Variance estimation of imputed estimators of change over time from repeated surveys

Yves G. BERGER and Emilio L. ESCOBAR

University of Southampton, Southampton, United Kingdom, y.g.berger@soton.ac.uk.

1 Introduction

Measuring change over time is a central problem for many users of social, economic and demographic data and is of interest in many areas of economics and social sciences. Smith *et al.* [17] recognised that assessing change is one of the most important challenges in survey statistics. The primary interest of many users is often in changes or trends from one time period to another. A common problem is to compare two cross-sectional estimates for the same study variable taken on two different waves or occasions. These cross-sectional estimates often include imputed values to compensate for item non-response. The estimation of the sampling variance of the estimator of change is useful to judge whether the observed change is statistically significant. Covariances play an important role in estimating the variance of a change.

We propose to use a multivariate linear regression approach to estimate covariances. The proposed estimator is not a model-based estimator, as this estimator is valid even if the model does not fit the data (Berger & Priam [4]). We show how this approach can be used to accommodate the effect of imputation. The regression approach gives design-consistent estimation of the variance of change when the sampling fraction is small and the finite population corrections are negligible. We illustrate the proposed approach for hot-deck imputation, although the proposed estimator can be also used for other imputation techniques.

The estimation of variance of change would be relatively straightforward if cross-sectional estimates were based upon the same sample. Furthermore, because of rotations used in repeated surveys, cross-sectional estimates are not independent. Let s_1 and s_2 denote respectively the first and second wave samples. The samples s_1 and s_2 are usually not completely overlapping sets of units, because repeated surveys use rotation designs which consist in selecting new units ($i \in s_2 \setminus s_1$) to replace old units ($i \in s_1 \setminus s_2$) that have been in the survey for a specified number of waves. We assume that s_1 and s_2 have the same sample size n. Let n_{12} denote the sample size of the common sample $s_1 \cap s_2$. The units sampled on wave 1 and 2 represent usually a large fraction of the sample s_1; that is, n_{12}/n is usually large. We denote the overall sample by $s = s_1 \cup s_2$. The size of the overall sample is denoted by $\tilde{n} = \#s$. We assume that the rotation sampling design is such that n and n_{12} are fixed quantities.

This class contains standard rotating sampling designs such as the rotating systematic sampling design (Holmes & Skinner [8]), the rotation groups sampling design (e.g. Kalton [11] and Gambino & Silva [7], the rotating design proposed by Tam [18] and the permanent random numbers rotating design (e.g. Ohlsson [14] and Nordberg [13]).

Let $y_{1;i}$ and $y_{2;i}$ denote respectively the value of the variable of interest for wave 1 and 2. Suppose, we wish to estimate the absolute change $\Delta = \tau_2 - \tau_1$ between the following two population totals

$$\tau_1 = \sum_{i \in U} y_{1;i} \qquad \text{and} \qquad \tau_2 = \sum_{i \in U} y_{2;i}$$

from wave 1 and wave 2 respectively, where U denotes the population which is assumed to be the same at both waves. The quantity N will denote the population size.

Some of the values of $y_{1;i}$ and $y_{2;i}$ can be missing due to nonresponse. We propose to impute these missing values.

In §2, we show how random hot-deck can be used to compensate for nonresponse. In §3, we propose to use a reverse approach proposed by Fay [6] to estimate the variance of the imputed estimator of change. The proposed variance estimator depends on a covariance matrix which will be estimated using a multivariate (general) linear regression approach described in §4.

2 Hot-deck imputation

For simplicity, we assume that random hot-deck imputation is used, although the proposed approach can be generalised for other imputation techniques. The main advantage of hot-deck imputation is the fact that it guarantees unbiased estimation of population distributions.

The missingness is represented by the following random variables

$$a_{1;i} = \begin{cases} 0 & \text{if } y_{1;i} \text{ is missing,} \\ 1 & \text{otherwise,} \end{cases} \qquad \text{and} \qquad a_{2;i} = \begin{cases} 0 & \text{if } y_{2;i} \text{ is missing,} \\ 1 & \text{otherwise.} \end{cases}$$

The observed values of $a_{1;i}$ and $a_{2;i}$ are known. For simplicity, we use the same notation for the random variables and their observed values. Note that we will not impute the values $y_{2;i}$ of the units $i \in s_2 \setminus s_1$ which rotate in and the values $y_{1;i}$ of the units $i \in s_1 \setminus s_2$ which rotate out, as these units are not sampled and their missingness is not due to nonresponse.

