Social Statistics Research Centre
University of Southampton

Variance Estimation for Measures of Change in Probability Sampling

Yves G. Berger

Abstract

We propose to estimate the design variance of absolute changes between two cross-sectional estimators under rotating sampling schemes. We show that the variance estimator proposed is generally positive. We also propose possible extensions for stratified samples, with dynamic stratification; that is, when units move between strata and new strata are created at the second waves.

SSRC Methodology Working Paper M03/10

Variance Estimation for Measures of Change in Probability Sampling

Yves G. Berger
Department of Social Statistics,
University of Southampton,
Southampton, SO17 1BJ
United Kingdom
E-mail: ygb@soton.ac.uk

May 20, 2003

Abstract

We propose to estimate the design variance of absolute changes between two cross-sectional estimators under rotating sampling schemes. We show that the variance estimator proposed is generally positive. We also propose possible extensions for stratified samples, with dynamic stratification; that is, when units move between strata and new strata are created at the second waves.

Keywords: Conditional Poisson Sampling, Design-Based Inference, Entropy, Inclusion Probabilities, Repeated Surveys, Rotating Scheme.
AMS Code: 62D05 sampling theory, sample surveys.

1 Introduction

The primary interest of many users of macroeconomic statistics is in the change from one period to another rather than the level of any particular variable (Smith *et al.*, 2003). Most surveys in government are continuing surveys; that is, repeated monthly, quarterly, annually or with some other fixed frequency. An important reason for doing this is to estimate the manner in which a population changes from one survey period (or wave) to the next. In this case, the population is sampled repeatedly and the same study variables are measured at each wave. For example, in many countries, labour-force surveys or business surveys are conducted monthly or quarterly to estimate the change in the number of employed or in the volume of retail sales. In many continuing surveys, it is more and more important to compare cross-sectional estimates for the same study variable taken on two different waves and to judge whether the observed absolute change is statistically significant (Smith *et al.*, 2003). Variance estimation would be relatively straightforward if the sample remained the same from one period to the next. Unfortunately, this is rarely the case, as at each wave, we have a new sample. Indeed, repeated surveys are usually selected with rotating schemes, where new units are selected to replace old units that have been in the sample for a specified number of waves. In this paper, we propose a variance estimator that copes with rotating sampling schemes.

Consider two population totals for the same study variable taken on two waves: $t = 0$ and $t = 1$.

$$\begin{aligned} Y_0 &= \sum_{i \in U} y_{0;i} \\ Y_1 &= \sum_{i \in U} y_{1;i} \end{aligned}$$

where $y_{0;i}$ and $y_{0;i}$ are respectively values of a study variable at time $t = 0$ and $t = 1$. The set U is the population frame. The aim is to estimate the absolute change of level:

$$\Delta = Y_1 - Y_0. \tag{1}$$

A reliable way to measure change is to measure successive levels and take differences. Suppose that Δ is estimated by the difference of cross-sectional estimators of level:

$$\widehat{\Delta} = \widehat{Y}_1 - \widehat{Y}_0, \tag{2}$$

where $\widehat{Y}_0$ and $\widehat{Y}_1$ are Horvitz-Thompson (HT) estimators (Narain, 1951; Horvitz and Thompson, 1952) given by

$$\begin{aligned} \widehat{Y}_0 &= \sum_{i \in s_0} \frac{y_{0;i}}{\pi_{0;i}} \\ \widehat{Y}_1 &= \sum_{i \in s_1} \frac{y_{1;i}}{\pi_{1;i}} \end{aligned} \tag{3}$$

where s_0 and s_1 denote samples reported on the first and second wave. The quantities $\pi_{0;i}$ and $\pi_{1;i}$ are the inclusion probabilities of unit i on $t = 0$ and $t = 1$. These probabilities are defined in Section 2.

There exist alternative point estimators for Δ. For example, changes can be estimated from the matched sample $s_0 \cap s_1$ or by more sophisticated estimators (Fuller and Rao, 2001). Taylor linearization can be even used to approximate alternative measures of change (Andersson and Nordberg, 1994). In this paper, we assume that Δ is estimated by (2).

Our aim is to estimate the variance of (2) given by

$$\operatorname{var}(\widehat{\Delta}) = \operatorname{var}(\widehat{Y}_1) + \operatorname{var}(\widehat{Y}_0) - 2\operatorname{cov}(\widehat{Y}_0, \widehat{Y}_1). \tag{4}$$

under rotating schemes. Standard methods can be used to estimate $\operatorname{var}(\widehat{Y}_1)$ and $\operatorname{var}(\widehat{Y}_0)$. However, variance estimators of $\widehat{\Delta}$ are complicated by the need to evaluate a design covariance in (4).

In practice, covariances are usually estimated from the matched sample (Holmes and Skinner, 2000); that is, by estimating

$$\mathrm{cov}(\widehat{Y}_0^m, \widehat{Y}_1^m) = \frac{-1}{2}\left[\mathrm{var}(\widehat{Y}_2^m) - \mathrm{var}(\widehat{Y}_1^m) - \mathrm{var}(\widehat{Y}_0^m)\right], \tag{5}$$

where $\widehat{Y}_1^m = \sum_{i\in s_0\cap s_1} y_{0;i}\pi_{0;i}^{-1}$, $\widehat{Y}_0^m = \sum_{i\in s_0\cap s_1} y_{1;i}\pi_{1;i}^{-1}$ and $\widehat{Y}_2^m = \widehat{Y}_1^m - \widehat{Y}_0^m$. This method can overestimate the covariance and produce negative variance estimates. Furthermore, as (5) is different from the covariance in (4), an estimator of (5) is therefore biased for $\mathrm{cov}(\widehat{Y}_0, \widehat{Y}_1)$. We will show that the estimator proposed in Section 5 generally gives positive variance estimates and will not be estimated only from the matched sample. Simulation in Section 7 will show that the estimator proposed is more accurate than an estimator based on (5).

We will assume fixed sample sizes for s_0, s_1 and the matched sample $s_0 \cap s_1$. This is the case for most rotating sampling schemes. The variance estimator proposed is based on the following idea. First, we estimate the variances and the covariances unconditionally; that is, assuming the sizes random. Secondly, in order to capture the fixed size feature of the sampling scheme, we derive variances and covariances conditionally on the given numbers of units caught in s_0, s_1 and $s_0 \cap s_1$ (see (17)). The final variance estimator estimates these conditional variances and covariances. We justify this approach (see Section 3) using a 'maximum entropy' model for the sampling scheme.

In Section 2, we define the class of rotating schemes considered in this paper and in Section 3, we propose a model for the rotating scheme. In Section 4, we derive the variance estimator proposed when the population is composed of a single stratum. In Section 6, the estimator proposed is generalized for stratified samples. In Section 7, a series of Monte-Carlo simulations supports our findings.

2 Rotating sampling schemes

In this section, we define the class of rotating sampling schemes considered in this paper. These rotating schemes are commonly used in practice (see Example 1 and Example 3). For simplicity, we suppose that the population is composed of a single stratum of size N. In Section 6, we generalise the variance estimator proposed for stratified samples.

As two samples (s_0 and s_1) are selected, we have two sampling designs: $p_0(s_0)$ the sampling design used at $t = 0$ and $p_1(s_1|s_0)$ the sampling design used at $t = 1$. The first-order inclusion probabilities of $p_0(s_0)$ and $p_1(s_1|s_0)$ are respectively denoted by $\pi_{0;i}$ and $\widetilde{\pi}_{1;i}$. A rotating sampling scheme is the combination of $p_0(s_0)$ and $p_1(s_1|s_0)$.

Assume that s_0 is an unequal probability sample without replacement with first-order inclusion probabilities $\pi_{0;i}$. The size of s_0 is denoted by n_0. Suppose that the sample s_1 is selected conditionally on s_0 in the following way: a simple random sample of n_{01} units is selected without replacement from s_0 and a simple random sample of $n_{1|0}$ units is selected without replacement form U/s_0; where U/s_0 is the set of units not selected on the first wave. With this scheme, the matched samples $s_0 \cap s_1$ contains n_{01} units. We have $n_1 = n_{01} + n_{1|0}$ units in the sample s_1. Unless otherwise stated, the sizes n_0, n_1 and n_{01} are assumed fixed.

