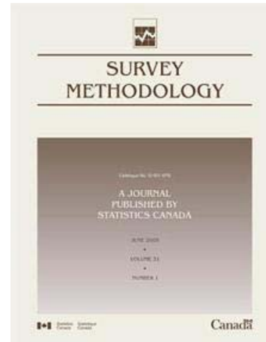
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Estimation in surveys using conditional inclusion probabilities: Complex design



by Yves Tillé

June 1999

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 1999
Vol. 25, No. 1, pp. 57-66
Statistics Canada, Catalogue No. 12-001

Estimation in surveys using conditional inclusion probabilities: Complex design

Yves Tillé [1]

Abstract

This paper investigates a repeated sampling approach to take into account auxiliary information in order to improve the precision of estimators. The objective is to build an estimator with a small conditional bias by weighting the observed values by the inverses of the conditional inclusion probabilities. A general approximation is proposed in cases when the auxiliary statistic is a vector of Horvitz-Thompson estimators. This approximation is quite close to the optimal estimator discussed by Fuller and Isaki (1981), Montanari (1987, 1997), Deville (1992) and Rao (1994, 1997). Next, the optimal estimator is applied to a stratified sampling design and it is shown that the optimal estimator can be viewed as an generalised regression estimator for which the stratification indicator variables are also used at the estimation stage. Finally, the application field of this estimator is discussed in the general context of the use of auxiliary information.

Key Words: Conditional estimation; Weighted observation; Generalised regression estimator; Complex survey.

1. Introduction

At the estimation stage, practitioners of survey sampling often have auxiliary information available. This information can be the knowledge of a set of population means or totals. Sometimes, the available information is detailed, for instance, when the values taken by a variable on all the units of the population are known. This information can be used to improve the precision of the estimators.

Our aim is to dealt with the use of auxiliary information based on a conditional principle. Conditional inference has been largely studied in the survey sampling literature. Indeed, the optimal estimator was discussed by Fuller and Isaki (1981), Montanari (1987, 1997), Deville (1992) and Rao (1994, 1997). The conditional properties of the post-stratified estimators has been studied by Casady and Valliant (1993). In an earlier paper (Tillé 1998), a general technique that allows to build a mean or total estimator that has a small conditional bias has been proposed for simple random sampling. This technique is based on the use of conditional inclusion probabilities and allows one to take into account auxiliary information without any reference to a super-population model.

In this paper the use of conditional inclusion probabilities is generalised to any sampling design. It is shown that this technique allows to construct an estimator very similar to the optimal estimator discussed by Montanari (1987), Deville (1992) and Rao (1994). This family of estimators provides a valid conditional inference and can also be viewed as the optimal linear estimator. Next, these estimators are applied in the stratification case and are compared to the GREG-estimator. The GREG-estimator is generally conditionally biased. Nevertheless, it is shown that, in regression, the optimal estimator is a particularly case of the GREG-estimator. Indeed, when the stratification variables are re-used as auxiliary variables in the GREG-estimator, it is equal to the optimal estimator. Next, a set of simulations is given that shows the interest of the optimal estimator in stratification. The gain of precision can be very important when the stratification variables are very correlated to the interest variable. Finally we discuss the general estimation problem in survey sampling that can be viewed as a third-order problem where three sets of variables interact: the planning variables, the calibration variables and the interest variables.

The paper is organised as follows. In section 2 the notation is defined. In section 3, the problem of conditional inference is presented. In section 4, an approximation of the SCW-estimator is given for complex designs under technical hypotheses. These hypotheses are discussed in section 5. In section 6 the optimal estimator and the SCW-estimator are compared to the generalised regression (GREG) estimator in the stratification framework. It is shown that the optimal estimator can be viewed as a GREG-estimator for which the stratification indicator variables are also used *a posteriori*. Next a set of simulations is presented in section 7 in order to compare the discussed estimators. Finally, the problem of interaction between the design and the auxiliary variables is discussed in section 8.

2. Problem and notation

Consider a finite population $U = \{1, \ldots, k, \ldots, N\}$ and suppose that a random sample S is drawn without replacement from this population following a sampling design $p(.)$. The probability of selecting the sample s is $\Pr(S = s) = p(s)$,

1. Yves Tillé, CREST–ENSAI, École Nationale de la Statistique et de l'Analyse de l'Information, rue Blaise Pascal, Campus de Ker Lann, 35170 Bruz, France. E-mail: tille@ensai.fr.

for all $s \subset U$. The indicator variables I_k take the value 1 if unit k is in the sample and 0 otherwise, for all $k \in U$. The inclusion probability of unit k is $\pi_k = E(I_k)$, where symbol $E(.)$ is the expectation with respect to the sampling design. The joint inclusion probability for unit k and l is $\pi_{kl} = E(I_k I_l)$. Let y_k denote the value of the variable y for the k^{th} unit of the population. The aim is to estimate the population mean of y:

$$\bar{y} = \frac{1}{N} \sum_{k \in U} y_k.$$

If $\pi_k > 0$, for all $k \in U$ the Horvitz-Thompson estimator (1952) given by

$$\hat{\bar{y}} = \frac{1}{N} \sum_{k \in S} \frac{y_k}{\pi_k}$$

provides an unbiased estimator of $\bar{y}$.

Let T be a statistic. The objective is to estimate $\bar{y}$ with a conditional bias as small as possible with respect to statistic T. Define the first-order conditional inclusion probabilities to be $\pi_{k|T} = E(I_k \mid T)$ for all $k \in U$ and the conditional joint inclusion probabilities to be $\pi_{kl|T} = E(I_k I_l | T)$ for all $k \in U, l \in U, k \neq l$. The simple conditionally weighted estimator (SCW) is defined by

$$\hat{\bar{y}}_{|T} = \frac{1}{N} \sum_{k \in S} \frac{y_k}{\pi_{k|T}}. \tag{2.1}$$

This estimator is not exactly conditionally unbiased. Indeed, a conditionally unbiased estimator exists if and only if $\pi_{k|T} > 0$ for all $k \in U$. For this reason, it is useful to enlarge the definition of conditional unbiasedness: an estimator is said to be virtually conditionally unbiased (VCU) if the conditional bias only depends on the units having null conditional inclusion probabilities. The SCW-estimator is VCU, indeed:

$$B(\hat{\bar{y}}_{|T} | T) = E(\hat{\bar{y}}_{|T} | T) - \bar{y} = -\frac{1}{N} \sum_{k | \pi_{k|T} = 0} y_k.$$

This estimator generalises some classic results (see Tillé 1998) like post-stratification. Moreover, it allows us to build an original estimator for a contingency table when the population marginal totals are known. Unfortunately, the computation of the $\pi_{k|T}$ becomes very difficult in complex sampling designs. A general approximation for the SCW-estimator will however be given when using a vector of Horvitz-Thompson estimators as auxiliary statistic.

