Journal of Official Statistics, Vol. 16, No. 4, 2000, pp. 363–378

On Variance Estimation for Measures of Change When Samples are Coordinated by the Use of Permanent Random Numbers

Lennart Nordberg[1]

A common objective in business surveys is to compare two estimates $\hat{\theta}_0$ and $\hat{\theta}_1$ of the same characteristic taken on two occasions, e.g., the level of production the same month in two consecutive years, and to judge whether the observed change is statistically significant or merely subject to random variation.

Business surveys often use samples at separate occasions that are positively coordinated, i.e., overlapping, in order to increase the precision in estimates of change over time. Such sample coordination will make $\hat{\theta}_0$ and $\hat{\theta}_1$ become correlated. Some systems used for sample coordination rely on Permanent Random Numbers (PRNs). However, the use of PRNs brings an additional component of randomness to the rotation pattern as compared to ordinary panel rotation. This makes the estimation of the correlation more difficult.

The SAMU system which is used by Statistics Sweden for sample coordination of business surveys is such a system. The purpose of the present paper is to show how to estimate the variance for measures of change such as $\hat{\psi} = \hat{\theta}_1 - \hat{\theta}_0$ or $\hat{\psi} = \hat{\theta}_1/\hat{\theta}_0$ when $\hat{\theta}_0$ and $\hat{\theta}_1$ are estimated from two separate SAMU samples.

Key words: Survey sampling; variance estimation; estimates of change; panel designs; permanent random numbers.

1. Introduction

Often in business surveys one wants to compare two estimates $\hat{\theta}_0$ and $\hat{\theta}_1$ of the same characteristic taken on two occasions 0 and 1, e.g., the level of production the same month in two consecutive years, and to judge whether the observed change is statistically significant or merely subject to random variation.

Business surveys often use samples at separate occasions that are positively coordinated, i.e., overlapping, in order to increase precision in estimates of change over time. Such sample coordination will make $\hat{\theta}_0$ and $\hat{\theta}_1$ become correlated.

Under common rotating panel designs this correlation can often be estimated in a straightforward way. However, some systems used for sample coordination are designed not only to generate samples that are positively coordinated between consecutive occasions but also to obtain negative coordination between different surveys in order to spread the response burden. One way to create positive and negative coordination simultaneously is to use Permanent Random Number (PRN) techniques. However, the use of PRNs brings

[1] Statistics Sweden, 70189 Örebro, Sweden. E-mail: lennart.nordberg@scb.se
Acknowledgments: I am indebted to Dr. Tiina Orusild for letting me use a result from her forthcoming paper Orusild (2000). I also want to thank Dr. Sixten Lundström for helpful comments on an earlier version of this article. The constructive criticism of two referees and the editor has led to substantial improvements of the article.

an additional component of randomness to the rotation pattern as compared to ordinary panel rotation. This makes the estimation of the correlation more difficult.

The SAMU system for sample coordination of business surveys at Statistics Sweden utilises PRNs. The word SAMU (*SAMordnade Urval*) is an acronyme in Swedish for co-ordinated samples. The purpose of the article is to show how to estimate the variances for measures of change such as $\hat{\psi} = \hat{\theta}_1 - \hat{\theta}_0$ or $\hat{\psi} = \hat{\theta}_1/\hat{\theta}_0$ when θ_0 and θ_1 are estimated from two separate SAMU samples. This problem was addressed by several people at Statistics Sweden in the 1970s and 80s, unfortunately without reaching a complete and workable solution. This work was summarised in Garås (1989).

Although the focus of the article is on the SAMU system, it is hoped that the proposed approach can be of interest also in the context of other PRN systems. Related approaches are found in Tam (1984), Laniel (1988) and Hidiroglou, Särndal, and Binder (1995).

SAMU applies to a variety of designs but I will confine the discussion here to the STSI design which is the simple and common case of stratified sampling of elements (businesses) with simple random sampling without replacement within strata. Next I give a brief presentation of SAMU. For more complete descriptions, see Ohlsson (1992, 1995).

Every sample in the SAMU system is drawn from an up-to-date version of the Business Register. The co-ordination of samples is obtained in the following way. A uniformly distributed random number is assigned to every element (enterprise or local unit) as soon as it enters the Business Register, and this association is kept as long as the element remains in the register. All generated random numbers are to be independent. Suppose that one wants to sample ten elements in a particular stratum. All the frame elements in the current stratum are ordered by the size of their random numbers. An arbitrary starting point is chosen and the first ten elements to the right (say) of this starting point are included in the sample. It can be shown (see Ohlsson 1992), that this sampling mechanism is equivalent to simple random sampling.

I will in the following consider SAMU sampling on two occasions, time 0 and time 1, and hence apply the PRN technique to two different versions of the Business Register. By this I will obviously introduce some additional randomness compared to the case of common rotating panels. Whether a certain element (business) which was included in the sample at time 0 will remain in the sample on the next sampling occasion at time 1 depends not only on the element itself but also on the behaviour of other elements, notably the random numbers associated with the births and deaths in the frame.

If $\hat{\psi} = \hat{\theta}_1 - \hat{\theta}_0$ we can write the variance of $\hat{\psi}$ as follows:

$$V(\hat{\psi}) = V(\hat{\theta}_0) + V(\hat{\theta}_1) - 2 \cdot C(\hat{\theta}_0, \hat{\theta}_1) \tag{1.1}$$

If $\hat{\psi} = \hat{\theta}_1/\hat{\theta}_0$ we have by Taylor linearisation,

$$\frac{V(\hat{\psi})}{\psi^2} \approx \frac{V(\hat{\theta}_0)}{\theta_0^2} + \frac{V(\hat{\theta}_1)}{\theta_1^2} - 2 \cdot \frac{C(\hat{\theta}_0, \hat{\theta}_1)}{\theta_0 \theta_1} \tag{1.2}$$

Although θ_0 and θ_1 may be more complex parameters than population totals (see relation (1.5) ahead), the main problem concerns the covariance term, not the variance components $V(\hat{\theta}_0)$ and $V(\hat{\theta}_1)$.

Estimation at time 0: Consider a set of variables $y_1, \ldots, y_j, \ldots, y_J$ and let y_{jk} be the value of the variable y_j for element k in the finite population U at time 0. We associate a population total $t_j = \sum_{k \in U} y_{jk}$ with every variable y_j. The population U is stratified into H strata, $U_1, \ldots, U_h, \ldots, U_H$, and a simple random sample is drawn from each stratum. Let s denote the chosen sample and let N_h and n_h be the number of population- and sample- elements respectively in stratum $U_h, h = 1, 2, \ldots, H$.

As estimator for the total t_j we consider the Horvitz-Thompson (H-T) or the Generalised Regression (GREG) estimator. We begin by assuming that the H-T estimator is used, and then later show how the achieved results can be modified to cover the GREG case.

