Use of functionals in linearization and composite estimation with application to two-sample survey data

C. Goga[*] J.-C. Deville [†] A. Ruiz-Gazen [‡]

June 22, 2009

Abstract

An important problem associated with two-sample surveys is estimation of nonlinear functions of finite population totals such as ratios, correlation coefficients or measures of income inequality. Computation and estimation of the variance of such complex statistics are made more difficult by the existence of overlapping units. In one-sample surveys, the linearization method based on the influence function approach is a powerful tool for variance estimation. We introduce a two-sample linearization technique which can be viewed as a generalization of the one-sample influence function approach. Our technique is based on expressing the parameters of interest as multivariate functionals of finite and discrete measures and then using partial influence functions to compute the linearized variables. Under broad assumptions, the asymptotic variance of the substitution estimator, derived from [8], is shown to be the variance of a weighted sum of the linearized variables. The paper then focuses on a general class of composite substitution estimators, and from this class the optimal estimator for minimizing the asymptotic variance is obtained. Finally, the efficiency of the optimal composite estimator is demonstrated through an empirical study.

Keywords: Gini index change; Partial influence function; Substitution estimator; Two-dimensional sampling design; Variance estimation; Variance optimization.

1 Introduction

The study and the comparison across time or space of income distribution and income inequality measures are of increasing current interest. Most of the properties of measures such as the Lorenz curve or the Gini index have been investigated. However, the variance estimation problem for sample survey data has only recently been addressed. Difficulties arise because these measures are nonlinear functions of population values.

There exist two approaches to variance estimation for complex statistics: resampling methods and linearization methods. Various resampling methods [24] exist such as the jackknife, the balanced repeated replication method and the bootstrap. The jackknife [3] is the most often used procedure and consists of computing the estimator repeatedly leaving out one unit. These methods can be very computing intensive. Besides and unlike linearization methods, resampling methods can only be applied to specific sampling designs. For unequal probability sampling designs, they may run into great difficulties [30].

In the following, the focus is on linearization methods. The well-known Taylor linearization method can be used for nonlinear but continuously differentiable functions of totals, but the method is not adapted for the estimation of quantiles, for example. For nonregular functions of totals, [18] propose an approach based on the estimating equations technique. A functional approach is also proposed in [8]. It uses the

[*] IMB, Université de Bourgogne, 9 Avenue Alain Savary, 21078 Dijon, France, camelia.goga@u-bourgogne.fr

[†] Laboratoire de Statistique d'Enquête, ENSAI/CREST, rue Blaise Pascal, Campus de Ker Lann 35170 Bruz, France, deville@ensai.fr

[‡] Toulouse School of Economics, Université Toulouse 1, 21 allée de Brienne, 31000 Toulouse, France, ruiz@cict.fr

influence function concept and provides a theoretical justification for the linearization proposal of [7] that gives practical rules for linearising complex statistics. Non-differentiable functions of totals like quantiles or the Gini index can be handled either by the influence function approach or by the estimating equation technique. More complex parameters such as eigenelements of functional data have been considered recently by the influence function approach, in an unpublished University of Burgundy technical report by H. Cardot, M. Chaouch, C. Goga and C. Labruère. All the linearization methods consist of computing the 'linearized variable' u_k associated with the parameters of interest for all the units k from the population U of size N and give a first-order expansion formula of the complex statistics which contains the Horvitz-Thompson estimator $\sum_{k\in s} u_k/\pi_k$ for the total of u_k. Here, $\pi_k = \mathrm{pr}(k \in s)$ is the first-order inclusion probability of k in the sample s. We consider the influence function approach, introduced in robust statistics by [13]. [5] uses the influence function for estimating the variance of complex statistics and compares it with a jackknife variance estimator. [8] uses a slightly modified definition of the influence function and provides a powerful variance-estimation tool for complex survey statistics. He gives computing rules and applies the technique to different examples such as quantiles, concentration indices and estimators of eigenvalues in principal component analysis in the one-sample case.

In Deville's approach, a population parameter of interest Φ can be written as a functional T with respect to a finite and discrete measure M, namely $\Phi = T(M)$. The substitution estimator $\hat{\Phi} = T(\hat{M})$ is the functional T of a random measure $\hat{M}$ that is associated with sampling weights w_k, $k \in U$, and is 'close' to M. Suppose that T is homogeneous of degree α, so that $T(rM) = r^{\alpha}T(M)$, and $\lim_{N\to\infty} N^{-\alpha}T(M) < \infty$. Under broad assumptions, Deville shows that

$$\begin{aligned}\sqrt{n}N^{-\alpha}\{T(\hat{M}) - T(M)\} &= \sqrt{n}N^{-\alpha}\int I_T(M,z)d(\hat{M}-M)(z) + o_p(1) \\ &= \sqrt{n}N^{-\alpha}\sum_{k=1}^{N} u_k(w_k - 1) + o_p(1). \qquad (1)\end{aligned}$$

The linearized variables u_k are the influence functions $I_T(M, z_k)$, where z_k is the value of the variable of interest for the kth unit and

$$I_T(M,z) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\{T(M+\varepsilon\delta_z) - T(M)\},$$

where δ_z is the unit mass at point $z \in R^p$. This definition is slightly different from the one used in robust statistics [13] which is based on a probability distribution instead of a finite measure M. A nonstandardised measure M is used in survey sampling because the total mass may be an unknown quantity. The influence function is a Gâteaux differential for $T(M)$ in the direction of the Dirac mass at z. As a consequence of (1) and under broad assumptions, the asymptotic variance of $T(\hat{M})$ is the variance of $\sum_{k=1}^{N} u_k(w_k - 1)$. For the Horvitz-Thompson weights $w_k = 1/\pi_k$, this variance is equal to

$$\sum_{k=1}^{N}\sum_{l=1}^{N}(\pi_{kl} - \pi_k\pi_l)\frac{u_k}{\pi_k}\frac{u_l}{\pi_l},$$

where the π_{kl} are the second-order probabilities. Deville estimates the variance $\mathrm{var}\{T(\hat{M})\}$ by the Horvitz-Thompson variance estimator

$$\sum_{k\in s}\sum_{l\in s}\frac{\pi_{kl} - \pi_k\pi_l}{\pi_{kl}}\frac{\hat{u}_k}{\pi_k}\frac{\hat{u}_l}{\pi_l} \qquad (2)$$

using the sample estimators $\hat{u}_k = I_T(\hat{M}, z_k)$ for the linearized variables u_k, $k \in s$. The main advantage of this method is that the variance estimators can be implemented in any survey software capable of calculating the Horvitz-Thompson variance estimator.

All previous methods concern variance estimation for one-sample survey data, but interest may lie in studying how statistics change over time or between different population subgroups. Estimating the change in the Gini index between two periods of time is one particular example. Difficulties arise from the existence of overlapping samples. Work concerning temporal change mainly deals with the estimation of simple statistics such as the population mean or total under the hypothesis of independence of the selection procedure. The first studies are by [16], [21] and [9]. Cochran (1977, §12.11), gives the most important ideas concerning repeated sampling and a more thorough discussion is found in [17]. All these studies are conducted for simple random sampling without replacement. More general sampling designs are considered in [26], [14] and [19] but they still assume the independence of successive samples. Recent works are dedicated to composite estimators with applications to specific types of survey [1, 12, 27]. We also mention the review in an unpublished Institut National de la Statistique et des Etudes Economiques (INSEE) working paper by N. Caron and P. Ravalet, the paper by [6] and the recent work by [31] and [2].

We propose an extension of the influence function approach to the two-sample case. In classical statistics, the partial influence function is introduced for estimators based on more than one sample [22] following the analogy with derivatives and partial derivatives. In the survey-sampling context, we also propose to extend the influence function approach to the multiple-sample case by considering partial influence functions. In the two-sample case, estimators are based on three disjoint samples which naturally lead us to consider three-variate functionals and their associated partial influence functions. These partial influence functions equal the linearized variables and, under broad assumptions, the asymptotic variance of the complex statistics is equal to the variance of a weighted linear sum of the linearized variables. The proposed methodology has already been applied to compute the precision of change estimators in the French employment survey [23].

2 Extension of the asymptotic results to two dimensions

2.1 Partial influence functions

Consider the finite population U of size N. Let $\mathcal{Z}_1$ and $\mathcal{Z}_2$ be two variables of interest measured on two different samples s_1 and s_2 selected from the same population U according to the sampling designs p_1 and p_2. The objective is to estimate a nonlinear function Φ of totals of $\mathcal{Z}_1$ and $\mathcal{Z}_2$. The sample s_1, respectively s_2, is of size n_1, respectively n_2. We consider that the matched sample $s_3 = s_1 \cap s_2$ is nonempty and of size n_3. Let $s_{1*} = s_1 - s_2$, respectively $s_{2*} = s_2 - s_1$, be the complementary sample of s_2 in s_1, respectively of s_1 in s_2, of size n_{1*}, respectively n_{2*}, and let $n = n_{1*} + n_3 + n_{2*}$. Let $\mathcal{D} = \{1*, 3, 2*\}$ be the set of the disjoint samples' indices and let $\mathcal{T} = \{1, 2, 3\}$ be the set of the matched samples' indices. Apart from particular cases, we assume from now on that $d \in \mathcal{D}$ and $t \in \mathcal{T}$. On the matched sample s_3, we know both $\mathcal{Z}_1$ and $\mathcal{Z}_2$ and we denote $(\mathcal{Z}_1, \mathcal{Z}_2)$, by $\mathcal{Z}_3$.

Each unit $k \in U$ is associated with a vector $z_{k,t} \in R^{p_t}$, $t \in \mathcal{T}$, where $z_{k,t} = \mathcal{Z}_t(k)$ is the value of the p_t-dimensional variable of interest $\mathcal{Z}_t$ for the kth unit and $p_3 = p_1 + p_2$. We consider the discrete and finite measures $M_t = \sum_{k=1}^{N} \delta_{z_{k,t}}$ defined on R^{p_t} to R taking the mass 1 for each $z_{k,t}$ with $k \in U$ and zero elsewhere. The measures M_t are of total mass equal to N, the population size, and take into account the units k in U together with the variable of interest $\mathcal{Z}_t$. Henceforth, defining an estimator $\hat{M}_t$ of M_t leads to definition of an estimator of the total of $\mathcal{Z}_t$ since the total of $\mathcal{Z}_t$ equals $\int \mathcal{Z}_t dM_t$ and is a functional of M_t.

Consideration of three different measures is justified because the variables $\mathcal{Z}_t$ are measured on different samples s_t, $t = 1, 2, 3$, and the measures M_t may be estimated in different ways. In particular, M_3 is useful if one wishes to estimate covariance terms of the form $\sum_{k=1}^{N} z_{k,1} z_{k,2}$ that cannot be expressed directly from M_1 and M_2. Therefore, and by analogy with the one-sample situation, we introduce the three-variate functional $T(M_1, M_2, M_3) = T(M)$ with the vector $M = (M_t)_{t \in \mathcal{T}}$ and consider as parameters of interest any population total function $\Phi = T(M)$. Let us consider three illustrative examples.

