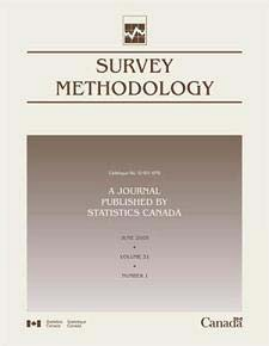
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Efficient bootstrap for business surveys



by James Chipperfield and John Preston

December, 2007

Statistics Canada Statistique Canada

Canada

Survey Methodology, December 2007
Vol. 33, No. 2, pp. 167-172
Statistics Canada, Catalogue no. 12-001-X

Efficient bootstrap for business surveys

James Chipperfield and John Preston [1]

Abstract

The Australian Bureau of Statistics has recently developed a generalized estimation system for processing its large scale annual and sub-annual business surveys. Designs for these surveys have a large number of strata, use Simple Random Sampling within Strata, have non-negligible sampling fractions, are overlapping in consecutive periods, and are subject to frame changes. A significant challenge was to choose a variance estimation method that would best meet the following requirements: valid for a wide range of estimators (*e.g.*, ratio and generalized regression), requires limited computation time, can be easily adapted to different designs and estimators, and has good theoretical properties measured in terms of bias and variance. This paper describes the Without Replacement Scaled Bootstrap (WOSB) that was implemented at the ABS and shows that it is appreciably more efficient than the Rao and Wu (1988)'s With Replacement Scaled Bootstrap (WSB). The main advantages of the Bootstrap over alternative replicate variance estimators are its efficiency (*i.e.*, accuracy per unit of storage space) and the relative simplicity with which it can be specified in a system. This paper describes the WOSB variance estimator for point-in-time and movement estimates that can be expressed as a function of finite population means. Simulation results obtained as part of the evaluation process show that the WOSB was more efficient than the WSB, especially when the stratum sample sizes are sometimes as small as 5.

Key Words: Variance; Bootstrap; Stratified sampling.

1. Introduction

In 2000, the Australian Bureau of Statistics (ABS) first obtained a register of businesses containing taxation data from the Australian Taxation Office (ATO). The data items included turnover, sales, and other expense items. In 2001, the ABS used this register as a sampling frame for some surveys in order to improve the efficiency of its sample designs. This data is updated for each business at least annually. To make maximum use of these administrative data items in estimation the ABS developed a generalized estimation system called ABSEST, with the capability of supporting generalized regression estimation (GREG) and variance estimation. ABSEST has been routinely used for the monthly ABS Retail Survey since July 2005.

A generalized estimation system is highly desirable for statistical agencies as it supports a variety of survey output requirements at high levels of statistical rigor for an acceptable cost. The ABS has invested considerable resources into its generalized estimation system for business surveys. Prior to 1998, the ABS's generalized estimation system was capable of Horvitz-Thompson, ratio, and two-phase estimation with variance estimates based on Taylor Series (TS) approximations. In 1999, the Taylor Series method was replaced with the Jackknife method. Subsequent feedback about the computer design and usability were that changes to the generalized estimation system made it increasingly complex to maintain and develop and that processing time could be undesirably long. These key features were important when choosing the variance estimation method for ABSEST.

Core survey output statistics for ABS business surveys are estimates at a point in time, estimates of movement between two time points, and estimates of rates. Business surveys are equal probability designs within stratum, are highly stratified (100s of strata), can be either single or two phase sample designs, and for surveys that sample on more than one occasion the overlapping sample can range from 0 to 100%. The sample size for business surveys range from less than 1,000 to 15,000; stratum level sample sizes can be as low as 3 and as high as several hundred.

Section 2 introduces the GREG estimator. Section 3 discusses alternative variance estimators for GREG and justifies why the Bootstrap variance estimator was chosen for ABSEST. Section 4 describes the Without Replacement Scaled Bootstrap (WOSB) and Rao and Wu (1988)'s With Replacement Bootstrap (WSB) variance estimators for point-in-time estimates under single-phase designs. Section 5 describes the WOSB for movement estimates. Section 6 measures the bias and variance properties of WOSB and WSB in a simulation study. Section 7 gives some concluding remarks.