3. Use of a complex auxiliary statistic

Suppose that the auxiliary information is represented by the vector $\mathbf{x}_k = (x_{k1}, ..., x_{kj}, ..., x_{kJ})'$ of values taken by the J auxiliary variables on the k^{th} unit of U. In a first step, it is supposed that the $\mathbf{x}_k$ are known for each unit of the population. Later, it will be considered the more restrictive case where only one function of the $\mathbf{x}_k$ such as

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{k \in U} \mathbf{x}_k$$

is known. Consider also the Horvitz-Thompson estimator of $\bar{\mathbf{x}}$ given by

$$\hat{\bar{\mathbf{x}}}_\pi = \frac{1}{N} \sum_{k \in S} \frac{\mathbf{x}_k}{\pi_k}.$$

If $\pi_k > 0$, for all $k \in U$, $\hat{\bar{\mathbf{x}}}_\pi$ is an unbiased estimator of $\bar{\mathbf{x}}$

$$E(\hat{\bar{\mathbf{x}}}_\pi) = \bar{\mathbf{x}}. \tag{3.2}$$

The variance of $\hat{\bar{\mathbf{x}}}_\pi$ is given by

$$\begin{aligned}\mathbf{\Sigma} = \mathrm{Var}(\hat{\bar{\mathbf{x}}}_\pi) &= \frac{1}{N^2} \sum_{l \in U} \mathbf{x}_l \frac{\mathbf{x}_{l'}}{\pi_l}(1 - \pi_l) \\ &+ \frac{1}{N^2} \sum_{l \in U} \sum_{\substack{m \in U \\ m \neq l}} \frac{\mathbf{x}_l \mathbf{x}'_m}{\pi_l \pi_m}(\pi_{lm} - \pi_l \pi_m).\end{aligned} \tag{3.3}$$

Suppose now that vector $(\hat{\bar{y}}_\pi \; \hat{\bar{\mathbf{x}}}'_\pi)'$ has a multinormal distribution. Under this hypothesis, it can be derived a conditional unbiased estimator (see for instance, Deville 1992). First the conditional bias is computed:

$$\begin{aligned}B(\hat{\bar{y}}_\pi | \hat{\bar{\mathbf{x}}}_\pi) &= E(\hat{\bar{y}}_\pi | \hat{\bar{\mathbf{x}}}_\pi) - \bar{y} \\ &= (\hat{\bar{\mathbf{x}}}_\pi - \bar{\mathbf{x}}) \mathrm{Var}(\hat{\bar{\mathbf{x}}}_\pi)^{-1} \mathrm{Cov}(\hat{\bar{\mathbf{x}}}_\pi, \hat{\bar{y}}_\pi).\end{aligned}$$

If an estimator of $B(\hat{\bar{y}}_\pi | \hat{\bar{\mathbf{x}}}_\pi)$ is available, the Horvitz-Thompson estimator can be corrected in the following way:

$$\begin{aligned}\hat{\bar{y}}_c &= \hat{\bar{y}}_\pi - \hat{B}(\hat{\bar{y}}_\pi | \hat{\bar{\mathbf{x}}}_\pi) \\ &= \hat{\bar{y}}_\pi + (\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}}_\pi) \widehat{\mathrm{Var}}(\hat{\bar{\mathbf{x}}}_\pi)^{-1} \widehat{\mathrm{Cov}}(\hat{\bar{\mathbf{x}}}_\pi, \hat{\bar{y}}_\pi).\end{aligned}$$

This estimator is related to the optimal linear estimator discussed by Fuller and Isaki (1981), Montanari (1987) and Rao (1994). Indeed, Montanari showed that the best estimator in the sense of the smallest mean square error (MSE) of the form

$$\hat{\bar{y}}_\beta = \hat{\bar{y}}_\pi + (\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}}_\pi)' \beta \tag{3.4}$$

occurs when β takes the value:

$$\beta_{\mathrm{OPT}} = \mathbf{\Sigma}^{-1} \mathrm{Cov}(\hat{\bar{\mathbf{x}}}_\pi, \hat{\bar{y}}_\pi).$$

The optimal linear estimator presented by Montanari leads thus to a very similar result to the conditional approach, although Montanari did not start with a conditional point of view. In Montanari's approach, the optimal estimator is found in a class of linear estimators defined by (3.4) without any reference to conditional properties. Nevertheless, Rao (1994) has pointed out that this estimator leads to valid conditional inference. The general problem of the optimal estimator is that β_{OPT} is not known and must thus be estimated. By estimating β_{OPT}, the optimal properties of the estimator are lost.

In order to estimate β_{OPT}, (or $B(\hat{\bar{y}}_\pi | \hat{\bar{\mathbf{x}}}_\pi)$) two cases can be distinguished. In the first one, the values taken by the auxiliary variable on all the units of the population are known. In this case, Σ is thus known and

$$\mathrm{Cov}(\hat{\bar{\mathbf{x}}}_\pi, \hat{\bar{y}}_\pi) = \frac{1}{N^2}\sum_{k\in U}\frac{\mathbf{x}_k y_k}{\pi_k}(1-\pi_k) + \frac{1}{N^2}\sum_{k\in U}\sum_{\substack{l\in U\\ l\neq k}}\frac{\mathbf{x}_k y_l}{\pi_k\pi_l}(\pi_{kl}-\pi_k\pi_l)$$

$$= \frac{1}{N}\sum_{k\in U}(\bar{\mathbf{x}}_{|k} - \bar{\mathbf{x}})\, y_k$$

can be unbiasedly estimated by

$$\widehat{\mathrm{Cov}}_1\,(\hat{\bar{\mathbf{x}}}_\pi, \hat{\bar{y}}_\pi) = \frac{1}{N}\sum_{k\in S}\frac{(\bar{\mathbf{x}}_{|k} - \bar{\mathbf{x}})\, y_k}{\pi_k} \tag{3.5}$$

where

$$\bar{\mathbf{x}}_{|k} = E(\hat{\bar{\mathbf{x}}}_\pi | k\in S) = \frac{1}{N}\sum_{\substack{l\in U\\ l\neq k}}\frac{\mathbf{x}_l\pi_{kl}}{\pi_k\pi_l} + \frac{\mathbf{x}_k}{\pi_k N}. \tag{3.6}$$

By using (3.5), a first asymptotically optimal estimator can be constructed

$$\hat{\bar{y}}_{\mathrm{AOPT1}} = \hat{\bar{y}}_\pi + (\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}}_\pi)'\,\Sigma^{-1}\frac{1}{N}\sum_{k\in S}\frac{\bar{\mathbf{x}}_{|k} - \bar{\mathbf{x}}}{\pi_k}y_k. \tag{3.7}$$

In the second case, only the population mean $\bar{\mathbf{x}}$ is known, Σ must thus be estimated and $\mathrm{Cov}(\hat{\bar{\mathbf{x}}}_\pi, \hat{\bar{y}}_\pi)$ can not be estimated using (3.5). Montanari proposes to estimate Σ and $\mathrm{Cov}(\hat{\bar{\mathbf{x}}}_\pi, \hat{\bar{y}}_\pi)$ by the classic Horvitz-Thompson estimator:

$$\hat{\Sigma} = \frac{1}{N^2}\sum_{k\in S}\frac{\mathbf{x}_k\mathbf{x}_k'}{\pi_k^2}(1-\pi_k) + \frac{1}{N^2}\sum_{k\in S}\sum_{\substack{l\in S\\ l\neq k}}\frac{\mathbf{x}_k\mathbf{x}_l'}{\pi_k\pi_l}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_{kl}}$$

and

$$\widehat{\mathrm{Cov}}_2\,(\hat{\bar{\mathbf{x}}}_\pi, \hat{\bar{y}}_\pi) = \frac{1}{N^2}\sum_{k\in S}\frac{\mathbf{x}_k y_k}{\pi_k^2}(1-\pi_k) + \frac{1}{N^2}\sum_{k\in S}\sum_{\substack{l\in S\\ l\neq k}}\frac{\mathbf{x}_k y_l}{\pi_k\pi_l}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_{kl}}.$$

β_{OPT} is thus estimated by ($\hat{\Sigma}$ is supposed non singular)

$$\hat{\beta}_{\mathrm{OPT}} = \hat{\Sigma}^{-1}\widehat{\mathrm{Cov}}_2\,(\hat{\bar{\mathbf{x}}}_\pi, \hat{\bar{y}}_\pi).$$

By estimating β_{OPT}, another asymptotically optimal (AOPT) estimator can be given:

$$\hat{\bar{y}}_{\mathrm{AOPT2}} = \hat{\bar{y}}_\pi + (\hat{\bar{\mathbf{x}}}_\pi - \bar{\mathbf{x}})'\hat{\beta}_{\mathrm{OPT}}. \tag{3.8}$$

The difference between the AOPT1 and AOPT2 estimator is the way we estimate $\mathrm{Cov}(\hat{\bar{\mathbf{x}}}_\pi, \hat{\bar{y}}_\pi)$ and Σ. However, the AOPT1-estimator needs more complete auxiliary information.