Hence

$$\hat{t}_j = \sum_h \frac{N_h}{n_k} \sum_{k \in U_h} y_{jk} \delta_k \tag{1.3}$$

where

$$\delta_k = \begin{cases} 1 & \text{if element } k \in s \\ 0 & \text{otherwise} \end{cases} \tag{1.4}$$

We assume that the estimator $\hat{\theta}_0$ can be expressed in the following form:

$$\hat{\theta}_0 = f(\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_J). \tag{1.5}$$

where f is an arbitrary rational function. A common estimator for a population mean or a difference between two domain totals, a ratio estimator or a poststratified ratio estimator for a population total are examples of such functions.

Estimation at time 1: The population U' at time 1 consisting of N' elements is stratified into L strata, $U'_1, \ldots, U'_l, \ldots, U'_L$. The stratification at time 1 does not have to be the same as the one at time 0. Let s' be the sample and let N'_l and n'_l be the population- and sample-size in stratum $U'_l, l = 1, 2, \ldots, L$.

The population totals at time 1 are estimated in analogy with (1.3).

$$\hat{t}'_j = \sum_l \frac{N'_l}{n'_l} \sum_{r \in U'_l} y'_{jr} \delta'_r \tag{1.6}$$

where

$$\delta'_r = \begin{cases} 1 & \text{if element r} \in \text{s}' \\ 0 & \text{otherwise} \end{cases} \tag{1.7}$$

We assume that the estimator $\hat{\theta}_1$ can be written in the form

$$\hat{\theta}_1 = f(\hat{t}'_1, \hat{t}'_2, \ldots, \hat{t}'_J) \tag{1.8}$$

Notice that the same function f is assumed in (1.5) and (1.8). This should be the most common case in practice. Generalisation to the case of different functions f_0 and f_1 is straightforward.

2. The Covariance

By standard Taylor linearisation the covariance in (1.1) or (1.2) may be written as follows.

$$C(\hat{\theta}_0, \hat{\theta}_1) \approx \sum_i \sum_j f_i'(t_1, t_2, \ldots, t_J) f_j'(t_1', t_2', \ldots, t_J') C(\hat{t}_i, \hat{t}_j') \tag{2.1}$$

f_i' being the partial derivative $\partial f / \partial t_i$ and $C(\hat{t}_i, \hat{t}_j')$ the covariance for the pair $(\hat{t}_i, \hat{t}_j')$. The covariance $C(\hat{t}_i, \hat{t}_j')$ is very complex under the SAMU framework, and it will be approached in several steps in the following way:

The union of the sampling frames for time 0 and time 1 respectively can be divided into three nonoverlapping parts. The first part consists of the elements that were included in the frame at time 0 but not at time 1, i.e., those elements that have disappeared between time 0 and time 1. We call this group D (''Deaths''). The second part consists of the elements that were included in both frames (time 0 and time 1). We call this group P (''Persistors''). The third part consists of the elements that are included in the frame at time 1 but not at time 0. We call this group B (''Births'').

The division into the three groups D, P and B is symbolised by the following figure:

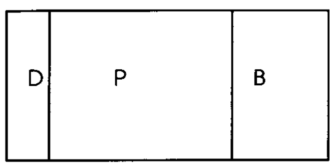

Fig. 1. Division of sampling frames into ''Deaths'', ''Persistors'' and ''Births''.

The set D can be split into the nonoverlapping subsets $\{D_h, h = 1, 2, \ldots H\}$ where D_h is the set of frame elements that belonged to stratum U_h at time 0 and had left the frame before time 1.

Correspondingly, the set B can be split into the nonoverlapping subsets $\{B_l, l = 1, 2, \ldots, L\}$ where B_l is the set of frame elements in stratum U_l' time 1, $l = 1, 2, \ldots, L$ which were not found in the frame at time 0.

The set P can be further divided into the nonoverlapping sets $\{P_{hl}, h = 1, 2, \ldots, H, l = 1, 2, \ldots, L\}$ where P_{hl} is the group of frame elements that belonged to stratum U_h at time 0 and stratum U_l' at time 1.

We now need some further notation. Let G_{hl}, G_{h+} and G_{+l} be the number of frame elements in P_{hl}, D_h and B_l respectively. Furthermore, suppose that among the n_h elements sampled from stratum U_h at time 0, a_{h+} belong to D_h, and a_{hl} belong to the stratum combination P_{hl}. Hence $n_h = a_{h+} + \sum_l a_{hl}$.

Among the n_l' elements sampled from stratum U_l' at time 1 we assume that a_{+l}' belong to B_l, and a_{hl}' belong to the stratum combination P_{hl}. Hence $n_l' = a_{+l}' + \sum_h a_{hl}'$.

Finally, let g_{hl} be the number of elements that belong to P_{hl} *and* are included in *both* samples (time 0 and time 1). We illustrate this by the following figure.

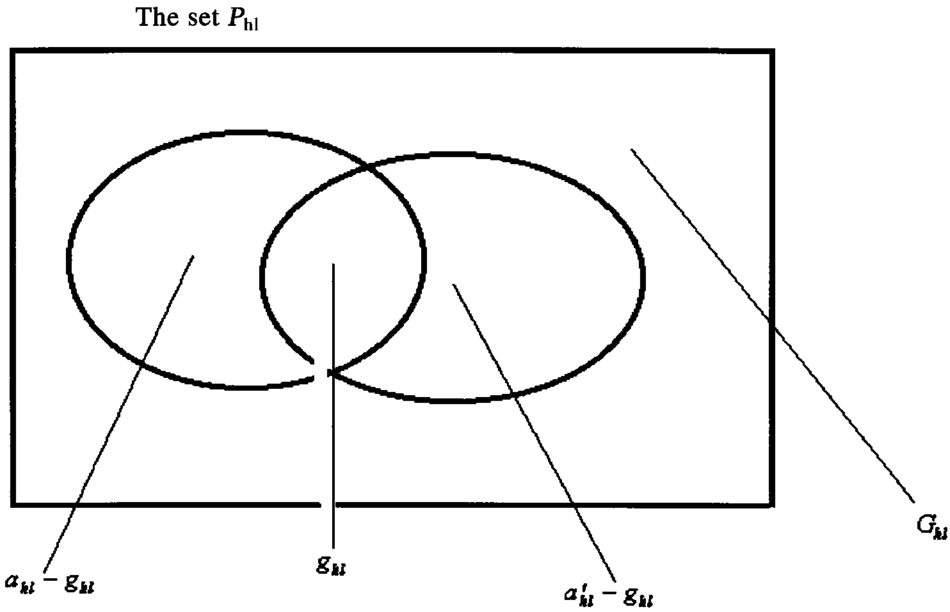

Fig. 2. The number of frame elments in P_{hl} and the number elements from P_{hl} appearing in the sample at time 0 only, in both samples and in the sample at time 1 only.