Ex 2.1 *Let $\mathcal{Z}_1$ and $\mathcal{Z}_2$ be the same variable of interest but measured on two occasions with totals $Z_t = \sum_{k=1}^{N} z_{k,t}$. The finite population total change $\Phi = Z_2 - Z_1$ can be written as $T(M) = \int \mathcal{Z}_2 dM_2 - \int \mathcal{Z}_1 dM_1$.*

Ex 2.2 *Consider two bivariate variables $\mathcal{Z}_t = (\mathcal{X}_t, \mathcal{Y}_t)$ for $t = 1, 2$ that may also correspond to two occasions. The functional*

$$T(M) = \Delta R = R_2 - R_1 = \frac{\int \mathcal{Y}_2 dM_2}{\int \mathcal{X}_2 dM_2} - \frac{\int \mathcal{Y}_1 dM_1}{\int \mathcal{X}_1 dM_1}$$

is the ratio change. Change of more complex statistics such as the Gini index or the Lorenz curve can also be considered.

Ex 2.3 *Consider the product of two variables $\mathcal{Z}_1$ and $\mathcal{Z}_2$, with $T(M) = \int \mathcal{Z}_1 \mathcal{Z}_2 dM_3 / \int dM_3$. This example illustrates the need to introduce M_3.*

We now introduce the partial influence functions of the functional $T(M)$ [25, 22].

Definition 2.1 *The first partial influence function $I_{1T}(M; z)$ of $T(M)$ is defined as the first partial Gâteaux derivative of T with respect to M_1 in the direction of Dirac mass at z,*

$$I_{1T}(M; z) = \lim_{\varepsilon \to 0} \frac{T(M_1 + \varepsilon \delta_z, M_2, M_3) - T(M_1, M_2, M_3)}{\varepsilon} \tag{3}$$

when this limits exists. The second, respectively third, partial influence function $I_{2T}(M; z)$, respectively $I_{3T}(M; z)$, is defined in a similar way.

Definition 2.2 *The linearized variables $u_{k,t}$ for $k \in U$ and $t \in \mathcal{T}$ are obtained by computing $I_{tT}(M; z)$ at $z = z_{k,t} \in R^{p_t}$, namely $u_{k,t} = I_{tT}(M; z_{k,t})$.*

The partial influence functions of $T = \Delta R = R_2 - R_1 = R(M_2) - R(M_1)$, see Example 2, are computed as partial derivatives of a function. Since $R_2 = R(M_2)$, respectively $R_1 = R(M_1)$, is constant with respect to M_1, respectively M_2, the first, respectively second, partial influence function consists of taking the linearized variable of the ratio R_1, respectively R_2. To be more precise, we have

$$\begin{aligned} u_{k,1} &= I_{1T}\{M; (x_{k,1}, y_{k,1})\} = -\frac{1}{X_1}(y_{k,1} - R_1 x_{k,1}), \\ u_{k,2} &= I_{2T}\{M; (x_{k,2}, y_{k,2})\} = \frac{1}{X_2}(y_{k,2} - R_2 x_{k,2}), \\ u_{k,3} &= 0. \end{aligned} \tag{4}$$

For Example 3, $u_{k,1} = u_{k,2} = 0$ and $u_{k,3} = (1/N)(z_{k,1} z_{k,2} - \sum_{k=1}^{N} z_{k,1} z_{k,2}/N)$. The $u_{k,t}$ depend on unknown quantities and cannot be calculated.

2.2 The substitution estimator and its asymptotic variance

By analogy with [8], we define $\hat{M}_t = \sum_{k=1}^{N} v_{k,t} \delta_{z_{k,t}}$ as an estimator of M_t which associates a weight $v_{k,t}$ with each vector $z_{k,t}$, for $k \in s_t$, and zero elsewhere. The weights $v_{k,t}$ will be derived in the next section.

Definition 2.3 *The substitution estimator of $T(M)$ is $T(\hat{M})$ where $\hat{M} = (\hat{M}_1, \hat{M}_2, \hat{M}_3)$.*

The estimator $\hat{M}$ defines the estimator $T(\hat{M})$. In §3.4, we give three different estimators of M which lead to three different estimators of the ratio change.

In the following, we give sufficient conditions for the asymptotic expansion of T to be valid. We need both the population and the samples sizes N and n_t to go to infinity with $n_t < N$. As in the one-sample case [15], we consider a sequence of populations and associated sequences of samples s_t of increasing sizes with

$\int \mathcal{Z}_t d\hat{M}_t$ as an estimator of $\int \mathcal{Z} dM_t$. By analogy with [8] we make the following assumptions, for $t \in \mathcal{T}$.

Assumption 1. We assume that $\lim_{N\to\infty} n_t^{-1} n_3 \in (0,1)$ and $\lim_{N\to\infty} N^{-1} n_t \in (0,1)$.

Assumption 2. We assume that $\lim_{N\to\infty} N^{-1} \int \mathcal{Z}_t dM_t$ exists.

Assumption 3. As $N \to \infty$, $N^{-1}(\int \mathcal{Z}_t d\hat{M}_t - \int \mathcal{Z}_t dM_t) \to 0$ in probability.

Assumption 4. As $N \to \infty$, $\{{n_t}^{1/2} N^{-1}(\int \mathcal{Z}_t d\hat{M}_t - \int \mathcal{Z}_t dM_t)\}_{t=1}^3 \to N(0,\Sigma)$ in distribution.

Let the functional T also satisfy the following smoothness assumptions.

Assumption 5. We assume that T is homogeneous, in that there exists a real number $\beta > 0$ dependent on T such that $T(rM) = r^\beta T(M)$ for any real $r > 0$.

Assumption 6. We assume that $\lim_{N\to\infty} N^{-\beta} T(M) < \infty$.

Assumption 7. We assume that T is Fréchet differentiable.

Theorem 1 is the most important result of the paper; it gives the first-order [29] expansion of the functional T at $\hat{M}/N$ and around M/N.

Theorem 1 *Let Assumptions* 1 *to* 7 *hold. Then*

$$\begin{aligned}\frac{\sqrt{n}}{N^\beta}\{T(\hat{M}) - T(M)\} &= \frac{\sqrt{n}}{N^\beta}\sum_{t=1}^{3}\int I_t T(M; z) d(\hat{M}_t - M_t)(z) + o_p(1) \\ &= \frac{\sqrt{n}}{N^\beta}\sum_{t=1}^{3}\{\sum_{k=1}^{N} u_{k,t}(v_{k,t} - 1)\} + o_p(1)\end{aligned}$$

and the asymptotic variance of $T(\hat{M})$ *is equal to the variance of* $\displaystyle\sum_{t=1}^{3}\{\sum_{k=1}^{N} u_{k,t}(v_{k,t} - 1)\}$.

The proof is given in the Appendix. The strong assumption of Fréchet differentiability for T ensures that the remainder of the first-order von Mises expansion is negligible. Moreover, when they exist, the Fréchet partial derivatives equal the Gâteaux partial derivatives, which are the partial influence functions. However, the result can be obtained if T is only Gâteaux or compact differentiable [10] but with some additional assumptions [22]. For particular functionals T, one may study the remainder term directly and prove that it is of order $o_p(n^{-1/2})$; see the unpublished report of H. Cardot and others for the one-sample case.

3 A general class of composite estimators

3.1 Preamble

In this section, we derive the weights $v_{k,t}$ defining the measures $\hat{M}_t$. The $v_{k,t}$ are expected to satisfy the unbiasedness conditions $E(\hat{M}_t) = M_t$, so that

$$E\{\sum_{t=1}^{3}\sum_{k=1}^{N} u_{k,t}(v_{k,t} - 1)\} = 0. \tag{5}$$

The variables of interest are known on different samples. Consequently, we propose unbiased composite estimators of M_t that combine information from s_1 and s_2 considering the interaction between them through the matched sample s_3. First we introduce the two-dimensional sampling design described in an unpublished INSEE working paper of F. Cotton and C. Hesse, and its corresponding inclusion probabilities. Next, we determine the weights $v_{k,t}$ which satisfy the unbiasedness conditions through a kind of two-sample Horvitz-Thompson estimation method.

3.2 Two-dimensional sampling design

Definition 3.1 *A two-dimensional sampling design is a probability measure $p\{s = (s_1, s_2)\}$ of selecting a two-sample $s = (s_1, s_2) \in \{\mathcal{P}(U)\}^2$. We have $p(s) \geq 0$ and $\sum_{s \in \{\mathcal{P}(U)\}^2} p(s) = 1$.*

As described in detail in C. Goga's unpublished 2003 Ph. D. thesis from the University of Rennes 2, marginal sampling designs and the distribution of any algebraic combination of s_1 and s_2 can be deduced from $p(s)$. Each unit $k \in U$ may belong to one of the disjointed samples s_d for $d \in \mathcal{D} = \{1*, 3, 2*\}$ or in the complementary set of $s_1 \cup s_2$. The sample membership dummy variables $I_k^d = 1_{\{k \in s_d\}}$ form a basis $\mathcal{B}$ in the algebra spanned by I_k^1 and I_k^2 and the following definition gives the inclusion probabilities with respect to $\mathcal{B}$.

Definition 3.2 *Let $p(s)$ be a two-dimensional sampling design. For all $k, l \in U$ and $d, d' \in \mathcal{D}$, we define the first- and second-order two-dimensional inclusion probabilities computed with respect to $\mathcal{B}$ as*

$$\pi_k^d = \mathrm{pr}(k \in s_d) = E(I_k^d), \quad \pi_{kl}^{d,d'} = \mathrm{pr}(k \in s_d \,\&\, l \in s_{d'}) = E(I_k^d I_l^{d'}),$$

where the expectation is considered with respect to $p(s)$.

There are therefore three, respectively six, sets of first-order, respectively second-order two-dimensional inclusion probabilities. We mention now some of the properties of $\pi_{kl}^{d,d'}$. First of all, for $d \neq d'$, the commutative property with respect to two units k and l no longer holds as in the one-sample selection case. Thus, $\pi_{kl}^{d,d'} \neq \pi_{lk}^{d,d'}$ but we have $\pi_{kl}^{d,d'} = \pi_{lk}^{d',d}$. When $d = d'$, $\pi_{kl}^{d,d} = \pi_{kl}^d$, the usual one-sample second-order inclusion probabilities, and there are six different sets of $\pi_{kl}^{d,d'}$. Finally, for $k = l$ and $d \neq d'$, we have $\pi_{kl}^{d,d'} = 0$.

Differently from the one-sample case, the algebra spanned by I_k^1 and I_k^2 contains 7 elements and we have 29 ways of choosing a basis with its corresponding inclusion probabilities; see C. Goga's thesis for more details. Note that changing from one basis to another is possible by linear transformations. By analogy with the one-sample case, let us define the size of a two-dimensional sample.

