not necessarily hold for the variance of change.

Taking into account the whole variability means that the sampling variation of the poverty threshold is also considered. However, the poverty rate is more complex than a ratio and cannot be expressed as a function of totals. We propose to use the linearisation approach proposed by Deville (1999). The implementation of this approach for the poverty rate and the inequality measures can be found in the literature (e.g. Berger and Skinner 2003; Verma and Betti 2005; Osier 2009; Muennich and Zins 2011; Verma and Betti 2011).

The linearised variable $L_{t;i}$ for individual i at wave t for the poverty rate is given by (see Osier 2009)

$$L_{t;i} = \frac{1}{\widehat{N}_t}\left(\delta\{y_{t;i} \leq 0.6\widehat{Y}_{t;0.5}\} - \widehat{\theta}_t\right) - \frac{0.6}{\widehat{N}_t}\frac{\widehat{f}_t(0.6\widehat{Y}_{t;0.5})}{\widehat{f}_t(\widehat{Y}_{t;0.5})}\left(\delta\{y_{t;i} \leq \widehat{Y}_{t;0.5}\} - 0.5\right), \tag{4}$$

where $\widehat{f}_t(\cdot)$ is an estimator of the density function, which is defined in (5). The second term in (4) is an additional term which reflects the sample variation originated from the randomness of the estimated median income.

The density functions can be estimated on the basis of the Gaussian kernel function as follows (e.g. Preston 1995):

$$\widehat{f}_t(x) = \frac{1}{\widehat{N}_t \widehat{h}_t} \sum_{i \in s_t} \frac{1}{\pi_{t;i}} K\left(\frac{x - y_{t;i}}{\widehat{h}_t}\right) \tag{5}$$

where $K(\eta) = (\sqrt{2\pi})^{-1}\exp(-\eta^2/2)$ is the Gaussian Kernel, $\widehat{N}_t = \sum_{i \in s_t} \pi_{t;i}^{-1}$ is the Horvitz and Thompson (1952) estimator of the population size at wave t ($t = 1, 2$), and $\widehat{h}_t$ is the bandwidth parameter that can be defined in several ways (Silverman 1986, p.45-48). For a normally distributed population and smooth densities, the following bandwidth parameter is recommended by Silverman (1986, p.46).

$$\widehat{h}_t = 1.06 \widehat{\sigma}_{t;\widehat{Y}} \widehat{N}_t^{-1/5}, \tag{6}$$

where

$$\widehat{\sigma}_{t;\widehat{Y}} = \sqrt{\frac{1}{\widehat{N}_t}\left\{\sum_{i \in s_t} \frac{1}{\pi_{t;i}} y_{t;i}^2 - \frac{1}{\widehat{N}_t}\left(\sum_{j \in s_t} \frac{1}{\pi_{t;j}} y_{t;j}\right)^2\right\}},$$

is the estimated standard deviation of the income distribution. However, for skewed and long-tailed distributions, Silverman (1986, p.47) proposed to use the inter-quartile range instead of the standard deviation of the distribution; that is,

$$\widehat{h}_t = 0.79 \widehat{Y}_{t;iqr} \widehat{N}_t^{-1/5}, \tag{7}$$

where $\widehat{Y}_{t;iqr} = \widehat{Y}_{t;0.75} - \widehat{Y}_{t;0.25}$ is the weighted inter-quartile-range of the income distribution. Another bandwidth, which is very suitable for many densities, and even for the modest bimodal ones, is suggested by Silverman (1986, p.48) as follows:

$$\widehat{h}_t = 0.9\widehat{A}_t\widehat{N}_t^{-1/5}, \tag{8}$$

where $\widehat{A}_t = \min(\widehat{\sigma}_{t;\widehat{Y}}, \widehat{Y}_{t;iqr}/1.34)$. It should be noted that the bandwidth in (8) is smaller than the other bandwidths in (6) and (7). Thus, we are likely to obtain less smooth densities with the bandwidth (8).

It is worth mentioning that choosing a bandwidth parameter is a crucial step in applications (e.g. Verma and Betti 2005; Graf 2013; Graf and Tillé 2014). For example, Verma and Betti (2005) showed that probability density functions are quite sensible to the chosen bandwidth parameter. A large value for the bandwidth parameter results in a smoother densities. Graf (2013, p.26-28) points out the potential danger of using standard deviation when estimating densities that might be arised from extreme values in the observed data (for example, with income data). In such cases, Graf (2013) proposed to use the logarithm to reduce the adverse impact of extreme values. He also remarks a fixed-bandwidth parameter might be problematic when observations are heaped up around some values. To avoid this problem, a more robust technique to estimate density involving nearest neighbours with minimal bandwidth is suggested by Graf (2013).

5 Estimation of change within domains

In practice, we are often interested in change within domains of interest. For example, we may be interested in change in poverty within different age groups. According to the definition given by the Eurostat (2003), the poverty threshold is calculated based on the overall estimated median income rather than the estimated median income over the domains. Hence, when we are interested in a domain, the threshold will be the same for all domains.

Consider $d_{t;i}$ to be a domain indicator for individual i at wave t defined by

$$d_{t;i} = \begin{cases} 1 & \text{if } i \in D \text{ at wave } t, \\ 0 & \text{if } i \notin D \text{ at wave } t, \end{cases}$$

where D refers to the domain of interest. The poverty rate over a domain is defined by

$$\widehat{\theta}_{Dt} = \frac{\sum_{i \in s_t} d_{t;i} \delta\{y_{t;i} \leq \widehat{Y}_{t;0.5}\} \pi_{t;i}^{-1}}{\sum_{t \in s_1} d_{t;i} \pi_{t;i}^{-1}}.$$

To estimate the variance of change within domains under the ratio approach (see (2)), we substitute $\breve{p}_{t;i}$ by $\breve{p}_{Dt;i} = d_{t;i}\breve{p}_{t;i}$, and $w_{t;i}$ by $w_{Dt;i} = d_{t;i}w_{t;i}$ into the model in (3). Note that the values of the response variables will be equal to zero for the units not included in the domain of interest.