Let $g = n_{01}/n_0$ denotes the fraction of the matched sample or the probability for a unit $i \in s_0$ to remain in the sample at the second wave. Let $q = n_{1|0}(N - n_0)^{-1}$ denotes the probability for a non-sampled unit to join the sample at the second wave. The inclusion probability $\widetilde{\pi}_{1;i}$ for a unit $i \in U_1$ to be selected in s_1 conditionally on s_0 is

$$\widetilde{\pi}_{1;i} = g\delta_{0;i} + q(1 - \delta_{0;i}) \tag{6}$$

where $\delta_{0;i} = 1$ if $i \in s_0$ and $\delta_{0;i} = 0$ otherwise. The probability $\widetilde{\pi}_{1;i}$ is random, as it depends on s_0. The probability $\pi_{1;i}$ in (3) is the design expectation of $\widetilde{\pi}_{1;i}$; that is,

$$\pi_{1;i} = E(\widetilde{\pi}_{1;i}) = g\pi_{0;i} + q(1 - \pi_{0;i}). \tag{7}$$

Example 1 *Business surveys in Statistics Canada are selected by rotation group sampling in which each stratum is randomly divided into mutually exclusive rotation groups of the same size. Suppose that we have a single stratum and suppose that the population is randomly divided into P rotation groups (for simplicity, we assume NP^{-1} integer). For the first wave, the first p groups are selected. On the second wave, group 1 rotates out and group $p+1$ rotates in. s_0 and s_1 are therefore simple random samples and $\pi_{0;i} = pP^{-1}$, $n_0 = pNP^{-1}$, $n_{01} = (p-1)NP^{-1}$ and $n_{1|0} = NP^{-1}$. This implies $g = (p-1)/p$ and $q = 1/(P-p)$.*

3 A model for the rotating sampling scheme

We propose to approximate the actual rotating sampling scheme by the maximum entropy rotating sampling scheme (MERS). In this section, we define the entropy and the MERS scheme. We will see that the MERS scheme is conditional on the sizes n_0, n_1 and n_{01} (see Theorem 1). This important characteristic will be used in Section 4 to derive a variance estimator.

Definition 1 *(Hájek, 1959) The entropy of a sampling design $p(s)$ is defined by*

$$H(p(s)) = -\sum_s p(s) \log(p(s)),$$

where $\sum_s$ represents the sum over all possible samples and $\log(.)$ is the Napierian Logarithm, where $0\log(0) = 0$.

Let Ω_0 and Ω_1 denote the sets of all possible samples s_0 and s_1; that is,

$$\begin{aligned} \Omega_0 &= \{s_0 : \#s_0 = n_0\} \\ \Omega_1 &= \{s_1 : \#s_1 = n_1;\ \#(s_0 \cap s_1) = n_{01}\}. \end{aligned}$$

Several schemes can be used to select s_0 and s_1 and the actual scheme is one of them. Let us approximate the actual sampling scheme by the MERS scheme.

Definition 2 *The* MERS *scheme is such that (i) it can select a sample* $s_0 \in \Omega_0$ *and a sample* $s_1 \in \Omega_1$ *and (ii) it is such that* $p_0(s_0)$ *and* $p_1(s_1|s_0)$ *have the largest entropy with first-order inclusion probabilities* $\pi_{0;i}$ *and* $\widetilde{\pi}_{1;i}$*.*

The MERS scheme is such that $p_0(s_0)$ and $p_1(s_1|s_0)$ have the largest entropy among all possible sampling designs that can select samples s_0 and s_1 from Ω_0 and Ω_1. Theorem 1 gives an analytic expression for the MERS scheme and shows that the MERS scheme is a Poisson sampling scheme conditional on n_0, n_1 and n_{01}.

Theorem 1 *The* MERS *scheme is defined by the combination of two conditional Poisson sampling designs*

$$p_0(s_0) = \frac{P_0(s_0)}{P_0(s_0 \in \Omega_0)}\delta\{s_0 \in \Omega_0\} \tag{8}$$

$$p_1(s_1|s_0) = \frac{P_1(s_1|s_0)}{P_1(s_1 \in \Omega_1|s_0)}\delta\{s_1 \in \Omega_1\} \tag{9}$$

where $\delta\{A\} = 1$ *if* A *is true and* $\delta\{A\} = 0$ *otherwise. The sampling designs* $P_0(s_0)$ *and* $P_1(s_1|s_0)$ *are Poisson sampling designs defined by*

$$P_0(s_0) = \prod_{i \in s_0} p_{0,i} \prod_{j \notin s_0} (1 - p_{0,j}) \tag{10}$$

$$P_1(s_1|s_0) = \prod_{i \in s_1} \widetilde{p}_{1,i} \prod_{j \notin s_1} (1 - \widetilde{p}_{1,j}). \tag{11}$$

The $p_{0,i}$ *are such that the first-order inclusion probabilities of* $p_0(s_0)$ *are given by* $\pi_{0,i}$*. The* $\widetilde{p}_{1,i}$ *are such that the first-order inclusion probabilities of* $p_1(s_1|s_0)$ *are given by* $\widetilde{\pi}_{1,i}$ *defined by (6). The probability* $P_0(s_0 \in \Omega_0)$ *is the probability of selecting a sample* $s_0 \in \Omega_0$ *under* $P(s_0)$*. The probability* $P_1(s_1 \in \Omega_1|s_0)$ *is the probability of selecting a sample* $s_1 \in \Omega_1$ *under* $P_1(s_1|s_0)$*.*

The proof is given in Appendix A.

As (8) and (9) are conditional probabilities, we see that the MERS scheme is conditional on n_0, n_1 and n_{01}. In Section 4, we derive the variance of $\widehat{\Delta}$ with respect to MERS. The variance will be not derived with respect to (10) and (11), as n_0, n_1 and n_{01} are not fixed with (10) and (11).

The MERS scheme is not necessarily the actual sampling scheme implemented. However, MERS can be implemented exactly if s_0 is a simple random sample without replacement (SRSWR). Indeed, if $\pi_{0;i} = n/N$, (8) is the SRSWR design. The MERS scheme approximates the actual scheme, as long as we have unequal probabilities. For highly randomised (with high entropy) sampling designs, (8) is a good approximation (Berger, 1998). Rao-Sampford, successive, rejective or simple random sampling designs are highly randomised (Berger, 1998).

The $p_{0;i}$ and $\widetilde{p}_{1;i}$ can be computed exactly (Hájek, 1980 page 139 and Chen *et al.*, 1994). Using Theorem 5.1 in Hájek (1964), we can show that $p_{0;i} \approx \pi_{0;i}$ and $\widetilde{p}_{1;i} \approx \widetilde{\pi}_{1;i}$. For simplicity, we assume

$$\begin{aligned} p_{0;i} &= \pi_{0;i} \\ \widetilde{p}_{1;i} &= \widetilde{\pi}_{1;i}. \end{aligned}$$

4 Estimator for the variance of change under MERS

In this section, assuming normality, we will derive a simple approximation for the variance of $\widehat{\Delta}$ with respect to MERS, using the fact that this scheme is conditional on n_0, n_1 and n_{01}. First, we will estimate the variances and the covariances unconditionally; that is, assuming that the sample sizes are random. Secondly, in order to capture the fixed sizes feature of the sampling scheme, we derive variances and covariances conditionally on the given numbers of units caught in the first sample, in the second sample and in both. Finally, we will estimate these conditional variances and covariances.

First, consider the following Poisson scheme: where s_0 is selected according to (10) and s_1 is selected according to (11). The Poisson scheme and the MERS scheme are two different schemes. It is important to note that under this Poisson sampling scheme, the sizes n_0, n_1 and n_{01} are now random.

Let us assume that the vector

$$\mathbf{u} = (\widehat{Y}_1, \widehat{Y}_0, n_0, n_1, n_{01})' \tag{12}$$

has a normal distribution under the Poisson sampling scheme; that is,

$$\mathbf{u} \sim N(\boldsymbol{\mu}_{\mathbf{u}}, \boldsymbol{\Sigma}_{\mathbf{u}}) \tag{13}$$

where $\boldsymbol{\mu}_{\mathbf{u}}$ and $\boldsymbol{\Sigma}_{\mathbf{u}}$ are respectively the mean and the variance-covariance matrix of $\mathbf{u}$ under the Poisson scheme. In Section 4.3, we derive the expression for $\boldsymbol{\Sigma}_{\mathbf{u}}$.

This normality assumption is not an assumption about the distribution of the study variable. It concerns the sampling distribution of $\mathbf{u}$ under the Poisson scheme. This assumption is natural, as each unit is selected independently under Poisson sampling. Moreover, using Lemma 3.2 in Hájek (1964), we can derive the Lindeberg conditions for asymptotic normality.

4.1 An approximation for the variance

The variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{u}}$ in (13) can be partitioned as

$$\boldsymbol{\Sigma}_{\mathbf{u}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{yy}} & \boldsymbol{\Sigma}_{\mathbf{yn}} \\ \boldsymbol{\Sigma}'_{\mathbf{yn}} & \boldsymbol{\Sigma}_{\mathbf{nn}} \end{bmatrix}$$

where $\boldsymbol{\Sigma}_{\mathbf{yy}}$ is the variance-covariance of the vector $(\widehat{Y}_1, \widehat{Y}_0)'$, $\boldsymbol{\Sigma}_{\mathbf{nn}}$ is the variance-covariance matrix of $(n_0, n_1, n_{01})'$ and $\boldsymbol{\Sigma}_{\mathbf{yn}}$ is the covariance between $(\widehat{Y}_1, \widehat{Y}_0)'$

and $(n_0, n_1, n_{01})'$. All the variances and covariances are with respect to the unconditional Poisson sampling scheme (10) and (11).