The generalised regression (GREG) estimator defined by Cassel, Särndal and Wretman (1976), Wright (1983), Särndal, Swensson and Wretman (1992, page 225) is also an estimator of the linear class given by expression (3.4). For the GREG-estimator β is defined by

$$\beta_{\mathrm{GREG}} = \left[\sum_{k\in U}\frac{\mathbf{x}_k\mathbf{x}_k'}{c_k}\right]^{-1}\sum_{k\in U}\frac{\mathbf{x}_k y_k}{c_k}$$

and can be estimated by

$$\hat{\beta}_{\mathrm{GREG}} = \left[\sum_{k\in S}\frac{\mathbf{x}_k\mathbf{x}_k'}{\pi_k c_k}\right]^{-1}\sum_{k\in S}\frac{\mathbf{x}_k y_k}{\pi_k c_k},$$

where quantities $c_k > 0$, $k\in U$, are weights defined for all the population units. The GREG-estimator does not have good conditional properties. It is generally conditionally biased (Rao 1994).

4. Approximation of the SCW-estimator

Another way to construct a conditionally unbiased estimator is to find an approximation of the SCW-estimator given in (2.1). Indeed this estimator has good unbiasedness properties because it is VCU. If $\hat{\bar{\mathbf{x}}}_\pi$ is used as an auxiliary statistic, we shall seek an approximation of

$$\pi_{k|T} = E(I_k | \hat{\bar{\mathbf{x}}}_\pi).$$

If the random vector $\hat{\bar{\mathbf{x}}}_\pi$ takes for instance the value $\mathbf{z}$, we get by Bayes's theorem that

$$E(I_k | \hat{\bar{\mathbf{x}}}_\pi = \mathbf{z}) = \Pr(k\in S | \hat{\bar{\mathbf{x}}}_\pi = \mathbf{z}) = \pi_k\frac{\Pr(\hat{\bar{\mathbf{x}}}_\pi = \mathbf{z} | k\in S)}{\Pr(\hat{\bar{\mathbf{x}}}_\pi = \mathbf{z})}.$$

In order to compute the conditional inclusion probabilities, it is thus necessary to know the probability distribution of $\hat{\bar{\mathbf{x}}}_\pi$ unconditionally and conditionally on the presence of each unit in the sample. Except for some particular case, this probability distribution is very complex; for this reason an approximation will be constructed.

It is possible to derive the means and variances of $\hat{\hat{\mathbf{x}}}_\pi$ unconditionally and conditionally on the presence of each unit in the sample. Indeed $E(\hat{\hat{\mathbf{x}}}_\pi)$ is given in (3.2), $\operatorname{Var}(\hat{\hat{\mathbf{x}}}_\pi)$ in (3.3), $E(\hat{\hat{\mathbf{x}}}_\pi | k \in S)$ in (3.6), and

$$\begin{aligned}
\boldsymbol{\Sigma}_k &= \operatorname{Var}(\hat{\hat{\mathbf{x}}}_\pi | k \in S) \\
&= \frac{1}{N^2} \sum_{\substack{l \in U \\ l \neq k}} \frac{\mathbf{x}_l \mathbf{x}_l' \pi_{kl}}{\pi_k \pi_l^2} \left(1 - \frac{\pi_{kl}}{\pi_k}\right) \\
&+ \frac{1}{N^2} \sum_{\substack{l \in U \\ l \neq k}} \sum_{\substack{m \in U \\ m \neq l \\ m \neq k}} \frac{\mathbf{x}_l \mathbf{x}_m'}{\pi_k \pi_l \pi_m} \left(\pi_{klm} - \frac{\pi_{kl}\pi_{km}}{\pi_k}\right) \qquad (4.9)
\end{aligned}$$

where π_{klm} is the third-order inclusion probability. Matrixes $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_k$ are assumed to be non singular.

As the probability distribution of $\hat{\hat{\mathbf{x}}}_\pi$ is generally unknown, the following three assumptions will be used to construct an approximation of conditional inclusion probabilities.

(i) If the sample size n is large, $\hat{\hat{\mathbf{x}}}_\pi$ has a multivariate normal distribution unconditionally and conditionally on the presence of each unit in the sample.

(ii) $\mathbf{R}_k^{-1} - \mathbf{R}^{-1} = O_{J \times J}(n^{-1})$ for all $k \in U$ where $\mathbf{R} = \mathbf{V}^{-1/2} \boldsymbol{\Sigma} \mathbf{V}^{-1/2}$, $\mathbf{R}_k = \mathbf{V}^{-1/2} \boldsymbol{\Sigma}_k \mathbf{V}^{-1/2}$, $\mathbf{V}$ denotes a $J \times J$ diagonal matrix having the elements of the diagonal of $\boldsymbol{\Sigma}$ on its diagonal and $O_{J \times J}(n^{-\alpha})$ denotes a matrix of quantities that when multiplied by n^α remains bounded as $n \to \infty$.

(iii) $\gamma_k = \mathbf{V}^{-1/2}(\overline{\mathbf{x}}_{|k} - \overline{\mathbf{x}}) = O_J(n^{-1/2})$ where $O_J(n^{-\alpha})$ denotes a vector of quantities that when multiplied by n^α remains bounded as $n \to \infty$.

These three hypotheses are made on the sample size. It is thus supposed that when n increases, N increases at least as quickly as n. Nevertheless, no hypothesis are made on $f = n/N$. Assuming that the hypotheses given in section 3 are verified, the following result gives an approximation of the SCW-estimator:

Result 1: Assuming (i), (ii) and (iii), and if the auxiliary statistic used is $\hat{\hat{\mathbf{x}}}_\pi$, then

$$\begin{aligned}
\hat{\hat{y}}_{|T} &= \hat{\hat{y}}_\pi + (\overline{\mathbf{x}} - \hat{\hat{\mathbf{x}}}_\pi)' \boldsymbol{\Sigma}^{-1} \frac{1}{N} \sum_{k \in S} \frac{\overline{\mathbf{x}}_{|k} - \overline{\mathbf{x}}}{\pi_k} y_k \\
&+ O_p(n^{-1}) \approx \hat{\hat{y}}_{\text{AOPT1}}. \qquad (4.10)
\end{aligned}$$

where $n \times O_p(n^{-1})$ is a quantity bounded in probability. Proof of Result 1 is given in the appendix.

