VARIANCE ESTIMATION OF CHANGE FOR ROTATING REPEATED SURVEYS

Yves G. Berger [1]

[1] *University of Southampton, SO17 1BJ, United Kingdom, y.g.berger@soton.ac.uk.*

Abstract. Smith et al. (2003) recognised that assessing change is one of the most important challenges in survey statistics. A common problem is to compare two cross-sectional estimates for the same study variable taken on two different waves or occasions, and to judge whether the observed change is statistically significant. The estimation of covariances plays an important role in the estimation of the variance of a change. Berger and Priam (2010, 2013) proposed a simple approach based upon a multivariate linear regression model. The proposed estimator is not a model-based estimator. Indeed, the proposed estimator is design-consistent when the sampling fractions are negligible. The proposed estimator can accommodate stratified and two-stage sampling designs, and can be extended for complex estimators of change. The main advantage of the proposed approach is its simplicity and flexibility. It can be applied to a wide class of sampling designs, and can be implemented with standard statistical regression techniques. Because of its flexibility, the proposed approach is well suited for the estimation of variance for the EU-SILC surveys. It allows to use a common approach for variance estimation for the different type of designs. The proposed approach does not rely on a specialised computer package and can be used with standard statistical software. The proposed approach can be used to produced variance estimates of change for the indicator of poverty and social exclusion (AROPE) based upon the European Union Statistics on Income and Living Conditions (EU-SILC) survey (Eurostat, 2012*a*). This poverty indicator is used to monitor change in poverty within the European Union (Eurostat, 2012*a*).

Résumé. Smith et al. (2003) ont reconnu que l'estimation du changement est l'un des défis les plus importants dans les statistiques de l'enquête. Les utilisateurs sont souvent intéressé à estimer des changements ou des tendances d'une période à l'autre. Un problème courant est de comparer deux estimations transversales pour la même variable mesurée à deux vagues différentes. L'estimation de la variance d'échantillonnage d'un estimateur du changement est utile pour juger si un changement observé est statistiquement significatif. Les covariances jouent un rôle important dans l'estimation de la variance d'un changement (Tam, 1984). Berger and Priam (2010, 2013) proposent d'utiliser une approche de régression linéaire multivariée (régression généralisée) pour estimer les covariances. L'estimateur proposé n'est pas un estimateur basé sur un modèle de superpopulation, et cet estimateur est valable même si le modèle multivariée n'ajuste pas les données. Nous utiliserons l'enquête Européenne EU-SILC (Eurostat, 2012*a*) pour illustrer l'approche proposée. En raison de sa flexibilité, l'approche proposée est bien adaptée pour l'estimation de la variance pour les enquêtes EU-SILC. Il permet d'utiliser une approche commune pour l'estimation de la variance pour les différents types d'échantilonnages. L'approche proposée ne repose pas sur un logiciel spécialisé et peut être utilisé avec des logiciels statistiques standards. L'approche proposée peut être utilisée pour produire des estimés de variance pour l'indice de pauvreté et d'exclusion sociale (AROPE) utilisé par l'union Européenne pour l'évaluation du changement de pauvreté au sein de l'Union Européenne (Eurostat, 2012*a*).

1 Introduction

The primary interest of many users is often in changes or trends from one time period to another. Suppose, we wish to estimate the absolute change

$$\Delta = \tau_2 - \tau_1,$$

between two population totals $\tau_1 = \sum_{i \in U} y_{1;i}$ and $\tau_2 = \sum_{i \in U} y_{2;i}$, of wave 1 and 2; where U denotes the population of interest. The quantities $y_{1;i}$ and $y_{2;i}$ denote respectively the values of the variable of interest at wave 1 and 2. For simplicity, we assume that U is the same at wave 1 and 2. The estimator proposed in this paper can be also used when the population at wave 1 is different from the population at wave 2. In §5, we show that the proposed approach can be adapted for relative change or more complex measures of change. We adopt a design-based approach where the sampling distribution is specified by the sampling design. The change Δ can be estimated by

$$\hat{\Delta} = \hat{\tau}_2 - \hat{\tau}_1;$$

where $\hat{\tau}_1$ and $\hat{\tau}_2$ are two cross-sectional Horvitz and Thompson (1952) estimators given by

$$\hat{\tau}_1 = \sum_{i \in s_1} \frac{y_{1;i}}{\pi_{1;i}} \quad \text{and} \quad \hat{\tau}_2 = \sum_{i \in s_2} \frac{y_{2;i}}{\pi_{2;i}}\,; \tag{1}$$

where s_1 and s_2 denote respectively the first and second wave samples. The quantities $\pi_{1;i}$ and $\pi_{2;i}$ are the first-order inclusion probabilities at wave 1 and 2. These probabilities are defined in §2. In this paper, s denotes the union of s_1 and s_2; that is, $s = s_1 \cup s_2$.

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

$$var(\hat{\Delta}) = var(\hat{\tau}_1) + var(\hat{\tau}_2) - 2\,cov(\hat{\tau}_1, \hat{\tau}_2) \tag{2}$$

$$= \boldsymbol{\nabla}'\boldsymbol{\Sigma}\boldsymbol{\nabla}; \tag{3}$$

where $var(\hat{\tau}_1)$ and $var(\hat{\tau}_2)$ denote respectively the design-based variances of $\hat{\tau}_1$ and $\hat{\tau}_2$. The quantity $cov(\hat{\tau}_1, \hat{\tau}_2)$ denotes the covariance between $\hat{\tau}_1$ and $\hat{\tau}_2$ with respect to the sampling design. The matrix $\boldsymbol{\Sigma}$ is the design-based covariance matrix of the vector $(\hat{\tau}_1, \hat{\tau}_2)'$ and $\boldsymbol{\nabla} = (-1,\ 1)'$.