As (8) and (9) are conditional probabilities, the MERS scheme is conditional on $(n_0, n_1, n_{01})'$. Thus, the variance-covariance matrix of $(\widehat{Y}_1, \widehat{Y}_0)'$ under MERS is the conditional variance-covariance matrix conditionally on $(n_0, n_1, n_{01})'$ given by

$$\mathbf{\Sigma_{yy|n}} = \mathbf{\Sigma_{yy}} - \mathbf{\Sigma_{yn}}\mathbf{\Sigma_{nn}^{-}}\mathbf{\Sigma_{yn}'} = \begin{bmatrix} \mathrm{var}_M(\widehat{Y}_0) & \mathrm{cov}_M(\widehat{Y}_0, \widehat{Y}_1) \\ \mathrm{cov}_M(\widehat{Y}_0, \widehat{Y}_1) & \mathrm{var}_M(\widehat{Y}_1) \end{bmatrix}, \qquad (14)$$

as the distribution of $\mathbf{u}$ is normal under the Poisson scheme. $\mathbf{\Sigma_{nn}^{-}}$ denotes a generalised inverse of $\mathbf{\Sigma_{nn}}$. Lawley (1943) shows that (14) holds under assumption weaker than (13). The variance of $\widehat{\Delta}$ with respect to the MERS is therefore approximated by

$$\mathrm{var}_M(\widehat{\Delta}) = \mathrm{var}_M(\widehat{Y}_0) + \mathrm{var}_M(\widehat{Y}_1) - 2\ \mathrm{cov}_M(\widehat{Y}_0, \widehat{Y}_1). \qquad (15)$$

As $\mathbf{\Sigma_{yy|n}}$ is a function of $\mathbf{\Sigma_u}$, the approximation (15) depends on $\mathbf{\Sigma_u}$. In Section 4.3, we will derive an analytic expression for $\mathbf{\Sigma_u}$.

4.2 The estimator proposed for the variance

Suppose that we have an estimator $\widehat{\mathbf{\Sigma}}_\mathbf{u}$ of $\mathbf{\Sigma_u}$. A natural estimator of (15) is

$$\widehat{\mathrm{var}}_M(\widehat{\Delta}) = \widehat{\mathrm{var}}_M(\widehat{Y}_0) + \widehat{\mathrm{var}}_M(\widehat{Y}_1) - 2\ \widehat{\mathrm{cov}}_M(\widehat{Y}_0, \widehat{Y}_1) \qquad (16)$$

where

$$\widehat{\mathbf{\Sigma}}_\mathbf{yy|n} = \widehat{\mathbf{\Sigma}}_\mathbf{yy} - \widehat{\mathbf{\Sigma}}_\mathbf{yn}\widehat{\mathbf{\Sigma}}_\mathbf{nn}^{-}\widehat{\mathbf{\Sigma}}_\mathbf{yn}' = \begin{bmatrix} \widehat{\mathrm{var}}_M(\widehat{Y}_0) & \widehat{\mathrm{cov}}_M(\widehat{Y}_0, \widehat{Y}_1) \\ \widehat{\mathrm{cov}}_M(\widehat{Y}_0, \widehat{Y}_1) & \widehat{\mathrm{var}}_M(\widehat{Y}_1) \end{bmatrix} \qquad (17)$$

where $\widehat{\mathbf{\Sigma}}_\mathbf{yy}$, $\widehat{\mathbf{\Sigma}}_\mathbf{yn}$ and $\widehat{\mathbf{\Sigma}}_\mathbf{nn}$ are the submatrices of $\widehat{\mathbf{\Sigma}}_\mathbf{u}$.

The variance estimator (16) should be asymptotically unbiased. Indeed, the variance (15) is a function of population totals given by the components of $\mathbf{\Sigma_u}$ (see Appendix C). In (16), these totals are replaced by their unbiased HT estimators (see Appendix C). If we assume that this function of total is smooth enough, the estimator (16) is then asymptotically unbiased.

When s_0 and s_1 are selected the permanent random number sampling or with collocated sampling (Brewer et al., 1972), the sizes n_0, n_1 and $n_{1|0}$ are not fixed. Sequential Poisson sampling scheme (Ohlson, 1990) is another example of sampling scheme with random sizes. If the variability of these sizes is small, on can assume them fixed. However, if we want to take this variability into account, we can use Nordberg's approach (Nordberg, 2000 page 367). Note that if the actual sampling scheme is the Poisson sampling scheme, the transformation (17) is not necessary and a variance estimator is obtained by using $\widehat{\mathbf{\Sigma}}_\mathbf{yy|n}$ instead of $\widehat{\mathbf{\Sigma}}_\mathbf{yy}$ in (16).

4.3 The unconditional variance-covariance matrix $\mathbf{\Sigma_u}$

It can be shown (see Appendix B) that

$$\mathbf{\Sigma_u} = \breve{\mathbf{A}}_U' \; \mathbf{C}_U \; \breve{\mathbf{A}}_U \tag{18}$$

where

$$\breve{\mathbf{A}}_U = \begin{bmatrix} \breve{\mathbf{y}}_0 & \mathbf{0} & \breve{\mathbf{z}}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \breve{\mathbf{y}}_1 & \mathbf{0} & \breve{\mathbf{z}}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \breve{\mathbf{z}}_0 \end{bmatrix}$$

$\breve{\mathbf{y}}_0 = (\breve{y}_{0;1}, ..., \breve{y}_{0;N})'$, $\breve{y}_{0;i} = y_{0;i}\pi_{0;i}^{-1}$, $\breve{\mathbf{y}}_1 = (\breve{y}_{1;1}, ..., \breve{y}_{1;N})'$, $\breve{y}_{1;i} = y_{1;i}\pi_{1;i}^{-1}$, $\breve{\mathbf{z}}_0$ is a $N \times 1$ vector of 1 and

$$\mathbf{C}_U = \begin{bmatrix} \mathbf{C}_{0;0} & \mathbf{C}_{0;1} & \mathbf{C}_{0;01} \\ \mathbf{C}_{0;1} & \mathbf{C}_{1;1} & \mathbf{C}_{1;01} \\ \mathbf{C}_{0;01} & \mathbf{C}_{1;01} & \mathbf{C}_{01;01} \end{bmatrix}. \tag{19}$$

The sub-matrices of (19) are $N \times N$ diagonal matrices given by

$$\begin{aligned} \mathbf{C}_{0;0} &= diag\{\pi_{0;i}(1-\pi_{0;i}) : i \in U\} && (20) \\ \mathbf{C}_{1;1} &= diag\{\pi_{1;i}(1-\pi_{1;i}) : i \in U\} && (21) \\ \mathbf{C}_{01;01} &= diag\{g\pi_{0;i}(1-g\pi_{0;i}) : i \in U\} && (22) \\ \mathbf{C}_{0;1} &= diag\{\pi_{0;i}(g-\pi_{1;i}) : i \in U\} && (23) \\ \mathbf{C}_{0;01} &= diag\{g\pi_{0;i}(1-\pi_{0;i}) : i \in U\} && (24) \\ \mathbf{C}_{1;01} &= diag\{g\pi_{0;i}(1-\pi_{1;i}) : i \in U\}. && (25) \end{aligned}$$