The quantity $\Omega = \{a_{hl}, a'_{hl}, g_{hl}, a_{h+}, a'_{+l}, h = 1, 2, \ldots, H, l = 1, 2, \ldots, L\}$, which is random in the present context will be used to split the covariance $C(\hat{\theta}_0, \hat{\theta}_1)$ of (2.1) into two components as follows:

$$C(\hat{\theta}_0, \hat{\theta}_1) = E_\Omega(\mathrm{C}(\hat{\theta}_0, \hat{\theta}_1|\Omega)) + \mathrm{C}_\Omega(E(\hat{\theta}_0|\Omega), E(\hat{\theta}_1|\Omega)) \tag{2.2}$$

The first term of (2.2) – the conditional covariance – is usually the dominating term while the second term can be characterised as a remainder term. Next we show how the conditional covariance can be estimated. Section 4 treats the second term.

3. The Conditional Covariance Term

In analogy with (2.1), we have

$$C(\hat{\theta}_0, \hat{\theta}_1|\Omega) \approx \sum_i \sum_j f'_i(t_1, t_2, \ldots, t_J) f'_j(t'_1, t'_2, \ldots, t'_J) C(\hat{t}_i, \hat{t}'_j|\Omega) \tag{3.1}$$

By combination of (1.3) and (1.6) we have

$$C(\hat{t}_i, \hat{t}'_j|\Omega) = \sum_h \sum_l \sum_{k \in U_h} \sum_{r \in U'_l} \frac{N_h \cdot N'_l}{n_h \cdot n'_l} \cdot y_{ik} \cdot y'_{jr} \cdot C(\delta_k, \delta'_r|\Omega) \tag{3.2}$$

The covariances $C(\delta_k, \delta'_r|\Omega)$ vanish for all pairs (k, r) where k and r belong to different strata on both occasions 0 and 1. Hence we only need to consider the cases where k and r are in the same stratum on one or both occasions, i.e., $(k \in P_{hl}, r \in P_{hl})$ or $(k \in P_{hl}, r \in P_{hl'})$ or $(k \in P_{hl}, r \in P_{h'l})$ or $(k \in D_h, r \in P_{hl})$ or $(k \in P_{hl}, r \in B_l)$

It follows from relation (v) in appendix A that $C(\delta_k, \delta'_r|\Omega)$ actually vanishes in all cases except when the elements (k, r) both belong to the same stratum combination: $(k \in P_{hl}, r \in P_{hl})$.

Hence the relation (3.2) can be expressed as follows:

$$C(\hat{t}_i, \hat{t}'_j|\Omega) = \sum_h \sum_l \sum_{k \in P_{hl}} \sum_{r \in P_{hl}} \frac{N_h \cdot N'_l}{n_h \cdot n'_l} \cdot y_{ik} \cdot y'_{jr} \cdot C(\delta_k, \delta'_r|\Omega) \tag{3.3}$$

or, equivalently

$$C(\hat{t}_i, \hat{t}'_j|\Omega) = \sum_h \sum_l \sum_{k \in P_{hl}} \sum_{r \in P_{hl}} \frac{N_h \cdot N'_l}{n_h \cdot n'_l} \cdot y_{ik} \cdot y'_{jr} \cdot (E(\delta_k \cdot \delta'_r|\Omega) - E(\delta_k|\Omega) \cdot E(\delta'_r|\Omega)) \tag{3.4}$$

The expectations appearing in (3.4) can be computed from Relations (*i*), (*iii*), (*vi*) and (*vii*) in Appendix A. We have to distinguish between three types of stratum combinations: $P_{hl}, h = 1, 2, \ldots, H, l = 1, 2, \ldots, L$.

Type 1: This type involves all stratum combinations P_{hl} where $a_{hl} >= 1, a'_{hl} >= 1$ and $g_{hl} >= 1$.

Type 2: This type involves all stratum combinations P_{hl} where $a_{hl} = 0$ or $a'_{hl} = 0$, and, as a consequence, $g_{hl} = 0$ (see Figure 2 above). Stratum combinations of this type do not contribute to the conditional covariance as seen when relations (*i*), (*iii*), (*vi*) and (*vii*) in Appendix A are inserted into (3.4).

Type 3: This type involves all stratum combinations P_{hl} where $a_{hl} >= 1$, $a'_{hl} >= 1$ and $g_{hl} = 0$. Considering the sampling mechanism used in SAMU as described in Section 1 above, this type should not exist if all elements keep their random numbers over time. However, in practice, the random numbers are in fact changed by a constant for a small proportion of the elements every year. The reason for this is to increase the sample rotation among small businesses. Hence stratum combinations of Type 3 can appear in practice although they are very unusual. Since there are no overlapping elements in this case we cannot estimate the contributions to the covariance from these stratum combinations. However, since these contributions are likely to be very small we will neglect this case in the following and estimate by a zero the contributions from stratum combinations of Type 3.

The following quantity is unbiased for $C(\hat{t}_i, \hat{t}'_j|\Omega)$:

$$\hat{C}(\hat{t}_i, \hat{t}'_j|\Omega) = \sum_h \sum_l \sum_{k \in P_{hl}} \sum_{r \in P_{hl}} \frac{N_h \cdot N'_l}{n_h \cdot n'_l} \cdot y_{ik} \cdot y'_{jr} \cdot \left(1 - \frac{E(\delta_k|\Omega) \cdot E(\delta'_r|\Omega)}{E(\delta_k \cdot \delta'_r|\Omega)}\right) \cdot \delta_k \cdot \delta'_r \tag{3.5}$$

where it is assumed that the stratum combinations P_{hl} are all of Type 1.

From relations (*i*), (*iii*), (*vi*) and (*vii*) in appendix A and some algebra we have,

$$\left(1 - \frac{E(\delta_k|\Omega)E(\delta'_r|\Omega)}{E(\delta_k \cdot \delta'_r|\Omega)}\right) = 1 - \frac{\tilde{a}_{hl}}{G_{hl}}, k = r \tag{3.6}$$

$$\left(1 - \frac{E(\delta_k|\Omega)E(\delta'_r|\Omega)}{E(\delta_k \cdot \delta'_r|\Omega)}\right) = \frac{\tilde{a}_{hl} - G_{hl}}{G_{hl}(\tilde{a}_{hl} - 1)}, k \neq r \tag{3.7}$$

where

$$\tilde{a}_{hl} = \frac{a_{hl} \cdot a'_{hl}}{g_{hl}} \tag{3.8}$$

By inserting (3.6)–(3.8) into (3.5) we arrive at the following expression.

$$\hat{C}(\hat{t}_i, \hat{t}'_j|\Omega) = \sum_h \sum_l A_{hl} \cdot \left\{ \sum_{k \in P_{hl}} y_{ik} \cdot y'_{jk} \delta_k \cdot \delta'_k - \frac{1}{\tilde{a}_{hl}} \left(\sum_{k \in P_{hl}} y_{ik} \delta_k \right) \left(\sum_{r \in P_{hl}} y'_{jr} \delta'_r \right) \right\}, \tag{3.9}$$

where

$$A_{hl} = \frac{N_h N'_l \tilde{a}_{hl}}{G^2_{hl} n_h n'_l} \left(\frac{G_{hl}(G_{hl} - \tilde{a}_{hl})}{(\tilde{a}_{hl} - 1)} \right)$$

Orusild (2000) has shown relation (3.9) in a similar context. The complete conditional covariance,

$$\hat{C}(\hat{\theta}_0, \hat{\theta}_1|\Omega) = \sum_i \sum_j f'_i(\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_J) f'_j(\hat{t}'_1, \hat{t}'_2, \ldots, \hat{t}'_J) \hat{C}(\hat{t}_1, \hat{t}'_j|\Omega) \tag{3.10}$$

is obtained by inserting the Expressions (1.3), (1.6) and (3.9) into (3.1).