Let $s_{1;r} = \{i : i \in s_1 \text{ and } a_{1;i} = 1\}$ denote the wave 1 sample of respondents and let $s_{2;r} = \{i : i \in s_2 \text{ and } a_{2;i} = 1\}$ denote the wave 2 sample of respondents. We propose to impute the missing values for the variables $y_{1;i}$ and $y_{2;i}$, using the random hot-deckimputed values given by

$$\begin{aligned} y^*_{1;i} &= z_{1;i} \left\{ (1 - a_{1;i}) \left(\hat{\mu}^r_1 + e_{1;i} \right) + a_{1;i} y_{1;i} \right\} \ , \\ y^*_{2;i} &= z_{2;i} \left\{ (1 - a_{2;i}) \left(\hat{\mu}^r_1 + e_{2;i} \right) + a_{2;i} y_{2;i} \right\} \ , \end{aligned}$$

where

$$z_{1;i} = \begin{cases} 1 & \text{if } i \in s_1, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{and} \qquad z_{2;i} = \begin{cases} 1 & \text{if } i \in s_2, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

and $\hat{\mu}^r_1$ and $\hat{\mu}^r_2$ are the respondents means given by

$$\hat{\mu}^r_1 = \frac{\hat{\tau}^r_1}{\hat{N}^r_1} \qquad \text{and} \qquad \hat{\mu}^r_2 = \frac{\hat{\tau}^r_2}{\hat{N}^r_2} \cdot$$

The quantities $\hat{\tau}_\ell^r$ and $\hat{N}_\ell^r$ are, respectively, the estimators of the respondents' totals and of the number of respondents in the population at wave ℓ. These estimators are given by

$$\hat{\tau}_\ell^r = \sum_{i\in s} \breve{y}^\circ_{\ell;i} \; , \qquad \hat{N}_\ell^r = \sum_{i\in s} \breve{a}^\circ_{\ell;i}$$

where $\breve{y}^\circ_{\ell;i} = y_{\ell;i}\breve{a}^\circ_{\ell;i}$, $\breve{a}^\circ_{\ell;i} = \breve{z}_{\ell;i}a_{\ell;i}$ and $\breve{z}_{\ell;i} = z_{\ell;i}\pi^{-1}_{\ell;i}$. The quantities $\pi_{1;i}$ and $\pi_{2;i}$ denote the first-order inclusion probabilities of the unit i at waves 1 and 2. The quantities $e_{1;i}$ and $e_{2;i}$ are the residuals

$$e_{1;i} = y_{1;j} - \hat{\mu}_1^r \qquad \text{and} \qquad e_{2;i} = y_{2;k} - \hat{\mu}_1^r \; ,$$

where j is a donor selected with replacement with probabilities $\{p_{1;i}\}$ within the set $s_{1;r}$, and k is a donor selected with replacement with probabilities $\{p_{2;i}\}$ within the set $s_{2;r}$. The selection probabilities $p_{\ell;i}$ are given by

$$p_{\ell;i} = \frac{\breve{a}^\circ_{\ell;i}}{\sum_{j\in s}\breve{a}^\circ_{\ell;j}}.$$

Note that we can also use $e_{\ell;i} = 0$, in this case the quantities $y^*_{\ell;i}$ are the imputed values under hot-deck mean imputation.

2.1 The imputed estimator of change

The imputed estimator for change is given by

$$\hat{\Delta}^* = \hat{\tau}_2^* - \hat{\tau}_1^* \; , \tag{2}$$

where $\hat{\tau}_1^*$ and $\hat{\tau}_2^*$ are two cross-sectional imputed Horvitz-Thompson [9] estimators defined by

$$\hat{\tau}_1^* = \sum_{i\in s}\frac{y^*_{1;i}}{\pi_{1;i}} \qquad \text{and} \qquad \hat{\tau}_2^* = \sum_{i\in s}\frac{y^*_{2;i}}{\pi_{2;i}}.$$

3 Variance of the imputed estimator of change

We propose to estimate the variance of (2) using a reverse approach (Fay [6] ; Shao & Steel [16]) for non-response. Let $U_{1;r}$ and $U_{2;r}$ be respectively the population of respondents at wave 1 and 2. In other words, at both waves, the population is randomly split into a population of respondents and a population of non-respondents according to an unknown response mechanism. Let E_r and V_r denote respectively the expectation and the variance operators with respect to that response mechanism.