5 Estimation of the matrix $\mathbf{\Sigma_u}$

All components of (43) are population totals that can be estimated by their corresponding unbiased HT estimators (see Appendix C). The estimators of $\mathbf{\Sigma_u}$ is (see Appendix C)

$$\widehat{\mathbf{\Sigma}}_\mathbf{u} = \breve{\mathbf{A}}_s' \; \breve{\mathbf{C}}_s \; \breve{\mathbf{A}}_s, \tag{26}$$

where $\breve{\mathbf{A}}_s$ is the $n \times 5$ sub-matrix of the $N \times 5$ matrix $\breve{\mathbf{A}}_U$; that is, $\breve{\mathbf{A}}_s$ is obtained by striking out the rows corresponding to the non-sampled units. The $n \times n$ matrix $\breve{\mathbf{C}}_s$ is the principal sub-matrix of the $N \times N$ matrix $\breve{\mathbf{C}}_U$ (defined by (27)); that is, $\breve{\mathbf{C}}_s$ is obtained by striking out the same rows and columns corresponding to the non-sampled units. The matrix $\breve{\mathbf{C}}_U$ is defined by

$$\breve{\mathbf{C}}_U = \begin{bmatrix} \breve{\mathbf{C}}_{0;0} & \breve{\mathbf{C}}_{0;1} & \breve{\mathbf{C}}_{0;0} \\ \breve{\mathbf{C}}_{0;1} & \breve{\mathbf{C}}_{1;1} & \breve{\mathbf{C}}_{1;1} \\ \breve{\mathbf{C}}_{0;0} & \breve{\mathbf{C}}_{1;1} & \breve{\mathbf{C}}_{01;01} \end{bmatrix} \tag{27}$$

where

$$\begin{aligned}\breve{\mathbf{C}}_{0;0} &= diag\left\{(1-\pi_{0;i}) : i \in s\right\} &(28)\\ \breve{\mathbf{C}}_{1;1} &= diag\left\{(1-\pi_{1;i}) : i \in s\right\} &(29)\\ \breve{\mathbf{C}}_{01;01} &= diag\left\{(1-\pi_{0;i}g) : i \in s\right\} &(30)\\ \breve{\mathbf{C}}_{0;1} &= diag\left\{(1-\pi_{1;i}g^{-1}) : i \in s\right\}. &(31)\end{aligned}$$

The sub-matrices of $\widehat{\mathbf{\Sigma}}_{\mathbf{u}}$ give $\widehat{\mathbf{\Sigma}}_{\mathbf{yy}}$, $\widehat{\mathbf{\Sigma}}_{\mathbf{yn}}$ and $\widehat{\mathbf{\Sigma}}_{\mathbf{nn}}$, which substituted into (17) give the variance estimator. The Matrix $\widehat{\mathbf{\Sigma}}_{\mathbf{u}}$ is a 5×5 matrix of HT totals. The expression of these totals can be derived from (26).

Remark 1 *$\widehat{\mathbf{\Sigma}}_{\mathbf{u}}$ can be computed using the S-Plus® and R® routines that can be downloaded free of charge from the author webpage. These routines also accommodate stratification (see Section 6).*

Example 2 *Consider a simple random sample $s_0 = \{1, 11, 15, 16, 19, 20\}$ of sizes $n_0 = 6$ in a population of size $N = 20$. Consider that $s_1 = \{1, 5, 7, 11, 16, 17\}$ has been selected with $g = 0.5$ and $q = 3/14$. This implies $\pi_{0;i} = \pi_{1;i} = 6/20$. The matrix $\widehat{\mathbf{\Sigma}}_{\mathbf{u}}$ is the following symmetric matrix*

$$\widehat{\mathbf{\Sigma}}_{\mathbf{u}} = \left[\begin{array}{cc|ccc} 0.7\omega^2\widehat{\sigma}_{00} & 0.4\omega^2\widehat{\sigma}_{01} & 0.7\omega\widehat{\tau}_0 & 0.4\omega\widehat{\tau}_0^m & 0.4\omega\widehat{\tau}_0^m \\ - & 0.7\omega^2\widehat{\sigma}_{11} & 0.4\omega\widehat{\tau}_1^m & 0.7\omega\widehat{\tau}_1 & 0.7\omega\widehat{\tau}_1^m \\ \hline - & - & 4.2 & 1.2 & 2.1 \\ - & - & - & 4.2 & 2.1 \\ - & - & - & - & 2.55 \end{array}\right]$$

where $\omega = \pi_{0;i}^{-1} = 20/6$ is the sampling weight and $\widehat{\sigma}_{00} = \sum_{i\in s_0} y_{0;i}^2$, $\widehat{\sigma}_{11} = \sum_{i\in s_0} y_{1;i}^2$, $\widehat{\sigma}_{01} = \sum_{i\in s_0\cap s_1} y_{0;i}y_{1;i}$, $\widehat{\tau}_0 = \sum_{i\in s_0} y_{0;i}$, $\widehat{\tau}_1 = \sum_{i\in s_1} y_{1;i}$, $\widehat{\tau}_0^m = \sum_{i\in s_0\cap s_1} y_{0;i}$ and $\widehat{\tau}_1^m = \sum_{i\in s_0\cap s_1} y_{1;i}$. Thus (17) implies

$$\widehat{\mathbf{\Sigma}}_{\mathbf{y}\mathbf{y}|\mathbf{n}} = \left[\begin{array}{cc} 0.7\omega^2\left(\widehat{\sigma}_{00} - \dfrac{\widehat{\tau}_0^2}{n}\right) & \widehat{cov}(\widehat{Y}_0, \widehat{Y}_1) \\ \widehat{cov}(\widehat{Y}_0, \widehat{Y}_1) & 0.7\omega^2\left(\widehat{\sigma}_{11} - \dfrac{\widehat{\tau}_1^2}{n}\right) \end{array}\right] \quad (32)$$

where

$$\begin{aligned}\widehat{cov}(\widehat{Y}_0, \widehat{Y}_1) &= \omega^2 [0.4\ \widehat{\sigma}_{01} - 0.1778\ \widehat{\tau}_0^m\ \widehat{\tau}_1^m \\ &\quad -0.0445\,(\widehat{\tau}_0^c\ \widehat{\tau}_1^c - \widehat{\tau}_0^m\ \widehat{\tau}_1^c - \widehat{\tau}_0^c\ \widehat{\tau}_1^m)] \qquad (33)\end{aligned}$$

and $\widehat{\tau}_1^c = \sum_{i\in s_1\setminus s_0} y_{1;i}$, $\widehat{\tau}_0^c = \sum_{i\in s_0\setminus s_1} y_{0;i}$. The constant $0.7 = 1 - \pi_{0;i}$ and $0.4 = 1 - \pi_{1;i}g^{-1}$ are finite population correction (see Section 5.2). It is worth noticing that the diagonal components of (32) are the usual variances under SRSWR. *Note that the covariance between $\widehat{Y}_0$ and $\widehat{Y}_1$ estimated from the matched* *sample is* $\widehat{cov}(\widehat{Y}_0^*, \widehat{Y}_1^*) = \omega^2(1-f^*)\widehat{\sigma}_{01} - (1-f^*)(n^*-1)\widehat{\tau}_0^m\ \widehat{\tau}_1^m$ *(See (5) and Särndal et al., 1992 page 170). If* $f^* = \pi_{0;i}g^{-1} = 0.6$ *and* n^* *is the size of the matched sample* ($n^* = 3$)*, we have*

$$\widehat{cov}(\widehat{Y}_0^*, \widehat{Y}_1^*) = \omega^2\,(0.4\ \widehat{\sigma}_{01} - 0.2\ \widehat{\tau}_0^m\ \widehat{\tau}_1^m)$$

which is approximately the first two components of (33). The third term of (33) can be therefore interpreted as a bias correction.

5.1 Non-negativeness of the variance estimator

Negative variance estimates of change is a common issue (Hidiroglou *et al.*, 1995). Theorem 2 shows that the estimator (16) for the variance generally gives positive estimates.

Theorem 2 *the matrix* $\widehat{\mathbf{\Sigma}}_{\mathbf{yy|n}}$ *is non-negative definite, if* $\eta_i > 0$ *and* $\tau_i > 0$ *for all* $i \in U$*; where*

$$\begin{aligned}\eta_i &= \varphi_{1;i}\varphi_{3;i} + \varphi_{2;i}\varphi_{3;i} + \varphi_{2;i}\varphi_{1;i} - \varphi_{1;i}^2 - \varphi_{2;i}^2 - \varphi_{3;i}^2\\ \tau_i &= 3\varphi_{1;i}\varphi_{2;i}\varphi_{3;i} - \varphi_{2;i}\varphi_{1;i}^2 - \varphi_{2;i}^2\varphi_{1;i} - \varphi_{4;i}^2\varphi_{3;i}\end{aligned}$$

$\varphi_{1;i} = 1 - \pi_{1;i}$*,* $\varphi_{2;i} = 1 - \pi_{0;i}$*,* $\varphi_{3;i} = 1 - \pi_{0;i}g$*,* $\varphi_{3;i} = 1 - \pi_{0;i}g$ *and* $\varphi_{4;i} = 1 - \pi_{1;i}g^{-1}$.

The proof is given in Appendix D.

Theorem 2 implies that the variance of any linear combination of $\widehat{Y}_0$ and $\widehat{Y}_1$ is always positive and in particular $\widehat{\text{var}}(\widehat{\Delta}) \geq 0$. Note that the conditions $\eta_i > 0$ and $\tau_i > 0$ are not necessary conditions; that is, $\widehat{\mathbf{\Sigma}}_{\mathbf{yy|n}}$ can be non-negative definite even if $\eta_i \leq 0$ or $\tau_i \leq 0$.