Definition 3.3 *The size of a two-dimensional sample $s = (s_1, s_2)$ is defined by $n_s = (n_{1*}, n_3, n_{2*})$ with $n_d = \sum_{k=1}^N I_k^d$ the size of s_d, for $d \in \mathcal{D}$.*

The size n_s may be random if at least one of the three components is random and fixed if all the components are fixed. In §4, we define the two-dimensional simple random sampling without replacement, which is a fixed-size design whereas the Bernoulli or the Poisson two-dimensional sampling designs in Goga's thesis are random size designs.

3.3 General composite estimation

The construction of the measure $\hat{M}_3$ depends only on the matched sample s_3 and so, by using the unbiasedness condition $E(\hat{M}_3) = M_3$, we have

$$v_{k,3} = I_k^3 / \pi_k^3, \quad \hat{M}_3 = \sum_{k=1}^N \frac{I_k^3 \delta_{z_{k,3}}}{\pi_k^3}. \tag{6}$$

Since the disjoint samples s_{1*}, s_3 and s_{2*} can be composed in different ways, there are several ways of defining the estimators $\hat{M}_t$, for $t = 1, 2$, which entail different substitution estimators $T(\hat{M})$. A general class of composite estimators is proposed if we define the weights $v_{k,t}$, $t = 1, 2$, as linear combinations of the basis elements I_k^{1*}, I_k^3 and I_k^{2*}. To be more precise, since $v_{k,t}$ is zero outside the sample s_t, $t = 1, 2$, we take $v_{k,1}$, respectively $v_{k,2}$, as a linear combination of I_k^{1*} and I_k^3, respectively of I_k^{2*} and I_k^3, as follows:

$$v_{k,1} = v_{k,1}^{1*} I_k^{1*} + v_{k,1}^3 I_k^3, \quad v_{k,2} = v_{k,2}^{2*} I_k^{2*} + v_{k,2}^3 I_k^3,$$

for some real numbers $v_{k,t}^d$, where $d \in \{1*, 3\}$ for $t = 1$ and $d \in \{2*, 3\}$ for $t = 2$. We propose to use the following weight sets where the π_k^d, for $d \in \mathcal{D}$, are given in Definition 3.2, and which satisfy the following unbiasedness conditions:

$$v_{k,1}^{1*} = \frac{a_k}{\pi_k^{1*}}, \quad v_{k,1}^3 = \frac{1 - a_k}{\pi_k^3}, \quad v_{k,2}^{2*} = \frac{b_k}{\pi_k^{2*}}, \quad v_{k,2}^3 = \frac{1 - b_k}{\pi_k^3}$$

for real numbers a_k, b_k and $k \in U$. We now apply Theorem 1 to the above $\hat{M}_t$.

Theorem 2 *Let the double sample $s = (s_1, s_2)$ be selected according to a two-dimensional sampling design $p(s)$. Define $\hat{M} = (\hat{M}_t)_{t \in \mathcal{T}}$ by*

$$\hat{M}_1 = \sum_{k=1}^N \left(\frac{a_k}{\pi_k^{1*}} I_k^{1*} + \frac{1 - a_k}{\pi_k^3} I_k^3 \right) \delta_{z_{k,1}}, \quad \hat{M}_2 = \sum_{k=1}^N \left(\frac{b_k}{\pi_k^{2*}} I_k^{2*} + \frac{1 - b_k}{\pi_k^3} I_k^3 \right) \delta_{z_{k,2}},$$

$$\hat{M}_3 = \sum_{k=1}^N \frac{I_k^3}{\pi_k^3} \delta_{z_{k,3}},$$

for some real numbers a_k and b_k , and consider the general composite estimator $T(\hat{M})$.

Let Assumptions 1 to 7 hold. Then $\sqrt{n}N^{-\beta}\{T(\hat{M}) - T(M)\}$ is approximated by $\sqrt{n}N^{-\beta}(\hat{Z}_{\{(a_k,b_k)_{k\in U}\}} - Z)$ with $Z = \sum_{k=1}^N (u_{k,1} + u_{k,2} + u_{k,3})$ and

$$\hat{Z}_{\{(a_k,b_k)_{k\in U}\}} = \sum_{k=1}^N a_k u_{k,1} \left(\frac{I_k^{1*}}{\pi_k^{1*}} - \frac{I_k^3}{\pi_k^3} \right) + \sum_{k=1}^N b_k u_{k,2} \left(\frac{I_k^{2*}}{\pi_k^{2*}} - \frac{I_k^3}{\pi_k^3} \right) + \sum_{k=1}^N (u_{k,1} + u_{k,2} + u_{k,3}) \frac{I_k^3}{\pi_k^3}.$$

The asymptotic variance of $T(\hat{M})$ is the variance of $\hat{Z}_{\{(a_k,b_k)_{k\in U}\}}$.

Theorem 2 is an immediate consequence of Theorem 1, given that $\sum_{t=1}^3 \sum_{k=1}^N u_{k,t}(v_{k,t} - 1) = \hat{Z}_{\{(a_k,b_k)_{k\in U}\}} - Z$. The estimator $\hat{Z}_{\{(a_k,b_k)_{k\in U}\}}$ can be interpreted as a Horvitz-Thompson estimator of the linearized variables total, based on the matched sample, added to weighted unbiased estimators of zero means, based on the unmatched samples. This addition improves the estimation by making use of the correlation of the units from the matched and unmatched samples. Goga's thesis determines a_k and b_k, $k \in U$, that minimize the variance of $\hat{Z}_{\{(a_k,b_k)_{k\in U}\}}$. These optimal values, a_k^{opt} and b_k^{opt}, have rather complicated expressions and depend on the unknown $u_{k,t}$ for all $k \in U$ and $t \in \mathcal{T}$. In the following, we consider three particular cases of a_k and b_k, $k \in U$.

3.4 Some particular cases

Let $\hat{t}_{u_t}^d = \sum_{k \in s_d} u_{k,t}/\pi_k^d$, for $t \in \mathcal{T}$ and $d \in \mathcal{D} \cup \mathcal{T}$, the Horvitz-Thompson estimators of the population total $\sum_{k=1}^N u_{k,t}$ using the sample s_d. For example, if $d = 1*$, $\hat{t}_{u_1}^{1*} = \sum_{k \in s_{1*}} u_{k,1}/\pi_k^{1*}$ and, if $d = 1$, $\hat{t}_{u_1}^1 = \sum_{k \in s_1} u_{k,1}/\pi_k^1$.

Case 1: *The 'union' estimator*. Let us consider $a_k = \pi_k^{1*}/\pi_k^1$ and $b_k = \pi_k^{2*}/\pi_k^2$ for all $k \in U$. In this case,

$$\hat{M}_t^{\text{uni}} = \sum_{k=1}^N \frac{I_k^t}{\pi_k^t} \delta_{z_{k,t}}, \quad t = 1, 2,$$

are the Horvitz-Thompson estimators of M_t based on the whole samples s_t, and $T(\hat{M}^{\text{uni}})$ is called the union substitution estimator. From Theorem 2, the asymptotic variance of $T(\hat{M}^{\text{uni}})$ is the variance of

$$\hat{Z}_{\{(\pi_k^{1*}/\pi_k^1, \pi_k^{2*}/\pi_k^2)_{k\in U}\}} \quad = \quad \sum_{k\in s_1} \frac{u_{k,1}}{\pi_k^1} + \sum_{k\in s_2} \frac{u_{k,2}}{\pi_k^2} + \sum_{k\in s_3} \frac{u_{k,3}}{\pi_k^3} = \hat{t}_{u_1}^1 + \hat{t}_{u_2}^2 + \hat{t}_{u_3}^3. \tag{7}$$

Consider the ratio change $\Delta R = R_2 - R_1$ from Example 2 in §2.1. We have $R_t = \left(\int \mathcal{Y}_t dM_t\right) / \left(\int \mathcal{X}_t dM_t\right)$ and we estimate M_t by $\hat{M}_t^{\text{uni}}$, $t = 1, 2$. We obtain $\hat{\Delta} R^{\text{uni}} = \hat{R}_2^{\text{uni}} - \hat{R}_1^{\text{uni}}$ with

$$\hat{R}_t^{\text{uni}} = \frac{\int \mathcal{Y}_t d\hat{M}_t^{\text{uni}}}{\int \mathcal{X}_t d\hat{M}_t^{\text{uni}}} = \frac{\sum_{k\in s_t} y_{k,t}/\pi_k^t}{\sum_{k\in s_t} x_{k,t}/\pi_k^t}, \quad t = 1, 2,$$

and the asymptotic variance of $\hat{\Delta} R^{\text{uni}}$ equals the variance of $\hat{t}_{u_1}^1 + \hat{t}_{u_2}^2$ where the linearized variables $u_{k,t}$ are given by (4).

Case 2: *The 'intersection' estimator*. Let $a_k = b_k = 0$ for all $k \in U$. Then $\hat{M}_t^{\text{int}}$ is the Horvitz-Thompson estimator of M_t based on s_3:

$$\hat{M}_t^{\text{int}} = \sum_{k=1}^N \frac{I_k^3}{\pi_k^3} \delta_{z_{k,t}}, \quad t = 1, 2,$$

From Theorem 2, the asymptotic variance of the intersection substitution estimator $T(\hat{M}^{\text{int}})$ is equal to the variance of

$$\hat{Z}_{(0,0)} = \sum_{k\in s_3} \frac{u_{k,1} + u_{k,2} + u_{k,3}}{\pi_k^3} = \sum_{t=1}^3 \hat{t}_{u_t}^3. \tag{8}$$

The ratio change ΔR is estimated by $\hat{\Delta} R^{\text{int}} = \hat{R}_2^{\text{int}} - \hat{R}_1^{\text{int}}$ with $\hat{R}_t^{\text{int}} = \left(\sum_{k\in s_3} \frac{y_{k,t}}{\pi_k^3}\right) / \left(\sum_{k\in s_3} \frac{x_{k,t}}{\pi_k^3}\right)$ and its asymptotic variance equals $\operatorname{var} \sum_{k\in s_3} \{(u_{k,1} + u_{k,2})/\pi_k^3\}$ with $u_{k,1}$ and $u_{k,2}$ given by (4).

Case 3: *The 'composite' estimator*. If we consider $a_k = a \in R$ and $b_k = b \in R$, then

$$v_{k,1} = \frac{a}{\pi_k^{1*}} I_k^{1*} + \frac{1-a}{\pi_k^3} I_k^3 \text{ and } v_{k,2} = \frac{b}{\pi_k^{2*}} I_k^{2*} + \frac{1-b}{\pi_k^3} I_k^3.$$

The measures M_t are estimated by the composite estimators,

$$\hat{M}_1^{\text{co}} \quad = \quad \sum_{k=1}^N \left(a \frac{I_k^{1*}}{\pi_k^{1*}} + (1-a) \frac{I_k^3}{\pi_k^3} \right) \delta_{z_{k,1}}, \quad \hat{M}_2^{\text{co}} = \sum_{k=1}^N \left(b \frac{I_k^{2*}}{\pi_k^{2*}} + (1-b) \frac{I_k^3}{\pi_k^3} \right) \delta_{z_{k,2}}. \tag{9}$$

From Theorem 2, the asymptotic variance of the composite substitution estimator $T(\hat{M}^{\text{co}})$ is given by the variance of

$$\hat{Z}_{(a,b)} \quad = \quad a \left(\hat{t}_{u_1}^{1*} - \hat{t}_{u_1}^3 \right) + b \left(\hat{t}_{u_2}^{2*} - \hat{t}_{u_2}^3 \right) + \sum_{t=1}^3 \hat{t}_{u_t}^3. \tag{10}$$

By taking $a = b = 0$ in (10), we obtain $\hat{Z}_{(0,0)}$ given by (8) and $T(\hat{M}^{co}) = T(\hat{M}^{\text{int}})$. The union estimator, defined by (7), belongs to the class defined by (10) if and only if the sampling design is an equal-probability two-dimensional design with constant weights π_k^{1*}, π_k^3 and π_k^{2*} for all $k \in U$. Section 4 provides an example of such a design.