For the linearisation approach, the linearised variables $L_{Dt;i}$ for individual i in domain D at wave t derived in Appendix B (B.5) is given by

$$L_{Dt;i} = \frac{d_{t;i}}{\widehat{N}_{Dt}} \left(\delta\{y_{t;i} \leq 0.6\widehat{Y}_{t;0.5}\} - \widehat{\theta}_{Dt} \right) - \frac{0.6}{\widehat{N}_t} \frac{\widehat{f}_{Dt}(0.6\widehat{Y}_{t;0.5})}{\widehat{f}_t(\widehat{Y}_{t;0.5})} \left(\delta\{y_{t;i} \leq \widehat{Y}_{t;0.5}\} - 0.5 \right),$$

where

$$\widehat{N}_{Dt} = \sum_{i \in s_t} \frac{d_{t;i}}{\pi_{t;i}},$$

$$\widehat{f}_{Dt}(x) = \frac{1}{\widehat{N}_{Dt}\widehat{h}_{Dt}} \sum_{i \in s_t} \frac{d_{t;i}}{\pi_{t;i}} K_D \left(\frac{x - y_{t;i}}{\widehat{h}_{Dt}} \right),$$

where $\widehat{h}_{Dt}$ can be (6), (7), or (8) with $\widehat{N}_{Dt}$, $\widehat{Y}_{Dt;iqr} = \widehat{Y}_{Dt;0.75} - \widehat{Y}_{Dt;0.25}$,

$$\widehat{\sigma}_{Dt;\widehat{Y}} = \sqrt{\frac{1}{\widehat{N}_{Dt}} \left\{ \sum_{i \in s_t} \frac{d_{t;i}}{\pi_{t;i}} y_{t;i}^2 - \frac{1}{\widehat{N}_{Dt}} \left(\sum_{j \in s_t} \frac{d_{t;j}}{\pi_{t;j}} y_{t;j} \right)^2 \right\}},$$

and $\widehat{A}_{Dt} = \min(\widehat{\sigma}_{Dt;\widehat{Y}}, \widehat{Y}_{Dt;iqr}/1.34)$. Thus, the variance of domain change is estimated by

$$\widehat{\text{var}}(\widehat{\Delta}) = \widehat{\text{var}}(\widehat{\theta}_1^L) + \widehat{\text{var}}(\widehat{\theta}_2^L) - 2\widehat{\text{corr}}(\widehat{\theta}_1^L, \widehat{\theta}_2^L)\sqrt{\widehat{\text{var}}(\widehat{\theta}_1^L)\widehat{\text{var}}(\widehat{\theta}_2^L)}, \tag{9}$$

with

$$\widehat{\theta}_t^L = \sum_{i \in s_t} \frac{L_{Dt;i}}{\pi_{t;i}} \, . \tag{10}$$

We use the approach proposed by Berger and Priam (2010, 2013) by treating $\widehat{\theta}_1^L$ and $\widehat{\theta}_2^L$ in (10) as the estimators of totals. The correlation term $\widehat{\text{corr}}(\widehat{\theta}_1^L, \widehat{\theta}_2^L)$ in (9) is computed from the residual covariance matrix $\mathbf{\Sigma}$ of the following model,

$$\begin{pmatrix} \breve{L}_{D1;i} \\ \breve{L}_{D2;i} \end{pmatrix} = \begin{pmatrix} \alpha_{1;1} z_{1;i} + \alpha_{1;2} z_{2;i} + \alpha_{1;3} z_{1;i} \times z_{2;i} \\ \alpha_{2;1} z_{1;i} + \alpha_{2;2} z_{2;i} + \alpha_{2;3} z_{1;i} \times z_{2;i} \end{pmatrix} + \boldsymbol{e}_i,$$

with $\breve{L}_{Dt;i} = L_{Dt;i}\pi_{t;i}^{-1}$.

It should be noted that the domain information is incorporated into the model through the response variables, in contrast to the stratification (see Section 3). Note that the proposed approach can be used for strata domains (with fixed sample size) and unplanned domains (with random sample size).

6 Simulation study

In this section, the variance estimators from the ratio and the linearisation approaches are compared in terms of the relative bias (RB) and the root mean square error (RRMSE), respectively defined by (11) and (12). Additionally, we investigate whether the ratio approach gives more conservative estimates.

The income variables at wave 1 and wave 2 are generated according to different probability distributions (see Appendix A). For each wave, a Gamma distribution (shape=2.5, rate=1), a Lognormal distribution (mean=1.119, standard deviation=0.602) and a Weibull distribution (shape=0.8, scale=1) is used to generate populations with a size of $N = 20\ 940$. As stated by Salem and Mount (1974) and McDonald (1984), these distributions are good approximations of income distributions. The correlation coefficient between the variables of the first and the second wave is given by $\rho = 0.94$, which is the correlation observed from the common sample of the Turkish EU-SILC survey data. Note that this correlation and the correlation in (1) are different; in other words, the correlation $\rho = 0.94$ is the correlation between the variables of interest, whereas the correlation in (1) is the correlation between the point estimators.