The conditions $\eta_i > 0$ and $\tau_i > 0$ can be verifies prior sampling. For example, with equal probability sampling, $\pi_{0;i} = \pi_{1;i} = f = n_0/N$ and η_i is always positive as

$$\begin{aligned}\eta_i &= 2(1-f)(1-fg) - (1-f)^2 - (1-fg^{-1})^2\\ &> 2(1-f)(1-fg) - 2(1-f)^2 = 2(1-f)(f-fg) > 0.\end{aligned}$$

The values of τ_i are positive when g is not too small; that is, when the size of the matched sample is not too small. In Table 1, we have the minimum value of g that guarantees $\tau_i > 0$ for f ranging from 0.01 to 0.20 and for SRSWR. We see that g must be larger than 0.56. Thus, when the matched sample is composed of more than 56% of s_0, $\widehat{\mathbf{\Sigma}}_{\mathbf{yy|n}}$ is non-negative definite and $\widehat{\text{var}}(\widehat{\Delta}) \geq 0$. This can be interpreted by the fact than small matched samples imply inaccurate covariance estimates. For smaller value of g, (16) is not necessarily negative, as $\eta_i > 0$ and $\tau_i > 0$ are not necessary conditions. However, Theorem 2 cannot guarantee that (16) will be always positive.

f	$\min\{g : \tau_i > 0\}$	f	$\min\{g : \tau_i > 0\}$
0.01	0.50	0.12	0.53
0.03	0.51	0.14	0.54
0.05	0.51	0.16	0.54
0.07	0.52	0.18	0.55
0.09	0.52	0.20	0.56

Table 1: Minimum value for g that guarantees positive variance estimates with SRSWR.

Note that the minimum value of g in Table 1 only increase slightly with f. Thus, we do not expect different conclusion with unequal probabilities. In Table 2, we have minimum of g that guarantees $\tau_i > 0$ and $\eta_i > 0$ for all i under unequal probability sampling with a population of size $N = 1000$ and a skewed distribution for $\pi_{0;i}$ given by $\pi_{0;i} \propto (i/N)^5 + 1/5$. We see that g must be larger than 0.62. The skewness of the $\pi_{0;i}$ can explain why we obtain this larger value.

f	$\min\{g : \tau_i > 0 \text{ and } \eta_i > 0\}$	f	$\min\{g : \tau_i > 0 \text{ and } \eta_i > 0\}$
0.01	0.60	0.12	0.61
0.03	0.60	0.14	0.61
0.05	0.60	0.16	0.61
0.07	0.60	0.18	0.62
0.09	0.60	0.20	0.62

Table 2: Minimum value for g that guarantees positive variance estimates with unequal probability sampling.

5.2 Finite population correction involved

The estimator (16) proposed for the variance involves different finite population corrections (FPC). In fact, these FPC are given by the non null components of $\check{\mathbf{C}}_U$. We have four different sort of FPC.

- $(1 - \pi_{0;i})$ and $(1 - \pi_{1;i})$ are the FPC for the variance of HT totals based respectively on s_0 and s_1. These FPC varie between units because of the unequal probabilities. These FPC are small for units selected with a large probability. Note that if s_0 is a SRSWR, $(1 - \pi_{0;i})$ equals $(1 - n_0/N)$.
- $(1 - g\pi_{0;i})$ are the FPC of the matched sample $s_0 \cap s_1$. The quantity $g\pi_{0;i}$ is the probability for $i \in s_0 \cap s_1$. These FPC are small for units having a large probability to be selected in the matched sample.
- We can interpret $(1 - \pi_{1;i}g^{-1})$ as generalised FPC for covariance between HT totals based on s_0 and s_1. As $\pi_{1;i}g^{-1} = pr(i \in s_0)pr(i \in s_1) / pr(i \in s_0 \text{ and } i \in s_1)$, these FPC are small when s_0 and s_1 are independent. These FPC can be interpreted as FPC due to the dependence of s_0 and s_1.

6 Generalisation to stratification

In this section, we show how the estimator (16) can be generalised to take stratification into account. We will show that the variance estimator (16) can still be used; if we modify the matrix $\breve{\mathbf{A}}_s$ or the matrix $\breve{\mathbf{C}}_s$.

6.1 Fixed stratification

Suppose that the population is split into the same strata $\{U_1, ..., U_H\}$ on both waves. Suppose that at $t = 1$, units are selected from s_0 with the same probability g and from $U \backslash s_0$ with the same probability q. Consider a set of H_0 stratification variables given by the indicator variables for the strata; that is, $z_{0;ih} = 1$ if i belongs to the h-th stratum and $z_{0;ih} = 0$ otherwise. This information is summarized in the following matrix

$$\breve{\mathbf{Z}}_0 = \{z_{0;ih}\}_{i=1,..,N;h=1,..,H_0} \tag{34}$$

Consider a vector $\mathbf{u}$ containing the strata sample sizes.

$$\mathbf{u} = \left(\widehat{Y}_1, \widehat{Y}_0, n_{0;1}, ..., n_{0;H}, n_1, n_{01}\right)' \tag{35}$$

where $n_{0;h} = \#(U_h \cap s_0)$. As (12) and (35) are different, the matrix $\breve{\mathbf{A}}_U$ is now given by

$$\breve{\mathbf{A}}_U = \begin{bmatrix} \breve{\mathbf{y}}_0 & \mathbf{0} & \breve{\mathbf{Z}}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \breve{\mathbf{y}}_1 & \mathbf{0} & \breve{\mathbf{z}}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \breve{\mathbf{z}}_0 \end{bmatrix}. \tag{36}$$

The MERS is still defined as in Section 3 but now

$$\begin{aligned} \Omega_0 &= \{s_0 : \#(U_h \cap s_0) = n_{0;h};\ h = 1, .., H\} \\ \Omega_1 &= \{s_1 : \#(U \cap s_1) = n_1;\ \#(U \cap s_1 \cap s_0) = n_{01};\ h = 1, .., H\}. \end{aligned}$$

Using the same methodology as in Section 5, we obtain a variance approximation (15) based on the matrix $\mathbf{\Sigma_u}$ defined by (18) with $\breve{\mathbf{A}}_U$ given by (36) and $\mathbf{C}_U$ given by (19). The final variance estimator is given by (16), where $\breve{\mathbf{A}}_s$ is the sub-matrix of (36). It is worth noticing that $\breve{\mathbf{A}}_s$ specifies the stratification and the survey variables. On the other hand, $\breve{\mathbf{C}}_s$ does not depend on the stratification.

The variance estimator (16) gives also positive variance estimates under the condition of Theorem 2, as $\widehat{\mathbf{\Sigma}}_\mathbf{u}$ is still given by (26) with $\breve{\mathbf{A}}_s$ a sub-matrix of (36) and as (27) is non-negative definite under the condition of Theorem 2 (see Appendix D). Indeed, $\breve{\mathbf{C}}_s$ is non-negative, as $\breve{\mathbf{C}}_U$ is non-negative (see Appendix D and Harville, 1997 pp. 214). Thus, the quadratic form (26) is non-negative (see Harville, 1997 pp. 213).

Example 3 *With the British Labour Force Survey, a systematic samples of addresses (clusters) is selected each wave (quarter). Each systematic sample is* *retained in the survey for five consecutive waves. Systematic sampling implies an implicit stratification where approximately the same number of units rotate in and out. This stratification is fixed if we assume that the population of addresses is the same on both waves. The probability g is the fraction of addresses retained in s_0. The estimator in this Section can be used with this survey.*

6.2 Fixed stratification and varying rotation within strata

Suppose that at $t = 1$, the probabilities to join and leave the sample are not constant, as previously assumed; that is, suppose that these probabilities varies from one stratum to another and are given by $g_h = n_{01h}/n_{0h}$ and $q_h = n_{1|0h}(N_h - n_{0h})^{-1}$ for $i \in U_h$. In this case, the matrix $\breve{\mathbf{C}}_U$ is given by (19) after replacing g by g_h and q by q_h. The matrix $\breve{\mathbf{A}}_s$ is the sub-matrix of (36) where, $\breve{\mathbf{z}}_0$ is replaced by $\breve{\mathbf{Z}}_0$, as the sizes $n_{1;h} = \#(U_h \cap s_1)$ and $n_{01;h} = \#(U_h \cap s_1 \cap s_0)$ are fixed. The conditions of Theorem 2 applied for each stratum guarantee (16) positive.

6.3 Dynamic stratification

By dynamic stratification, we mean that the stratification at $t = 0$ is different from the stratification at $t = 1$; that is, new strata are created and units move between strata. If at $t = 1$, units are selected from s_0 with the same probability g and from $U \backslash s_0$ with the same probability q, the probabilities g and q are constant across strata and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}$ is still given by (26) with $\breve{\mathbf{A}}_s$ a sub-matrix of (36).