Computation by Formula (3.10) poses no problem in principle. However, in practical situations typically occurring at statistical agencies (3.10) must often be applied many times simultaneously, for many domains and possibly for several different functions *f*. How to organise the computations then becomes a nontrivial issue. In such cases it is practical to be able to rely on some standard software for variance/covariance estimation. In Appendix B, I present a procedure for computation of (3.10). The main ingredient of this procedure is a data transformation which makes it possible to rewrite (3.10) in the ''standard form'' for variances/ covariances under the STSI design. The estimate (3.10) can then be computed by application of standard software to the transformed data. It is also shown in Appendix B how the suggested method for estimation of the conditional covariance can be extended to the case when H–T estimation is replaced by GREG.

4. The Remainder Covariance Term

By Taylor linearisation we have

$$C_\Omega((E(\hat{\theta}_0|\Omega), E(\hat{\theta}_1|\Omega)) \approx \sum_i \sum_j f'_i(t_1, t_2, \ldots, t_J) f'_j(t'_1, t'_2, \ldots, t'_J) C_\Omega((E(\hat{t}_i|\Omega), E(\hat{t}'_j|\Omega)) \tag{4.1}$$

It follows from (1.3) and relations (*i*) and relations (*ii*) in Appendix A that

$$\begin{aligned} E(\hat{t}_i|\Omega) &= \sum_h \frac{N_h}{n_h} \left\{ \left(\frac{a_{h+}}{G_{h+}} \sum_{k \in D_h} y_{ik} \right) + \left(\sum_l \frac{a_{hl}}{G_{hl}} \sum_{k \in P_{hl}} y_{ik} \right) \right\} \\ &= \sum_h \left\{ \left(a_{h+} \left(\frac{N_h}{n_h} \cdot \bar{Y}_{i,h+} \right) + \sum_l a_{hl} \left(\frac{N_h}{n_h} \cdot \bar{Y}_{i,hl} \right) \right) \right\} \end{aligned} \tag{4.2}$$

where

$$\bar{Y}_{i,h+} = \frac{1}{G_{h+}} \sum_{k \in D_h} y_{i,k} \quad \text{and} \quad \bar{Y}_{i,hl} = \frac{1}{G_{hl}} \cdot \sum_{k \in P_{hl}} y_{i,k}$$

and correspondingly from (1.6) and relations (*iii*) and (*iv*) in Appendix A:

$$\begin{aligned} E(t'_j|\Omega) &= \sum_l \frac{N'_l}{n'_l} \left\{ \left(\frac{a'_{+l}}{G_{+l}} \sum_{k \in B_l} y'_{jk} \right) + \left(\sum_h \frac{a'_{hl}}{G_{hl}} \sum_{k \in P_{hl}} y'_{jk} \right) \right\} \\ &= \sum_l \left\{ \left(a'_{+l} \left(\frac{N'_l}{n'_l} \cdot \bar{Y}'_{j,+l} \right) + \sum_h a'_{hl} \left(\frac{N'_l}{n'_l} \cdot \bar{Y}'_{j,hl} \right) \right) \right\} \end{aligned} \tag{4.3}$$

where

$$\bar{Y}'_{j,+l} = \frac{1}{G_{+l}} \sum_{k \in B_l} y'_{jk} \quad \text{and} \quad \bar{Y}'_{j,hl} = \frac{1}{G_{hl}} \cdot \sum_{k \in P_{hl}} y'_{jk}$$

Considering the Expressions (4.2) and (4.3) it seems difficult to find a suitable analytical expression for $C_\Omega((E(\hat{t}_i|\Omega), E(\hat{t}'_j|\Omega))$ and consequently for $C_\Omega((E(\hat{\theta}_0|\Omega), E(\hat{\theta}_1|\Omega))$. However, keeping in mind that $a_{h+}, a_{hl}, a'_{+l}, a'_{hl}$ are the random variables, the following computer intensive procedure, which includes simulation of the sampling mechanism of SAMU, can be used for estimation of $C_\Omega((E(\hat{\theta}_0|\Omega), E(\hat{\theta}_1|\Omega))$.

Procedure 4.1:

Repeat the following steps 1–4 M times (M = 1000, say).

1) *For repetition m, $m = 1, 2, \ldots, M$: Generate and assign a uniformly distributed random number $R_k^{(m)}$ to every element k included in the union of the sample frames U and U'. All such random numbers are presumed to be independent.*
2) *For every (time 0−) stratum U_h, assign at repetition m a value $v_k(m) = 1$ to each of those n_h frame elements which have the smallest random numbers and assign a value $v_k(m) = 0$ to every other element in the stratum.*
3) *For every (time 1−) stratum U'_l, assign at repetition m a value $v'_k(m) = 1$ to each of those n'_l frame elements which have the smallest random numbers and assign a value $v'_k(m) = 0$ to every other element in the stratum.*
4) *Compute the following quantities*

$$a_{h+}(m) = \sum_{k \in D_h} v_k(m), \quad a_{hl}(m) = \sum_{k \in P_{hl}} v_k(m)$$

$$a'_{hl}(m) = \sum_{k \in P_{hl}} v'_k(m), \quad a'_{+l}(m) = \sum_{k \in B_l} v'_k(m)$$

$$\hat{u}_i(m) = \sum_h \left\{ \left(a_{h+}(m) \cdot \left(\frac{N_h}{n_h} \cdot \bar{y}_{i,h+} \right) + \sum_l a_{hl}(m) \cdot \left(\frac{N_h}{n_h} \cdot \bar{y}_{i,hl} \right) \right) \right\} \tag{4.4}$$

$$\hat{u}'_j(m) = \sum_l \left\{ \left(a'_{+l}(m) \cdot \left(\frac{N'_l}{n'_l} \cdot \bar{y}'_{j,+l} \right) + \sum_h a'_{hl}(m) \cdot \left(\frac{N'_l}{n'_l} \cdot \bar{y}'_{j,hl} \right) \right) \right\} \tag{4.5}$$

where the quantities $\bar{y}_{i,h+}, \bar{y}_{i,hl}, \bar{y}'_{j,+l}$ *and* $\bar{y}'_{j,hl}$ *are sample means (from the original SAMU samples) corresponding to* $\bar{Y}_{i,h+}, \bar{Y}_{i,hl}, \bar{Y}'_{j,+l}$ *and* $\bar{Y}'_{j,hl}$ *appearing in (4.2) and (4.3).*