Standard design-based estimators can be used to estimate the variances $var(\hat{\tau}_1)$ and $var(\hat{\tau}_2)$ (e.g. Wolter, 2007). The covariance $cov(\hat{\tau}_1, \hat{\tau}_2)$ is the most difficult part to estimate because with overlapping samples, $\hat{\tau}_1$ and $\hat{\tau}_2$ are estimated from different samples. Several estimators have been proposed for the covariance in (2) (e.g. Kish, 1965; Tam, 1984; Nordberg, 2000; Holmes and Skinner, 2000; Berger, 2004*b*; Qualité and Tillé, 2008; Wood, 2008; Goga et al., 2009; Muennich and Zins, 2011; Knottnerus and van Delden, 2012). In a series of simulations based on the Swedish Labour Force Survey, Andersson et al. (2011*a*,*b*) showed that the estimator proposed by Berger (2004*b*) gives more accurate estimates than standard variance estimators (e.g. Tam, 1984; Qualité and Tillé, 2008) when we are interested in change within strata domains. The estimator proposed in this paper has the same property when the sampling fractions are small, because in §3.1, we show that the proposed estimator is approximately equivalent to the Berger (2004*b*) estimator, in this situation.

The proposed approach is based upon the residual matrix of a multivariate regression model. The multivariate regression is not a super-population approach, as it gives design-consistent covariance estimates. However, it relies on the assumption that the sampling fractions are negligible, which is usually the case for social surveys, such as the EU-SILC surveys (Eurostat, 2012*a*). The proposed approach has the advantage of not requiring joint-inclusion probabilities which can be unknown with rotating designs (e.g. Wood, 2008; Muennich and Zins, 2011, p. 20).

In §2, we defined the class of rotating sampling designs considered. The proposed estimator for the covariance is defined in §3.1. In §4, we show how the proposed estimator can be extended to include stratification, multi-stage sampling and more complex measures of change. In §6, we support our result with a simulation study based upon the Italian EU-SILC survey. In §7, we show how the proposed approach can be used to estimate the variance of change of the EU-SILC indicator of poverty and social exclusion (AROPE).

2 Fixed sizes rotating sampling designs

With 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 both on wave 1 and 2 usually represent a large fraction of the sample s_1. This fraction is called the fraction of the common sample and is denoted by g. For example, for the EU-SILC surveys, $g = 75\%$. For the Canadian labour force survey and the British labour force survey, $g = 80\%$. For the Finish labour force survey, $g = 60\%$.

The class of fixed sizes rotating sampling designs is defined as follows. Assume that s_1 is a probability sample of size n_1 selected without replacement with first-order inclusion probabilities $\pi_{1;i} = pr\{i \in s_1\}$. Suppose that s_2 is a sample of size n_2 selected with conditional inclusion probabilities $\pi_{2;i}(s_1) = pr\{i \in s_2|s_1\}$ such that s_2 contains n_c units from s_1; where $0 \leq n_c \leq n_1$. The wave 2 inclusion probabilities are given by $\pi_{2;i} = E_1[\pi_{1;i}(s_1)]$; where $E_1[\,\cdot\,]$ denotes the design expectation with respect to the first wave design. We consider that the sizes n_1, n_2 and n_c are given quantities which are fixed (non random). Note that the fraction of the common sample is given by $g = n_c/n_1$. The units from $s_1 \setminus s_2$ are the units that rotate out and the units from $s_2 \setminus s_1$ are the units that rotate in. In principle, we can have $g = 0$ (when we have two independent samples) or $g = 1$ (when we have two indentical samples). The proposed approach is valid when $g = 0$ or 1, as it gives standards estimates in this situation.

This class contains standard rotating sampling designs such as the rotating randomised systematic sampling design (e.g. Holmes and Skinner, 2000), the rotation groups sampling design (e.g. Kalton, 2009; Gambino and Silva, 2009, p. 415) used for the EU-SILC surveys (Muennich and Zins, 2011; Eurostat, 2012*a*) and the rotating design proposed by Tam (1984). Other examples can be found in Berger and Priam (2013) and in Christine and Rocher (2012).

3 Estimation of the covariance matrix

The estimation of the covariance matrix $\boldsymbol{\Sigma}$ in (3) would be relatively straightforward if s_1 and s_2 were the same sample ($g = 1$). Unfortunately, s_1 and s_2 are usually not completely overlapping sets of units ($g \leq 1$), because rotations are usually used in repeated surveys (e.g. Nordberg, 2000; Gambino and Silva, 2009; Kalton, 2009). Berger (2004*b*) showed that under the assumption of high entropy,

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\tau\tau}} - \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\tau n}} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{nn}}^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\tau n}}' \; ; \tag{4}$$

is a consistent estimator for the covariance matrix $\boldsymbol{\Sigma}$ (defined in (3)); with

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\tau\tau}} = \begin{pmatrix} \sum_{i\in s} \breve{c}_{1;i}\, \breve{y}_{i;1}^2 & \sum_{i\in s} \breve{c}_{12;i}\, \breve{y}_{i;1}\, \breve{y}_{i;2} \\ \sum_{i\in s} \breve{c}_{12;i}\, \breve{y}_{i;1}\, \breve{y}_{i;2} & \sum_{i\in s} \breve{c}_{2;i}\, \breve{y}_{i;2}^2 \end{pmatrix}, \tag{5}$$

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{nn}} = \begin{pmatrix} \sum_{i\in s} \breve{c}_{1;i}\, z_{1;i} & \sum_{i\in s} \breve{c}_{12;i}\, z_{1;i}\, z_{2;i} & \sum_{i\in s} \breve{c}_{1;i}\, z_{1;i}\, z_{2;i} \\ \sum_{i\in s} \breve{c}_{12;i}\, z_{1;i}\, z_{2;i} & \sum_{i\in s} \breve{c}_{2;i}\, z_{2;i} & \sum_{i\in s} \breve{c}_{2;i}\, z_{1;i}\, z_{2;i} \\ \sum_{i\in s} \breve{c}_{1;i}\, z_{1;i}\, z_{2;i} & \sum_{i\in s} \breve{c}_{2;i}\, z_{1;i}\, z_{2;i} & \sum_{i\in s} \breve{c}_{i}\, z_{1;i}\, z_{2;i} \end{pmatrix}, \tag{6}$$