If at $t = 1$, these probabilities varies from one stratum to another (see Section 6.2), we need to take into account of the fixed number of units that rotate in and out at $t = 1$; that is, we need to take the stratification at $t = 1$ into account in $\breve{\mathbf{A}}_U$. Suppose that U is composed of H_0 strata at $t = 0$ and of H_1 strata at $t = 1$. At $t = 0$, we have a set of H_0 stratification variables given by (34). Equivalently, the stratification variables at $t = 1$ are given by an $N \times H_1$ matrix $\breve{\mathbf{Z}}_1 = \{z_{1;ih}\}$. As $n_{0;h}$, $n_{1;h}$ and $n_{01;h}$ are fixed, $\breve{\mathbf{A}}_s$ is the sub-matrix of

$$\breve{\mathbf{A}}_U = \begin{bmatrix} \breve{\mathbf{y}}_0 & \mathbf{0} & \breve{\mathbf{Z}}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \breve{\mathbf{y}}_1 & \mathbf{0} & \breve{\mathbf{Z}}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \breve{\mathbf{Z}}_0 \end{bmatrix}$$

and $\breve{\mathbf{C}}_s$ are the sub-matrix of (19) after replacing g by g_h and q by q_h. The final variance estimator is still given by (16).

7 An empirical study

Consider a population composed of a single stratum of size $N = 500$. The study variable $y_{0;i}$ will be generated according to a log-normal distribution $LogN(\mu, \sigma)$ with $\mu = 3$. Different values for σ will be considered. The $\pi_{0;i}$ are proportional to a size variable correlated with the $y_{0;i}$ with a coefficient of correlations of 0.6. The study variable $y_{1;i}$ is generated randomly from $y_{0;i}$ with a linear model that gives a correlation of 0.9 between the $y_{1;i}$ and the $y_{0;i}$. The values of $y_{1;i}$ and $y_{0;i}$ are standardized such that the mean of $y_{1;i}$ and $y_{0;i}$ are respectively given by 10 and $(10+\Delta/N)$; and such that the standard deviation of $y_{1;i}$ and $y_{0;i}$ are 5 and 7. We will consider different values for Δ varying from 1 to 10.

At $t=0$, the S-Plus® function `sample()` is used to select s_0. At $t=1$, we select a simple random sample from s_0 and U/s_0 with $g=0.8$. On both wave, the sample sizes are the same and given by $n=n_0=n_1$. Different values for n will be considered.

We will compare the accuracy of the variance estimator proposed (16) with the estimator based on (5) where the variances are computed with the standard variance estimator based on the assumption of with replacement sampling; that is

$$\widehat{\text{var}}(\widehat{Y}_\ell^*) = \frac{n}{n-1} \sum_{i \in s_0 \cap s_1} \left(\frac{y_{\ell;i}}{\pi_{\ell;i}} - n^{-1} \sum_{j \in s_0 \cap s_1} \frac{y_{\ell;j}}{\pi_{\ell;j}} \right)^2 \tag{37}$$

where $\ell = 0,\, 1,\, 2$. By substituting $\widehat{\text{var}}(\widehat{Y}_0^*)$, $\widehat{\text{var}}(\widehat{Y}_1^*)$ and $\widehat{\text{var}}(\widehat{Y}_2^*)$ into (5), we obtain an estimates for the covariance that can be used to estimate the variance (4).

The accuracy of the variance estimators are measured by the relative bias (RB) and the root mean square error ($RMSE$)

$$RB = \left[E(\widehat{\text{var}}(\widehat{\Delta})) - \text{var}(\widehat{\Delta}) \right] \text{var}(\widehat{\Delta})^{-1} \tag{38}$$

$$RMSE = MSE(\widehat{\text{var}}(\widehat{\Delta}))^{1/2} \tag{39}$$

where $E(\cdot)$, $\text{var}(\cdot)$ and $MSE(\cdot)$ represents the empirical expectation variance and mean square error operator based on 10000 samples.

The results are shown in Table 3. The column σ gives the standard deviations of $y_{0;i}$ considered. The column f gives the different sampling fractions $f=n/N$ considered. The values of Δ/N are given in the second column. The column "RB with (37)" gives the RB of the variance estimator based on (37). We observe a large negative bias. The column "RB of (16)" gives the RB of (16). We see that the bias of (16) is smaller. However, the bias is still slightly negative. The column "Coverage" gives the coverage of the 95% confidence interval using (16) and the quantile of a normal distribution. The coverage is good except for $f=0.05$. This is not surprising as $n=25$ in this case. In addition, the column "Ratio $RMSE$", gives ratios between the $RMSE$ of (16) and the $RMSE$ of the estimator based on (37). A value less than 1 means that (16) is more accurate than estimator based on (37). This is what we observe.

σ	Δ/N	f	RB with (37)	RB of (16)	Ratio $RMSE$	Coverage
0.2	1	0.05	-0.32	-0.12	0.73	0.92
0.2	1	0.10	-0.33	-0.06	0.54	0.94
0.2	1	0.15	-0.42	-0.03	0.37	0.94
0.2	1	0.20	-0.54	-0.02	0.24	0.94
0.2	5	0.05	-0.33	-0.13	0.71	0.92
0.2	5	0.10	-0.34	-0.05	0.52	0.94
0.2	5	0.15	-0.44	-0.03	0.34	0.94
0.2	5	0.20	-0.57	-0.03	0.22	0.94
0.2	10	0.05	-0.31	-0.11	0.71	0.92
0.2	10	0.10	-0.33	-0.05	0.53	0.93
0.2	10	0.15	-0.44	-0.04	0.35	0.94
0.2	10	0.20	-0.57	-0.04	0.23	0.94
0.9	1	0.05	-0.35	-0.12	0.80	0.92
0.9	1	0.10	-0.38	-0.05	0.61	0.94
0.9	1	0.15	-0.51	-0.04	0.41	0.95
0.9	1	0.20	-0.67	-0.03	0.27	0.95
0.9	5	0.05	-0.35	-0.11	0.75	0.93
0.9	5	0.10	-0.41	-0.07	0.54	0.94
0.9	5	0.15	-0.53	-0.03	0.36	0.94
0.9	5	0.20	-0.69	-0.03	0.24	0.95
0.9	10	0.05	-0.35	-0.12	0.70	0.92
0.9	10	0.10	-0.37	-0.05	0.51	0.94
0.9	10	0.15	-0.52	-0.07	0.32	0.94
0.9	10	0.20	-0.64	0.00	0.22	0.94

Table 3: Measure of accuracy (%) for the estimators of variance.

8 Conclusion

The variance estimator proposed generally gives positive estimates. This estimator can be generalised for stratified populations. We have assumed that we have a fixed number of units rotating in and out as well as a fixed number of units in the matched sample. The variance estimator proposed is based on these assumptions which hold in most rotating sampling schemes. First, we estimate the variances and the covariances unconditionally; that is, assuming the sample sizes random. Secondly, using the multivariate normal theory, we derive variances and covariances conditionally on the given sizes (see (13) and (14)) to capture the fixed sizes feature of the sampling scheme.

Acknowledgements

This work was supported by the European research project entitled "Data Quality in Complex Surveys within the New European Information Society" (DAC-

SEIS). The author is grateful to Chris Skinner (University of Southampton, UK) and to the associate editor and the referee for helpful comments.

Appendix A - Proof of Theorem 1

First, let us show that (8) is the maximum entropy sampling design. Lagrange multipliers gives the maximum entropy sampling design (see Hájek, 1959)

$$P_{\max}(s_0) = \delta\{s_0 \in \Omega_0\} \prod_{k \in s_0} \theta_k \tag{40}$$

where θ_k are such that

$$\sum_{k \ni s_0} P_{\max}(s_0) = \pi_{0;k}. \tag{41}$$

It is clear that $\theta_k \geq 0$ otherwise $P_{\max}(s_0)$ is negative. Now (8) implies

$$p_0(s_0) = \delta\{s_0 \in \Omega\} \prod_{i \in U}(1 - p_{0;i}) \prod_{j \in s_0} \frac{p_{0;j}}{1 - p_{0;j}}. \tag{42}$$

We can see that (40) equals (42) when

$$\theta_k = \frac{p_{0;k}}{1 - p_{0;k}} \prod_{i \in U}(1 - p_{0;i}).$$

As the first-order inclusion probabilities of $p_0(s_0)$ are given by the $\pi_{0;k}$, (41) holds when $p_0(s_0) = P_{\max}(s)$. Hájek (1959) shows that $P_{\max}(s)$ is unique. Thus, $p_0(s_0)$ is the maximum entropy sampling design. Moreover, $\theta_k \geq 0$ implies $p_{0;k} < 1$. The $p_{0;k}$ are therefore the first-order inclusion probabilities of a Poisson sampling design. Similar proof yields to the expression for $p_1(s_1|s_0)$. □