5) *Use for* $m = 1, 2, \ldots, M$, *(4.4) and (4.5) as M independent estimates of* $E(\hat{t}_i|\Omega)$ *and* $E(\hat{t}'_j|\Omega))$ *respectively. Then estimate the covariance* $C_\Omega((E(\hat{t}_i|\Omega), E(\hat{t}'_j|\Omega))$ *by:*

$$\tilde{C}_{ij} = \frac{1}{M} \cdot \left\{ \sum_{m=1}^{M} (\hat{u}_i(m) \cdot \hat{u}'_j(m)) - \frac{\left(\sum_{m=1}^{M} \hat{u}_i(m)\right) \cdot \left(\sum_{m=1}^{M} \hat{u}'_j(m)\right)}{M} \right\} \qquad (4.6)$$

6) *Finally, estimate the desired covariance* $C_\Omega((E(\hat{\theta}_0|\Omega), E(\hat{\theta}_1|\Omega))$ *of (4.1) by*

$$\hat{C}_\Omega((E(\hat{\theta}_0|\Omega), E(\hat{\theta}_1|\Omega)) = \sum_i \sum_j f'_i(\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_J) f'_j(\hat{t}'_1, \hat{t}'_2, \ldots, \hat{t}'_J) \cdot \tilde{C}_{ij}. \qquad (4.7)$$

End of Procedure 4.1

Remark 4.1: A SAS program for implementation of Procedure 4.1 can be obtained from the author. When this program was applied to a case where the union $U \cup U'$ of the sampling frames included about 50,000 businesses – the size of the whole manufacturing sector in Sweden – it used about four hours of running time on a PC for $M = 1,000$ replications. Renewal of the SAMU samples is at present done only once or twice a year, though possibly it will be four times a year in the future. Hence Steps 1 to 4, which are the time-consuming parts of the procedure, need to be performed only a few times a year, which means that the procedure is workable.

Remark 4.2: Some stratum combinations P_{hl} with few frame elements may be unrepresented in the sample. It would then be impossible to compute the sample means $\bar{y}_{i,hl}$ and $\bar{y}'_{j,hl}$ which are to appear in (4.4) and (4.5) respectively. However, this problem is likely to appear only for stratum combinations where G_{hl}, and hence $a_{hl}(m)$ and $a'_{hl}(m)$, $m = 1, 2, \ldots, M$, are small. We then use the means $\bar{y}_{i,h}$ and $\bar{y}'_{j,l}$ in the original strata U_h and U'_l as approximations to $\bar{y}_{i,hl}$ and $\bar{y}'_{j,hl}$, and this approximation should not affect $\hat{u}_i(m)$ and $\hat{u}'_j(m)$ very much.

Next we show how Procedure 4.1 for estimation of the remainder term can be modified to cover more general estimation weights, such as the weights connected with the GREG estimator (see Särndal et al. 1992). Notice that the STSI design is still assumed in the following. Suppose that $\hat{t}_i$ can be written $\hat{t}_i = \sum_{k \in s} w_k y_{ik}$. Relation (4.2) now takes the following form.

$$\begin{aligned} E(\hat{t}_i|\Omega) &= \sum_h \left\{ \left(\frac{a_{h+}}{G_{h+}} \sum_{k \in D_h} w_k y_{ik} \right) + \left(\sum_l \frac{a_{hl}}{G_{hl}} \sum_{k \in P_{hl}} w_k y_{ik} \right) \right\} \\ &= \sum_h \left\{ (a_{h+} \cdot \bar{Y}^*_{i,h+}) + \sum_l a_{hl} \cdot \bar{Y}^*_{i,hl} \right\} \end{aligned} \qquad (4.8)$$

where

$$\bar{Y}^*_{i,h+} = \frac{1}{G_{h+}} \sum_{k \in D_h} w_k y_{i,k} \quad \text{and} \quad \bar{Y}^*_{i,hl} = \frac{1}{G_{hl}} \cdot \sum_{k \in P_{hl}} w_k y_{i,k}$$

The modification of Procedure 4.1 amounts to the following modification of (4.4) (and a completely analogous modification of (4.5)):

$$\hat{u}_i(m) = \sum_h \left\{ (a_{h+}(m) \cdot \bar{y}^*_{i,h+}) + \sum_l a_{hl}(m) \cdot \bar{y}^*_{i,hl})) \right\} \tag{4.9}$$

where

$$\bar{y}^*_{i,h+} = \frac{1}{a_{h+}} \sum_{k \in D_h} w_k \, y_{ik} \delta_k \quad \text{and} \quad \bar{y}^*_{i,hl} = \frac{1}{a_{hl}} \sum_{k \in P_{hl}} w_k \, y_{ik} \delta_k$$

δ_k being the sample inclusion indicator as defined in (1.4).

APPENDIX A: Some Conditional First and Second Order Inclusion Probabilities

Let assumptions and notation be as in Sections 2 and 3 above.

Lemma:

(i) $E(\delta_k|\Omega) = \dfrac{a_{hl}}{G_{hl}}, k \in P_{hl}$

(ii) $E(\delta_k|\Omega) = \dfrac{a_{h+}}{G_{h+}}, k \in D_h$

(iii) $E(\delta'_r|\Omega) = \dfrac{a'_{hl}}{G_{hl}}, r \in P_{hl}$

(iv) $E(\delta'_r|\Omega) = \dfrac{a'_{+l}}{G_{+l}}, r \in B_l$

(v) $E(\delta_k \delta'_r|\Omega) = E(\delta_k|\Omega) \cdot E(\delta'_r|\Omega)$, *for* $k \neq r$, if $(k \in P_{hl}, r \in P_{hl'})$ or $(k \in P_{hl}, r \in P_{h'l})$ or $(k \in D_h, r \in P_{hl})$ or $(k \in P_{hl}, r \in B_l)$

(vi) $E(\delta_k \delta'_k|\Omega) = \dfrac{g_{hl}}{G_{hl}}, k \in P_{hl}$

(vii) $E(\delta_k \delta'_r|\Omega) = \left(\dfrac{a_{hl}a'_{hl} - g_{hl}}{G_{hl}(G_{hl}-1)} \right), k \neq r, k, r \in P_{hl}$

Proof: (i) By symmetry $E(\delta_k|\Omega)$ must be a constant for all k $\in P_{hl}$. Set this constant to τ_{hl}. Then $\sum_{k \in P_{hl}} E(\delta_k|\Omega) = \tau_{hl} \cdot G_{hl}$

But $\sum_{k \in P_{hl}} E(\delta_k|\Omega) = E\left(\sum_{k \in P_{hl}} \delta_k|\Omega\right) = a_{hl}$ since $\sum_{k \in P_{hl}} \delta_k = a_{hl}$

Hence $\tau_{hl} = \dfrac{a_{hl}}{G_{hl}}$

(ii)–(iv) are proved analogously.