5. Discussion about the hypotheses

These three hypotheses are verified for simple random sampling without replacement when only one auxiliary variable is available. Indeed, in this case, we have $J = 1$, $\mathbf{x}_k = x_k$, $\overline{\mathbf{x}} = \overline{x}$, $\overline{\mathbf{x}}_{|k} = \overline{x}_{|k}$ $\hat{\hat{\mathbf{x}}}_\pi = \hat{\hat{x}}_\pi$. We get

$$\pi_k = \frac{n}{N}, \ \pi_{kl} = \frac{n}{N}\frac{n-1}{N-1} \quad \text{and} \quad \pi_{klm} = \frac{n}{N}\frac{n-1}{N-1}\frac{n-2}{N-2}.$$

By (3.6), (3.3), (4.9),

$$\overline{x}_{|k} = \overline{x} + \frac{N-n}{N-1}\frac{x_k - \overline{x}}{n}, \qquad (5.11)$$

$$\operatorname{Var}(\hat{\hat{x}}_\pi) = \frac{N-n}{N-1}\frac{\sigma_x^2}{n}, \qquad (5.12)$$

$$\operatorname{Var}(\hat{\hat{x}}_\pi | k \in S) = \frac{N(N-n)(n-1)}{(N-2)(N-1)n^2}\left\{\sigma_x^2 - \frac{(x_k - \overline{x})^2}{N-1}\right\} \qquad (5.13)$$

where

$$\sigma_x^2 = \frac{1}{N}\sum_{k \in U}(x_k - \overline{x})^2.$$

Now, consider the three hypotheses for this particular case.

- Hypothesis (i) was proved by Madow (1948) under some conditions.
- Hypothesis (ii) becomes

$$\frac{\operatorname{Var}(\hat{\hat{x}}_\pi | k \in S)}{\operatorname{Var}(\hat{\hat{x}}_\pi)} - 1 = O(n^{-1}).$$

By (5.12) and (5.13), we get

$$\begin{aligned}
&\frac{\operatorname{Var}(\hat{\hat{x}}_\pi | k \in S)}{\operatorname{Var}(\hat{\hat{x}}_\pi)} \\
&= \frac{N(n-1)}{(N-2)n}\left\{1 - \frac{(x_k - \overline{x})^2}{(N-1)\sigma_x^2}\right\} \\
&= 1 - \frac{1}{n}\left\{\frac{N-2n}{(N-2)} + \frac{N(n-1)}{(N-2)}\frac{(x_k - \overline{x})^2}{(N-1)\sigma_x^2}\right\} \\
&= 1 + O\left(\frac{1}{n}\right).
\end{aligned}$$

- Hypothesis (iii) becomes

$$\frac{\overline{x}_{|k} - \overline{x}}{\sqrt{\operatorname{Var}(\hat{\hat{x}}_\pi)}} = O(n^{-1/2}).$$

By (5.11) (5.12), we get

$$\gamma_k = \frac{\bar{x}_{|k} - \bar{x}}{\sqrt{\mathrm{Var}(\hat{\bar{x}}_\pi)}} = \sqrt{\frac{N-n}{N-1}}\,\frac{x_k - \bar{x}}{\sigma_x \sqrt{n}} = O\left(\frac{1}{\sqrt{n}}\right).$$

In simple random sampling, these hypotheses can better be interpreted. Hypothesis (i) is the classic assumption of normality that was also needed for the construction of the optimal estimator. In simple random sampling, it is easy to verify that Hypothesis (iii) implies hypothesis (ii) both technical hypotheses simply imply that a particular unit cannot take a $|x_k - \bar{x}|$ value much more important than the other ones.

The three hypotheses are thus valid under simple random sampling when only one variable is available. This result can also be extended to stratified sampling when the number of strata is fixed and the sample size within each stratum is large. In cluster sampling, when the number of clusters is large and the clusters are selected with a simple random sampling design, these hypotheses are still applicable. Hypothesis (i) was also partially showed by Rosén (1972) for sampling with unequal probabilities. Actually, the proof of Rosén is restricted to a rejective sampling design.

The proposed hypotheses are generally less restrictive than a superpopulation model. Indeed, a superpopulation model is a set of hypotheses on the interest variables while the three hypotheses presented only affect the auxiliary variables. In a superpopulation model, the relation between the interest variable and the auxiliary variables are the most extensive contribution of the model. In the conditional approach, no hypothesis is made on the interest variable. If the hypotheses presented are debatable, it is thus clear that a superpopulation model is a set of hypotheses much more restrictive than those used in the conditional approach.

6. Application to stratified sampling

6.1 The problem

In stratification, auxiliary information is used it *a priori* to improve the estimation. In this case, three sets of variables interact: the stratification variables, the auxiliary variables used *a posteriori* and the interest variable. Suppose that the population is partitioned into H strata $U_h, h = 1, ..., H,$ of size $N_h, h = 1, ..., H.$ The population means of the strata are denoted $\bar{y}_h = N_h^{-1} \sum_{k \in U_h} y_k$ and $\bar{\mathbf{x}}_h = N_h^{-1} \sum_{k \in U_h} \mathbf{x}_k$. A simple random sample S_h of fixed size n_h $(\sum_{h=1}^{H} n_h = n)$ is selected without replacement independently in each stratum. From the general theory of stratification (see for instance Särndal, Swensson and Wretman 1992, page 100), we get

$$\hat{\bar{y}}_\pi = \frac{1}{N}\sum_{h=1}^{H} N_h \hat{\bar{y}}_h \quad \text{and} \quad \hat{\bar{\mathbf{x}}}_\pi = \frac{1}{N}\sum_{h=1}^{H} N_h \hat{\bar{\mathbf{x}}}_h$$

where

$$\hat{\bar{y}}_h = \frac{1}{n_h}\sum_{k \in S_h} y_k \quad \text{and} \quad \hat{\bar{\mathbf{x}}}_h = \frac{1}{n_h}\sum_{k \in S_h} \mathbf{x}_k.$$

Moreover, we have that

$$\mathrm{Cov}(\hat{\bar{\mathbf{x}}}_\pi, \hat{\bar{y}}_\pi) = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 \frac{1-f_h}{n_h}\frac{1}{N_h - 1}\sum_{k \in U_h} (\mathbf{x}_k - \bar{\mathbf{x}}_h)(y_k - \bar{y}_h)'$$

and

$$\boldsymbol{\Sigma} = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2 \frac{1-f_h}{n_h}\frac{1}{N_h - 1}\sum_{k \in U_h} (\mathbf{x}_k - \bar{\mathbf{x}}_h)(\mathbf{x}_k - \bar{\mathbf{x}}_h)'$$

where $f_h = n_h / N_h,\ h = 1, ..., H.$

6.2 AOPT1-estimator

If $k \in U_\alpha$, by extending expression (3.6) to stratified sampling, we get

$$\bar{\mathbf{x}}_{|k} = E(\hat{\bar{\mathbf{x}}}_\pi | k \in S) = \bar{\mathbf{x}} + \frac{N_\alpha^2}{N(N_\alpha - 1)}\frac{1-f_\alpha}{n_\alpha}(\mathbf{x}_k - \bar{\mathbf{x}}_\alpha)$$

and

$$\frac{1}{N}\sum_{k \in S}\frac{(\bar{\mathbf{x}}_{|k} - \bar{\mathbf{x}})\, y_k}{\pi_k} = \frac{1}{N^2}\sum_{h=1}^{H}\frac{N_h^3}{N_h - 1}\frac{1-f_h}{n_h}\frac{1}{n_h}\sum_{k \in S_h}(\mathbf{x}_k - \bar{\mathbf{x}}_h)\, y_k.$$

From (3.7) the AOPT1-estimator of can be derived as

$$\hat{\bar{y}}_{\mathrm{AOPT1}} = \hat{\bar{y}}_\pi + (\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}}_\pi)' \left\{\sum_{h=1}^{H} N_h^2 \frac{1-f_h}{n_h}\frac{1}{N_h - 1}\sum_{k \in U_h}(\mathbf{x}_k - \bar{\mathbf{x}}_h)(\mathbf{x}_k - \bar{\mathbf{x}}_h)'\right\}^{-1} \times \sum_{h=1}^{H}\frac{N_h^3}{N_h - 1}\frac{1-f_h}{n_h}\frac{1}{n_h}\sum_{k \in S_h}(\mathbf{x}_k - \bar{\mathbf{x}}_h)\, y_k.$$

The use of this estimator requires the knowledge of very substantial auxiliary information. The population means $\bar{\mathbf{x}}_h$ of the auxiliary variables must be known for each stratum as well as the stratum sizes N_h. Moreover, the values taken by the auxiliary variables must be known for each unit of U.