The population is assumed fixed and the same sample size is used for both waves. We have 1047 primary sampling units in the Turkish EU-SILC survey data. For this reason, we use $n_1 = n_2 = 1047$ units for each wave. The fraction of the common sample is 75%. Hence, the number of units in the common sample is $n_c = 785$. Unequal and equal probabilities are used to select the samples. For unequal probability sampling (πps) design, the Chao (1982) sampling design is used. The first wave samples are selected without replacement with the inclusion probabilities proportional to a size variable x_i, which is generated by the model $x_i = \alpha + \rho y_{1;i} + e_i$, with $e_i \sim N(0, (1-\rho^2)\sigma^2_{y_1})$, $\alpha = 5$, and $\rho = 0.7$. For the second wave, a simple random sample of n_c units are selected from the sample s_1; and, $n_2 - n_c$ units are selected with the probabilities proportional to size $q_i = \pi_{1;i}/(1-\pi_{1;i})$ from the population $U \backslash s_1$. It can be shown that $\pi_{2;i} \approx \pi_{1;i}$ (Christine and Rocher 2012). For equal probability sampling designs, $\pi_{2;i} = \pi_{1;i} = n_1/N$.

We did six simulation studies for three populations and two sampling designs. For each simulation, 10 000 samples are selected. For each sample, the RB and the RRMSE are computed for the cross-sectional variance estimators, the variance estimator of change and the estimator of the correlation. The RB and the RRMSE are defined by

$$RB(\widehat{\sigma}) = \frac{E(\widehat{\sigma}) - \sigma}{\sigma} 100\%, \tag{11}$$

$$RRMSE(\widehat{\sigma}) = \frac{\sqrt{(B-1)^{-1}\sum_{b=1}^{B}(\widehat{\sigma}_b - \sigma)^2}}{\sigma} 100\%, \tag{12}$$

where $E(\widehat{\sigma}) = B^{-1}\sum_{b=1}^{B}\widehat{\sigma}_b$, with $B = 10\,000$, is the empirical expectation; σ is either the empirical variances or the empirical correlation in (1); $\widehat{\sigma}$ is the estimator of the quantity σ; $\widehat{\sigma}_b$ is the estimate of the quantity σ for the b-th sample. For the linearisation, we consider three bandwidth parameters (see (6), (7) and (8)). The linearisation based on (6), (7) and (8) are respectively labelled as Lin_Sd, Lin_Iqr, and Lin_A in Table 1 and Table 2.

For a Gamma distribution, the poverty rates are 24.2% and 23.6% for the first and the second wave respectively. Hence, we have -0.59% point change between two waves. For a Lognormal distribution, the poverty rates are 19.4% and 19.9%. Thus, there is a 0.54% point change for this case. For a Weibull distribution, we have the highest poverty rates, which are 36.6% and 37.3% respectively. Hence, the change is 0.66% point.

Table 1 shows the RB (%) of the variance and the correlation estimators for several distributions and sampling designs. Overall, the linearisation approach has lower RB comparing the ratio one. Thus, we have more accurate estimates with the linearisation. Differences between two approaches in terms of the RB is much more pronounced for the Weibull distribution, which is the most skewed distribution. For all situations except the Lognormal distribution, the ratio approach overestimates all the variances and the correlations. Therefore, the ratio approach may not always give more conservative estimates. However, note that whenever we have a positive bias, we obtain relatively larger variance estimates with

Table 1: Empirical RB (%) of the Variance and Correlation Estimators for the Poverty Rates for Three Distributions and Two Sampling Designs

	Relative Bias (%)							
	Gamma							
	SRS				πps			
	Ratio	Lin_Sd	Lin_Iqr	Lin_A	Ratio	Lin_Sd	Lin_Iqr	Lin_A
Var Wave1	41.3	2.4	2.6	3.1	50.9	7.2	7.4	7.8
Var Wave2	42.8	5.1	5.3	5.8	41.1	2.9	3.0	3.5
Var Change	8.1	1.0	1.2	1.8	13.0	2.1	2.4	2.9
Correlation	23.2	2.6	2.6	2.5	22.0	2.7	2.6	2.5
	LogNormal							
	SRS				πps			
	Ratio	Lin_Sd	Lin_Iqr	Lin_A	Ratio	Lin_Sd	Lin_Iqr	Lin_A
Var Wave1	15.6	0.9	2.2	2.9	22.7	-0.5	0.5	1.0
Var Wave2	24.1	6.4	7.6	8.2	28.9	4.2	5.1	5.6
Var Change	-14.1	1.3	2.6	3.4	-8.7	0.5	1.7	2.4
Correlation	38.1	3.1	2.9	2.8	35.5	1.6	1.3	1.1
	Weibull							
	SRS				πps			
	Ratio	Lin_Sd	Lin_Iqr	Lin_A	Ratio	Lin_Sd	Lin_Iqr	Lin_A
Var Wave1	140.1	4.3	6.5	6.7	132.9	2.9	4.8	5.1
Var Wave2	146.0	1.9	4.0	4.2	137.6	1.0	2.9	3.1
Var Change	26.6	0.9	4.2	4.4	28.3	2.0	5.2	5.5
Correlation	152.0	6.6	3.3	3.3	132.1	-0.4	-3.8	-3.8

the ratio approach. When we compare the three linearisation methods based on different bandwidth, we obtained the largest RB with the smallest bandwidth (see (8)).