(v): Suppose that $k \in P_{hl}$ and $r \in P_{hl'}$:

By symmetry, $E(\delta_k \cdot \delta'_r|\Omega)$ must be the same ($= \tau_{(hl,hl')}$, say) for all such pairs of elements. Hence $E\left(\sum_{r \in P_{hl'}} \delta_k \cdot \delta'_r|\Omega\right) = \tau_{(hl,hl')} \cdot G_{hl'}$

But $E\left(\sum_{r \in P_{hl'}} \delta_k \cdot \delta'_r|\Omega\right) = E(\delta_k|\Omega) \cdot \left(\sum_{r \in P_{hl'}} \delta'_r\right)$, the latter term being a constant equal to $a'_{hl'}$.

Then $\tau_{(hl,hl')} = E(\delta_k|\Omega)\dfrac{a'_{hl'}}{G_{hl'}}$ and, due to (iii), $\tau_{(hl,hl')} = E(\delta_k|\Omega) \cdot E(\delta'_r|\Omega)$

An analogous argument can be used for the cases when $(k \in P_{hl}, r \in P_{h'l})$ or $(k \in D_h, r \in P_{hl})$ or $(k \in P_{hl}, r \in B_l)$. This completes the proof of (v).

(vi): Let $E(\delta_k \delta'_k|\Omega) = \tau_{hl}$. Hence $\sum_{k \in P_{hl}} E(\delta_k \delta'_k|\Omega) = \tau_{hl} \cdot G_{hl}$

But $\sum_{k \in P_{hl}} \delta_k \delta'_k = g_{hl}$ i.e., $\tau_{hl} = \dfrac{g_{hl}}{G_{hl}}$

(vii): Let $E(\delta_k \delta'_r|\Omega) = \tau_{hl}, k \neq$ r, since this expectation due to symmetry must be the same for all pairs k, r; $k \neq r$ in P_{hl}

Now, $E\left(\delta_k \sum_{r \in P_{hl}} \delta'_r|\Omega\right) = E(\delta_k \cdot \delta'_k|\Omega) + \tau_{hl}(G_{hl} - 1)$

But the left-hand side is $E(\delta_k \cdot a'_{hl}|\Omega) = a'_{hl} E(\delta_k|\Omega) = \dfrac{a_{hl} \cdot a'_{hl}}{G_{hl}}$

Hence with assistance of (vi) above: $\dfrac{a_{hl} a'_{hl}}{G_{hl}} = \dfrac{g_{hl}}{G_{hl}} + \tau_{hl}(G_{hl} - 1)$ which proves (vii).

End of proof

APPENDIX B: A Procedure for Computation of the Conditional Covariance Estimate

We derive a data transformation which will make it possible to write the covariance estimator (3.10) in the main text in a ''standard form'' which can be handled by a ''normal'' variance/covariance formula under STSI sampling.

We first show how the covariance formula (3.9) can be transformed so that it can be written in the usual form under the STSI design.

Set $z_{jk} = y_{jk} \cdot \dfrac{N_h \cdot \tilde{a}_{hl}}{n_h \cdot G_{hl}}$ and $z'_{jk} = y'_{jk} \cdot \dfrac{N'_l \cdot \tilde{a}_{hl}}{n'_l \cdot G_{hl}}$, $j = 1, 2, \ldots, J$, for $k \in P_{hl}$

$h = 1, 2, \ldots, H$, $l = 1, 2, \ldots, L$

Then (3.9) can be written as follows, i.e., by the usual formula for the estimator of the covariance between two π-weighted totals under STSI where strata comprise every combination of (h, l) in P_{hl}, the population size ''Capital N'' equals G_{hl} and the sample size ''small n'' equals $\tilde{a}_{hl}$.

$$\hat{C}(\hat{t}_i, \hat{t}'_j|\Omega) = \sum_h \sum_l \frac{G_{hl}(G_{hl} - \tilde{a}_{hl})}{\tilde{a}_{hl}(\tilde{a}_{hl} - 1)} \cdot \left\{ \sum_{k \in P_{hl}} z_{ik} \cdot z'_{jk} \delta_k \cdot \delta'_k - \frac{1}{\tilde{a}_{hl}} \left(\sum_{k \in P_{hl}} z_{ik} \delta_k \right) \left(\sum_{k \in P_{hl}} z'_{jk} \delta'_k \right) \right\} \tag{b.1}$$

However, to catch the whole covariance term (3.10) one must also compute the point estimates $\hat{t}_i$ and $\hat{t}'_j$ in accordance with (1.3) and (1.6) respectively. This means that those observations that are not included in P_{hl} but contribute to the partial derivatives in (3.10) must also be taken into account. We will now introduce the data transformation mentioned above.

Procedure B.1

1) *Put every sample element (which appeared in* ***either*** *one of the time 0 and time 1 samples) into the correct group among:*

$$\boldsymbol{q} = (q_1, q_{2,\ldots,} q_Q) = (D_h, P_{hl}, B_l, h = 1, 2, \ldots, H, l = 1, 2, \ldots, L)$$

where

- D_h *includes the elements in stratum* U_h *at time 0 which disappeared from the frame before time 1, i.e. the deaths in stratum* U_h.

- *P_{hl} (type 1, 2 or 3) includes the elements which belonged to stratum U_h at time 0 and stratum U'_l at time 1. Notice that elements that appear in only one of the two samples must also be included, not only the ones appearing in both samples.*
- *B_1 includes the elements that are not found in the frame at time 0 but that belong to stratum U'_l at time 1, i.e. the births in stratum U'_l.*

2) Consider group q where $q = P_{hl}$ (Type1), $h = 1, 2, \ldots H, l = 1, 2, \ldots, L$

- *Set $\tilde{N}_q = G_{hl}$ and $\tilde{n}_q = \tilde{a}_{hl}$*
- *Transform y_{jk} and y'_{jk} to z_{jk} and z'_{jk} for $j = 1, 2, \ldots, J$ as follows. If $k \in P_{hl}$, $h = 1, 2, \ldots H, l = 1, 2, \ldots, L$.:*

$$z_{jk} = \begin{cases} y_{jk} \cdot \dfrac{N_h \cdot \tilde{a}_{hl}}{n_h \cdot G_{hl}} & \text{if } k \in \text{s} \\ 0 & \text{if } \ k \notin \text{s} \end{cases}$$

$$z'_{jk} = \begin{cases} y'_{jk} \cdot \dfrac{N'_l \cdot \tilde{a}_{hl}}{n'_l \cdot G_{hl}} & \text{if } k \in \text{s}' \\ 0 & \text{if } \ k \notin \text{s}' \end{cases}$$

3) Consider group q where $q = P_{hl}$ (Type2)