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\tau n}} = \begin{pmatrix} \sum_{i\in s} \breve{c}_{1;i}\, \breve{y}_{i;1}\, z_{1;i} & \sum_{i\in s} \breve{c}_{12;i}\, \breve{y}_{i;1}\, z_{2;i} & \sum_{i\in s} \breve{c}_{1;i}\, \breve{y}_{i;1}\, z_{1;i}\, z_{2;i} \\ \sum_{i\in s} \breve{c}_{12;i}\, \breve{y}_{i;2}\, z_{1;i} & \sum_{i\in s} \breve{c}_{2;i}\, \breve{y}_{i;2}\, z_{2;i} & \sum_{i\in s} \breve{c}_{2;i}\, \breve{y}_{i;2}\, z_{1;i}\, z_{2;i} \end{pmatrix}; \tag{7}$$

where $s = s_1 \cup s_2$ denotes the overall sample. The quantities $\breve{c}_{\ell;i}$, $\breve{c}_i$ and $\breve{c}_{12;i}$ are finite population correction given by $\breve{c}_{\ell;i} = (1 - \pi_{\ell;i})$, $\breve{c}_i = (1 - \pi_{c;i})$ and $\breve{c}_{12;i} = (\pi_{c;i} - \pi_{1;i}\pi_{2;i})/\pi_{c;i}$; with $\pi_{c;i} =$

$pr\{i \in s_c\}$; where $pr\{\cdot\}$ denotes the probability with respect to the design. We consider that $\breve{c}_{12;i} = 1$ when $\pi_{c;i} = 0$. The variables $z_{1;i}$ and $z_{2;i}$ are design variables defined by

$$z_{1;i} = \delta\{i \in s_1\}, \qquad \text{and} \qquad z_{2;i} = \delta\{i \in s_2\} \cdot \tag{8}$$

The function $\delta\{A\}$ is the indicator function which is equal to one when A is true and zero otherwise. The variables $\breve{y}_{1;i}$ and $\breve{y}_{2;i}$ are defined by

$$\breve{y}_{i;1} = y_{1;i}\pi_{1;i}{}^{-1}\delta\{i \in s_1\} \qquad \text{and} \qquad \breve{y}_{i;2} = y_{2;i}\pi_{2;i}{}^{-1}\delta\{i \in s_2\} \; ; \tag{9}$$

with $\breve{y}_{\ell;i} = 0$ when $i \notin s_\ell$.

The estimator (4) is not easy to implement, because it requires a specialised package (Berger, 2004*a*; Andersson et al., 2011*a*,*b*). Berger and Priam (2013) showed that when the sampling fractions are negligible, the estimator (4) can be easily approximated using a multivariate linear regression approach described in the following §.

3.1 Estimation of the covariance using a multivariate regression approach

Berger and Priam (2010) proposed to use the residual covariance of the following multivariate (or general) linear regression model (see also Berger and Priam, 2013).

$$\begin{pmatrix} \breve{y}_{1;i} \\ \breve{y}_{2;i} \end{pmatrix} = \begin{pmatrix} \beta_1^{(1)} z_{1;i} + \beta_2^{(1)} z_{2;i} + \beta_{12}^{(1)} z_{1;i}\; z_{2;i} \\ \beta_1^{(2)} z_{1;i} + \beta_2^{(2)} z_{2;i} + \beta_{12}^{(2)} z_{1;i}\; z_{2;i} \end{pmatrix} + \boldsymbol{\epsilon}_i \; ; \tag{10}$$

where $i \in s = s_1 \cup s_2$ and the residuals $\boldsymbol{\epsilon}_i$ have a bivariate distribution with mean $\mathbf{0}$ and an unknown residual variance-covariance matrix $\boldsymbol{V}$. It is important to point out that as the distribution of $\boldsymbol{\epsilon}_i$ does not need to be specified and is not used for inference, as a least squares technique (or a projection in the space spanned by the design variables) will be used. The response variables in the regression (10) are given by (9). The covariates $z_{1;i}$ and $z_{2;i}$ are design variables defined by (8). Note the absence of intercept and the presence of an interaction in the regression (10).

Matrix notations can be used to define the model (10) in a more convenient way. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)})$ be the 3×2 matrix of parameters, where $\boldsymbol{\beta}^{(1)} = (\beta_1^{(1)}, \beta_2^{(1)}, \beta_{12}^{(1)})'$ and $\boldsymbol{\beta}^{(2)} = (\beta_1^{(2)}, \beta_2^{(2)}, \beta_{12}^{(2)})'$ are parameters of the model (10). The model (10) can be re-written as

$$\breve{\boldsymbol{Y}}_s = \boldsymbol{Z}_s\boldsymbol{\beta} + \boldsymbol{\epsilon} \; ; \tag{11}$$

where $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n)'$. The quantities $\breve{\boldsymbol{Y}}_s$ and $\boldsymbol{Z}_s$ are respectively defined by the following $n \times 2$ matrix $\breve{\boldsymbol{Y}}_s$ and the $n \times 3$ matrix $\boldsymbol{Z}_s$.

$$\breve{\boldsymbol{Y}}_s = (\breve{\boldsymbol{y}}_1, \breve{\boldsymbol{y}}_2) \, , \tag{12}$$

$$\boldsymbol{Z}_s = (\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_c) \; ; \tag{13}$$

where

$$\breve{\boldsymbol{y}}_\ell = (\breve{y}_{\ell;1}, \breve{y}_{\ell;2}, \cdots, \breve{y}_{\ell;n})' \, ,$$

$$\boldsymbol{z}_\ell = (z_{\ell;1}, z_{\ell;2}, \cdots, z_{\ell;n})' \, , \tag{14}$$

$$\boldsymbol{z}_c = (z_{1;1}z_{2;1}, z_{1;2}z_{2;2}, \cdots, z_{1;n}z_{2;n})' \cdot \tag{15}$$

The model (11) is also a multivariate analysis of variance (MANOVA) model, as the covariates are all dummy variables.