Rotating samples s_1 and s_2 are selected from the population U according to a rotation sampling design (see §1). The sample of respondents at wave ℓ is given by $s_{\ell;r} = U_{\ell;r} \cap s_\ell$. Let E_d and V_d denote respectively the expectation and the variance operators with respect to the sampling design. Furthermore, we suppose that the random hot-deck imputation described in §2 is used to impute the missing values. Let E_I and V_I denote respectively the expectation and the variance operators with respect to the random imputation.

The overall variance of the imputed estimator of change is given by the following three stage variance,

$$V(\hat{\Delta}^*) = E_r(V_d(\hat{\Delta}^*_{R,S}|R)) + V_r(E_d(\hat{\Delta}^*_{R,S}|R)) + E_r(E_d(V_I(\hat{\Delta}^*|S,R)|R)) \tag{3}$$

where

$$\hat{\Delta}^*_{R,S} = E_I(\hat{\Delta}^*|S,R)$$

and $S = \{s_1, s_2\}$ and $R = \{U_{1;r}, U_{2;r}\}\cdot$ Note that the overall variance includes the effect of the response mechanism, the sampling design and the randomness of the hot-deck imputation. As the second term of (3) is negligible when the sampling fraction is negligible (Shao & Steel [16]), we have that

$$V(\hat{\Delta}^*) \doteq E_r(V_d(\hat{\Delta}^*_{R,S}|R)) + E_r(E_d(V_I(\hat{\Delta}^*|S,R)|R))\cdot \tag{4}$$

Note that as $E_I(e_{\ell;i}|S,R) = 0$, and as the $e_{\ell;i}$ are selected independently, the conditional variance $V_I(\hat{\Delta}^*|S,R)$ is given by

$$V_I(\hat{\Delta}^*|S,R) = \sum_{\ell=1}^{2} V_I(e_{\ell;i}|S,R) \sum_{i\in s} \frac{z_{\ell;i}}{\pi^2_{\ell;i}}(1-a_{\ell;i}) \ ,$$

where

$$V_I(e_{\ell;i}|S,R) = \sum_{i\in s} a_{\ell;i} p_{\ell;i} e^2_{\ell;i}\cdot \tag{5}$$

Note that in (5), we use the same notation for the random variables $e_{\ell;i}$ and their observed values. Note that under mean imputation, we have that $V_I(\hat{\Delta}^*|S,R) = 0$, as $e_{\ell;i} = 0\cdot$

As $E_I(e_{\ell;i}|S,R) = 0$, we have that

$$\hat{\Delta}^*_{R,S} = \hat{N}_2 \hat{\mu}^r_2 - \hat{N}_1 \hat{\mu}^r_1 \ ,$$

where $\hat{N}_\ell = \sum_{i\in s} \breve{z}_{\ell;i}$ is an estimator of $N\cdot$ Note that $\hat{\Delta}^*_{R,S}$ is a function of six totals; that is, $\hat{\Delta}^*_{R,S} = f(\hat{\boldsymbol{\tau}})$ where $\hat{\boldsymbol{\tau}} = (\hat{N}_1, \hat{N}_2, \hat{\tau}^r_1, \hat{\tau}^r_2, \hat{N}^r_1, \hat{N}^r_2)'$ is a vector of Horvitz & Thompson [9] totals. Using Taylor approximation, we have that

$$\hat{\Delta}^*_{R,S} - \Delta \doteq \boldsymbol{\nabla}(\boldsymbol{\tau})'(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \ ,$$

where

$$\boldsymbol{\nabla}(\boldsymbol{\tau}) = \left(\frac{-\tau^r_1}{N^r_1}, \frac{\tau^r_2}{N^r_2}, \frac{-N}{N^r_1}, \frac{N}{N^r_2}, \frac{N\tau^r_1}{(N^r_1)^2}, \frac{-N\tau^r_2}{(N^r_2)^2} \right)'$$

is the gradient of $f(\boldsymbol{\tau})$ at $\boldsymbol{\tau} = (N, N, \tau^r_1, \tau^r_2, N^r_1, N^r_2)'$, where τ^r_1 and τ^r_2 are the population totals of the variable of interest over the respondents at waves 1 and 2; and N^r_1 and N^r_2 are the total number of respondents in the population at waves 1 and 2.