- *Set* $\tilde{N}_q = \begin{cases} N_h & \text{if } a_{hl} >= 1 \text{ and } a'_{hl} = 0 \\ N'_l & \text{if } a_{hl} = 0 \text{ and } a'_{hl} >= 1 \end{cases}$
- *Set* $\tilde{n}_q = \begin{cases} n_h & \text{if } a_{hl} >= 1 \text{ and } a'_{hl} = 0 \\ n'_l & \text{if } a_{hl} = 0 \text{ and } a'_{hl} >= 1 \end{cases}$
- *Transform y_{jk} and y'_{jk} to z_{jk} and z'_{jk} for $j = 1, 2, \ldots, J$ as follows. If $k \in P_{hl}$, $h = 1, 2, \ldots, H, l = 1, 2, \ldots, L.$:*

$$z_{jk} = \begin{cases} y_{jk} & \text{if } k \in \text{s} \\ 0 & \text{if } k \notin \text{s} \end{cases}$$

$$z'_{jk} = \begin{cases} y'_{jk} & \text{if } k \in \text{s}' \\ 0 & \text{if } k \notin \text{s}' \end{cases}$$

4) Consider group q where $q = P_{hl}$ (Type3)

- *Set $\tilde{N}_q = G_{hl}$ and $\tilde{n}_q = G_{hl}$*
- *Transform y_{jk} and y'_{jk} to z_{jk} and z'_{jk} for $j = 1, 2, \ldots, J$ as follows. If $k \in P_{hl}$, $h = 1, 2, \ldots, H, l = 1, 2, \ldots, L.$:*

$$z_{jk} = \begin{cases} y_{jk} \cdot \dfrac{N_h}{n_h} & \text{if } k \in \text{s} \\ 0 & \text{if } \ k \notin \text{s} \end{cases}$$

$$z'_{jk} = \begin{cases} y'_{jk} \cdot \dfrac{N'_l}{n'_l} & \text{if } k \in \text{s}' \\ 0 & \text{if } \ k \notin \text{s}' \end{cases}$$

5) Consider group q where $q = D_h, h = 1, 2, \ldots, H$

- *Set $\tilde{N}_q = N_h$ and $\tilde{n}_q = n_h$*

- *Transform* y_{jk} and y'_{jk} *to* z_{jk} and z'_{jk} *for* $j = 1, 2, \ldots, J$ *as follows:*
  If $k \in D_h, h = 1, 2, \ldots, H.:$ $z_{jk} = y_{jk}$ *and* $z'_{jk} = 0$

6) *Consider group q where* $q = B_l$, $l = 1, 2, \ldots, L$

- *St* $\tilde{N}_q = N'_l$ *and* $\tilde{n}_q = n'_l$
- *Transform* y_{jk} and y'_{jk} *to* z_{jk} and z'_{jk} *for* $j = 1, 2, \ldots, J$ *as follows.*
  If $k \in B_l$, $l = 1, 2, \ldots, L.:$ $z_{jk} = 0$ *and* $z'_{jk} = y'_{jk}$

End of Procedure B.1

It only takes some elementary algebra to see that (b2) och (b3) below are equivalent to (1.3) and (1.6) respectively. Furthermore, the expression (b4) is equivalent to (3.9). Notice that the contributions to (b4) from D_h, B_l and P_{hl} of types 2 and 3 are zero as intended.

$$\hat{t}_j = \sum_q \frac{\tilde{N}_q}{\tilde{n}_q} \sum_{k \in q} z_{jk} \tag{b2}$$

$$\hat{t}'_j = \sum_q \frac{\tilde{N}_q}{\tilde{n}_q} \sum_{k \in q} z'_{jk} \tag{b3}$$

$$\hat{C}(\hat{t}_i, \hat{t}'_j|\Omega) = \sum_q \frac{\tilde{N}_q(\tilde{N}_q - \tilde{n}_q)}{\tilde{n}_q(\tilde{n}_q - 1)} \cdot \left\{ \sum_{k \in q} z_{ik} \cdot z'_{jk} \cdot \delta_k \cdot \delta'_k - \frac{1}{\tilde{n}_q} \left(\sum_{k \in q} z_{ik} \delta_k \right) \left(\sum_{k \in q} z'_{jk} \delta'_k \right) \right\} \tag{b4}$$

Hence by generating the pseudo data z and z' as in *Procedure B.1* and then applying standard formulas for point-, variance- and covariance-estimators under the STSI design with q serving as strata, $\tilde{N}_q$ serving as population size parameter (Capital N) and $\tilde{n}_q$ as sample size parameter (small n) we can compute the covariance (3.9) in the main text by (b4).

Furthermore, suppose that a software is available which – under this STSI design – can compute proper variance estimates for $\hat{\theta}_0, \hat{\theta}_1$ and $\hat{\theta}_0 + \hat{\theta}_1$ where $\hat{\theta}_0 = f(\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_J)$, $\hat{\theta}_1 = f(\hat{t}'_1, \hat{t}'_2, \ldots, \hat{t}'_J)$ and $\hat{t}_j$ and $\hat{t}'_j$ are defined by (b2) and (b3). Then the covariance estimate of (3.10) can be extracted from the following general relation:

$$\hat{V}(\hat{\theta}_0 + \hat{\theta}_1) = \hat{V}(\hat{\theta}_0) + \hat{V}(\hat{\theta}_1) + 2 \cdot \hat{C}(\hat{\theta}_0, \hat{\theta}_1) \tag{b5}$$

Notice that this software must also meet the following requirements:

– Since $\tilde{n}_q$ may not necessarily be an integer, the software must be able to accept arbitrary values $\tilde{n}_q >= 1$ for the sample size parameter (small n). Softwares that compute n by counting elements in the input data set will not be appropriate for this task. The software must be able to accept the value $\tilde{n}_q = 1$ which can possibly appear in some extreme strata.
– Even though $\tilde{N}_q$ is normally larger than $\tilde{n}_q$ there is no absolute guarantee for this. As a consequence the software must be able to accept negative variance contributions from some (extreme) strata. Notice that the quantities $\tilde{V}$ that appear in step c of *Procedure B2* below are based on the pseudo data and may include negative contributions from certain strata whenever $\tilde{a}_{hl} > G_{hl}$ (see (3.7)).

The software CLAN, developed at Statistics Sweden (see Andersson and Nordberg 1994, 1998), meets the above-mentioned requirements and can readily be used to implement *Procedure B2*.

Procedure B.2 (Estimation of the conditional covariance)

(i) Perform procedure B.1
(ii) Compute the estimates $\tilde{V}(\tilde{\theta}_0), \tilde{V}(\tilde{\theta}_1)$ *and* $\tilde{V}(\tilde{\theta}_0 + \tilde{\theta}_1)$ *for* $V(\tilde{\theta}_0), V(\tilde{\theta}_1)$ *and* $V(\tilde{\theta}_0 + \tilde{\theta}_1)$ *based on the pseudo data generated in (i). The tilde sign symbolizes the fact that computations are based on the pseudo data.*
(iii) Estimate the covariance (3.10) by the following relation:

$$\tilde{C} = 0.5\cdot(\tilde{V}(\tilde{\theta}_0 + \tilde{\theta}_1) - \tilde{V}(\tilde{\theta}_0) - \tilde{V}(\tilde{\theta}_1)) \tag{b6}$$

End of Procedure B.2

Remark 1: Procedure B.1 – which generates the input data for the covariance estimation – must be implemented separately. A SAS-program for implementation of Procedure B.1 can be obtained from the author.