However $\hat{\hat{y}}_{\text{AOPT1}}$ has in stratification an important drawback: it is not calibrated on the strata size N_h *i.e.*, when the objective consists in estimating the strata sizes N_h, generally $\hat{N}_{\text{AOPT1}} \neq N_h$. This drawback can easily be overcome by centring the interest variable. We thus get:

$$\hat{\hat{y}}_{\text{AOPT1C}} =$$

$$\hat{\bar{y}}_{\pi} + (\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}}_{\pi})' \left\{ \sum_{h=1}^{H} N_h^2 \frac{1-f_h}{n_h} \frac{1}{N_h - 1} \sum_{k \in U_h} (\mathbf{x}_k - \bar{\mathbf{x}}_h)(\mathbf{x}_k - \bar{\mathbf{x}}_h)' \right\}^{-1}$$

$$\times \sum_{h=1}^{H} \frac{N_h^3}{N_h - 1} \frac{1-f_h}{n_h} \frac{1}{n_h} \sum_{k \in S_h} (\mathbf{x}_k - \bar{\mathbf{x}}_h)(y_k - \hat{\bar{y}}_h).$$

6.3 AOPT2-estimator

The AOPT2-estimator can also be used in stratification. In this case, from (3.8) we get

$$\hat{\hat{y}}_{\text{AOPT2}} =$$

$$\hat{\bar{y}}_{\pi} + (\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}}_{\pi})' \left\{ \sum_{h=1}^{H} N_h^2 \frac{1-f_h}{n_h} \frac{1}{n_h - 1} \sum_{k \in S_h} (\mathbf{x}_k - \hat{\bar{\mathbf{x}}}_h)(\mathbf{x}_k - \hat{\bar{\mathbf{x}}}_h)' \right\}^{-1}$$

$$\times \sum_{h=1}^{H} N_h^2 \frac{1-f_h}{n_h} \frac{1}{n_h - 1} \sum_{k \in S_h} (\mathbf{x}_k - \hat{\bar{\mathbf{x}}}_h)(y_k - \hat{\bar{y}}_h).$$

The AOPT2-estimator only needs the knowledge of the population mean vector $\bar{\mathbf{x}}$ and of the stratum sizes N_h. It has however a drawback, the $\bar{\mathbf{x}}_h$ are estimated and thus $J \times H$ degrees of freedom are lost. If the number of strata is large, this loss of degrees of freedom could increase the instability of this estimator when $J \times H$ is large.

6.4 GREG-estimator

The GREG-estimator does not take into account the joint inclusion probabilities. It is given by

$$\hat{\hat{y}}_{\text{GREG}} =$$

$$\hat{\bar{y}}_{\pi} + (\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}}_{\pi})' \left\{ \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in S_h} \frac{\mathbf{x}_k \mathbf{x}_k'}{c_k} \right\}^{-1} \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in S_h} \frac{\mathbf{x}_k y_k}{c_k}.$$

Although this estimator is more stable, it is conditionally biased. Moreover, if we want to estimate the stratum sizes N_h by the GREG-estimator, we do not find exactly N_h. Indeed, if $y_k = 1$ when $k \in U_h$ and $y_k = 0$ when $k \notin U_h$, then

$$\hat{N}_{\text{GREG}} = N \hat{\hat{y}}_{\text{GREG}}$$

$$= \hat{N}_{h\pi} + N(\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}}_{\pi})' \left\{ \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in S_h} \frac{\mathbf{x}_k \mathbf{x}_k'}{c_k} \right\}^{-1}$$

$$\times \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in S_h} \frac{\mathbf{x}_k y_k}{c_k}.$$

Since, in stratified sampling, $\hat{N}_{h\pi} = N_h$, we get

$$\hat{N}_{\text{GREG}} = N_h + N(\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}}_{\pi})' \left\{ \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in S_h} \frac{\mathbf{x}_k \mathbf{x}_k'}{c_k} \right\}^{-1}$$

$$\times \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in S_h} \frac{\mathbf{x}_k}{c_k}. \tag{6.14}$$

Expression (6.14) shows that generally $\hat{N}_{\text{GREG}} \neq N_h$. Thus, the GREG-estimator destroys the stratification effect because it does not take the stratification into account. Indeed, the stratification is represented by the joint inclusion probabilities. In the GREG-estimator, only the first-order inclusion probabilities are used. On the other hand, it is easy to verify that the AOPT1 and AOPT2-estimator of N_h are exactly equal to N_h. The AOPT-estimators is thus calibrated on the N_h.

We propose to use the GREG-estimator in the following cases: when the sample size is small or if the number of strata is large and when the stratification gives poor auxiliary information on the interest variable. Indeed, in this case, the loss of precision due to the loss of degrees of freedom, will be more important than the precision benefit due to the optimality of the estimator. An interesting analysis of the benefit due to the optimal estimator is also given in Montanari (1998).

6.5 GREG-estimator with use of the stratification variables

A variant of use of the GREG-estimator consists in re-using the stratification variables at the estimation stage. Consider the column vector

$$\mathbf{w}_k = (z_{k1}, \ldots, z_{kh}, \ldots, z_{k(H)} \;\; \mathbf{x}_k')'$$

where $z_{kh} = 1$ if $k \in U_h$ and 0 if not. This vector is thus composed of the values taken by the indicator variables of the presence of unit k in the H strata and of the values taken by the x-auxiliary variables.

Now if $\bar{\mathbf{w}}$ denotes the population mean of vectors $\mathbf{w}_k$ and $\hat{\bar{\mathbf{w}}}_{\pi}$ its Horvitz-Thompson estimator, the GREG-estimator using the auxiliary information $\bar{\mathbf{w}}_{\pi}$ is given by

$$\hat{\hat{y}}_{\text{GREG}w} = \hat{\hat{y}}_{\pi} + (\bar{\mathbf{w}} - \hat{\hat{\mathbf{w}}}_{\pi})' \left\{ \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in S_h} \frac{\mathbf{w}_k \mathbf{w}_k'}{c_k} \right\}^{-1} \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{k \in S_h} \frac{\mathbf{w}_k y_k}{c_k}. \quad (6.15)$$

The presentation of expression (6.15) can be simplified. Indeed, the following result was proved by Tillé (1994) and generalised by Särndal (1996):

Result 2: When the stratification variables are re-used at the estimation stage, and if the c_k are equal into the strata $(c_k = c_h,\ k \in U_h)$ the GREG-estimator can be written

$$\hat{\hat{y}}_{\text{GREG}w} =$$

$$\hat{\hat{y}}_{\pi} + (\bar{\mathbf{x}} - \hat{\hat{\mathbf{x}}}_{\pi})' \left\{ \sum_{h=1}^{H} \frac{N_h}{n_h c_h} \sum_{k \in S_h} (\mathbf{x}_k - \hat{\hat{\mathbf{x}}}_h)(\mathbf{x}_k - \hat{\hat{\mathbf{x}}}_h)' \right\}^{-1}$$

$$\times \sum_{h=1}^{H} \frac{N_h}{n_h c_h} \sum_{k \in S_h} (\mathbf{x}_k - \hat{\hat{\mathbf{x}}}_h)(y_k - \hat{\hat{y}}_h). \quad (6.16)$$

A proof of Result 2 is given in the Appendix. Note that expression (6.16) is equal to the AOPT2-estimator when

$$c_h = C \frac{n_h}{N_h} \frac{1 - 1/n_h}{1 - f_h},$$

for $h = 1, \ldots, H,$ where $C > 0$ is a constant. When the f_h are small and the n_h are large and proportional to the N_h both estimators are equivalent. This result shows that with the conditional approach, the fact that the sampling design is stratified is automatically taken into account in the estimation method. The GREG-estimator does not take into account the stratification effect and thus it is necessary to reintroduce the stratification variables at the estimation stage so as not to lose the stratification effect.