The Taylor approximation of $V_d(\hat{\Delta}^*_{R,S}|R)$ is given by

$$V_d(\hat{\Delta}^*_{R,S}|R) \doteq \boldsymbol{\nabla}(\boldsymbol{\tau})' \boldsymbol{V}_d(\hat{\boldsymbol{\tau}}|R) \boldsymbol{\nabla}(\boldsymbol{\tau}) \ ,$$

where $\boldsymbol{V}_d(\hat{\boldsymbol{\tau}}|R)$ is the design covariance matrix of the vector $\hat{\boldsymbol{\tau}}\cdot$ Thus, an approximately design-based unbiased estimator for $V_d(\hat{\Delta}^*_{R,S}|R)$ is given by the following linearised variance estimator

$$\hat{V}_d(\hat{\Delta}^*_{R,S}|R) = \boldsymbol{\nabla}(\hat{\boldsymbol{\tau}})' \hat{\boldsymbol{V}}_d(\hat{\boldsymbol{\tau}}|R) \boldsymbol{\nabla}(\hat{\boldsymbol{\tau}}) \ , \tag{6}$$

where $\hat{\boldsymbol{V}}_d(\hat{\boldsymbol{\tau}}|R)$ is the approximately design-unbiased estimator (9) of the covariance matrix $\boldsymbol{V}_d(\hat{\boldsymbol{\tau}}|R)\cdot$ Note that in (6), the $a_{1;i}$ and the $a_{2;i}$ are treated as fixed quantities, as it is a conditional variance $V_d(\hat{\Delta}^*_{R,S}|R)$ given $R\cdot$

The proposed estimator of the variance of the imputed estimator of change is given by

$$\hat{V}_I(\hat{\Delta}^*) = \hat{V}_d(\hat{\Delta}^*_{R,S}|R) + V_I(\hat{\Delta}^*|S,R)\cdot \tag{7}$$

The proposed estimator (7) can be generalised for other types of imputation, as long $\hat{\Delta}^*_{R,S}$ is a function of Horvitz-Thompson totals. In this situation the gradient will have a different expression which depends on the used imputation.

The proposed estimator (7) is an approximately unbiased estimator of the variance (3), as the expectation of (7) is given by

$$\begin{aligned}
E_r(E_d(E_I(\hat{V}_I(\hat{\Delta}^*)|S,R)|R)) &= E_r(E_d(E_I(\hat{V}_d(\hat{\Delta}^*_{R,S}|R)|S,R)|R)) \\
&\quad + E_r(E_d(E_I(V_I(\hat{\Delta}^*|S,R))|R)) \\
&\simeq E_r(V_d(\hat{\Delta}^*_{R,S}|R)) + E_r(E_d(V_I(\hat{\Delta}^*|S,R)|R)) \\
&\simeq V(\hat{\Delta}^*)\ ,
\end{aligned}$$

by using (4) and the fact that (6) does not depends on the $e_{\ell;i}\cdot$ The main advantage of the proposed estimator is the fact that it is approximately unbiased under the unknown response mechanism without making assumptions about it for the estimation of the variance .

4 Estimation of the covariance matrix by the multivariate regression approach

In this section, we derive an expression for the estimator of the covariance matrix $\boldsymbol{V}_d(\hat{\boldsymbol{\tau}}|R)$ under rotation sampling. This covariance is not straightforward to estimate because it involves covariance between the components of $\hat{\boldsymbol{\tau}}$ defined from the different samples, s_1 and $s_2\cdot$ Several methods can be used to estimate these covariances (e.g. Kish [12] ; Tam [18], Holmes & Skinner [8], Nordberg [13], Berger [3] ; Qualité & Tillé [15] ; Wood [19]). We propose to use a multivariate (or general) linear regression model to estimate this covariance matrix.