Appendix B - Proof of (19)

First, using matrix notation, we have

$$\mathbf{u} = \breve{\mathbf{A}}'_U \ \boldsymbol{\delta}$$

where $\boldsymbol{\delta} = \left(\boldsymbol{\delta}'_0, \boldsymbol{\delta}'_1, \boldsymbol{\delta}'_{01}\right)'$. The vectors $\boldsymbol{\delta}_0$, $\boldsymbol{\delta}_1$, $\boldsymbol{\delta}_{01}$ are respectively the indicator of s_0, s_1 and of the matched sample $s_0 \cap s_1$; that is, $\boldsymbol{\delta}_0 = (\delta_{0;1}, ..., \ \delta_{0;N})'$, $\boldsymbol{\delta}_1 = (\delta_{1;1}, ..., \ \delta_{1;N})'$ and $\boldsymbol{\delta}_{01} = (\delta_{0;1}\delta_{1;1}, ..., \ \delta_{0;N}\delta_{1;N})'$; where the sample indicators $\delta_{0;i}$ and $\delta_{1;i}$ are defined by $\delta_{0;i} = 1$ if $i \in s_0$ and $\delta_{0;i} = 0$ otherwise; $\delta_{1;i} = 1$ if $i \in s_1$ and $\delta_{1;i} = 0$ otherwise.

As the matrix $\boldsymbol{\Sigma}_{\mathbf{u}}$ is the variance-covariance matrix of $\mathbf{u}$ under the Poisson scheme, we have

$$\boldsymbol{\Sigma}_{\mathbf{u}} = \breve{\mathbf{A}}'_U \ \mathrm{var}_P(\boldsymbol{\delta}) \ \breve{\mathbf{A}}_U$$

where $\mathrm{var}_P(\boldsymbol{\delta})$ is the variance of $\boldsymbol{\delta}$ under the Poisson scheme (10) and (11). In the rest of Appendix B, we show that $\mathrm{var}_P(\boldsymbol{\delta}) = \mathbf{C}_U$ defined by (19).

The matrix (20) is the variance-covariance matrix of the vector $\boldsymbol{\delta}_0$ under the Poisson scheme. The extra-diagonal components are null, as the $\delta_{0;i}$ are independent. Indeed, the i-th diagonal component of (20) is the variance

$$\text{var}_P(\delta_{0;i}) = E_P(\delta_{0;i}^2) - E_P(\delta_{0;i})^2 = \pi_{0;i}(1 - \pi_{0;i})$$

which gives (20). An analogous proof gives (21).

The matrix (22) is the variance-covariance matrix of $\boldsymbol{\delta}_{01}$. The extra-diagonal components are null, as the $\delta_{0;i}\delta_{1;i}$ are independent. The i-th diagonal component of (22) is

$$\text{var}_P(\delta_{0;i}\delta_{1;i}) = E_P\left((\delta_{0;i}\delta_{1;i})^2\right) - E_P(\delta_{0;i}\delta_{1;i})^2 = g\pi_{0;i}(1 - g\pi_{0;i}).$$

The matrix (23) is the covariance matrix between $\boldsymbol{\delta}_0$ and $\boldsymbol{\delta}_1$. This matrix is diagonal, as $\delta_{0;i}$ is independent of $\delta_{1;j}$ if $i \neq j$. The i-th diagonal component of (23) is

$$\text{cov}_P(\delta_{0;i}, \delta_{1;i}) = E_P\left(\delta_{0;i}\delta_{1;i}\right) - E_P(\delta_{0;i})E_P(\delta_{1;i}) = \pi_{0;i}(g - \pi_{1;i}).$$

The matrix (24) is the covariance matrix between $\boldsymbol{\delta}_0$ and $\boldsymbol{\delta}_{01}$. This matrix is diagonal, as $\delta_{0;i}$ is independent of $\delta_{0;j}\delta_{1;j}$ if $i \neq j$. The diagonal components of (24) are

$$\text{cov}_P(\delta_{0;i}, \delta_{0;i}\delta_{1;i}) = E_P\left(\delta_{0;i}\delta_{0;i}\delta_{1;i}\right) - E_P(\delta_{0;i})E_P(\delta_{0;i}\delta_{1;i}) = g\pi_{0;i}(1 - \pi_{0;i}).$$

An analogous proof gives (25). Finally, $\text{var}_P(\boldsymbol{\delta}) = \mathbf{C}_U$. This complete the proof. □

Appendix C - Proof of (26)

With (18) and (19), we have

$$\boldsymbol{\Sigma}_{\mathbf{u}} = \begin{bmatrix} \breve{\mathbf{y}}_0'\mathbf{C}_{0;0}\breve{\mathbf{y}}_0 & \breve{\mathbf{y}}_0'\mathbf{C}_{0;1}\breve{\mathbf{y}}_1 & \breve{\mathbf{y}}_0'\mathbf{C}_{0;0}\breve{\mathbf{z}}_0 & \breve{\mathbf{y}}_0'\mathbf{C}_{0;1}\breve{\mathbf{z}}_0 & \breve{\mathbf{y}}_0'\mathbf{C}_{0;01}\breve{\mathbf{z}}_0 \\ \breve{\mathbf{y}}_0'\mathbf{C}_{0;1}\breve{\mathbf{y}}_1 & \breve{\mathbf{y}}_1'\mathbf{C}_{1;1}\breve{\mathbf{y}}_1 & \breve{\mathbf{y}}_0'\mathbf{C}_{0;1}\breve{\mathbf{z}}_0 & \breve{\mathbf{y}}_0'\mathbf{C}_{1;1}\breve{\mathbf{z}}_0 & \breve{\mathbf{y}}_0'\mathbf{C}_{1;01}\breve{\mathbf{z}}_0 \\ \breve{\mathbf{y}}_0'\mathbf{C}_{0;0}\breve{\mathbf{z}}_0 & \breve{\mathbf{y}}_0'\mathbf{C}_{0;1}\breve{\mathbf{z}}_0 & \breve{\mathbf{z}}_0'\mathbf{C}_{0;0}\breve{\mathbf{z}}_0 & \breve{\mathbf{z}}_0'\mathbf{C}_{0;1}\breve{\mathbf{z}}_0 & \breve{\mathbf{z}}_0'\mathbf{C}_{0;01}\breve{\mathbf{z}}_0 \\ \breve{\mathbf{y}}_0'\mathbf{C}_{0;1}\breve{\mathbf{z}}_0 & \breve{\mathbf{y}}_0'\mathbf{C}_{1;1}\breve{\mathbf{z}}_0 & \breve{\mathbf{z}}_0'\mathbf{C}_{0;1}\breve{\mathbf{z}}_0 & \breve{\mathbf{z}}_0'\mathbf{C}_{1;1}\breve{\mathbf{z}}_0 & \breve{\mathbf{z}}_0'\mathbf{C}_{1;01}\breve{\mathbf{z}}_0 \\ \breve{\mathbf{y}}_0'\mathbf{C}_{0;01}\breve{\mathbf{z}}_0 & \breve{\mathbf{y}}_0'\mathbf{C}_{1;01}\breve{\mathbf{z}}_0 & \breve{\mathbf{z}}_0'\mathbf{C}_{0;01}\breve{\mathbf{z}}_0 & \breve{\mathbf{z}}_0'\mathbf{C}_{1;01}\breve{\mathbf{z}}_0 & \breve{\mathbf{z}}_0'\mathbf{C}_{01;01}\breve{\mathbf{z}}_0 \end{bmatrix}. \quad (43)$$

Consider $\breve{\mathbf{v}} = (\breve{v}_1, ..., \breve{v}_N)'$ and $\breve{\mathbf{x}} = (\breve{x}_1, ..., \breve{x}_N)'$ be any $N \times 1$ vectors and $\mathbf{C} = diag\{c_i : i \in U\}$ an $N \times N$ diagonal matrix. It can be easily shown that

$$\breve{\mathbf{v}}' \ \mathbf{C} \ \breve{\mathbf{x}} = \sum_{i \in U} c_i \breve{v}_i \breve{x}_i.$$

Thus, all components of (43) are population totals that can be estimated by their corresponding unbiased HT estimators.