Consider again the ratio change of Example 2. Replace M_t with $\hat{M}_t^{\text{co}}$ and obtain the composite estimator $\hat{\Delta}R^{\text{co}} = \hat{R}_2^{\text{co}} - \hat{R}_1^{\text{co}}$ with $\hat{R}_t^{\text{co}} = \int R_t d\hat{M}_t^{\text{co}}$. To be more precise,

$$\begin{aligned}\hat{R}_1^{\text{co}} &= \frac{a\sum_{k\in s_{1*}} y_{k,1}/\pi_k^{1*} + (1-a)\sum_{k\in s_3} y_{k,1}/\pi_k^3}{a\sum_{k\in s_{1*}} x_{k,1}/\pi_k^{1*} + (1-a)\sum_{k\in s_3} x_{k,1}/\pi_k^3},\\ \hat{R}_2^{\text{co}} &= \frac{b\sum_{k\in s_{2*}} y_{k,2}/\pi_k^{2*} + (1-b)\sum_{k\in s_3} y_{k,2}/\pi_k^3}{b\sum_{k\in s_{2*}} x_{k,2}/\pi_k^{2*} + (1-b)\sum_{k\in s_3} x_{k,2}/\pi_k^3}.\end{aligned}$$

The asymptotic variance of $\hat{\Delta}R^{\text{co}}$ is the variance of $\hat{Z}_{(a,b)} = a\left(\hat{t}_{u_1}^{1*} - \hat{t}_{u_1}^3\right) + b\left(\hat{t}_{u_2}^{2*} - \hat{t}_{u_2}^3\right) + \sum_{t=1}^2 \hat{t}_{u_t}^3$ with $u_{k,1}, u_{k,2}$ given by (4).

In an unpublished University of Burgundy technical report by C. Goga, J.-C. Deville and A. Ruiz-Gazen, composite estimators are developed for other parameters of interest such as the changes of the population total and of the Gini index.

To calculate $\text{var}(\hat{Z}_{(a,b)})$, each estimator $\hat{t}_{u_t}^d$ is written as a function of the sample membership I_k^d, namely $\hat{t}_{u_t}^d = \sum_{k=1}^N u_{k,t} I_k^d/\pi_k^d$. We have $\text{cov}(I_k^d, I_l^{d'}) = \pi_{kl}^{d,d'} - \pi_k^d\pi_l^{d'} = \Delta_{kl}^{d,d'}$. For example,

$$\text{var}(\hat{t}_{u_1}^{1*}) = \sum_{k=1}^N\sum_{l=1}^N \Delta_{kl}^{1*}\frac{u_{k,1}}{\pi_k^{1*}}\frac{u_{l,1}}{\pi_l^{1*}}, \quad \text{cov}(\hat{t}_{u_1}^{1*}, \hat{t}_{u_1}^3) = \sum_{k=1}^N\sum_{l=1}^N \Delta_{kl}^{1*,3}\frac{u_{k,1}}{\pi_k^{1*}}\frac{u_{l,1}}{\pi_l^3}.$$

The variance of $\hat{Z}_{(a,b)}$ may be considered as a two-sample Horvitz-Thompson variance formula. It is the sum of variance terms computed according to a one-sample Horvitz-Thompson variance formula and of covariance terms which contain the covariance between I_k^d and $I_l^{d'}$ for $d \neq d'$ and are not common in survey sampling theory.

3.5 Variance estimator of the composite substitution estimator

Consider the composite substitution estimator $T(\hat{M}^{\text{co}})$ with $\hat{M}^{\text{co}} = (\hat{M}_1^{\text{co}}, \hat{M}_2^{\text{co}}, \hat{M}_3)$ given by (6) and (9) and assume that a and b are fixed real numbers. We propose to estimate the variance of $T(\hat{M}^{\text{co}})$ by an estimator, $\hat{\text{var}}\hat{Z}_{(a,b)}$. In order to derive such an estimator, we write

$$\text{var}(\hat{Z}_{(a,b)}) = \text{var}(A) + \text{var}(B) + \text{var}(C) + 2\,\text{cov}(A,B) + 2\,\text{cov}(A,C) + 2\,\text{cov}(B,C), \tag{11}$$

$$\text{with } \hat{Z}_{(a,b)} = A + B + C, \text{where } A = \hat{t}_{u_1}^{1*} + (1-a)\,\hat{t}_{u_1}^3,\ B = \hat{t}_{u_2}^{2*} + (1-b)\,\hat{t}_{u_2}^3,\ C = \hat{t}_{u_3}^3.$$

The linearized variables $u_{k,t}$ and the variance and covariance terms are to be estimated. The linearized variables depend on the unknown variables of interest $\mathcal{Z}_t$ and several estimators are possible. Furthermore, explicit expressions for $u_{k,t}$ cannot be derived so long as the functional T is not given precisely. In these conditions, finding the most suitable estimators of $u_{k,t}$ is not a simple issue. In the following, we simply estimate $u_{k,t}$ based on the matched sample s_3 by

$$\hat{u}_{k,t}^{\text{int}} = I_{tT}(\hat{M}^{\text{int}}, z_{k,t}),$$

but other estimators may be advisable, in particular if the sample sizes n_1^* and n_2^* are much larger than n_3. Consider Example 2 of §2.1. We have

$$\hat{u}_{k,1}^{\text{int}} = -\left(1/\sum_{k\in s_3}\frac{x_{k,1}}{\pi_k^3}\right)(y_{k,1} - \hat{R}_1^{\text{int}}x_{k,1}),\ \hat{u}_{k,2}^{\text{int}} = \left(1/\sum_{k\in s_3}\frac{x_{k,2}}{\pi_k^3}\right)(y_{k,2} - \hat{R}_2^{\text{int}}x_{k,2}),$$

for $\hat{R}_1^{\rm int}, \hat{R}_2^{\rm int}$ as given in §3.4. However, other possible estimators are

$$\hat{u}_{k,1}^{\rm uni} = -\left(1/\sum_{k\in s_1}\frac{x_{k,1}}{\pi_k^1}\right)(y_{k,1} - \hat{R}_1^{\rm uni}x_{k,1}),\ \hat{u}_{k,2}^{\rm uni} = \left(1/\sum_{k\in s_2}\frac{x_{k,2}}{\pi_k^2}\right)(y_{k,2} - \hat{R}_2^{\rm uni}x_{k,2}),$$

for $\hat{R}_1^{\rm uni}$ and $\hat{R}_2^{\rm uni}$ given in §3.4. We estimate ${\rm var}(C)$, respectively ${\rm var}(A)$ and ${\rm var}(B)$, by Horvitz-Thompson variance estimators (2) based on the matched sample s_3, respectively on s_1 and s_2, with $u_{k,t}$ replaced by $\hat{u}_{k,t}^{\rm int}$, $t \in \mathcal{T}$. To be more precise, we have

$$\hat{\rm var}(C) = \sum_{k\in s_3}\sum_{l\in s_3}\frac{\Delta_{kl}^3}{\pi_{kl}^3}\frac{\hat{u}_{k,3}^{\rm int}}{\pi_k^3}\frac{\hat{u}_{l,3}^{\rm int}}{\pi_l^3},$$

$$\hat{\rm var}(A) = \sum_{k\in s_1}\sum_{l\in s_1}\hat{u}_{k,1}^{\rm int}\hat{u}_{l,1}^{\rm int}\frac{1}{\pi_{kl}^1}\left\{a^2\frac{\Delta_{kl}^{1*}}{\pi_k^{1*}\pi_l^{1*}} + 2a(1-a)\frac{\Delta_{kl}^{1*,3}}{\pi_k^{1*}\pi_l^3} + (1-a)^2\frac{\Delta_{kl}^3}{\pi_k^3\pi_l^3}\right\},$$

$$\hat{\rm var}(B) = \sum_{k\in s_2}\sum_{l\in s_2}\hat{u}_{k,2}^{\rm int}\hat{u}_{l,2}^{\rm int}\frac{1}{\pi_{kl}^2}\left\{b^2\frac{\Delta_{kl}^{2*}}{\pi_k^{2*}\pi_l^{2*}} + 2b(1-b)\frac{\Delta_{kl}^{2*,3}}{\pi_k^{2*}\pi_l^3} + (1-b)^2\frac{\Delta_{kl}^3}{\pi_k^3\pi_l^3}\right\}.$$

The covariance term

$${\rm cov}(A,C) = \sum_{k=1}^N\sum_{l=1}^N u_{k,1}u_{l,3}\left\{a\frac{\Delta_{kl}^{1*,3}}{\pi_k^{1*}\pi_l^3} + (1-a)\frac{\Delta_{kl}^3}{\pi_k^3\pi_l^3}\right\}$$

is estimated by

$$\hat{\rm cov}(A,C) = \sum_{k\in s_3}\sum_{l\in s_3}\hat{u}_{k,1}^{\rm int}\hat{u}_{l,3}^{\rm int}\frac{1}{\pi_{kl}^3}\left\{a\frac{\Delta_{kl}^{1*,3}}{\pi_k^{1*}\pi_l^3} + (1-a)\frac{\Delta_{kl}^3}{\pi_k^3\pi_l^3}\right\},$$

and ${\rm cov}(A,C)$ and ${\rm cov}(B,C)$ are estimated in a similar way. Note that the proposed variance estimator $\hat{\rm var}\hat{Z}_{(a,b)}$ is no longer unbiased for ${\rm var}\hat{Z}_{(a,b)}$ since $\hat{u}_{k,t}^{\rm int}$ is generally biased for $u_{k,t}$. However, $\hat{u}_{k,t}^{\rm int}$ is a function of Horvitz-Thompson estimators and is consistent for $u_{k,t}$ as N tends to infinity, implying $n_3 \to \infty$ by Assumption 1.

Theorem 3 *Under the Assumptions* 1 *to* 7 *and* A1 *and* A2 *given in the Appendix,* $\hat{\rm var}(\hat{Z}_{(a,b)})$ *is a consistent estimator of* ${\rm AV}\{T(\hat{M}^{\rm co})\} = {\rm var}(\hat{Z}_{(a,b)})$.