Consider the following $\tilde{n} \times 6$ matrix $\breve{\boldsymbol{Y}}^\circ = (\breve{\boldsymbol{y}}^\circ_1, \ldots, \breve{\boldsymbol{y}}^\circ_{\tilde{n}})'$, where $\tilde{n} = \#\{s_1 \cup s_2\}$ and $\breve{\boldsymbol{y}}^\circ_i = (\breve{z}_{1;i}, \breve{z}_{2;i}, \breve{y}^\circ_{1;i}, \breve{y}^\circ_{2;i}, \breve{a}^\circ_{1;i}, \breve{a}^\circ_{2;i})'\cdot$ Consider the following multivariate (general) regression model

$$\breve{\boldsymbol{Y}}^\circ = \boldsymbol{Z}_s\ \boldsymbol{\alpha} + \boldsymbol{\epsilon}\ , \tag{8}$$

where $\boldsymbol{\alpha}$ is the 3×6 matrix of regression parameters, the residuals $\boldsymbol{\epsilon}$ have a 6×6 covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{Z}_s$ is the $\tilde{n}\times 3$ design matrix which specifies the fixed sizes constraints of the rotation design. The matrix $\boldsymbol{Z}_s$ is defined by $\boldsymbol{Z}_s = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{\tilde{n}})'$, with $\boldsymbol{z}_i = (z_{1;i}, z_{2;i}, z_{1;i} \times z_{2;i})'$ where $z_{1;i}$ and $z_{2;i}$ are defined by (1). The model (8) belongs to the class of general linear model. In fact, (8) is also a multivariate analysis of variance (MANOVA) model, as the independent variables are all dummy variables.

Note that we have the following fixed size constraints $\sum_{i\in s} z_{\ell;i} = n$, $\sum_{i\in s} z_{1;i}z_{2;i} = n_{12}$, because only samples with these sample sizes can be selected. Thus, by using the design variables as independent variables, we are conditioning on them. This takes into account of the fixed size constraints in the estimation of the covariance (see Berger & Priam [4]). Note that the model (8) 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 $\sum_{i\in s} z_{1;i}z_{2;i} = n_{12}$.

Berger & Priam [4] proposed the following estimator for the covariance matrix of the vector $\hat{\boldsymbol{\tau}}$·

$$\hat{\boldsymbol{V}}_d(\hat{\boldsymbol{\tau}}|R) = \hat{\boldsymbol{D}}'\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{D}} \ , \tag{9}$$

where the matrix $\hat{\boldsymbol{\Sigma}}$ is the *Ordinary Least Squares* residual covariance matrix estimate of the model (8) and $\hat{\boldsymbol{D}}$ is a diagonal matrix with diagonal elements $\{\hat{V}(\hat{\tau}_q|R)\hat{\Sigma}_{qq}^{-1}\}^{1/2}$, where $\hat{V}(\hat{\tau}_q|R)$ is a design-based variance estimator of the q-th component of $\hat{\boldsymbol{\tau}}$, and $\hat{\Sigma}_{qq}$ is the q-th diagonal component of $\hat{\boldsymbol{\Sigma}}$· Any unbiased standard variance estimator can be used to calculate $\hat{V}(\hat{\tau}_q|R)$· Note that (9) is positive definite, as $\hat{\boldsymbol{\Sigma}}$ is always positive definite. Hence the proposed variance estimator (7) is always positive.

Berger & Priam [4] showed that the estimator $\hat{\boldsymbol{V}}_d(\hat{\boldsymbol{\tau}}|R)$ defined in (9) is an approximately design unbiased estimator for the covariance matrix when the finite population corrections are negligible. The estimator (9) is a design-consistent estimator for the covariance matrix even when model (8) does not fit the data (Berger & Priam [4]). Note that the estimator (9) takes into account of the unequal probabilities.

In a series of simulations based on the Swedish Labour Force Survey, Andersson *et al.* [2, 1] showed that the estimator (9) gives more accurate estimates than standard variance estimators (Tam [18], Qualité & Tillé [15]) when we are interested in change between strata domains.