The component $(1,1)$ of $\mathbf{\Sigma_u}$ and $\widehat{\mathbf{\Sigma}}_\mathbf{u}$ are respectivelly

$$\begin{aligned}
\Sigma_{11} &= \breve{\mathbf{y}}_0' \mathbf{C}_{0;0} \breve{\mathbf{y}}_0 = \sum_{i \in U} \pi_{0;i}(1-\pi_{0;i})\ \breve{y}_{0;i}\ \breve{y}_{1;i} \\
\widehat{\Sigma}_{11} &= \sum_{i \in s_0} \delta_{0;i}\ \pi_{0;i}(1-\pi_{0;i})\pi_{0;i}^{-1}\ \breve{y}_{0;i}\ \breve{y}_{1;i}.
\end{aligned}$$

It is then clear that $\widehat{\Sigma}_{11}$ is an unbiased estimator of Σ_{11}, as $E(\delta_{0;i}) = \pi_{0;i}$. Analogously we can show that the component $(2,2)$, $(3,3)$, $(4,4)$ of $\widehat{\mathbf{\Sigma}}_\mathbf{u}$ are unbiased. The remaining components (k,ℓ) of $\mathbf{\Sigma_u}$ and $\widehat{\mathbf{\Sigma}}_\mathbf{u}$ can be written as

$$\begin{aligned}
\Sigma_{k\ell} &= \sum_{i \in U} c_i\ \breve{v}_i\ \breve{x}_i \\
\widehat{\Sigma}_{k\ell} &= \sum_{i \in U} \delta_{0;i}\ \delta_{1;i}\ c_i\ (g\pi_{0;i})^{-1}\ \breve{v}_i\ \breve{x}_i.
\end{aligned}$$

The estimator $\widehat{\Sigma}_{k\ell}$ is unbiased as $E(\delta_{0;i}\delta_{1;i}) = g\pi_{0;i}$. □

Appendix D - Proof of Theorem 2

If $\breve{\mathbf{C}}_U$ is non-negative definite, it implies the Theorem. Indeed $\breve{\mathbf{C}}_s$ is a principal sub-matrix of $\breve{\mathbf{C}}_U$ and by Corollary 14.2.12 (Harville, 1997, pp. 214), $\breve{\mathbf{C}}_s$ is non-negative definite. Thus by Theorem 14.2.9 (Harville, 1997, pp. 213), $\widehat{\mathbf{\Sigma}}_\mathbf{u} = \breve{\mathbf{A}}_s'$ $\breve{\mathbf{C}}_s$ $\breve{\mathbf{A}}_s$ is non-negative definite. The Schur complement $\widehat{\mathbf{\Sigma}}_{\mathbf{yy}|\mathbf{n}}$ of $\widehat{\mathbf{\Sigma}}_\mathbf{u}$ is therefore non-negative definite by Theorem 14.8.4 (Harville, 1997, pp. 242).

$\breve{\mathbf{C}}_U$ is positive definite if all its eigen-values are positive (Harville, 1997, pp. 543). This is the case if for all $\lambda < 0$,

$$\det(\breve{\mathbf{C}}_U - \lambda \mathbf{I}_{3N}) > 0$$

where $\mathbf{I}_{3N}$ is the $(3N \times 3N)$ identity matrix. As $\breve{\mathbf{C}}_U$ is a partitioned matrix (see (27)), we have

$$\begin{aligned}
\det(\breve{\mathbf{C}}_U - \lambda \mathbf{I}_{3N}) &= \det(\mathbf{A}) \det(\breve{\mathbf{C}}_{1;1} - \lambda\ \mathbf{I}_N - \breve{\mathbf{C}}_{1;1}\mathbf{A}\breve{\mathbf{C}}_{1;1}) \\
&\quad \det(\breve{\mathbf{C}}_{0;0} - \lambda\ \mathbf{I}_N - \breve{\mathbf{C}}_{0;0}\mathbf{A}\breve{\mathbf{C}}_{0;0} - (\breve{\mathbf{C}}_{0;1} - \breve{\mathbf{C}}_{0;0}\mathbf{A}\breve{\mathbf{C}}_{1;1}) \\
&\quad (\breve{\mathbf{C}}_{1;1} - \lambda\ \mathbf{I}_N - \breve{\mathbf{C}}_{1;1}\mathbf{A}\breve{\mathbf{C}}_{1;1})^{-1}(\breve{\mathbf{C}}_{0;1} - \breve{\mathbf{C}}_{1;1}\mathbf{A}\breve{\mathbf{C}}_{0;0})) \\
&= \prod_{i \in U} F_i(\lambda)
\end{aligned}$$

where $\mathbf{A} = \breve{\mathbf{C}}_{01;01} - \lambda\ \mathbf{I}_N$, $\mathbf{I}_N$ is the $(N \times N)$ identity matrix and

$$\begin{aligned}
F_i(\lambda) &= (1-\pi_{0;i}-\lambda)(1-\pi_{1;i}-\lambda)(1-\pi_{0;i}g-\lambda) \\
&\quad +2(1-\pi_{1;i}g^{-1})(1-\pi_{0;i})(1-\pi_{1;i}) \\
&\quad -(1-\pi_{0;i}-\lambda)(1-\pi_{1;i})^2 \\
&\quad -(1-\pi_{0;i})^2(1-\pi_{1;i}-\lambda) \\
&\quad -(1-\pi_{1;i}g^{-1})^2(1-\pi_{0;i}g-\lambda)
\end{aligned}$$

The proof is complete if we can show that $F_i(\lambda) > 0$ for all $\lambda < 0$. The second derivative of $F_i(\lambda)$ is

$$\frac{\partial^2 F_i(\lambda)}{\partial^2 \lambda} = 6(1-\lambda) - 2(\pi_{0;i} g + \pi_{1;i} + \pi_{0;i}) \qquad (44)$$

is clearly always positive when $\lambda < 0$. Moreover, (44) increases as λ decreases. This imply that $\partial F_i(\lambda)/\partial\lambda$ decreases as λ decreases. Now, the first derivative of $F_i(\lambda)$ at $\lambda = 0$ is negative, as

$$\left.\frac{\partial F_i(\lambda)}{\partial \lambda}\right|_{\lambda=0} = -\eta_i < 0.$$

Thus the first derivative of $F_i(\lambda)$ is negative for $\lambda < 0$. Moreover $F_i(0) = \tau_i > 0$. Thus $F_i(\lambda) > 0$ for all $\lambda < 0$. This completes the proof. □

References

Andersson, C. and Nordberg, L. (1994). A method for variance estimation of non-linear function of totals in surveys - Theory and a software implementation. *Journal of Official Statistics* **10**, 395-405.

Berger, Y.G. (1998) Rate of Convergence for Asymptotic Variance for the Horvitz-Thompson Estimator. *J. Statist. Plann. Inference* **74**, 149-168.

Chen, X.H., Dempster, A.P. & Liu, S.L. (1994). Weighted Finite Population Sampling to Maximise Entropy. *Biometrika* **81**, 457-469.

Brewer, K.W.R, Early, L.J. and Joyce, S.F. (1972) Selecting Several Samples from a Single Population. *Australian Journal of Statistics,* **14**, 231-239.

Fuller, W.A. and Rao, J.N.K. (2001) A Regression Composite Estimator with Application to the Canadian Labour Force Survey. *Survey Methodology* **27**, 45-52.

Gabler, S. (1984) On Unequal Probability Sampling: Sufficient Conditions for the Superiority of Sampling Without Replacement. *Biometrika* **71**, 171-175.

Hájek, J. (1964). Asymptotic Theory of Rejective Sampling With Varying Probabilities from a Finite Population. *Ann. Math. Statist.* **35**, 1491-1523.

Hájek, J. (1981). *Sampling from a Finite Population*, New York, Marcel Dekker.

Harville, D.A. (1997). *Matrix Algebra From a Statistician's Perspective*, New York, Springer.

HOLMES, D.J. & SKINNER, C.J. (2000). Variance Estimation for Labour Force Survey Estimates of Level and Change. *Government Statistical Service Methodology Series.* **21**.

HORVITZ, D.G. & THOMPSON, D.J. (1952). A Generalization of Sampling Without Replacement From a Finite Universe. *J. Amer. Statist. Assoc.* **47**, 663-685.

HIDIROGLOU, M.A., SÄRNDAL, C.E. AND BINDER, D.A. (1995). Weighting and Estimation in Business Surveys. In *Business Survey Methods,* Wiley.

LAWLEY, D.N. (1943). A Note on Karl Pearson's Selection Formula. In *Proceedings of the Royal Society of Edinburgh* **A62**, 28-30.

NARAIN, R.D. (1951). On Sampling without Replacement with Varying Probabilities. *Journal of the Indian Society of Agricultural Statistics* **3**, 169–174.

NORDBERG, L. (2000). On Variance Estimation for Measure of Change when Samples are Coordinated by the Use of Permanent Random Numbers. *Journal of Official Statistics* **16**, 363-378.

OHLSON, L. (1990). Sequential Sampling from Business Register and Its Application to the Swedish Consumer Price Index. R&D Report 1990:6, Stockholm: Statistics Sweden.

SÄRNDAL, C.E., SWENSSON, 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*, J. Roy. Statist. Soc. Ser. D, **52**, 1-30.