For the proof, see the Appendix. In §5, a small simulation study confirms that the variance estimator $\hat{\rm var}\{T(\hat{M}^{\rm co})\} = \hat{\rm var}(\hat{Z}_{(a,b)})$ does not differ very much from the asymptotic variance ${\rm AV}\{T(\hat{M}^{\rm co})\}$ in large samples.

3.6 Optimal asymptotic variance composite estimator

In this section, we derive real numbers a and b such that the asymptotic variance of the composite substitution estimator $T(\hat{M}^{\rm co})$ is minimum. Let $\theta = (a,b)' \in \mathcal{R}^2$ and rewrite (10) as

$$\hat{Z}_{(a,b)} \quad = \quad \theta'\begin{pmatrix}\hat{t}_{u_1}^{1*} - \hat{t}_{u_1}^3\\ \hat{t}_{u_2}^{2*} - \hat{t}_{u_2}^3\end{pmatrix} + \sum_{t=1}^3\hat{t}_{u_t}^3. \qquad (12)$$

The asymptotic variance of $T(\hat{M}^{\rm co})$ is

$${\rm AV}\{T(\hat{M}^{\rm co})\} = {\rm var}(\hat{Z}_{(a,b)}) = \theta'\Gamma\theta + 2\theta'\gamma + {\rm var}\left(\sum_{t=1}^3\hat{t}_{u_t}^3\right) \qquad (13)$$

with

$$\Gamma = \mathrm{var}\begin{pmatrix} \hat{t}^{1*}_{u_1} - \hat{t}^{3}_{u_1} \\ \hat{t}^{2*}_{u_2} - \hat{t}^{3}_{u_2} \end{pmatrix}, \quad \gamma = \mathrm{cov}\left(\begin{pmatrix} \hat{t}^{1*}_{u_1} - \hat{t}^{3}_{u_1} \\ \hat{t}^{2*}_{u_2} - \hat{t}^{3}_{u_2} \end{pmatrix}, \textstyle\sum_{t=1}^{3} \hat{t}^{3}_{u_t} \right). \tag{14}$$

Theorem 4 *Consider a general two-dimensional sampling design $p(s)$ and suppose that Assumptions 1 to 7 hold. The asymptotic variance of $T(\hat{M}^{\mathrm{co}})$ is minimum for $\theta_{\mathrm{opt}} = (a_{\mathrm{opt}}, b_{\mathrm{opt}})' = -\Gamma^{-1}\gamma$ with Γ and γ given by (14) and if Γ is assumed nonsingular. This minimum asymptotic variance is the variance of $\hat{Z}_{(a_{\mathrm{opt}}, b_{\mathrm{opt}})}$ and is equal to*

$$\mathrm{AV}_{\mathrm{opt}}\{T(\hat{M}^{\mathrm{co}}_{\mathrm{opt}})\} = \mathrm{var}(\hat{Z}_{(a_{\mathrm{opt}}, b_{\mathrm{opt}})}) = \mathrm{var}\left(\sum_{t=1}^{3} \hat{t}^{3}_{u_t}\right) - \gamma'\Gamma^{-1}\gamma. \tag{15}$$

The proof is given together with the proof of Corollary 5 in the Appendix. The optimal variance is obtained whatever the two-dimensional sampling design is. Explicit expressions for the optimal θ and the asymptotic variance are given in C. Goga's thesis for several two-dimensional sampling designs. Expression (8) leads to $\mathrm{AV}\{T(\hat{M}^{\mathrm{int}})\} = \mathrm{var}(\hat{Z}_{(0,0)}) = \mathrm{var}(\sum_{t=1}^{3} \hat{t}^{3}_{u_t})$, which means that, whatever the sampling design may be, $T(\hat{M}^{\mathrm{co}}_{\mathrm{opt}})$ has a smaller asymptotic variance than $T(\hat{M}^{\mathrm{int}})$.

Unfortunately, the optimal variance (15) depends on unknown population variances and covariances and cannot be calculated. We propose to estimate all the unknown quantities in (13) using the estimators described in the above section.

Corollary 5 (i) *The variance estimator $\hat{\mathrm{AV}}\{T(\hat{M}^{\mathrm{co}})\} = \theta'\hat{\Gamma}\theta + 2\theta'\hat{\gamma} + \hat{\mathrm{var}}(\sum_{t=1}^{3} \hat{t}^{3}_{u_t})$ is minimum for $\hat{\theta}_{\mathrm{opt}} = -\hat{\Gamma}^{-1}\hat{\gamma}$, if $\hat{\Gamma}$ is assumed nonsingular.*

For (ii) and (iii), let Assumptions 1 to 7 hold. Suppose also that $\hat{\theta}_{\mathrm{opt}}$ is a consistent estimator of θ_{opt}, that is, for any fixed $\varepsilon > 0$, $\lim_{N\to\infty} \mathrm{pr}(||\hat{\theta}_{\mathrm{opt}} - \theta_{\mathrm{opt}}|| > \varepsilon) = 0$, where $||\cdot||$ is the Euclidian norm.
(ii) *Consider the estimator $\hat{Z}_{(\hat{a}_{\mathrm{opt}}, \hat{b}_{\mathrm{opt}})}$ given by (12) for $\hat{\theta}_{\mathrm{opt}} = (\hat{a}_{\mathrm{opt}}, \hat{b}_{\mathrm{opt}})$. The asymptotic variance of $\hat{Z}_{(\hat{a}_{\mathrm{opt}}, \hat{b}_{\mathrm{opt}})}$ is equal to the variance of $\hat{Z}_{(a_{\mathrm{opt}}, b_{\mathrm{opt}})}$.*
(iii) *Consider now the estimator $T(\tilde{M}^{\mathrm{co}}_{\mathrm{opt}})$ with $\tilde{M}^{\mathrm{co}}_{t,\mathrm{opt}}$, $t = 1, 2$, obtained from (9) for $a = \hat{a}_{\mathrm{opt}}$ and $b = \hat{b}_{\mathrm{opt}}$. The asymptotic variance of $T(\tilde{M}^{\mathrm{co}}_{\mathrm{opt}})$ is equal to the variance of $\hat{Z}_{(a_{\mathrm{opt}}, b_{\mathrm{opt}})}$.*

The proof is given in the Appendix. Part (i) gives the estimator $\hat{\theta}_{\mathrm{opt}}$ that minimizes the asymptotic variance estimator for a constant θ. [20] and [11] obtained a similar result concerning the optimality of the regression coefficient. The drawback of Theorem 4 is that θ_{opt} is assumed to be known but in practice it has to be estimated. Corollary 5 (iii) takes the estimation of θ_{opt} into account and states that, if θ_{opt} is estimated consistently, the asymptotic variance of the substitution estimator $T(\tilde{M}^{\mathrm{co}}_{\mathrm{opt}})$ with estimated θ_{opt} is the minimum variance $\mathrm{var}(\hat{Z}_{(a_{\mathrm{opt}}, b_{\mathrm{opt}})})$ given by (15).

4 Two-dimensional simple random sampling without replacement

Let us focus now on a particular two-dimensional sampling design, namely two-dimensional simple random sampling without replacement defined in the working paper by F. Cotton and C. Hesse and used for two-sample coordination. In what follows, we consider functionals Φ not depending on M_3 and we assume the two-dimensional simple random sampling without replacement design for estimating $\Phi = T(M)$. This design can be described as follows.

Definition 4.1 *A two-dimensional simple random sampling without replacement of fixed size (n_{1*}, n_3, n_{2*}) is a two-dimensional sampling design $p(s)$ which assigns equal selection probability to all samples $s = (s_1, s_2)$ for which s_{1*}, respectively s_3 and s_{2*}, have the fixed sizes n_{1*}, respectively n_3 and n_{2*}.*

In this case, the design $p(s)$ is a discrete uniform probability distribution on the set of

$$\binom{N}{n_{1*}+n_3+n_{2*}}\binom{n_{1*}+n_3+n_{2*}}{n_{1*}}\binom{n_3+n_{2*}}{n_3}\binom{n_{2*}}{n_{2*}}$$

possible samples of fixed size (n_{1*}, n_3, n_{2*}), which implies that

$$p\{s=(s_1,s_2)\}=\frac{n_{1*}!n_3!n_{2*}!(N-n_{1*}-n_3-n_{2*})!}{N!}.$$

In their working paper, Cotton and Hesse study this design and give some of its properties. The most important of them is the fact that the marginal sampling designs are simple random sampling without replacement from U. This property makes the design very attractive. The first-order two-dimensional inclusion probabilities are $\pi_k^d = n_d/N$ and the second-order probabilities are

$$\pi_{kl}^d=\frac{n_d(n_d-1)}{N(N-1)},\quad \pi_{kl}^{d,d'}=\frac{n_d n_{d'}}{N(N-1)},$$

for $d \neq d'$. From a practical point of view, this design can be implemented by selecting the simple random samples $s_1 \subset U$ and $s_3 \subset s_1$ and next by selecting s_{2*} from $U - s_1$ also according to a simple random design. Such a sampling design can be found in repeated sampling [28] when a matched sample of fixed size is desired in order to improve the estimation of the absolute change of the parameter of interest. Another way of implementing the two-dimensional simple random design is by selecting three nonoverlapping simple random samples. We select s_{1*} from U, s_3 from $U - s_{1*}$ and s_{2*} from $U - s_1$, each time using simple random designs. Such a design is also of interest for reducing the response burden (Särndal et al., 1992, p. 67). Note that the selection of two, not necessarily independent, simple random samples from U cannot be considered as a two-dimensional simple random design since the matched sample is of random size. Nevertheless, conditioning on n_3, we obtain a two-dimensional simple random design.

We consider a functional $\Phi = T(M)$ estimated by the composite substitution estimator $T(\hat{M}^{\text{co}})$ with asymptotic variance equal to the variance of

$$\hat{Z}_{(a,b)}=a\left(\hat{t}^{1*}_{u_1}-\hat{t}^{3}_{u_1}\right)+b\left(\hat{t}^{2*}_{u_2}-\hat{t}^{3}_{u_2}\right)+\sum_{t=1}^{2}\hat{t}^{3}_{u_t}. \tag{16}$$

We compute the optimal values of a and b by using Theorem 4. Let $h_1 = n_{1*}/n_1$ and $h_2 = n_{2*}/n_2$ be the nonoverlapping rates and ρ the correlation coefficient of the linearized variables $u_{k,1}$ and $u_{k,2}$. We denote by $m(u_t)$ the population mean of u_t and by $S^2_{u_t} = \sum_{k=1}^N \{u_{k,t} - m(u_t)\}^2/(N-1)$ the population variances of u_t, for $t = 1, 2$, estimated by $\hat{S}^2_{\hat{u}_t} = \sum_{k\in s_t}\{\hat{u}_{k,t} - \hat{m}(u_t)\}^2/(n_t - 1)$ and by $S_{u_1u_2} = \sum_{k=1}^N\{u_{k,1} - m(u_1)\}\{u_{k,2} - m(u_2)\}/(N-1)$ the population covariance between u_1 and u_2 estimated by $\hat{S}_{\hat{u}_1\hat{u}_2} = \sum_{k\in s_3}\{\hat{u}_{k,1} - \hat{m}(u_1)\}\{\hat{u}_{k,2} - \hat{m}(u_2)\}/(n_3 - 1)$, where $\hat{u}_{k,t} = \hat{u}^{\text{int}}_{k,t}$. Let $S = S_{u_2}/S_{u_1}$ and let $f_3 = n_3/N$ be the overlapping sampling fraction. We have the following result.