The proposed estimator for the variance of change (2) is given by the following substitution estimator.

$$\widehat{var}(\hat{\Delta}) = \widehat{var}(\hat{\tau}_1) + \widehat{var}(\hat{\tau}_2) - 2\left[\widehat{var}(\hat{\tau}_1)\widehat{var}(\hat{\tau}_2)\right]^{1/2}\ \widehat{\rho}\,; \tag{16}$$

where $\widehat{var}(\hat{\tau}_1)$ and $\widehat{var}(\hat{\tau}_2)$ denote respectively standard design-based variance estimators of $\hat{\tau}_1$ and $\hat{\tau}_2$ (e.g. Wolter, 2007) and

$$\widehat{\rho} = \widehat{V}_{12}\left(\widehat{V}_{11}\widehat{V}_{22}\right)^{-1/2}; \tag{17}$$

where $\widehat{V}_{k\ell}$ is the component (k, ℓ) of the ordinary least squares residual matrix $\widehat{\boldsymbol{V}}$ of the model (11). The matrix $\widehat{\boldsymbol{V}}$ is an estimator of $\boldsymbol{V}$. Berger and Priam (2013) showed that $\widehat{\boldsymbol{V}} \doteq \alpha^{-1}\hat{\boldsymbol{\Sigma}}$, when

$$\breve{c}_{\ell;i} \doteq 1, \quad \breve{c}_i \doteq 1 \quad \text{and} \quad \breve{c}_{12;i} \doteq 1, \tag{18}$$

where $\alpha = (n - r)$ is a constant scale factor, where $n = \#s$ is the number of units in the sample $s = s_1 \cup s_2$ and r is the number of linearly independent columns of $\boldsymbol{Z}_s$. This implies that (16) is design-consistent, because (4) is design-consistent. The assumptions (18) hold when sampling fractions are negligible (see Example 1 in Berger and Priam (2013)). Note that g can be large even when the sampling fractions are negligible; that is, the assumptions (18) may hold even when g is large.

Berger and Priam (2013) showed that $\alpha\widehat{V}_{12}$ is the standard estimator of covariance under pps with replacement sampling. Berger and Priam (2013) also showed that $\alpha\widehat{V}_{12}$ reduces to Tam (1984) estimator under equal probabilities and when the fraction $n_1 n_2/N n_c = \pi_{2;i} g^{-1}$ is negligible and when n_c is sufficiently large.

It is preferable to have an estimator of covariance based upon a correlation (see (16) and (17)) to avoid negative variance estimates. Berger and Priam (2013) showed that (16) is always positive because the residual variance-covariance matrix is non-negative definit.

We have the following fixed size constraints $\sum_{i\in s} z_{1;i} = n_1$, $\sum_{i\in s} z_{2;i} = n_2$ and $\sum_{i\in s} z_{1;i}z_{2;i} = n_c$, because only samples with these sample sizes can be selected. Thus, we are conditioning on variables which have their totals fixed by design. Note that there is a clear analogy between Birch (1963) approach and the proposed conditioning approach. This regression includes interactions between the variable $z_{1;i}$ and $z_{2;i}$. These interactions capture the rotation of the sampling design which is represented by the constraint $\sum_{i\in s} z_{1;i}z_{2;i} = n_c$.

The proposed approach requires the creation of design variables (13) which are used as covariates. The interactions (see (15)) take account of the rotation of the design. The weighted variables of interest (see (12)) measured at each waves are used as response variables. The proposed estimator takes into account of all the data, as it utilises the data of the units from the common sample and the units that rotate in and out.

The proposed estimator is easier to implement than Nordberg (2000), Wood (2008), Goga et al. (2009) and Muennich and Zins (2011, p. 20) estimators because it does not rely on joint-inclusion probabilities. Furthermore, the proposed estimator is based on a multivariate regression approach which can be implemented with most statistical software, without the need of a specialised statistical package.

The matrix $\widehat{\boldsymbol{V}}$ can be easily calculated, as the multivariate regression (11) can be easily fitted by most statistical software. It is only necessary to create the variables $\breve{y}_{i;1}$, $\breve{y}_{i;2}$, $z_{1;i}$ and $z_{2;i}$. For example, the SAS procedure `REG` can be used to fit the multivariate regression. The multivariate regression can be also fitted using the `GLM Multivariate` procedure in SPSS. With Stata, the output `e(Sigma)` of the function `mvreg()` gives the estimates $\widehat{\boldsymbol{V}}$. With the statistical software R (R Development Core Team, 2006), the command `estVar(lm(formula=Y~-1+Z1*Z2))` gives the estimate $\widehat{\boldsymbol{V}}$; where `Z1` and `Z2` denote the $n \times 1$ vectors $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ (see (14)) and `Y` is the matrix $\breve{\boldsymbol{Y}}_s$ given by (12). Note that Berger (2005) showed that $\widehat{var}(\hat{\tau}_1)$ and $\widehat{var}(\hat{\tau}_2)$ can also be calculated using a regression approach (see also Deville and Tillé, 2005).

4 Extensions to stratified and two-stage sampling design

The proposed estimator can be easily extended to accommodate stratification. Suppose that we have H strata $U_1, U_2, \cdots, U_H$ such that $\cup_{h=1}^{H} U_h = U$. Let s_{1h} and s_{2h} denote respectively the samples of U_h for wave 1 and 2. Let n_{1h}, n_{2h} and n_{ch} be respectively the sample sizes of s_{1h}, s_{2h} and $s_{ch} = s_{1h} \cap s_{2h}$. Suppose that a fixed sizes rotating design (see §2) is implemented with each stratum. We have the following covariates $z_{1h;i} = \delta\{i \in s_{1h}\}$ and $z_{2h;i} = \delta\{i \in s_{2h}\}$ which specifies the stratification.

The multivariate regression model is still given by (11) with the same response variables $\breve{\boldsymbol{Y}}_s$ defined by (12) with $\breve{y}_{i;\ell}$ is given by (9). However, the matrix $\boldsymbol{Z}_s$ is different as it now contains the stratification variables $z_{1h;i}$ and $z_{2h;i}$ and suitable interactions. As we have a rotation within each stratum, the sample sizes $n_{ch} = \#s_{ch}$ are fixed and we need to include the interactions $z_{1h;i} \times z_{2h;i}$ in $\boldsymbol{Z}_s$. The ordinary least squares estimate of residuals variance-covariance of model (11) is used to estimate the covariance between $\hat{\tau}_1$ and $\hat{\tau}_2$ (see (16)).

Berger and Priam (2013) showed that $\hat{S}_{12} = \sum_{h=1}^{H} \hat{S}_{12h}$ and $\hat{S}_{\ell\ell} = \sum_{h=1}^{H} \hat{S}_{\ell\ell h}$ where $\hat{S}_{12h}$ (resp. $\hat{S}_{\ell\ell h}$) denotes the within stratum covariance (resp. variance). Note that $\hat{S}_{12}$ and $\hat{S}_{\ell\ell}$ are natural estimators of covariance and variances under stratified designs. Consequently, the proposed estimator for the covariance is consistent when the assumptions (18) hold within each strata and when the number of strata H is asymptotically bounded.