7. Simulations

A set of simulations was carried out in order to compare the four following estimators: $\hat{\hat{y}}_{\pi}$, $\hat{\hat{y}}_{\text{AOPT1}C}$, $\hat{\hat{y}}_{\text{AOPT2}}$, $\hat{\hat{y}}_{\text{GREG}}$. The population is made up of 4 strata of 250 units ($N =$ 1,000). A stratified sampling design is applied with proportional allocation. For each simulation, 10,000 samples of size $n = 100$ are selected and the following ratios has been estimated:

$$M_1 = \text{MSE}(\hat{\hat{y}}_{\pi}) / \text{MSE}(\hat{\hat{y}}_{\pi}) = 1,$$

$$M_2 = \text{MSE}(\hat{\hat{y}}_{\text{GREG}}) / \text{MSE}(\hat{\hat{y}}_{\pi}),$$

$$M_3 = \text{MSE}(\hat{\hat{y}}_{\text{AOPT1}C}) / \text{MSE}(\hat{\hat{y}}_{\pi}),$$

$$M_4 = \text{MSE}(\hat{\hat{y}}_{\text{AOPT2}}) / \text{MSE}(\hat{\hat{y}}_{\pi}).$$

The populations are generated by means of the following models: $\mathcal{P}_1$: $x_k = a_k,\ y_k = e_k,\ k \in U,$ (total independence), $\mathcal{P}_2$: $x_k = a_k,\ y_k = 3x_k + e_k,\ k \in U,$ (dependence between x and y), $\mathcal{P}_3$: $x_k = a_k,\ y_k = x_k + 2h(k) + e_k,\ k \in U,$ (dependence between x, y and the strata), $\mathcal{P}_4$: $x_k = a_k,\ y_k = \exp(10 + 2x_k + 10h(k) + e_k),\ k \in U,$ (non-linearity and dependence between x, y and the strata), $\mathcal{P}_5$: $x_k = a_k,\ y_k = \exp(e_k + 3x_k) + 3h(k),\ k \in U,$ (non-linearity and dependence between x, y and the strata), $\mathcal{P}_6$: $x_k = a_k,\ y_k = 3h(k) + e_k,\ k \in U,$ (strong dependence between y and the strata), $\mathcal{P}_7$: $x_k = a_k,\ y_k = 50h(k) + e_k,\ k \in U,$ (very strong dependence between y and the strata), where a_k and e_k are independent normal variable with mean equal to 0 and variance equal to 1, and $h(k)$ is the number of the stratum of unit k. Results of the simulations is given in Table 1.

Table 1
Results of 10,000 simulations

	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$	$\mathcal{P}_7$
M_1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
M_2	1.0070	0.0906	0.5180	0.9261	0.9263	1.1047	38.5104
M_3	1.0069	0.0906	0.4835	0.9277	0.9269	1.0015	1.0123
M_4	1.0060	0.0936	0.4850	0.9257	0.9239	1.0006	1.0111

Table 1 shows that the GREG-estimator provides a good estimation when the stratification variables are not correlated to the interest variable. Nevertheless, the more is the dependence between the stratification variable and the interest variable, the more is the gain of precision of $\hat{\hat{y}}_{\text{AOPT1}C}$ and $\hat{\hat{y}}_{\text{AOPT2}}$. The loss of degrees of freedom of the optimal estimator does not seem to affect the precision for this sample size. Moreover, the gain obtained by the knowledge of the population stratum is not significant for this sample size. For all these cases, the optimal estimator is thus clearly preferable to the GREG-estimator.

8. A third-order problem

The complexity of determining the conditional weights is not a specific problem of the SCW-estimator. It is due to the general problem of estimation with auxiliary information used *a posteriori* when an auxiliary variable is already used *a priori* in the sampling design. This problem can be presented as a third-order interaction problem among

– the interest variables;
– the sampling design and thus the auxiliary variables used *a priori*;
– the auxiliary variables used *a posteriori*.

Indeed, the use of auxiliary information at the estimation stage leads to the following problem: how do these auxiliary variables used *a posteriori* interact with the interest variable through a given sampling design? The problem being complex, we have to take into account the relationships between each set of variables above as well as the third-order interactions among these three sets of variables.

It is very difficult to find a really operational estimator which uses the three second-order interactions and the third-order interaction. For this reason, one can attempt to simplify the problem. The neutralisation of one of the aspects of this problem significantly simplifies the research of an estimator. Most of the possible simplifications have already been studied. We can cite some of these:

- If no auxiliary information is used *a posteriori* (except the population size N) we can only construct the Horvitz-Thompson estimator or Hájek's ratio (1971).
- Searching general solutions using auxiliary information for simple random sampling does not pose major problems. In this case, no auxiliary information is used *a priori*.
- Using a superpopulation model allows one to fix a relation existing between the interest variable and the auxiliary variables used *a posteriori*. In this case, it is possible to determine the optimal estimator (under the model).
- For the GREG-estimator and also for the calibration methods (see Deville and Särndal 1992), in the design-based inference framework, only the first-order probabilities are retained from the sampling design. A simple random sampling is thus treated in the same way as a stratified design for which the first-order inclusion probabilities are all equal. For this reason, a regression estimator applied to a stratified design generally destroys the calibration on the stratum frequencies given by the *a priori* stratification. In this case, the simplification arises because all the contributions of the auxiliary variables used *a priori* to the sampling design can be described only by the first-order probabilities.
- Finally, for the optimal linear estimator, it is implicitly supposed that the dependence between Horvitz-Thompson estimators of the variables $\mathbf{x}$ and y is linear. Obviously, these estimators neglect the non-linear dependence between the estimators. Nevertheless, it takes into account the joint inclusion probabilities. When the sampling design is stratified, the estimator remains calibrated on the population stratum frequencies.

The CW-estimator takes into account this third-order interaction. Moreover, in this case, auxiliary information does not necessarily intervene in a linear way. The weights depend on both the sampling design and the auxiliary statistic. These weights applied to the values taken by the interest variable take into account all the interactions between the three variable groups.

The methods using conditional inclusion probabilities are interesting for different reasons: they give a general frame allowing to search and conceive estimators using auxiliary information without reference to a superpopulation model and lead to valid conditional inference. They bring into prominence all the complexity of the estimation problem with auxiliary information. According to the known auxiliary information, we can find either known results (as for example post-stratification) or very complex and not really operational estimators. However, a first approximation leads to a known result, *i.e.* the optimal linear estimator.

Acknowledgements

The author is grateful to two Anonymous Referees and an Associate Editor for constructive suggestions and to Professor Carl Särndal for interesting comments on a previous version of this paper.

Appendix

Proof of Result 1 and 2

Lemma 1 will be used in the proof of Result 1.

Lemma 1: If $\mathbf{R}_k^{-1} - \mathbf{R}^{-1} = O_{J\times J}(n^{-1})$, then $|\mathbf{R}_k|^{-1}|\mathbf{R}| = 1 + O(n^{-1})$, where $\mathbf{R}$ and $\mathbf{R}_k$ are defined as in hypothesis (ii).