Remark 2: Notice that the above procedure is intended only for the conditional covariance term (3.1) in the main text which contributes to the covariance appearing in (1.1) and (1.2). Estimates of the variance components $V(\hat{\theta}_0)$ and $V(\hat{\theta}_1)$ appearing in (1.1) and (1.2) are computed in a conventional manner.

We now show how the procedure presented above can be modified to incorporate the case when the H-T estimators of (1.3) and (1.6) respectively are replaced by GREG estimators. See Särndal, Swensson, and Wretman (1992) for a comprehensive presentation of the GREG estimator.

The variance- and covariance-estimators for the H-T estimator presented above can be extended to the GREG case. The only modification needed is that the y_k:s are replaced by the product $(g_{ks}\cdot e_{ks})$ where

$$g_{ks} = \left\{ 1 + (\mathbf{t}_x - \hat{\mathbf{t}}_x)^T \cdot \left(\sum_{k \in s} \frac{\mathbf{x}_k \mathbf{x}_k^T \cdot \tau_k}{\pi_\mathbf{k}} \right)^{-1} \cdot \mathbf{x}_k \cdot \tau_k \right\} \tag{b7}$$

$$e_{ks} = y_k - \mathbf{x}_k^T \cdot \hat{\mathbf{B}} \tag{b8}$$

and

$$\hat{\mathbf{B}} = \left(\sum_{k \in s} \frac{\mathbf{x}_k \mathbf{x}_k' \tau_k}{\pi_k} \right)^{-1} \sum_{k \in s} \frac{\mathbf{x}_k y_k \tau_k}{\pi_k} \tag{b9}$$

π_k being the inclusion probability for element k while $\tau_k > 0$ is a scale parameter which can be set by the user. The quantity $\mathbf{x}_k$ is the vector of covariates involved in the GREG estimator, while $\mathbf{t}_\mathbf{x}$ is the vector of population totals (assumed known) for $\mathbf{x}$ and $\hat{\mathbf{t}}_\mathbf{x}$ is the corresponding H-T estimator for $\mathbf{x}$ from the sample.

Procedure B.1, and consequently Procedure B.2, can be applied to the GREG case if the data transformation in procedure B.1 is applied to y and to every covariate x. However, one problem then arises: It follows from (b7) and (b9) that $\hat{\mathbf{B}}$ and g_{ks} will be affected (which they should not be). However, this can be adjusted for in the following way.

Suppose that $k \in P_{hl}$ $(Type1)$. The data transformation of y follows from Procedure B.1, and takes the form

$$\tilde{y}_k = \begin{cases} y_k \cdot \dfrac{N_h \cdot \tilde{a}_{hl}}{n_h \cdot G_{hl}} & \text{if } k \in \mathrm{s} \\ 0 & \text{if } k \notin \mathrm{s} \end{cases} \tag{b10}$$

The corresponding transformation for the covariate x_j is

$$\tilde{x}_{jk} = \begin{cases} x_{jk} \cdot \dfrac{N_h \cdot \tilde{a}_{hl}}{n_h \cdot G_{hl}} & \text{if } k \in \mathrm{s} \\ 0 & \text{if } k \notin \mathrm{s} \end{cases} \tag{b11}$$

Furthermore, $\pi_k = \dfrac{n_h}{N_h}$ under the STSI design while the corresponding transformed value should be $\tilde{\pi}_k = \dfrac{\tilde{a}_{hl}}{G_{hl}}$ (see Procedure B.1)

Hence $\tilde{\pi}_k = \pi_k \cdot \dfrac{N_h \cdot \tilde{a}_{hl}}{n_h \cdot G_{hl}}$. Furthermore, modify the scale factor τ_k as follows: $\tilde{\tau}_k = \tau_k \cdot \dfrac{n_h \cdot G_{hl}}{N_h \cdot \tilde{a}_{hl}}$. We then have the following relations

$$\frac{\tilde{\mathbf{x}}_k \tilde{\mathbf{x}}_k^T \tilde{\tau}_k}{\tilde{\pi}_k} = \frac{\mathbf{x}_k \mathbf{x}_k^T \tau_k}{\pi_k}, \quad \frac{\tilde{\mathbf{x}}_k \tilde{y}_k \tilde{\tau}_k}{\tilde{\pi}_k} = \frac{\mathbf{x}_k y_k \tau_k}{\pi_k} \quad \text{and} \quad \tilde{\mathbf{x}}_k \tilde{\tau}_k = \mathbf{x}_k \tau_k$$

Hence $\hat{\mathbf{B}}$ and g_{ks} are unaffected if the scale factor τ_k is modified in such a way that τ_k is replaced by $\tau_k \cdot \dfrac{n_h \cdot G_{hl}}{N_h \cdot \tilde{a}_{hl}}$ if $k \in s$ and by $\tau_k \cdot \dfrac{n_l' \cdot G_{hl}}{N_l' \cdot \tilde{a}_{hl}}$ if $k \in s'$. After these modifications the procedure for the estimation of the conditional covariance applies also when the GREG estimator is used as estimator for the population totals involved.

5. References

Andersson, C. and Nordberg, L. (1994): A Method for Variance Estimation of Non-Linear Functions of Totals in Surveys – Theory and a Software Implementation. Journal of Official Statistics, 10, 395–405.

Andersson, C. and Nordberg, L. (1998): A User's Guide to CLAN 97 – SAS Program for Computation of Point and Standard Error Estimates in Sample Surveys. Statistics Sweden.

Garås, T. (1989): Estimators of Change in Dynamic Populations. Memo, Statistics Sweden. (In Swedish).

Hidiroglou, M., Särndal, C. E., and Binder, D. (1995): Weighting and Estimation in Business Surveys. In Business Survey Methods, eds. B.G. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M. Colledge, and P.S. Kott, Wiley.

Laniel, N. (1988). Variances for a Rotating Sample from a Changing Population. Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp. 246–250.

Ohlsson, E. (1992). SAMU – The System for Co-Ordination of Samples from the Business Register at Statistics Sweden. R&D Report, 1992: 18, Statistics Sweden.

Ohlsson, E. (1995). Coordination of Samples Using Permanent Random Numbers. In Business Survey Methods, eds. B. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M. Colledge, and P.S. Kott, New York: Wiley & Sons.

Orusild, T. (2000). Confidence Intervals for Functions of Quantiles under Finite Population Sampling. Technical Report, Department of Mathematical Statistics, University of Stockholm, (forthcoming).

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

Tam, S.M. (1984). On Covariances from Overlapping Samples. The American Statistician, 38, pp. 288–292.

Received January 1999
Revised March 2000