Note that the proposed estimator accommodates dynamic stratification. In other words, the strata at wave 1 can be different from the strata at wave 2; i.e. new strata can be created, and units can move between strata. For example, a unit may belong to stratum U_1 at wave 1 and to stratum U_2 at wave 2.

Suppose that we have overlapping stratified samples of primary sampling units (PSU), and that the rotation consists in rotating PSUs rather than secondary sampling units. As we assume that the sampling fractions are negligible, we suggest using an ultimate cluster strategy to estimate the covariance; that is, we consider that the variance between PSUs captures most of the variability of the estimator of change. This usually holds in practice (e.g. Särndal et al., 1992, §4.3.2). Berger and Priam (2013) showed how the approach described in §3.1 can be adapted under an ultimate cluster strategy.

5 Extension to more complex measure of change

Suppose that we are interested in the variance of the absolute change $\hat{\Delta}_\theta = \hat{\theta}_2 - \hat{\theta}_1$ or the relative change $\hat{\Delta}_\theta = \hat{\theta}_2 / \hat{\theta}_1$, where $\hat{\theta}_2$, $\hat{\theta}_1$ are two smooth (differentiable) functions of estimators of totals.

Therefore, in both cases, $\hat{\Delta}_\theta$ is a smooth function of totals; that is,

$$\hat{\Delta}_\theta = f(\hat{\boldsymbol{\tau}}) \tag{19}$$

where $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \hat{\tau}_2 \cdots, \hat{\tau}_p, \cdots, \hat{\tau}_P)'$ and P is the number of totals. The quantity $\hat{\tau}_p$ is the Horvitz and Thompson (1952) estimator $\hat{\tau}_p = \sum_{i \in s_\ell} y_{p;i}/\pi_{\ell;i}$ of a variable y_p; where $p = 1, \ldots, Q, Q+1, \ldots, P$, $\ell = 1$ if $p \leq Q$ and $\ell = 2$ if $p > Q$. The constant Q is the number of totals calculated from the first wave.

Suppose that $\hat{\Delta}_\theta$ is an approximately unbiased estimator of $\Delta_\theta = f(\boldsymbol{\tau})$; where $\boldsymbol{\tau} = E(\hat{\boldsymbol{\tau}})$. Using the delta method (Taylor linearisation), we have that an approximation of $\hat{\Delta}_\theta$ in the neighbourhood of $\boldsymbol{\tau}$ is given by $\hat{\Delta}_\theta - \Delta_\theta \doteqdot \boldsymbol{\nabla}(\boldsymbol{\tau})'(\hat{\boldsymbol{\tau}} - \hat{\boldsymbol{\tau}})$; where $\boldsymbol{\nabla}(\boldsymbol{\tau})$ is the gradient of $f(\boldsymbol{\tau})$ at $\boldsymbol{\tau}$. Therefore, the linearisation estimator for the variance is

$$\widehat{var}(\hat{\Delta}_\theta) = \boldsymbol{\nabla}(\hat{\boldsymbol{\tau}})'\widehat{\boldsymbol{var}}(\hat{\boldsymbol{\tau}})\boldsymbol{\nabla}(\hat{\boldsymbol{\tau}})\cdot \tag{20}$$

Note that (20) is an estimator of the mean square error. However, this estimator is also a consistent estimator for the variance, because $\hat{\Delta}_\theta$ is asymptotically unbiased (Wolter, 2007, p. 232). Berger and Priam (2013) showed how the approach described in §3.1 can be used to estimate the covariance matrix $\widehat{\boldsymbol{var}}(\hat{\boldsymbol{\tau}})$.

Another approach consists in substituting $y_{\ell;i}$ by linearised variables in (1) (e.g. Deville, 1999; Demnati and Rao, 2004), and using the estimator of covariance (16) where $\hat{\tau}_\ell$ is the Horvitz and Thompson (1952) sum of the linearised variable at wave ℓ. For example, this approach is recommended when $\hat{\theta}_2$, $\hat{\theta}_1$ are not functions of totals. For example, when $\hat{\theta}_2$ and $\hat{\theta}_1$ are Gini coefficients. It can be shown that this approach is equivalent to (20) when $\hat{\theta}_\ell$ is a smooth function of totals computed only from s_ℓ. This would not be the case if $\hat{\theta}_\ell$ depends on totals computed from s_1 and s_2. For example, if $\hat{\theta}_1 = \hat{\tau}_1/(\hat{\tau}_2 + \hat{\tau}_1)$ and $\hat{\theta}_2 = \hat{\tau}_2/(\hat{\tau}_2 + \hat{\tau}_1)$, where $\hat{\tau}_1$ and $\hat{\tau}_2$ are given by (1). In this case, the linearised variable approach gives a biased estimator for the variance and the estimator (20) is still approximatelly unbiased.

6 Empirical simulation studies

In a series of simulations based on the Swedish Labour Force Survey, Andersson et al. (2011*b*) (see also Andersson et al. (2011*a*)) showed that for estimation of change within strata domains, the estimator proposed by Berger (2004*b*) is more accurate than standard estimators of variance of change (e.g. Tam, 1984; Qualité and Tillé, 2008). Therefore, based on Andersson et al. (2011*b*) simulation studies, the estimator proposed by Berger (2004*b*) is recommended when we are interested in change within strata domains. The estimator proposed in this paper has the same property, as it reduces to the Berger (2004*b*) estimator when the sampling fractions are small.

In this §, we report the results of a series of simulation based upon the Italian EU-SILC survey. For each simulation, 1000 samples were selected to compute the empirical relative bias (RB)

$$\text{RB} = \frac{E[v\hat{a}r(\hat{\Delta})] - var(\hat{\Delta})}{var(\hat{\Delta})} \ \%$$

and the empirical relative root mean square error (RRMSE)

$$\text{RRMSE} = \frac{mse[v\hat{a}r(\hat{\Delta})]^{1/2}}{var(\hat{\Delta})} \ \%\cdot$$

Table 1: Observed RB and RRMSE for several variables of interest and several domain of interest. Italian EU-SILC data (2008, 2009). The column "Prop." gives the results for the proposed approach. The columns "Corr. (23)" and "Corr. (22)" give the results for the approaches based respectivelly on the correlations (23) and (22).