As far as the RRMSE is concerned (see Table 2), we have more precise estimates for the linearisation approach. We observe the smallest RRMSE with the bandwidth (6) and the largest RRMSE with the bandwidth (8). The ratio approach gives less accurate point estimates. However, the differences between the two approaches can be negligible for the variance of change, except the Weibull distribution.

Table 2: Empirical RRMSE (%) of the Variance and Correlation Estimators for the Poverty Rates for Three Distributions and Two Sampling Designs

	Relative Root Mean Square Error (%)							
	Gamma							
	SRS				πps			
	Ratio	Lin_Sd	Lin_Iqr	Lin_A	Ratio	Lin_Sd	Lin_Iqr	Lin_A
Var Wave1	41.5	4.8	5.1	5.9	51.2	8.2	8.4	9.0
Var Wave2	36.8	6.8	7.1	7.9	41.4	4.9	5.1	5.9
Var Change	10.9	7.9	8.1	8.6	15.4	8.7	8.8	9.4
Correlation	20.0	6.0	6.5	6.5	22.7	7.3	7.3	7.3
	LogNormal							
	SRS				πps			
	Ratio	Lin_Sd	Lin_Iqr	Lin_A	Ratio	Lin_Sd	Lin_Iqr	Lin_A
Var Wave1	16.4	4.9	6.2	7.2	30.6	7.8	8.2	8.7
Var Wave2	24.6	8.1	9.8	10.8	35.2	8.9	9.8	10.5
Var Change	15.1	7.0	8.0	8.8	18.5	10.6	11.2	11.7
Correlation	38.5	7.1	7.0	7.0	37.4	11.1	11.0	11.0
	Weibull							
	SRS				πps			
	Ratio	Lin_Sd	Lin_Iqr	Lin_A	Ratio	Lin_Sd	Lin_Iqr	Lin_A
Var Wave1	140.1	6.5	8.5	9.0	133.0	6.5	8.0	8.5
Var Wave2	146.0	5.5	7.1	7.7	137.7	6.2	7.4	7.9
Var Change	27.1	5.7	7.8	8.4	29.1	7.0	9.3	9.9
Correlation	152.2	16.7	16.3	16.6	132.4	16.2	17.2	17.5

7 An application to the Turkish EU-SILC survey

The 2007 and 2008 cross-sectional Turkish EU-SILC survey data was used. The Turkish EU-SILC survey has a stratified two-stage cluster probability sampling design. For the first stage, address blocks are selected within each stratum with a probability proportional to size (πps) without replacement sampling design. Each block is composed of approximately 100 addresses. Households within the selected address blocks are selected with a systematic sampling design. All individuals within the selected households participate to the survey.

The cross-sectional survey weights in the 'personal register' file (RB050) were used as an inverse of the inclusion probabilities. The effect of calibration was not taken into account, because we did not have any information about the auxiliary variables. The effect of imputation was ignored for the same reason.

In Table 3, we have the estimates for several domains when the poverty threshold is treated as fixed (see (2)). We observe a significant change for the domain tenant at the 5% level.

In Table 4, we have the estimates obtained with the linearisation approach based on the bandwidth in (6) described in Section 4. We also observe a highly significant change for the domain tenant. We do not observe major differences in the p-values between Table 3 and Table 4. We observe a slight decrease in the p-values when the sampling variation of the poverty threshold is taken into account. This is due to the fact that the variances of changes are larger in Table 3.

The correlations in Table 4 are smaller than in Table 3 overall. Hence, the estimated correlations are smaller when the variability of the poverty threshold is taken into account.

By comparing Table 3 and Table 4, we also found that all variances were estimated more conservatively when the threshold is treated as fixed. Preston (1995), Berger and Skinner (2003), and Verma and Betti (2011) demonstrated that the cross-sectional variances are more conservative when the poverty threshold is treated as fixed. This finding is explained by Preston (1995) by the fact that the two sources of variability offset each other. This is more pronounced when the high fractions of the median are used.

For the variance of change, we cannot anticipate an increase in the variance when the poverty threshold is treated as fixed for the following reason. Let us assume that the cross-

Table 3: Estimates when the poverty threshold is treated as fixed (see (2))

Domain	Pov'07(%)	Var'07	Pov'08	Var'08(%)	Change(in % point)	Var Change	Corr	p-value
Turkey	23.4	0.616	24.1	0.644	0.7	0.447	0.65	0.297
Male	23.0	0.650	23.7	0.665	0.7	0.494	0.62	0.328
Female	23.8	0.639	24.6	0.678	0.7	0.465	0.65	0.299
Owner	24.9	0.739	23.8	0.872	-1.1	0.593	0.63	0.140
Tenant	18.5	1.395	25.3	1.511	6.7	1.522	0.48	0.000
0_14	33.5	1.164	34.5	1.258	1.1	0.882	0.64	0.263
15_24	24.2	1.162	25.3	1.181	1.1	1.118	0.52	0.296
25_49	19.8	0.527	20.7	0.548	0.9	0.405	0.62	0.178
50_64	14.4	0.568	15.0	0.719	0.6	0.569	0.56	0.404
65+	17.7	1.077	16.2	0.929	-1.5	0.988	0.51	0.120

Source: 2007 and 2008 cross-sectional data of the EU-SILC survey for Turkey conducted by TurkStat.

Table 4: Estimates when the sampling variation of the poverty threshold taken into account (see Sections 4 and 5). The bandwidth parameter is based on the standard deviation of the income distribution (see (6)).