5 Discussion

The variance estimator proposed is applicable for unequal rotation sampling designs when random hot-deck imputation is used at both waves and the sampling fractions are negligible. The variance estimator proposed may be extended in various ways. Point estimators, such as calibration estimators (Huang & Fuller [10] ; Deville & Särndal [5]) which employ auxiliary population information may often be expressible as functions of totals. The proposed variance estimator (7) can be modified to accommodate this situation.

The proposed approach is not limited to hot-deck imputation, as the proposed approach can be extended to other imputation techniques, as long as the expectation of the imputed estimator of change under the random imputation method can be expressed as a function of totals. For simplicity, we assumed that we have only one imputation class. The proposed estimator can be extended to several imputation classes which are defined as homogeneous sub-groups of the overall sample s· At the second wave, it is also a common practice to impute using observations from the first wave. It would be useful to generalise the proposed estimator for this method of imputation.

Résumé

Mesurer le changement temporel est un problème central pour de nombreux utilisateurs de données sociales, économiques et démographiques et suscite beaucoup d'intérêt dans de

nombreux domaines des sciences économiques et sociales. Smith et al. [17] ont reconnu que l'estimation du changement est l'un des défis les plus importants dans les statistiques d'enquêtes. Les utilisateurs sont souvent interessés 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 étudiée à deux vagues différentes. Ces estimations transversales comprennent souvent des valeurs imputées pour compenser la non-réponse. L'estimation de la variance d'échantillonnage de l'estimateur du changement imputé est utile pour juger si le changement observé est statistiquement significatif. Les covariances jouent un rôle important dans l'estimation de la variance d'un changement. Nous proposons d'utiliser une approche de régression linéaire multivariée (Berger et Priam [4]) pour estimer les covariances. L'estimateur proposé n'est pas un estimateur basé sur un modèle, et cet estimateur est valable même si le modèle n'ajuste pas les données. Nous montrons comment cette approche peut être utilisée pour tenir compte de l'effet de l'imputation. L'approche par la régression multivariée donne des estimateurs consistents pour la variance des changements lorsque la fraction d'échantillonnage est faible et que les corrections de population finie sont négligeables.

Acknowledgments

This work was supported by the grant RES-000-22-3045 of the Economic and Social Research Council (UK).

References

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

[2] Andersson, C., Andersson, K., and Lundquist, P. Variansskattningar avseende förändringsskattningar i panelundersökningar (variance estimation of change in panel surveys). *Methodology reports from Statistics Sweden (Statistiska centralbyrån)* (2011).

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

[4] Berger, Y. G., and Priam, R. 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* (2010).

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

[6] Fay, B. E. A design-based perspective on missing data variance. *Proceeding of the 1191 Annual Research Conference. U.S. Bureau of the Census* (1991), 429–440.

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

[8] Holmes, D. J., and Skinner, C. J. Variance estimation for labour force survey estimates of level and change. *Government Statistical Service Methodology Series*, 21 (2000).

[9] Horvitz, D. G., and Thompson, D. J. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association 47*, 260 (1952), 663–685.

[10] Huang, E. T., and Fuller, W. A. Nonnegative regression estimation for survey data. *Proceedings Social Statistics Section American Sta-tistical Association* (1978), 300303.

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

[12] Kish, L. *Survey Sampling.* Wiley, 1965.

[13] Nordberg, L. On variance estimation for measures of change when samples are coordinated by the use of permanent random numbers. *Journal of Official Statistics 16* (2000), 363–378.

[14] Ohlsson, E. Coordination of samples using permanent random numbers. *In Business Survey Methods, eds. B. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M. Colledge, and P.S. Kott, New York : Wiley and Sons* (1995).

[15] Qualité, L., and Tillé, Y. Variance estimation of changes in repeated surveys and its application to the swiss survey of value added. *Survey Methodology 34*, 2 (2008), 173–180.

[16] Shao, J., and Steel, P. Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association 94* (1999), 254–265.

[17] Smith, P., Pont, M., and Jones, T. 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 (2003), 257–295.

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

[19] Wood, J. On the covariance between related horvitz-thompson estimators. *Journal of Official Statistics 24*, 1 (2008), 53–78.