Variables of interest	Domains of interest	RB (%)			RRMSE (%)		
		Prop.	Corr. (23)	Corr. (22)	Prop.	Corr. (23)	Corr. (22)
Afford holiday	Population	-4	20	14	442	466	461
	Detached	-6	7	0	495	508	501
	Semi-detached	-4	4	0	229	220	224
	Home owner	-4	16	11	255	276	271
	Males	-5	13	8	424	442	437
	Females	-5	16	10	901	922	916
Own a car	Population	-9	9	1	156	174	165
	Detached	0	19	6	112	129	117
	Semi-detached	-3	12	2	33	37	29
	Home owner	-6	11	4	217	200	207
	Males	-5	11	3	309	325	317
	Females	-10	9	-1	77	95	85
Equivalised disposable income	Population	-3	14	11	101	117	114
	Detached	-2	7	3	107	112	108
	Semi-detached	-5	-6	1	102	99	106
	Home owner	-5	10	8	99	113	111
	Males	1	11	14	107	114	118
	Females	-7	9	2	97	112	104
At risk of poverty	Population	-9	15	1	1713	1738	1723
	Detached	-3	19	5	1746	1768	1754
	Semi-detached	0	16	3	1581	1597	1584
	Home owner	-4	17	5	981	1002	990
	Males	-9	11	-2	2631	2652	2639
	Females	-5	21	4	3018	3044	3028

The quantity $var(\hat{\Delta})$ denotes the empirical variance of $\hat{\Delta}$. The quantities $E[v\hat{a}r(\hat{\Delta})]$ and $mse[v\hat{a}r(\hat{\Delta})]$ denote respectively the empirical expectation and the empirical mean square error of an estimator $v\hat{a}r(\hat{\Delta})$ of variance of change.

The Chao (1982) unequal probability design is used to select samples. For wave 1, a sample is selected with inclusion probabilities $\pi_{1;i}$. For wave 2, we use the sampling design described in the Example 1 of Berger and Priam (2013). This rotating design belongs to the class of designs described in §2 and is such that $\pi_{1;i} \simeq \pi_{2;i}$. The statistical software R is used to fit the multivariate regression model. The variances $var(\hat{\tau}_1)$ and $var(\hat{\tau}_2)$ are estimated by the Hájek (1964) variance estimator.

The common sample of waves 2008 and 2009 of of the Italian Statistics on Income and Living Conditions (EU-SILC) survey is treated as a population from which stratified samples are selected. This gives a population size $N = 19644$. Stratified samples of size $n_1 = n_2 = 982$ are selected using the uni-stage Chao (1982) sampling design. The strata are five geographical regions. We consider $g = 75\%$.

We consider the change between the mean (or proportion) of several variables of interest. We consider three dummy variables of interest (afford holiday, own a car, at risk of poverty) and one quantitative variables (equivalised disposable income). The change between the means (or proportion) is estimated by $\hat{\tau}_2 N^{-1} - \hat{\tau}_1 N^{-1}$. We consider that the inclusion probabilities $\pi_{1;i}$ are proportional to the inverse of the cross-sectional sampling weights at wave 1. We also consider several domains of interest given by the type of accommodation (detached, semi-detached), the population of home owners, the population of males and the population of females. The households are the units, and the quantities $y_{1;i}$ and $y_{2;i}$ denote the household totals of the variables of interest ($y_{\ell;i} = 0$ if i does not belong to the domain of interest).

Berger and Priam (2013) showed that the proposed estimator is equivalent to the variance estimator proposed by Qualité (2009, p. 83) and Muennich and Zins (2011, p. 20). Thus, the proposed estimator will not be more accurate than these estimators. We propose to compare (16) with the estimator (21) which consists in using the same variance estimates as in (16) with a naïve correlation (see (22) below) based on Tam (1984) estimator for the covariance (under equal probability sampling).

$$\widehat{var}(\hat{\Delta})^* = \widehat{var}(\hat{\tau}_1) + \widehat{var}(\hat{\tau}_2) - 2\,\widehat{\rho}^* \quad [\widehat{var}(\hat{\tau}_1) \quad \widehat{var}(\hat{\tau}_2)]^{1/2}\,; \tag{21}$$

with

$$\widehat{\rho}^* = \widehat{cov}(\hat{\tau}_1, \hat{\tau}_2)_{Tam}\, [\widehat{var}(\hat{\tau}_1)_{SRS} \quad \widehat{var}(\hat{\tau}_2)_{SRS}]^{-\frac{1}{2}} \tag{22}$$

and

$$\widehat{var}(\hat{\tau}_\ell)_{SRS} = N^2 \left(1 - \frac{n_\ell}{N}\right) \frac{\hat{\sigma}_{\ell\ell}}{n_\ell},$$

where $\widehat{cov}(\hat{\tau}_1, \hat{\tau}_2)_{Tam}$ is the covariance proposed by Tam (1984) and $\hat{\sigma}_{\ell\ell}$ is defined by

$$\hat{\sigma}_{\ell\ell} = \frac{1}{n_\ell} \sum_{i \in s_\ell} (\breve{y}_{i;\ell} - \bar{y}_\ell)^2 \cdot$$

We also compare the proposed estimator with another estimator which consists in using (21) with a correlation based upon a stratified equal probability sampling design (e.g. Qualité, 2009, p. 79, Muennich and Zins, 2011, p. 26). This correlation is given by

$$\widehat{\rho}^*_{st} = \widehat{cov}(\hat{\tau}_1, \hat{\tau}_2)_{st} \;\; [\widehat{var}(\hat{\tau}_1)_{st} \quad \widehat{var}(\hat{\tau}_2)_{st}]^{-\frac{1}{2}}; \tag{23}$$

where

$$\begin{aligned}
\widehat{cov}(\hat{\tau}_1, \hat{\tau}_2)_{st} &= \sum_{h=1}^{H} \sum_{i \in s_{ch}} \left(1 - \frac{n_{1h} n_{2h}}{N_h n_{ch}}\right) \frac{n_{ch}}{n_{ch} - 1} \frac{N_h^2 n_{ch}}{n_{1h} n_{2h}} \hat{\sigma}_{ch}, \\
\widehat{var}(\hat{\tau}_\ell)_{st} &= \sum_{h=1}^{H} \sum_{i \in s_{\ell h}} \left(1 - \frac{n_{\ell h}}{N_h}\right) N_h^2 \frac{\hat{\sigma}^2_{\ell h}}{n_{\ell h}},
\end{aligned}$$

$$\hat{\sigma}_{ch} = \frac{1}{n_{ch}} \sum_{i \in s_{ch}} (y_{1;i} - \bar{y}_{1;ch})(y_{2;i} - \bar{y}_{2;ch}) \ ,$$
$$\hat{\sigma}^2_{\ell h} = \frac{1}{n_{\ell h}} \sum_{i \in s_{\ell h}} (y_{\ell;i} - \bar{y}_{\ell h})^2 \ .$$