Domain	Pov'07(%)	Var'07	Pov'08	Var'08(%)	Change(in % point)	Var Change	Corr	p-value
Turkey	23.4	0.292	24.1	0.290	0.7	0.372	0.36	0.252
Male	23.0	0.314	23.7	0.306	0.7	0.416	0.33	0.287
Female	23.8	0.327	24.6	0.327	0.7	0.390	0.40	0.257
Owner	24.9	0.417	23.8	0.495	-1.1	0.527	0.42	0.117
Tenant	18.5	1.121	25.3	1.238	6.7	1.435	0.39	0.000
0_14	33.5	0.796	34.5	0.793	1.1	0.787	0.50	0.236
15_24	24.2	0.790	25.3	0.919	1.1	1.050	0.39	0.281
25_49	19.8	0.255	20.7	0.252	0.9	0.362	0.29	0.154
50_64	14.4	0.403	15.0	0.491	0.6	0.476	0.47	0.361
65+	17.7	0.929	16.2	0.807	-1.5	0.978	0.44	0.118

Source: 2007 and 2008 cross-sectional data of the EU-SILC survey for Turkey conducted by TurkStat.

sectional variances are equal: $\widehat{\text{var}}(\widehat{\theta}_1) = \widehat{\text{var}}(\widehat{\theta}_2)$. Thus, the variance estimator of change is given by $\widehat{\text{var}}(\widehat{\Delta}) = 2\widehat{\text{var}}(\widehat{\theta}_1)(1 - \widehat{\text{corr}}(\widehat{\theta}_1, \widehat{\theta}_2))$. Hence, the variance of change is affected in the same direction by the variance term, and in the opposite direction by the correlation term. Thus, when both the variance and the correlation terms increase or decrease concurrently, the direction of the effect on the variance of change cannot be predicted. Therefore, we may not necessarily have more conservative estimates of the variance of change when the poverty threshold is treated as fixed. With the Turkish EU-SILC survey data, we found that the variances of changes were more conservative, although the differences between the two approaches were not as pronounced as the differences between the cross-sectional variances (see Table 3 and Table 4).

In Table 4, the bandwidth parameter is given by (6). We also investigate the situations when the bandwidth parameter is given by (7) and (8). The results are given in Table 5 and Table 6. By comparing Table 5 and Table 6 with Table 3, we also observed smaller cross-sectional variances, variance of change and correlation when the bandwidth parameter is (7) and (8). When we compare Table 4, Table 5, and 6, the estimates do not differ significantly between the three linearisation approaches based on different bandwidth parameters, although we observe slight differences among them in terms of the RB and the RRMSE in the simulation study (see Section 6).

Table 5: Estimates when the sampling variation of the poverty threshold taken into account (see Sections 4 and 5). The bandwidth parameter is based on the inter-quartile range of the income distribution (see (7)).

Domain	Pov'07(%)	Var'07	Pov'08	Var'08(%)	Change(in % point)	Var Change	Corr	p-value
Turkey	23.4	0.292	24.1	0.290	0.7	0.372	0.36	0.252
Male	23.0	0.316	23.7	0.306	0.7	0.416	0.33	0.287
Female	23.8	0.325	24.6	0.328	0.7	0.391	0.40	0.257
Owner	24.9	0.418	23.8	0.497	-1.1	0.530	0.42	0.118
Tenant	18.5	1.117	25.3	1.226	6.7	1.428	0.39	0.000
0_14	33.5	0.802	34.5	0.814	1.1	0.805	0.50	0.241
15_24	24.2	0.787	25.3	0.907	1.1	1.038	0.39	0.278
25_49	19.8	0.256	20.7	0.251	0.9	0.361	0.29	0.154
50_64	14.4	0.403	15.0	0.491	0.6	0.476	0.47	0.361
65+	17.7	0.946	16.2	0.791	-1.5	0.976	0.44	0.118

Source: 2007 and 2008 cross-sectional data of the EU-SILC survey for Turkey conducted by TurkStat.

Table 6: Estimates when the sampling variation of the poverty threshold taken into account (see Sections 4 and 5). The bandwidth parameter is based on the parameter A (see (8)).

Domain	Pov'07(%)	Var'07	Pov'08	Var'08(%)	Change(in % point)	Var Change	Corr	p-value
Turkey	23.4	0.291	24.1	0.291	0.7	0.372	0.36	0.253
Male	23.0	0.316	23.7	0.306	0.7	0.416	0.33	0.287
Female	23.8	0.324	24.6	0.329	0.7	0.392	0.40	0.258
Owner	24.9	0.419	23.8	0.498	-1.1	0.531	0.42	0.119
Tenant	18.5	1.114	25.3	1.223	6.7	1.425	0.39	0.000
0_14	33.5	0.802	34.5	0.823	1.1	0.812	0.50	0.243
15_24	24.2	0.787	25.3	0.903	1.1	1.034	0.39	0.277
25_49	19.8	0.255	20.7	0.251	0.9	0.361	0.29	0.154
50_64	14.4	0.403	15.0	0.491	0.6	0.476	0.47	0.361
65+	17.7	0.949	16.2	0.788	-1.5	0.977	0.44	0.118

Source: 2007 and 2008 cross-sectional data of the EU-SILC survey for Turkey conducted by TurkStat.

8 Conclusion

We applied a simple approach to estimate the variances of changes for the poverty rates over several domains by using the 2007-2008 Turkish EU-SILC survey data. It involves a multivariate linear regression model proposed by Berger and Priam (2010), which can be easily applied. Survey characteristics such as rotation, stratification, and cluster sampling are all taken into account. The proposed approach is flexible and can be implemented for most of the EU-SILC surveys as long as sampling fractions are negligible. This assumption implies that the second order inclusion probabilities are not needed.