Theorem 6 *For a two-dimensional simple random design and under Assumptions* 1 *to* 7, *the asymptotic variance of* $T(\hat{M}^{\text{co}})$ *is given by* (13) *with var*$(\hat{t}^3_{u_1} + \hat{t}^3_{u_2}) = N(1-f_3)f_3^{-1}S_{u_1}S_{u_2}(S + 2\rho + 1/S)$,

$$\Gamma=\frac{N}{f_3}S_{u_1}S_{u_2}\begin{pmatrix} S^{-1}h_1^{-1} & \rho \\ \rho & Sh_2^{-1}\end{pmatrix},\quad \gamma=-\frac{N}{f_3}S_{u_1}S_{u_2}\begin{pmatrix}\rho+S^{-1}\\ \rho+S\end{pmatrix}.$$

The optimal composite substitution estimator $T(\hat{M}^{\text{co}}_{\text{opt}})$ *is given by Theorem* 4 *with*

$$\theta_{\text{opt}}=(a_{\text{opt}},b_{\text{opt}})'=\frac{-h_1h_2}{1-\rho^2h_1h_2}\begin{pmatrix}\rho^2+\rho(1-1/h_2)S-1/h_2\\ \rho^2+\rho(1-1/h_1)S^{-1}-1/h_1\end{pmatrix} \tag{17}$$

and has the minimum asymptotic variance calculated according to (15).

The proof is given in the Appendix. The vector θ_{opt} is unknown and, according to Corollary 5, we obtain the expression for $\hat{\theta}_{\text{opt}}$ by replacing the unknown ρ and S with their estimators $\hat{\rho}$ and $\hat{S}$ in (17).

In §3.6, we proved that the substitution estimator $T(\hat{M}^{\text{int}})$ is always less competitive than $T(\hat{M}^{\text{co}}_{\text{opt}})$, whatever the sampling design is. For a two-dimensional simple random design, both estimators have the same asymptotic variance for $\rho = -1$ and $S = 1$.
The second natural competitor of $T(\hat{M}^{\text{co}}_{\text{opt}})$ is $T(\hat{M}^{\text{uni}})$ with asymptotic variance $\text{AV}\{T(\hat{M}^{\text{uni}})\} = \text{var}(\hat{t}^1_{u_1} + \hat{t}^2_{u_2})$. If $a = h_1$ and $b = h_2$ in (16), we have

$$\hat{Z}_{(h_1,h_2)} = h_1\left(\hat{t}^{1*}_{u_1} - \hat{t}^3_{u_1}\right) + h_2\left(\hat{t}^{2*}_{u_2} - \hat{t}^3_{u_2}\right) + \left(\hat{t}^3_{u_1} + \hat{t}^3_{u_2}\right) = \hat{t}^1_{u_1} + \hat{t}^2_{u_2}$$

which means that $\hat{t}^1_{u_1} + \hat{t}^2_{u_2}$ belongs to the class of composite estimators defined by (16). It follows that $\text{AV}\{T(\hat{M}^{\text{co}}_{\text{opt}})\} \leq \text{AV}\{T(\hat{M}^{\text{uni}})\} = \text{var}(\hat{t}^1_{u_1} + \hat{t}^2_{u_2})$ with equality for $\rho = 0$. In particular, one may obtain $\text{AV}\{T(\hat{M}^{\text{uni}})\}$ using (13) for $\theta = (h_1, h_2)'$ and Γ, and γ given by Theorem 6.

5 Empirical study

5.1 General framework

We consider the estimation of a nonlinear functional $\Phi = T(M_1, M_2)$ based on $s = (s_1, s_2)$ selected according to a two-dimensional simple random sampling design. The empirical studies presented below intend to give the gain of the optimal composite estimator $T(\hat{M}^{\text{co}}_{\text{opt}})$ defined in Theorem 6 over $T(\hat{M}^{\text{uni}})$, respectively $T(\hat{M}^{\text{int}})$. The gain is defined as the ratio between the asymptotic variance of $T(\hat{M}^{\text{uni}})$, respectively $T(\hat{M}^{\text{int}})$, and the asymptotic variance of $T(\hat{M}^{\text{co}}_{\text{opt}})$.

In this subsection, we consider a general functional Φ. Let u_1 and u_2 be the linearized variables of a functional $\Phi = T(M_1, M_2)$. We consider a population U of size $N = 3000$ and a two-dimensional simple random sample design such that $n = n_1 + n_2 - n_3 = 300$ and $n_{1*} = 100$. We assume that the variance ratio $S = S_{u_2}/S_{u_1}$ is equal to 1 and we consider different values of the correlation coefficient ρ between u_1 and u_2, namely $\rho = -0.8,\ -0.5,\ 0,\ 0.5,\ 0.8$. This correlation coefficient ρ depends on the form of the functional Φ and on the correlation coefficient between the variables of interest but we cannot give a general expression.

We plot in Fig. 1 (a) and (b) respectively the gain of $T(\hat{M}^{\text{co}}_{\text{opt}})$ over $T(\hat{M}^{\text{int}})$ and $T(\hat{M}^{\text{uni}})$ as a function of the overlapping rate n_3/n. Each curve corresponds to a different correlation coefficient.

As can be expected, concerning $T(\hat{M}^{\text{int}})$, the ratio of variances decreases to 1 when the overlapping rate increases and this ratio is small if the correlation coefficient is low. When the original variables are highly negatively correlated, $\rho \leq -0.8$, and as soon as the overlapping rate is greater than 10%, we do not gain anything by using the optimal estimator instead of using the estimator based on the intersection sample. In §4, we obtained that $T(\hat{M}^{\text{uni}}) = T(\hat{M}^{\text{co}}_{\text{opt}})$ for $\rho = -1$ and $S = 1$ and this is confirmed by the empirical study. When the correlation is greater than -0.5, the gain can be substantial at least when the overlapping rate is smaller than 30%.

With regard to comparison of the asymptotic variances of $T(\hat{M}^{\text{uni}})$ and $T(\hat{M}^{\text{co}}_{\text{opt}})$, Fig. 1 (b) shows that there is no great difference when the correlation coefficient between the linearized variables is low in absolute value, $|\rho| < 0.5$, and, for $\rho = 0$, the variance ratio is equal to unity; this confirms the theoretical result. However, for high values of $|\rho|$, the gain of the optimal estimator over the union estimator is more important especially when $\rho < 0$; the ratios increase as soon as the overlapping rate is less than say 30% and decrease when the rate is larger than 30%. For very low or very high overlapping rates the two estimators are not very different but, when the overlapping rate is, say, 30%, the optimal estimator is much superior.

5.2 Estimating the change of a Gini index

We consider data from the French employment surveys of 1999 and 2000, namely the wages of $N = 22\ 741$ wage-earners who have been sampled in both years. We are interested in estimating the variance of the change

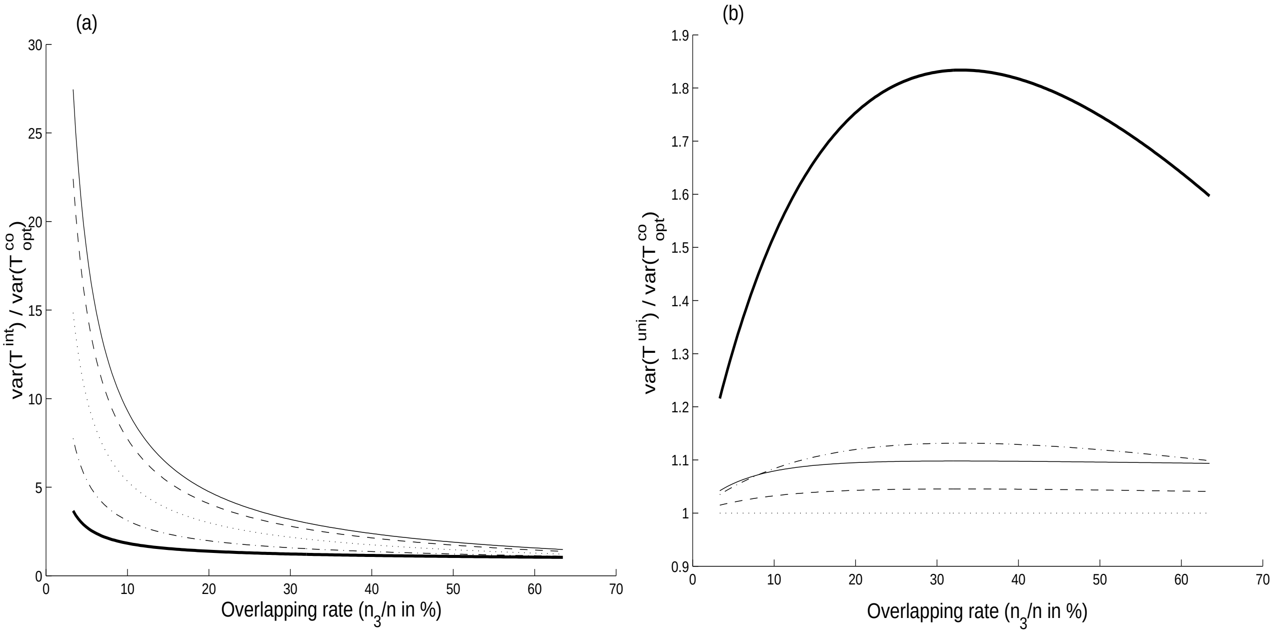

Figure 1: Simulation study for the general case. The ratio of the asymptotic variances of (a) the intersection estimator and the optimal estimator and (b) the union estimator and the optimal estimator, as functions of the overlapping rate and for different correlation coefficients ρ (heavy solid line for $\rho = -0.8$, dotted dashed line for $\rho = -0.5$, dotted line for $\rho = 0$, dashed line for $\rho = 0.5$, light solid line for $\rho = 0.8$).

in the Gini index between the two years, $\Delta G = G_2 - G_1$, where

$$G_t = \frac{\int_0^\infty \{2F_t(y) - 1\} y dM_t(y)}{\int_0^\infty y dM_t(y)} = \frac{1}{N\bar{Y}_t} \sum_{k=1}^{N} y_{k,t}\{2F_t(y_{k,t}) - 1\}$$

is the Gini index and $F_t(y) = (1/N)\int_0^\infty \mathbf{1}_{\{\xi \leq y\}} dM_t(\xi)$ is the distribution function in year $t = 1, 2$. Since G_t involves the step-function F_t, we cannot apply the Taylor linearization approach. In the one-sample case, the influence function approach [8] and the estimating equations approach [18] are two possible methodologies. In the two-sample situation, we propose to use the partial influence function approach. The linearized variables of ΔG are

$$u_{k,1} = -\{2F(y_{k,1})\frac{y_{k,1} - \bar{y}_{k,1<}}{Y_1} - y_{k,1}\frac{G_1 + 1}{Y_1} + \frac{1 - G_1}{N}\},$$

$$u_{k,2} = 2F(y_{k,2})\frac{y_{k,2} - \bar{y}_{k,2<}}{Y_2} - y_{k,2}\frac{G_2 + 1}{Y_2} + \frac{1 - G_2}{N},$$

where $\bar{y}_{k,t<}$ denotes the mean of the $y_{j,t}$ lower than $y_{k,t}$. The correlation of the linearized variables of ΔG between 1999 and 2000 is $\rho = -0.87$ and the population variance ratio is $S = 0.97$.