The quantities $\bar{y}_{1;ch}$, $\bar{y}_{2;ch}$ are the sample means of the common sample of the stratum h, and $\bar{y}_{\ell h}$ is the sample mean of the stratum h at wave ℓ.

The result of this simulation is given in Table 1. The observed bias of the proposed estimator is usually negligible. We observe that the estimator based upon (23) may have non negligible biases. This is due to the fact that this estimator overestimates the covariance, as it does not take the inclusion probabilities into account. However, the estimator based upon (22) may have negligible biases, but not for all the variables and domains considered. In most of the cases, we observed smaller RRMSEs for the proposed approach. The large values for the RRMSEs is due to the fact that the variances $var(\hat{\Delta})$ can be very small.

7 AN APPLICATION TO THE EU-SILC HOUSEHOLD SURVEYS

Suppose we are interested in the change of the *at-risk-of-poverty or social exclusion* (AROPE) indicator between two consecutive years (2009 and 2010). The AROPE indicator is a key indicator used to monitor poverty within the European Union (Eurostat, 2012*b*; Atkinson and Marlier, 2010). This indicator is calculated from the EU-SILC surveys (Eurostat, 2012*a*) which collect yearly information on income, poverty, social exclusion and living conditions from approximately 300 000 households across Europe. In this §, we show briefly how to estimate the variance of the net change of the AROPE indicator. The computations were made in SAS by Guillaume Osier (Statistics Luxembourg, STATEC and Luxembourg Income Study), Emilio Di Meglio (Eurostat Unit F4 Quality of Life) and Emanuela Di Falco (Eurostat Unit F4 Quality of Life). The EU-SILC production data base was used.

An ultimate cluster approach (see §4) was adopted, because the sampling fractions are small. The units are the primary sampling units (PSUs). For some countries, the PSUs are households (e.g. Austria, UK, Latvia). Scandinavian countries, use single stage design based on registers. In this case, the PSUs are sets containing one individual. The response variables of the multivariate model are given by

$$\breve{y}_{\ell;i} = \delta\{i \in s_\ell\} \sum_{j \in PSU_i} w_{\ell;j}\, y_{\ell;j} \ ,$$

where s_ℓ is the sample of PSUs at wave ℓ, PSU_i denotes the i-th PSU, $y_{\ell;j}$ is the value of the variable of interest for individuals j and $w_{\ell;j}$ is the survey cross-sectional weight of individuals j at wave ℓ. The variable $z_{\ell h;i}$ are dummy variable which specifies the stratification at PSU level. The variables $\breve{y}_{\ell;i}$ and $z_{\ell h;i}$ need to be defined for all $i \in s = s_1 \cup s_2$. A more detailed description of the computation can be found in Berger et al. (2012); Di Meglio et al. (2013).

Table 2: Estimates of the AROPE indicator for 2009 and 2010 based on the EU-SILC surveys' data. The estimates of change in bold face are statistically significant at 5%.

Country	AROPE 2009 (%)	AROPE 2010 (%)	Change (in % point)	Standard Error	Country	AROPE 2009 (%)	AROPE 2010 (%)	Change (in % point)	Standard Error
Iceland	11.6	13.7	**2.09**	0.34	Malta	20.2	20.3	0.09	0.42
Czech Rep.	14.0	14.4	0.36	0.30	UK	22.0	23.1	**1.18**	0.25
Netherlands	15.1	15.1	-0.07	0.14	Cyprus	22.9	23.6	0.67	0.55
Norway	15.2	14.9	-0.34	0.28	Estonia	23.4	21.7	**-1.69**	0.38
Sweden	15.9	15.0	**-0.90**	0.29	Spain	23.4	25.5	**2.16**	0.02
Finland	16.9	16.9	-0.01	0.33	Italy	24.7	24.5	-0.16	0.32
Austria	17.0	16.6	-0.44	0.27	Portugal	24.9	25.3	**0.40**	0.10
Slovenia	17.1	18.3	**1.17**	0.22	Ireland	25.7	29.9	**4.18**	0.93
Switzerland	17.2	17.2	-0.08	0.39	Greece	27.6	27.7	0.11	0.30
Denmark	17.6	18.3	0.74	0.40	Poland	27.8	27.8	-0.07	0.27
Luxembourg	17.8	17.1	-0.72	0.43	Lithuania	29.5	33.4	**3.90**	0.48
France	18.5	19.2	0.71	0.53	Hungary	29.6	29.9	0.32	0.41
Slovakia	19.6	20.6	**1.01**	0.17	Latvia	37.4	38.1	0.64	0.34
Germany	20.0	19.7	-0.26	0.24	Romania	43.1	41.4	**-1.66**	0.11
Belgium	20.2	20.8	**0.66**	0.07	Bulgaria	46.2	41.6	**-4.57**	0.75

The AROPE depends on a poverty threshold which is estimated. The estimation of the poverty threshold can be taken into account using a linearised variables technique (Osier, 2009). Berger and Oguz Alper (2013) showed how the proposed approach can be used to take into account the variability of the threshold. For simplicity, we assume that the poverty threshold is fixed which ensures conservative variances (Berger and Skinner, 2003). Therefore, we consider that the AROPE indicator is a ratio of two totals: an estimate of the total number of individuals in poverty and social exclusion divided by an estimate of the population size (or exposure if we are interested in a domain). Hence $\hat{\boldsymbol{\tau}}$ in (19) is a vector of four totals. The approach of §5 is used. The effect of calibration can be taken into account, by replacing the response variables by residuals (Deville and Särndal, 1992). However, The effect of calibration was ignored, because the calibration variables were not available. For multi-stage designs, the effect of re-weighting due to nonresponse adjustments does not need to be taken into account, because they are done within PSUs. For single stage designs, the effect of nonresponse adjustments is ignored. This is not crucial, because single stage designs are often based on registers (like in the Scandinavian countries) which usually have a small fraction of missing values. The effect of imputation was ignored. Note that some countries use a rotation within PSU. In this case, the proposed approach can still be used (see end of §4).