We have two ways to estimate the variances depending on whether we treat the poverty threshold as fixed or not. When we treated as fixed, we obtained more conservative variance estimates of change with the Turkish EU-SILC survey data. However, our simulation study shows that treating the threshold as fixed does not necessarily give more conservative variance estimates of change. For the Lognormal distribution, for example, variances of changes are underestimated with the ratio method. On the other hand, differences between the variance estimators of changes can be negligible in terms of the RB and the RRMSE, even though we observe significant differences between the cross-sectional variances and the correlations. For the latter, the linearisation approach gives more unbiased and more precise variance estimates. Thus, based upon our results and due to the fact that linearisation involves complex numerical computations, the simple ratio approach may sound preferable to estimate the variance of change for the poverty rates. Albeit, we should be careful with highly skewed distributions similar to a Weibull one. As in this case, the linearisation approach is significantly better.

The proposed approach can also be used to estimate the variances of the other poverty and income inequality measures like the relative median at-risk of poverty gap (RMPG), the quantile share ratio (QSR) and the GINI coefficient, which are included in the 'Laeken' indicators (Eurostat 2003), by using the linearisation (e.g. Berger 2008). The RMPG and the GINI coefficient can not be treated as a simple ratio, whereas the QSR can be. The linearised variables of many complex parameters are given by Verma and Betti (2005, 2011).

In this paper, we implemented the fixed-bandwidth kernel method for simplicity (Silverman 1986, p.95). Note that, the bandwidth in (8) is a suitable choice for a wide range of densities as pointed out by Silverman (1986). If there is a more concern in the tails, then, an adaptive kernel method can be applied (Silverman 1986, chap.5). This method uses a variable bandwidth; that is, for each observed data point, a different bandwidth is computed. It would be interesting to check if an adaptive bandwidth improved the variance estimation in the presence of outliers.

Appendix A Generation of the Income Variables for the Simulation Study

For the Gamma random variables, we used the algorithm proposed by Schmeiser and Lal (1982, p.358). First, three independent random variables are generated by a Gamma distribution as follows:

$$Y_1 \sim Gamma(\alpha_1 - \rho\sqrt{\alpha_1}\sqrt{\alpha_2}, 1),$$

$$Y_2 \sim Gamma(\alpha_2 - \rho\sqrt{\alpha_1}\sqrt{\alpha_2}, 1),$$

$$Y_3 \sim Gamma(\rho\sqrt{\alpha_1}\sqrt{\alpha_2}, 1),$$

with $\alpha_1 = 2.5$, $\alpha_2 = 2.5$, and $\rho = 0.94$. Then, the income variables are obtained by the following expressions: $y_{1;i} = Y_1 + Y_3$ and $y_{2;i} = Y_2 + Y_3$, so that $y_{1;i} \sim Gamma(2.5, 1)$, $y_{2;i} \sim Gamma(2.6, 1)$, and $\rho(y_{1;i}, y_{2;i}) \approx 0.94$.

The Cholesky decomposition is used to generate the correlated Lognormal variables. Hence, the log income variables with the correlation of $\rho = 0.95$, a mean of $\mu = 1.119$ and a standard deviation of $\sigma = 0.602$ are generated by

$$\log(y_{1;i}) = \mu + \sigma X_1,$$

$$\log(y_{2;i}) = \mu + \rho\sigma X_1 + \sqrt{1-\rho^2}\sigma X_2,$$

where X_1 and X_2 are independent standard normal variables. The correlation coefficient between the income variables is approximately 0.94.

For a correlated Weibull variables, we followed the algorithm proposed by Feiveson (2002, p.117). First, two correlated standard normal variables Y_1 and Y_2 with a correlation of $\rho = 0.95$ are generated by using the Cholesky decomposition: $Y_1 = X_1$ and $Y_2 = \rho X_1 + \sqrt{1-\rho^2} X_2$, where X_1 and X_2 are independent standard normal variables. Secondly, correlated uniform variables are obtained by the standard normal cumulative distribution function transformation; such that $U_1 = \Phi(Y_1)$ and $U_2 = \Phi(Y_2)$, where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. Finally, uniform random variables are transformed by the inverse of the Weibull cumulative distribution function to achieve the correlated income variables as follows: $y_{1;i} = \mathrm{F}_U^{-1}(U_1) = (-\ln(1-U_1))^{5/4}$ and $y_{2;i} = \mathrm{F}_U^{-1}(U_2) = (-\ln(1-U_2))^{5/4}$, so that $y_{1;i}, y_{2;i} \sim Weibull(0.8, 1)$ and $\rho(y_{1;i}, y_{2;i}) \approx 0.94$.

Appendix B Derivation of the influence function of the poverty rate over a domain

Let M be a measure that assigns a unit mass to each unit i in the population U. For example, the population size N can be written as $N = \int d\mathrm{M} = \sum_{i \in U} 1$ and the total of a variable y can be expressed as $N = \int y d\mathrm{M} = \sum_{i \in U} y_i$ (Deville 1999). Let $\mathrm{F}(\mathrm{M}, x)$ be the income distribution function at x over the population U, that is,

$$\mathrm{F}(\mathrm{M}, x) = \frac{1}{N} \sum_{i \in U} \delta\{y_i \leq x\}.$$

Then, the income distribution function at the median of the income distribution is given by $\mathrm{F}(\mathrm{M}, Med(\mathrm{M})) = 0.5$. Thus, the influence function of the functional $\mathrm{F}(\mathrm{M}, Med(\mathrm{M}))$ at i is equal to 0, that is, $\mathrm{IF}_i(\mathrm{M}, Med(\mathrm{M})) = 0$. By using the 'Rule 7' in Deville (1999, p.198), the influence function of F at i (see also Osier 2009, p.181-183) can be derived as follows:

$$\mathrm{IF}_i(\mathrm{M}, Med(\mathrm{M})) = \mathrm{IF}_i(\mathrm{M}, Med(\mathrm{M})|_{Med(\mathrm{M})\,\mathrm{fixed}}) + \frac{\partial \mathrm{F}(\mathrm{M}, x)}{\partial x}|_{x=Med(\mathrm{M})}\mathrm{IMed}_i(\mathrm{M}) = 0. \quad \text{(B.1)}$$