We consider a two-dimensional simple random sampling design of size $n = 1000$ and three different composite estimators: the 'intersection' $\hat{\Delta} G^{\text{int}}$, the 'union' $\hat{\Delta} G^{\text{uni}}$ and the 'optimal composite' estimator $\hat{\Delta} G^{\text{co}}_{\text{opt}}$ given by Theorem 6. We calculate the asymptotic variances of these estimators using (13) with $\theta = (0, 0)'$ for the 'intersection', $\theta = (h_1, h_2)'$ for the 'union' and θ given by 17 for the 'optimal composite' estimator. We give in Fig. 2 the gain of the optimal composite estimator $\hat{\Delta} G^{\text{co}}_{\text{opt}} = \hat{G}^{\text{co}}_{2,\text{opt}} - \hat{G}^{\text{co}}_{1,\text{opt}}$ over the two competitors $\hat{\Delta} G^{\text{int}}$ and $\hat{\Delta} G^{\text{uni}}$ as a function of the ratio n_3/n and for different sample sizes n_{1*}.

The approximate variance of the intersection estimator is quite similar to that of the optimal estimator when the overlapping rate is larger than 30% but can be larger for small overlapping rates. Except for very small or very large overlapping rates, the approximate variance of the union estimator is much higher than that of the optimal estimator.

In all the above examples we assume that the population variances and covariances are known. In order to verify the quality of the corresponding estimators, we carried out a small simulation study for the Gini example. We estimated the change in the Gini index using the ‘optimal composite’ estimator as defined in Corollary 5 (iii). Since we can compute the true change in the Gini index from the original sample of $22\,741$ earners, we calculated, as percentages, the relative bias and the relative root mean squared error of the change estimator using $10\,000$ simulations. We also calculated the relative difference between the asymptotic variance given by (15) and the empirical variance, and the relative bias of the asymptotic variance estimator, considering the empirical variance as the true variance. For the asymptotic variance estimation, the linearized variables are estimated on the overlapping sample s_3. Different values for n_3 and n_{1*} with $n_{1*} = n_{2*}$ are considered. Table 1 shows that the relative biases, the root mean squared errors and the relative differences are quite low in general and very low for large sample sizes.

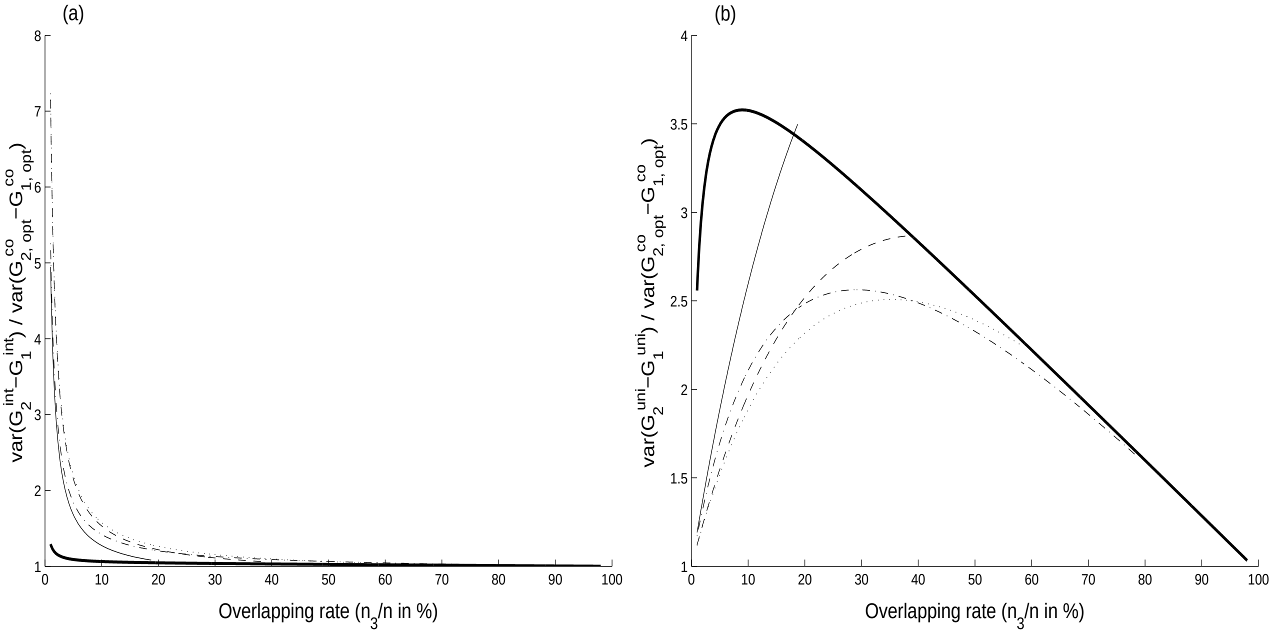

Figure 2: Gini example. (a) Ratio of the asymptotic variances of the intersection estimator and the optimal estimator and (b) ratio of the asymptotic variances of the union estimator and the optimal estimator as functions of the overlapping rate, for different sample sizes n_{1*} (heavy solid line for $n_{1*} = 10$, dotted dashed line for $n_{1*} = 210$, dotted line for $n_{1*} = 410$, dashed line for $n_{1*} = 610$, light solid line for $n_{1*} = 810$).

Acknowledgement

We would like to thank Hervé Cardot and Guillaume Chauvet for useful discussions, together with anonymous referees for remarks and comments that have improved the presentation and the contents of the paper.

	RB (%) for $\hat{\Delta}G^{\mathrm{co}}_{\mathrm{opt}}$			RRMSE (%) for $\hat{\Delta}G^{\mathrm{co}}_{\mathrm{opt}}$			RD (%) for AV$(\hat{\Delta}G^{\mathrm{co}}_{\mathrm{opt}})$			RB (%) for A$\hat{\mathrm{V}}$$(\hat{\Delta}G^{\mathrm{co}}_{\mathrm{opt}})$		
	n_3			n_3			n_3			n_3		
n_{1^*}	500	1000	3000	500	1000	3000	500	1000	3000	500	1000	3000
100	3.01	1.38	-0.55	0.61	0.42	0.23	-3.27	-0.25	0.18	-5.27	-1.02	-0.63
300	-4.23	-1.57	0.40	0.58	0.42	0.23	1.39	-0.25	-0.75	-4.37	-2.35	-0.79

Table 1: Gini example. Relative biases (RB), relative root mean squared errors (RRMSE) and relative differences (RD), as percentages, for different values of $n_{1^*} = n_{2^*}$ and n_3.

Appendix

Technical details

Proof 7 (of Theorem 1) *Let $t \in \mathcal{T}$. From Assumptions 5 and 6, we have that $N^{-\beta}T(M) = T\,(M/N) < \infty$. Following [8], let us provide the spaces (R^{p_t}, M_t) with metrics d_t, satisfying $d_t\,(Q_t/N, M_t/N) \to 0$ if and only if $N^{-1}\{\int \mathcal{Z}_t dQ_t(z) - \int \mathcal{Z}_t dM_t(z)\} \to 0$ for any variable of interest $\mathcal{Z}_t$, defined on R^{p_t}. In this way, studying the distance d_t between the Horvitz-Thompson measure $\hat{M}_t$ and the true unknown M_t is equivalent to studying the distance between the Horvitz-Thompson estimator for the population total of a variable of interest, $\sum_{k\in s} z_{k,t}/\pi^t_k = \int \mathcal{Z}_t d\hat{M}_t(z)$, and the true unknown total, $\sum_{k=1}^N z_{k,t} = \int \mathcal{Z}_t dM_t(z)$. We also consider a metric $\tilde{d}$ for the vectors $(\hat{M}/N, M/N)$ associated with the distances d_t. From Assumption 4, we have that $d_t(\hat{M}_t/N, M_t/N) = O_p(n_t^{-1/2})$ and Assumption 1 gives us that $\tilde{d}(\hat{M}/N, M/N) = O_p(n^{-1/2})$. Using a three-variate [29] expansion and the fact that T is Fréchet differentiable, see Huber (1981, p. 35), we have*

$$N^{-\beta}\{T(\hat{M}) - T(M)\} = \sum_{t=1}^{3} \int I_{tT}\left(\frac{M}{N}; z\right) d\left(\frac{\hat{M}_t}{N} - \frac{M_t}{N}\right)(z) + o\{\tilde{d}(\hat{M}/N, M/N)\},$$

where $I_{tT}(M; z)$ are the partial influence functions defined by (3). Finally, because the remainder term is $o_p(n^{-1/2})$ and the partial Fréchet derivatives are linear, Assumption 5 implies that $I_{tT}\,(M/N; z) = N^{-\beta+1} I_{tT}(M; z)$.

Proof 8 (of Theorem 3) *The variance var$\hat{Z}_{(a,b)}$, given by (11), is estimated unbiasedly by the Horvitz-Thompson variance estimator,*

$$\hat{\mathrm{var}}_{\mathrm{HT}}\hat{Z}_{(a,b)} = \sum_{t,t'\in\mathcal{T}} \sum_{k=1}^{N} \sum_{l=1}^{N} c^{t,t'}_{kl} u_{k,t} u_{l,t'}.$$

Since the linearized variables are unknown, the proposed estimator is

$$\hat{\mathrm{var}}\hat{Z}_{(a,b)} = \sum_{t,t'\in\mathcal{T}} \sum_{k=1}^{N} \sum_{l=1}^{N} c^{t,t'}_{kl} \hat{u}_{k,t} \hat{u}_{l,t'}$$

where $\hat{u}_{k,t} = \hat{u}^{\mathrm{int}}_{k,t}$ and $c^{t,t'}_{kl}$ depends on inclusion probabilities and sample membership indicators for $t, t' \in \mathcal{T} = \{1, 2, 3\}$. For any $t, t' \in \mathcal{T}$, we make the following assumptions.