Proof

$$[\mathbf{R}_k^{-1} - \mathbf{R}^{-1}]\mathbf{R} = O_{J\times J}(n^{-1})\,\mathbf{R}$$

and thus

$$|\mathbf{R}_k^{-1}\mathbf{R}| = |\mathbf{I} + O_{J\times J}(n^{-1})\,\mathbf{R}| = [1 + O(n^{-1})]^J$$

$$+ O(n^{-1}) = 1 + O(n^{-1})$$

where $\mathbf{I}$ is a $J \times J$ identity matrix. Thus,

$$\frac{|\mathbf{R}_k|}{|\mathbf{R}|} = \frac{1}{1 + O(n^{-1})} = 1 + O(n^{-1}).$$

Note that lemma 1 is a consequence of hypothesis (ii).

Proof of Result 1

If we define

$$d_k = \frac{n}{N\pi_k}, \text{ for all } k \in U.$$

by hypothesis (i), we get:

$$a_k(\hat{\hat{\mathbf{x}}}_\pi) = \frac{n}{\pi_{k|T} N} = \frac{n\Pr(\hat{\hat{\mathbf{x}}}_\pi)}{N\pi_k \Pr(\hat{\hat{\mathbf{x}}}_\pi \mid k \in S)} = d_k \frac{f(\hat{\hat{\mathbf{x}}}_\pi)}{f_k(\hat{\hat{\mathbf{x}}}_\pi)}$$

where f (resp. f_k) is the density function of a multivariate normal variable with mean $\bar{\mathbf{x}}$ (resp. $\bar{\mathbf{x}}_{|k}$) and variance-covariance matrix $\boldsymbol{\Sigma}$ (resp. $\boldsymbol{\Sigma}_k$). Thus,

$$a_k(\hat{\hat{\mathbf{x}}}_\pi) = d_k \frac{\frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp - \frac{1}{2}(\hat{\hat{\mathbf{x}}}_\pi - \bar{\mathbf{x}})' \boldsymbol{\Sigma}^{-1} (\hat{\hat{\mathbf{x}}}_\pi - \bar{\mathbf{x}})}{\frac{1}{|\boldsymbol{\Sigma}_k|^{1/2}} \exp - \frac{1}{2}(\hat{\hat{\mathbf{x}}}_\pi - \bar{\mathbf{x}}_{|k})' \boldsymbol{\Sigma}_k^{-1} (\hat{\hat{\mathbf{x}}}_\pi - \bar{\mathbf{x}}_{|k}).} \qquad (8.17)$$

If we also note

$$\mathbf{R} = \mathbf{V}^{-1/2} \boldsymbol{\Sigma} \mathbf{V}^{-1/2},$$

$$\mathbf{R}_k = \mathbf{V}^{-1/2} \boldsymbol{\Sigma}_k \mathbf{V}^{-1/2},$$

$$\hat{\hat{\mathbf{x}}}_\pi^c = \mathbf{V}^{-1/2} (\hat{\hat{\mathbf{x}}}_\pi - \bar{\mathbf{x}}),$$

$$\gamma_k = \mathbf{V}^{-1/2} (\bar{\mathbf{x}}_{|k} - \bar{\mathbf{x}}),$$

and

$$c_k = d_k \frac{|\mathbf{R}_k|^{1/2}}{|\mathbf{R}|^{1/2}} \exp \frac{1}{2} \hat{\hat{\mathbf{x}}}_\pi^{c\prime} (\mathbf{R}_k^{-1} - \mathbf{R}^{-1}) \hat{\hat{\mathbf{x}}}_\pi^c, \qquad (8.18)$$

we get

$$a_k(\hat{\hat{\mathbf{x}}}_\pi) = d_k \frac{|\mathbf{R}|^{-1/2} \exp - \frac{1}{2} \hat{\hat{\mathbf{x}}}_\pi^{c\prime} \mathbf{R}^{-1} \hat{\hat{\mathbf{x}}}_\pi^c}{|\mathbf{R}_k|^{-1/2} \exp - \frac{1}{2} (\hat{\hat{\mathbf{x}}}_\pi^c - \gamma_k)' \mathbf{R}_k^{-1} (\hat{\hat{\mathbf{x}}}_\pi^c - \gamma_k)}$$

$$= c_k \exp \frac{1}{2} \gamma_k \mathbf{R}_k^{-1} (\gamma_k - 2 \hat{\hat{\mathbf{x}}}_\pi^c). \qquad (8.19)$$

By using a Taylor development for the vector γ_k of (8.19), we get

$$a_k(\hat{\hat{\mathbf{x}}}_\pi) = c_k (1 - \gamma_k \mathbf{R}_k^{-1} \hat{\hat{\mathbf{x}}}_\pi^c) + R(\gamma_k^{(0)}). \qquad (8.20)$$

where

$$R(\gamma_k^{(0)}) = c_k \left[\exp \frac{1}{2} \gamma_k^{(0)} \mathbf{R}_k^{-1} (\gamma_k^{(0)} - 2 \hat{\hat{\mathbf{x}}}_\pi^c) \right]$$

$$\times \gamma_k' \{ [\mathbf{R}_k^{-1} (\gamma_k^{(0)} - \hat{\hat{\mathbf{x}}}_\pi^c)] [\mathbf{R}_k^{-1} (\gamma_k^{(0)} - \hat{\hat{\mathbf{x}}}_\pi^c)]' - \mathbf{R}_k^{-1} \} \gamma_k,$$

and $\gamma_k^{(0)}$ is a vector whose elements are included between the correspondent elements of γ_k and 0. By hypothesis (iii), we directly get

$$R(\gamma_k^{(0)}) = O_p(n^{-1}).$$

On the other hand, we have by hypothesis (ii), lemma 1 and (8.18) that

$$c_k = d_k [1 + O(n^{-1})]^{1/2} \exp \left\{ \frac{1}{2} \hat{\hat{\mathbf{x}}}_\pi^{c\prime} O_{J \times J}(n^{-1}) \hat{\hat{\mathbf{x}}}_\pi^c \right\}$$

$$= d_k [1 + O_p(n^{-1})]. \qquad (8.21)$$

By (8.20) and (8.21), we get

$$a(\hat{\hat{\mathbf{x}}}_\pi) = d_k [1 + O_p^{-1}(n^{-1})] \{1 - \gamma_k (\mathbf{R}^{-1} + O_{J \times J}(n^{-1})) \hat{\hat{\mathbf{x}}}_\pi^c + O_p(n^{-1})\}$$

$$= d_k \{1 - \gamma_k' \mathbf{R}^{-1} \hat{\hat{\mathbf{x}}}_\pi^c + O_p(n^{-1})\}$$

$$= d_k \{1 + (\bar{\mathbf{x}} - \hat{\hat{\mathbf{x}}}_\pi)' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}}_{|k} - \bar{\mathbf{x}}) + O_p(n^{-1})\}.$$

Finally, we get

$$\hat{\hat{y}}_{|T} = \frac{1}{n} \sum_{k \in S} a_k(\hat{\hat{\mathbf{x}}}_\pi) y_k$$

$$= \hat{\hat{y}}_\pi + (\bar{\mathbf{x}} - \hat{\hat{\mathbf{x}}}_\pi)' \Sigma^{-1} \frac{1}{N} \sum_{k \in S} \frac{(\bar{\mathbf{x}}_{|k} - \bar{\mathbf{x}}) y_k}{\pi_k} + O_p(n^{-1}).$$