The influence function of F, when the median is fixed, is given by

$$\mathrm{IF}_i(\mathrm{M}, Med(\mathrm{M})|_{Med(\mathrm{M})\,\mathrm{fixed}}) = \frac{1}{N}[\delta\{y_i \leq Med\} - 0.5].$$

Thus, the influence function of the functional $Med(\mathrm{M})$ is obtained as

$$\mathrm{IMed}_i(\mathrm{M}) = -\frac{1}{N}\frac{1}{f(Med)}[\delta\{y_i \leq Med\} - 0.5], \quad \text{(B.2)}$$

where

$$f(Med) = \frac{\partial \mathrm{F}(\mathrm{M}, x)}{\partial x}|_{x=Med(\mathrm{M})}$$

is the probability density function at the median of the income distribution.

Now define the income distribution function at x over a domain D as follows:

$$\mathrm{F}_D(\mathrm{M}, x) = \frac{1}{N_D}\sum_{i\in U} d_i\delta\{y_i \leq x\}.$$

Hence, the income distribution function over a domain D at the poverty threshold T is defined by

$$\mathrm{F}_D(\mathrm{M}, T(\mathrm{M})) = \frac{1}{N_D}\sum_{i\in U} d_i\delta\{y_i \leq T(\mathrm{M})\},$$

where $T(\mathrm{M}) = 0.6Med(\mathrm{M})$ and d_i is the domain indicator, that is, 1 when $i \in D$, and 0 otherwise. $\mathrm{F}_D(\mathrm{M}, T(\mathrm{M}))$ is equivalent to the povery rate over a domain D (i.e. R_D). Thus, we can obtain the influence function of the poverty rate analogously to (B.1), that is,

$$\mathrm{IF}_{D;i}(\mathrm{M}, T(\mathrm{M})) = \mathrm{IF}_{D;i}(\mathrm{M}, T(\mathrm{M}))|_{T(\mathrm{M})\,\mathrm{fixed}}) + \frac{\partial \mathrm{F}_D(\mathrm{M}, x)}{\partial x}|_{x=T(\mathrm{M})}\mathrm{IT}_i(\mathrm{M}) = \mathrm{IR}_{D;i}.$$

The influence function of F_D, when the threshold is fixed, is given by

$$\mathrm{IF}_{D;i}(\mathrm{M}, T(\mathrm{M}))|_{T(\mathrm{M})\,\mathrm{fixed}}) = \frac{d_i}{N_D}[\delta\{y_i \leq T\} - R_D].$$

Hence, the influence function of the poverty rate is obtained as follows:

$$\mathrm{IR}_{D;i} = \frac{d_i}{N_D}[\delta\{y_i \leq T\} - R_D] + f_D(T)\mathrm{IT}_i(\mathrm{M}), \tag{B.3}$$

where

$$f_D(T) = \frac{\partial \mathrm{F}_D(\mathrm{M}, x)}{\partial x}|_{x=T(\mathrm{M})}$$

is the probability density function at the poverty threshold. The influence function of the functional $T(\mathrm{M})$ at i is given by

$$\mathrm{IT}_i(\mathrm{M}) = 0.6\,\mathrm{IMed}_i(\mathrm{M}). \tag{B.4}$$

If we substitute $\mathrm{IMed}_i(\mathrm{M})$ in (B.2) into (B.4), we obtain the following:

$$\mathrm{IT}_i(\mathrm{M}) = -\frac{0.6}{N}\frac{1}{f(Med)}[\delta\{y_i \leq Med\} - 0.5].$$

Therefore, the influence function of the poverty rate at i over a domain D given in (B.3) can be re-written as follows:

$$\mathrm{IR}_{D;i} = \frac{d_i}{N_D}[\delta\{y_i \leq T\} - R_D] - \frac{0.6}{N}\frac{f_D(T)}{f(Med)}[\delta\{y_i \leq Med\} - 0.5]. \tag{B.5}$$

Note that we assume the derivatives of F and F_D exist and are strictly non-negative for all x.

References

Berger, Y. G. (2004), "Variance estimation for measures of change in probability sampling," *Canadian Journal of Statistics*, 32(4), 451–467.

Berger, Y. G. (2008), "A note on the asymptotic equivalence of Jackknife and linearization variance estimation for the Gini coefficient," *Journal of Official Statistics*, 24(4), 541–555.

Berger, Y. G., Goedemé, T., and Osier, G. (2013), *Handbook on standard error estimation and other related sampling issues in EU-SILC* Second Network for the Analysis of EU-SILC, EuroStat. http://www.cros-portal.eu/content/handbook-standard-error-estimation-and-other-related-sampling-issues-ver-29072013. [Online; accessed 06 February 2013].

Berger, Y. G., and Priam, R. (2010), "Estimation of correlations between cross-sectional Estimates from Repeated Surveys - an Application to the Variance of Change," *Proceeding of the 2010 Symposium of Statistics Canada*, .