Assumption A1. We assume that $N^{1-\beta}(\hat{u}_{k,t} - u_{k,t}) = o_p(1)$ and $N^{1-\beta}u_{k,t} = O(1)$ uniformly in k, *Assumption A2. We assume that $c_{kl}^{t,t'} = O(n^{-1})$ if $k \neq l$ and $c_{kl}^{t,t'} = O(1)$ if $k = l$ uniformly in k, l. We assume also that the Horvitz-Thompson variance estimators with true linearized variables are design-consistent for the Horvitz-Thompson variance terms.*

We show that $nN^{-2\beta}\{\hat{\text{var}}(\hat{Z}_{(a,b)}) - \text{var}(\hat{Z}_{(a,b)})\} = o_p(1)$, since $\hat{\text{var}}\{T(\hat{M}^{co})\} = \hat{var}(Z_{(a,b)})$ and, from Assumptions 1 to 7, $\text{AV}(T\{\hat{M}^{co}\}) = \text{var}(\hat{Z}_{(a,b)})$ with $\text{var}(\hat{Z}_{(a,b)})$ given by (11). The proofs of convergence are similar for the different variance and covariance terms of the sum in (11) and we concentrate on the first term, proving that $nN^{-2\beta}\{\hat{var}(A) - var(A)\} = o_p(1)$. We have $\hat{var}(A) - var(A) = \hat{var}(A) - \hat{var}_{\text{HT}}(A) + \hat{var}_{\text{HT}}(A) - var(A)$ with $\hat{var}_{\text{HT}}(A) = \sum_{k=1}^{N}\sum_{l=1}^{N} c_{kl}^{1,1} u_{k,1}u_{l,1}$. By Assumption A2, we have $nN^{-2\beta}\{\hat{var}_{\text{HT}}(A) - var(A)\} = o_p(1)$. As a result,

$$\hat{var}(A) - \hat{var}_{\text{HT}}(A) = \sum_{k=1}^{N}\sum_{l=1}^{N} c_{kl}^{1,1}(\hat{u}_{k,1} - u_{k,1})(\hat{u}_{l,1} - u_{l,1}) + 2\sum_{k=1}^{N}\sum_{l=1}^{N} c_{kl}^{1,1}(\hat{u}_{k,1} - u_{k,1})u_{l,1}$$

and we have that $nN^{-2\beta}\{\hat{var}(A) - \hat{var}_{\text{HT}}(A)\} = o_p(1)$ by Assumptions A1 and A2. The reader is referred to [4] for conditions under which Assumption A2 is available.

Proof 9 (of Corollary 1) *Part* (i). *The derivative of $\hat{\text{AV}}\{T(\hat{M}^{\text{co}})\}$ with respect to θ is equal to $2\hat{\Gamma}\theta + 2\hat{\gamma}$, which vanishes for $\theta = -\hat{\Gamma}^{-1}\hat{\gamma}$ assuming that $\hat{\Gamma}$ is non-singular.*

Part (ii). Following the same reasoning as in [11], we have that

$$\hat{Z}_{(\hat{a}_{\text{opt}},\hat{b}_{\text{opt}})} = (\hat{\theta}_{\text{opt}} - \theta_{\text{opt}})' \begin{pmatrix} \hat{t}_{u_1}^{1*} - \hat{t}_{u_1}^{3} \\ \hat{t}_{u_2}^{2*} - \hat{t}_{u_2}^{3} \end{pmatrix} + \theta'_{\text{opt}} \begin{pmatrix} \hat{t}_{u_1}^{1*} - \hat{t}_{u_1}^{3} \\ \hat{t}_{u_2}^{2*} - \hat{t}_{u_2}^{3} \end{pmatrix} + \sum_{t=1}^{3} \hat{t}_{u_t}^{3}.$$

Thus, $\sqrt{n}N^{-\beta}(\hat{Z}_{(\hat{a}_{\text{opt}},\hat{b}_{\text{opt}})} - Z) = \sqrt{n}N^{-\beta}(\hat{Z}_{(a_{\text{opt}},b_{\text{opt}})} - Z) + o_p(1)$ since $\hat{\theta}_{\text{opt}}$ is consistent for θ_{opt} and $\sqrt{n}N^{-\beta}(\hat{t}_{u_1}^{1*} - \hat{t}_{u_1}^{3}, \hat{t}_{u_2}^{2*} - \hat{t}_{u_2}^{3})'$ is bounded in probability by Assumption 4 and the fact that $\sum_{k=1}^{N} u_{k,t}$ is of degree β. This completes the proof.

Part (iii). From the proof of Theorem 1, we have that the reminder of the von-Mises expansion of $T(\tilde{M}_{\text{opt}}^{\text{co}})$ is $o\{\tilde{d}(\tilde{M}_{\text{opt}}^{\text{co}}/N, M/N)\}$. Assumptions 1 and 4 and the consistency of $\hat{\theta}_{\text{opt}}$ imply that the remainder is of order $o_p(n^{-1/2})$. Following the proof of (ii), we have

$$\frac{\sqrt{n}}{N^{\beta}}\{T(\tilde{M}_{\text{opt}}^{\text{co}}) - T(M)\} = \frac{\sqrt{n}}{N^{\beta}}(\hat{Z}_{(\hat{a}_{\text{opt}},\hat{b}_{\text{opt}})} - Z) + o_p(1) = \frac{\sqrt{n}}{N^{\beta}}(\hat{Z}_{(a_{\text{opt}},b_{\text{opt}})} - Z) + o_p(1)$$

and, as a consequence, the asymptotic variance of $T(\tilde{M}_{\text{opt}}^{\text{co}})$ is equal to the variance of $\hat{Z}_{(a_{\text{opt}},b_{\text{opt}})}$.

Proof 10 (of Theorem 5) *We have var$(\hat{t}_{u_t}^{d}) = N^2 n_d^{-1}(1 - n_d/N)S_{u_t}^2$ for $d \in \{1*, 3\}$ if $t = 1$ and $d \in \{2*, 3\}$ if $t = 2$. The covariance terms become cov$(\hat{t}_{u_1}^{1*}, \hat{t}_{u_1}^{3}) = -NS_{u_1}^2$, cov$(\hat{t}_{u_2}^{2*}, \hat{t}_{u_2}^{3}) = -NS_{u_2}^2$, cov$(\hat{t}_{u_1}^{3}, \hat{t}_{u_2}^{3}) = Nf_3^{-1}(1 - f_3)S_{u_1u_2}$ and cov$(\hat{t}_{u_1}^{1*}, \hat{t}_{u_2}^{2*}) =$ cov$(\hat{t}_{u_1}^{1*}, \hat{t}_{u_2}^{3}) =$ cov$(\hat{t}_{u_1}^{3}, \hat{t}_{u_2}^{2*}) = -NS_{u_1u_2}$. To conclude, we introduce these values in the expressions of Γ and γ given in Theorem 4.*

References

[1] P. Bell. Comparison of alternative labour force survey estimators. *Survey Methodol.*, 27:53–64, 2001.

[2] Y. Berger. Variance estimation for measures of change in probability sampling. *Can. J. Statist.*, 32:451–67, 2004.

[3] Y. Berger and C. Skinner. A jackknife variance estimator for unequal probability sampling. *J.R.S.S. B*, 67:79–89, 2005.

[4] F. J. Breidt and J. D. Opsomer. Local polynomial regression estimators in survey sampling. *Ann. Statist.*, 28:1026–53, 2000.

[5] C. Campbell. A different view of finite population estimation. In DC : American Statistical Association Washington, editor, *Proc. Survey Res. Meth. Sec. Am. Statist. Assoc.*, pages 319–24, 1980.

[6] A. J. Canty and A. C. Davison. Resampling-based variance estimation for labour force surveys. *The Statistician*, 48:379–91, 1999.

[7] A. Demnati and J. N. K. Rao. Linearization variance estimators for survey data. *Survey Methodol.*, 30:17–26, 2004.

[8] J.-C. Deville. Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodol.*, 25:193–203, 1999.

[9] A. R. Eckler. Rotation sampling. *Ann. Marh. Statist.*, 26:664–85, 1955.

[10] L. T. Fernholz. *Von Mises Calculus for Statistical Functionals*, volume 19 of *Lecture Notes in Statistics*. New York: Springer, 1983.

[11] W. Fuller. Regression estimation for survey samples. *Survey Methodol.*, 28:5–23, 2002.

[12] W. Fuller and J. N. K. Rao. A regression composite estimator with application to the canadian labour force survey. *Survey Methodol.*, 27:45–52, 2001.

[13] F. R. Hampel. The influence curve and its role in robust statistics. *J. Am. Statist. Assoc.*, 69:383–93, 1974.

[14] M. A. Hidiroglou. Double sampling. *Survey Methodol.*, 27:143–54, 2001.

[15] C. T. Isaki and W. A. Fuller. Survey design under the regression superpopulation model. *J. Am. Statist. Assoc.*, 77:89–96, 1982.

[16] R. J. Jessen. Statistical investigation of a farm survey for obtaining farm facts. *Iowa Agric. Station Res. Bul.*, 304:54–9, 1942.

[17] L. Kish. *Survey Sampling*. New York: Wiley, 1965.

[18] M. Kovačević and D. A. Binder. Variance estimation for measures of income inequality and polarization - the estimating equations approach. *J. Offic. Statist.*, 13:41–58, 1997.

[19] T. Merkouris. Cross-sectional estimation in multiple-panel household surveys. *Survey Methodol.*, 27:171–81, 2001.

[20] G. E. Montanari. Post-sampling efficient prediction in large scale surveys. *Int. Statist. Rev.*, 55:191–202, 1987.

[21] H. D. Patterson. Sampling on successive occasions with partial replacement of units. *J. R. Statist. Soc. B*, 12:241–55, 1950.

[22] A. M. Pires and J. A. Branco. Partial influence functions. *J. Mult. Anal.*, 83:451–68, 2002.

[23] D. Place. Calcul de la précision des estimations longitudinales dans l'enquête emploi en continu. In P. Guibert, D. Haziza, A. Ruiz-Gazen, and Y. Tillé, editors, *Méthodes de sondage : applications aux enquêtes longitudinales, à la santé, aux enquêtes électorales et aux enquêtes dans les pays en développement*, pages 53–7. Paris : Dunod, pp. 53-7, 2008.

[24] J. N. K. Rao, C. F. J. Wu, and K. Yue. Some recent works on resampling methods for complex surveys. *Survey Methodol.*, 18:209–17, 1992.

[25] N. Reid. Influence functions for censored data. *Ann. Statist.*, 9:78–92, 1981.

[26] C. E. Särndal, Swensson B., and J. H. Wretman. *Model Assisted Survey Sampling*. New York: Springer-Verlag, 1992.

[27] A. C. Singh, B. Kennedy, and S. Wu. Regression composite estimation for the canadian labour force survey with a rotating panel design. *Survey Methodol.*, 27:33–44, 2001.

[28] S. M. Tam. On covariances from overlapping samples. *Amer. Statistician*, 38:288–89, 1984.

[29] R. von Mises. On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.*, 18:309–48, 1947.

[30] K. Wolter. *Introduction to variance estimation*. Springer, 2007.

[31] C. Wu. Combining information from multiple surveys through the empirical likelihood method. *Can. J. Statist.*, 32:15–26, 2003.