The estimates based on the proposed approach are given in Table 2. We notice that the change can be significant for some countries (the values in bold face). However, these estimates need to be interpreted with caution. These estimates are for illustrative purpose only, and are not part of any results officially released by Eurostat. The quality of these estimates relies on the availability and quality of the design variables. These estimates are likely to overestimate the variance because the effect of calibration adjustment was not taken into account. This effect may be more pronounced for Scandinavian countries. One of the aim of the Net-SILC2 project (Atkinson and Marlier, 2010) is to provide recommendations regarding the quality of the design variables. These recommendations can be found in Goedeme (2010).

8 Discussion

In this paper, we propose a novel variance estimator for change under complex rotating sampling designs. We show that the proposed variance estimator gives a design-consistent estimator for the variance when the finite population corrections are negligible, and always gives positive variance estimates.

The advantage of the proposed estimator is its generality, simplicity and flexibility. The proposed estimator can be used for a large class of complex sampling designs, involving unequal probabilities, stratification and two-stage sampling. It can be also extended for complex measure of changes. The proposed estimator is simple to implement, as it is based upon a multivariate (general) linear regression technique which can be easily implement with standard statistical software. This regression technique requires the creation of design variables which are used as covariates. Interactions take account of the rotation of the design. The weighted variables of interest measured at each waves are used as response variables. The proposed estimator takes into account of all the data, as it utilises the data of the units from the common sample and the units that rotates in and out.

Acknowledgements

This work was supported by the grant RES-000-22-3045 of the Economic and Social Research Council (UK) and by consulting work for the Net-SILC2 project (Atkinson and Marlier, 2010). I am also grateful to Guillaume Osier (Statistics Luxembourg, STATEC and Luxembourg Income Study), Emilio Di Meglio (Eurostat Unit F4 Quality of Life), Emanuela Di Falco (Eurostat Unit F4 Quality of Life) for testing the approach on the EU-SILC survey data. I am also grateful to Dr. Rodolphe Priam, Dr. Emilio López Escobar (Instituto Technológico Autónomo de México, México), Melike Oguz Alper (University of Southampton) for helpful comments.

References

Andersson, C., Andersson, K., and Lundquist, P. (2011*a*), "Estimation of change in a rotation panel design," *Proceeding of the 58th session of International Statistical Institute, Dublin*, .

Andersson, C., Andersson, K., and Lundquist, P. (2011*b*), "Variansskattningar avseende förändringsskattningar i panelundersökningar (Variance Estimation of change in panel surveys)," *Methodology reports from Statistics Sweden (Statistiska centralbyrån)*, .

Atkinson, A. B., and Marlier, E. (2010), *Income and living conditions in Europe*, Luxembourg: Office for Official Publications `http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-31-10-555/EN/KS-31-10-555-EN.PDF`.

Berger, Y. G. (2004*a*), *A R library for estimation of variance-covariance matrices for rotating sampling schemes*, UK: http://www.yvesberger.co.uk/.

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

Berger, Y. G. (2005), "Variance Estimation with Highly Stratified Sampling Designs with Unequal Probabilities," *Australian and New Zealand Journal of Statistics*, 47(3), 365–373.

Berger, Y. G., Goedemé, T., and Osier, G. (2012), *Standard error estimation and related sampling issues*, Vienna: EU-SILC conference, Vienna, December 2012. www.statistik.at/web_en/static/conference_netsilc2_-_dec_2012_-_session_3_berger_et_al_-_web_dd_191112_068707.pdf.

Berger, Y. G., and Oguz Alper, M. (2013), *Variance estimation of change of poverty based upon the Turkish EU-SILC survey*, Brussels: Proceeding of the conference on New Techniques and Technologies for Statistics, Brussels. http://www.cros-portal.eu/sites/default/files/NTTS2013fullPaper_137.pdf.

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.

Birch, M. W. (1963), "Maximum likelihood in three-way contingency tables," *Journal of the Royal Statistical Society*, B(25), 220–233.

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 à 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.

Deville, J. C., and Särndal, C. E. (1992), "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, 87(418), 376–382.

Deville, J. C., and Tillé, Y. (2005), "Variance approximation under balanced sampling," *Journal of Statistical Planning and Inference*, 128, 569–591.

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 (2012*a*), "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].

Eurostat (2012*b*), "People at risk of poverty or social exclusion (EU-SILC),", `http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/dataset?p_product_code=T2020_50`. [Online; accessed 7 Jan. 2013]].

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 construction and use of sample design variables in EU-SILC. A user's perspective," *Report prepared for Eurostat, Antwerp: Herman Deleeck Centre for Social Policy, University of Antwerp* `http://www.centrumvoorsociaalbeleid.be/index.php?q=node/2157/en`, p. 16p.

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), 691–709.

Hájek, J. (1964), "Asymptotic Theory of Rejective Sampling with Varying Probabilities from a Finite Population," *The Annals of Mathematical Statistics*, 35(4), 1491–1523.

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* Wiley.

Knottnerus, P., and van Delden, A. (2012), "On variances of changes estimated from rotating panels and dynamic strata," *Survey Methodology*, 38(1), 43–52.

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.

Qualité, L. (2009), Unequal probability sampling and repeated surveys,. PhD Thesis, University of Neuchatel, Switzerland.

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.

R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing. `http://www.R-project.org`, Vienna, Austria. ISBN 3-900051-07-0.

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

Smith, P., Pont, M., and Jones, T. (2003), "Developments in Business Survey Methodology in the Office for National Statistics, 1994-2000," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 52(3), 257–295.

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

Wolter, K. (2007), *Introduction to Variance Estimation* Springer Series in Statistics, 2nd ed.

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