Berger, Y. G., and Priam, R. (2013), *A Simple Variance Estimator of Change for Rotating Repeated Surveys: an Application to the EU-SILC Household Surveys*, Southampton: Southampton Statistical Sciences Research Institute. http://eprints.soton.ac.uk/347142/.

Berger, Y. G., and Skinner, C. J. (2003), "Variance Estimation of a Low-Income Proportion," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52, 457–468.

Betti, G., and Gagliardi, F. (2007), Jackknife variance estimation of differences and averages of poverty measures,, Working Paper 68, Siena: Dipartimento di Metodi Quantitativi, Universit degli Studi.

Chao, M. T. (1982), "A General Purpose Unequal Probability Sampling Plan," *Biometrika*, 69, 653–656.

Christine, M., and Rocher, T. (2012), "Construction d'échantillons astreints à des conditions de recouvrement par rapport un échantillon antérieur et à des conditions d'équilibrage par rapport à des variables courantes," *Proceeding of the 10th Journée de Méthodologie Statistique de l'INSEE (Paris, 24-26 January 2012)*, .

Demnati, A., and Rao, J. N. K. (2004), "Linearization variance estimators for survey data," *Survey Methodology*, 30, 17–26.

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

Di Meglio, E., Osier, G., Goedemé, T., Berger, Y. G., and Di Falco, E. (2013), *Standard Error Estimation in EU-SILC - First Results of the Net-SILC2 Project*, Brussels: Proceeding of the conference on New Techniques and Technologies for Statistics, Brussels. http://www.cros-portal.eu/sites/default/files/NTTS2013fullPaper_144.pdf.

Eurostat (2003), "'Laeken' Indicators-Detailed Calculation Methodology,", http://www.cso.ie/en/media/csoie/eusilc/documents/Laeken%20Indicators%20-%20calculation%20algorithm.pdf. [Online; accessed 4 Feb. 2014].

Eurostat (2012), "European Union Statistics on Income and Living Conditions (EU-SILC),", http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc. [Online; accessed 7 Jan. 2013].

Feiveson, A. H. (2002), "Power by simulation," *The STATA Journal*, 2, 107–124.

Gambino, J. G., and Silva, P. L. N. (2009), "Sampling and estimation in household surveys," *Handbook of Statistics: Design, Method and Applications: D. Pfeffermann and C.R. Rao.(editors). Elsevier*, 29A, 407–439.

Goedeme, T. (2010), The standard error of estimates based on EU-SILC. An exploration through the Europe 2020 poverty indicators.,, Working paper 10/09. [Online; accessed 4 Feb. 2014].

Graf, E. (2013), *Variance estimation by linearization for indicators of poverty and social exclusion in a person and household survey context* Presented at New Techniques and Technologies for Statistics, Brussels. http://www.cros-portal.eu/content/14a01ericgraf. [Online; accessed 5 Feb. 2014].

Graf, E., and Tillé, Y. (2014), "Estimation de variance par linéarisation pour des indices de pauvreté et dexclusion sociale," *Techniques d'enquête, a paraître*, pp. 1–24.

Hansen, M. H., and Hurwitz, W. N. (1943), "On the Theory of Sampling from Finite Populations," *The Annals of Mathematical Statistics*, 14(4), pp. 333–362.

Holmes, D. J., and Skinner, C. J. (2000), "Variance estimation for Labour Force Survey estimates of level and change," *Government Statistical Service Methodology Series*, (21).

Horvitz, D. G., and Thompson, D. J. (1952), "A Generalization of Sampling Without Replacement From a Finite Universe," *Journal of the American Statistical Association*, 47(260), 663–685.

Kalton, G. (2009), "Design for surveys over time," *Handbook of Statistics: Design, Method and Applications: D. Pfeffermann and C.R. Rao.(editors). Elsevier*, 29A, 89–108.

Kish, L. (1965), *Survey Sampling* New York: John Wiley and Sons.

Laniel, N. (1987), "Variances for a rotating sample from a changing population," *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 496–500.

McDonald, J. B. (1984), "Some generalized functions for the size distribution of income," *Econometrica*, 52, 647–664.

Muennich, R., and Zins, S. (2011), "Variance Estimation for Indicators of Poverty and Social Exclusion,", Work-package of the European project on Advanced Methodology for European Laeken Indicators (AMELI) http://www.uni-trier.de/index.php?id=24676. [Online; accessed 4 Jan. 2013].

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, 363–378.

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

Osier, G., Berger, Y. G., and Goedemé, T. (2013), "Standard Error Estimation for the EU-SILC Indicators of Poverty and Social Exclusion," *Eurostat Methodologies and Working Papers series*, .

Preston, I. (1995), "Sampling distributions of relative poverty statistics," *Appl. Statist.*, 44, 91–99.

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), 173–181.

Salem, A. B. Z., and Mount, T. D. (1974), "A convenient descriptive model of income distribution: the Gamma density," *Econometrika*, 42, 1115–1127.

Schmeiser, B. W., and Lal, R. (1982), "Bivariate gamma random vectors," *Operations Research*, 30, 355–374.

Silverman, B. W. (1986), *Density estimation for statistics and data analysis* London: Chapman and Hall.

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

Verma, V., and Betti, G. (2005), Sampling errors and design effects for poverty measures and other complex statistics,, Working Paper 53, Siena: Dipartimento di Metodi Quantitativi, Universit degli Studi.

Verma, V., and Betti, G. (2011), "Taylor linearisation sampling errors and design effects for poverty measures and other complex statistics," *Journal of Applied Statistics*, 38, 1549–1576.

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