Proof of Result 2

In Särndal (1980), we see that the GREG-estimator presented in (6.15) can also be written:

$$\hat{\hat{y}}_{\text{GREG}w} = \hat{\hat{y}}_\pi + N^{-1} (\mathbf{1}_N' \mathbf{W}_N - \mathbf{1}_S' \Pi_S^{-1} \mathbf{W}_S)'$$

$$(\mathbf{W}_S' \mathbf{C}_S^{-1} \Pi_S^{-1} \mathbf{W}_S)^{-1} \mathbf{W}_S' \mathbf{C}_S^{-1} \Pi_S^{-1} \mathbf{y}_S$$

where $\mathbf{1}_N$ (resp. $\mathbf{1}_S$) is a column vector composed of N (resp. n) ones, Π_N (resp. Π_S) is a diagonal matrix having the inclusion probabilities of the population (resp. sample) units on its diagonal, $\mathbf{C}_S$ is a diagonal matrix having the c_k of the sample units on its diagonal, $\mathbf{y}_S$ is a column vector composed of the values taken by the interest variable y in the sample,

$$\mathbf{W}_N = \begin{pmatrix} z_{11} & \cdots & z_{1H} & \mathbf{x}_1' \\ \vdots & & \vdots & \vdots \\ z_{N1} & \cdots & z_{NH} & \mathbf{x}_N' \end{pmatrix}$$

and $\mathbf{W}_S$ is a $n \times (H + J)$ matrix composed of the n rows of $\mathbf{W}_N$ corresponding to the units selected in the sample.

The matrix to invert can be partitioned into four parts:

$$(\mathbf{W}_S' \mathbf{C}_S^{-1} \Pi_S^{-1} \mathbf{W}_S) = \begin{bmatrix} \mathbf{A} & \mathbf{D} \\ \mathbf{D}' & \mathbf{B} \end{bmatrix}$$

where $\mathbf{A}$ is an $H \times H$ matrix having N_h/c_h, $h = 1, ..., H$, on its diagonal,

$$\mathbf{B} = \sum_{h=1}^{H} \frac{N_h}{n_h\, c_h} \sum_{k \in S_h} \mathbf{x}_k\, \mathbf{x}_k'$$

and

$$\mathbf{D}' = \left[\frac{N_1 \hat{\hat{\mathbf{x}}}_1}{c_1} \; \ldots \; \frac{N_H \hat{\hat{\mathbf{x}}}_H}{c_H} \right].$$

By using the technique of matrix inversion by partition, we get

$$(\mathbf{W}_S' \mathbf{C}_S^{-1} \Pi_S^{-1} \mathbf{W}_S)^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{D}\mathbf{B}^{-1}\mathbf{D}')^{-1} & -\mathbf{A}^{-1}\mathbf{D}\mathbf{Q} \\ -\mathbf{Q}\mathbf{D}'\mathbf{A}^{-1} & \mathbf{Q} \end{bmatrix}$$

where $\mathbf{Q} = (\mathbf{B} - \mathbf{D}'\mathbf{A}^{-1}\mathbf{D})^{-1}$. Since

$$(\mathbf{1}_N' \mathbf{W}_N - \mathbf{1}_S' \Pi_S^{-1} \mathbf{W}_S) = \begin{bmatrix} \mathbf{0}_H \\ \hat{\mathbf{x}}_\pi - \overline{\mathbf{x}} \end{bmatrix}$$

where $\mathbf{0}_H$ is a column vector composed of H zeros, we get

$$(\mathbf{1}_N' \mathbf{W}_N - \mathbf{1}_S' \Pi_S^{-1} \mathbf{W}_S)' (\mathbf{W}_S' \mathbf{C}_S^{-1} \Pi_S^{-1} \mathbf{W}_S)^{-1}$$
$$= (\hat{\hat{\mathbf{x}}}_\pi - \overline{\mathbf{x}})' \mathbf{Q} [-\mathbf{D}'\mathbf{A}^{-1} \; \mathbf{I}_{(J \times J)}] \quad (8.22)$$

where $\mathbf{I}_{(J \times J)}$ is a $J \times J$ identity matrix. Since

$$\mathbf{Q} = \left\{ \sum_{h=1}^{H} \frac{N_h}{n_h c_h} \sum_{k \in S_h} (\mathbf{x}_k - \hat{\hat{\mathbf{x}}}_h)(\mathbf{x}_k - \hat{\hat{\mathbf{x}}}_h)' \right\}^{-1},$$

$$[-\mathbf{D}\mathbf{A}^{-1} \; \mathbf{I}_{(J \times J)}] = [-\hat{\hat{\mathbf{x}}}_1 \ldots -\hat{\hat{\mathbf{x}}}_H \; \mathbf{I}_{(J \times J)}]$$

and

$$\mathbf{W}_S' \mathbf{C}_S^{-1} \Pi_S^{-1} \mathbf{y}_S = \begin{bmatrix} \dfrac{N_1 \hat{\hat{y}}_1}{c_1} \\ \vdots \\ \dfrac{N_H \hat{\hat{y}}_H}{c_H} \\ \displaystyle\sum_{h=1}^{H} \frac{N_h}{n_h c_h} \sum_{k \in S_h} \mathbf{x}_k y_k \end{bmatrix}, \qquad (8.23)$$

we get Result 2 by multiplication of (8.22) and (8.23).

References

Casady, R.J., and Valliant R. (1993). Conditional properties of post-stratified estimators under normal theory. *Survey Methoddology*, 19, 183-192.

Cassel, C.-M., Särndal, C.-E. and Wretman, J.H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite population. *Biometrika*, 63, 615-620.

Deville, J.-C. (1992). Constrained samples, conditional inference, weighting: Three aspects of the utilisation of auxiliary information. *Proceeding of the Workshop*: *Auxiliary information in surveys*, Örebro.

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

Fuller, W.A., and Isaki, C.T. (1981). Survey design under superpopulation models. In *Current topics in survey sampling*, (Eds., D. Krewski, R. Platek, J.N.K. Rao and M.P. Singh). New York: Academic Press, 196-226.

Hájek, J. (1971). Comment on a essay of D. Basu. In *Foundations of statistical inference*, (Eds., V.P. Godambe and D.A. Sprott), Toronto, Holt, Rinehart et Winston, 236.

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

Madow, W.G. (1948). On the limiting distributions of estimates based on samples from finite universes. *Annals of Mathematical Statistics*, 19, 535-545.

Montanari, G.E. (1987). Post sampling efficient QR-prediction in large sample survey. *International Statistical Review*, 55, 191-202.

Montanari, G.E. (1997). On conditional properties of finite population mean estimators. *Proceeding of the 51st Session of the International Statistical Institute*, Contributed paper, 351-352.

Montanari, G.E. (1998). On regression estimation of finite population means. *Survey Methodology*, 24, 69-77.

Rao, J.N.K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *Journal of Official Statistics*, 10, 153-165.

Rao, J.N.K. (1997). Development in sample survey theory: An appraisal. *Canadian Journal of Statistics*, 25, 1-21.

Rosén, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement. *Annals of Mathematical Statistics*, 43, 373-397.

Särndal, C.-E. (1980). On π-inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika*, 67, 639-650.

Särndal, C.-E. (1996). Efficient estimators with simple variance in unequal probability sampling. *Journal of the American Statistical Association*, 91, 1289-1300.

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

Tillé, Y. (1994). *Utilisation d'information auxiliaire en théorie des sondages sans référence à un modèle*. Ph.D Thesis, Université Libre de Bruxelles, Institut de Statistique.

Tillé, Y. (1998). Estimation in surveys using conditional inclusion probabilities: Simple random sampling. *International Statistical Review*, 66, 303-322.

Wright, R.L. (1983). Finite population with multivariate auxiliary information. *Journal of the American Statistical Association*, 78, 879-883.