2. Generalised regression (GREG) estimator

In this section we briefly describe the GREG that is implemented in ABSEST. Consider a finite population U divided into H strata $U = \{U_1, U_2, ..., U_H\}$, where U_h is comprised of N_h units. The finite population total of interest is $Y = \sum_h Y_h$, where $Y_h = \sum_{i \in U_h} y_{hi}$ and $h = 1, ..., H$. Within stratum h, the sample s_h of n_h units is selected

1. James Chipperfield, Australian Bureau of Statistics. E-mail: james.chipperfield@abs.gov.au; John Preston, Australian Bureau of Statistics. E-mail: john.preston@abs.gov.au.

from U_h by Simple Random Sampling without Replacement (SRSWOR). The complete sample set is denoted by $s = \{s_1, s_2, ..., s_H\}$.

Consider the case where a K vector of auxiliary variables $\boldsymbol{x}_i = (x_{1i}, ..., x_{ki}, ..., x_{Ki})^T$ is available for $i \in s$ and the corresponding vector of population totals $\boldsymbol{X} = \sum_{i \in U} \boldsymbol{x}_i$ are known. The GREG estimator (Särndal, Swensson and Wretman 1992, page 227) is given by $\hat{Y}_{\text{reg}} = \sum_{i \in s} \tilde{w}_i y_i = \sum_{i \in s} w_i y_i + (\boldsymbol{X} - \hat{\boldsymbol{X}})^T \hat{\boldsymbol{B}}$, where $\tilde{w}_i = w_i g_i$, $w_i = N_h / n_h$, $\hat{\boldsymbol{B}} = \hat{\boldsymbol{T}}^{-1} \hat{\boldsymbol{t}}$ with $\hat{\boldsymbol{T}}^{-1}$ being the generalised inverse of $\hat{\boldsymbol{T}}$, $\hat{\boldsymbol{X}} = \sum_{i \in s} w_i \boldsymbol{x}_i$ and $g_i = (1 + \sigma_i^{-2} \boldsymbol{x}_i^T \hat{\boldsymbol{T}}^{-1} (\boldsymbol{X} - \hat{\boldsymbol{X}})^T)$, $\hat{\boldsymbol{T}} = \sum_{i \in s} w_i \boldsymbol{x}_i \boldsymbol{x}_i^T \sigma_i^{-2}$, $\hat{t} = \sum_{i \in s} w_i \boldsymbol{x}_i y_i \sigma_i^{-2}$, σ_i^2 is a constant motivated by the superpopulation model $y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i$ such that ε_i is independently and identically distributed with mean 0 and variance σ_i^2, and $E(\hat{\boldsymbol{B}}) = \boldsymbol{\beta}$. It is well known that $\hat{Y}_{\text{reg}}$ is unbiased to $O(n^{-1})$. The weights $\tilde{w}_i$ are stored for ready calculation of estimates. In practice bounds will be placed on the weights, $\tilde{w}_i$. If the weights, $\tilde{w}_i$, given by the above equation, are outside these bounds, they are calculated through iteration (see Method 5 of Singh and Mohl 1996).

The expression for $\hat{Y}_{\text{reg}}$ can be adapted to a range of estimates, including domains and multi-phase (see Estevao, Hidiroglou and Särndal 1995). For example, when $\boldsymbol{x}_i = 1$, $\hat{Y}_{\text{reg}}$ becomes the Horvitz-Thompson estimator given by $\hat{Y} = \sum_h w_h \sum_{i \in s_h} y_{hi}$ with estimated variance $\widehat{\text{Var}}(\hat{Y}) = \sum_h N_h^2 n_h^{-1} (1 - f_h) \hat{s}_h^2$, where $\hat{s}_h^2 = (n_h - 1)^{-1} \sum_{i \in s_h} (y_{hi} - \hat{\bar{y}}_h)^2$, $\hat{\bar{y}}_h = \sum_{i \in s_h} y_{hi} / n_h$ and $f_h = n_h / N_h$.

3. Comparison of alternative variance estimators

The ABSEST variance estimation method was required to have bias and variance properties that were competitive in simulation studies, when compared with alternatives in the literature. In order to simplify the maintenance and development of the system, the variance estimation system specifications were required to be generic such that all calculations were largely independent of the estimator. (ABSEST need only support SRSWOR within stratum and single stage sample designs). Also, strong consideration was given to minimise the computational costs.

Firstly, we also considered the Bootstrap, Jackknife and Balanced Repeated Replication (BRR) methods (Shao and Tu 1995, Rao and Wu 1988). Consider estimating the variance of a function $\hat{\theta} = \theta(\hat{\boldsymbol{Y}})$, where $\hat{\boldsymbol{Y}}$ is a P vector of estimates $\hat{\boldsymbol{Y}} = \sum_{i \in s} w_i y_i$, $\boldsymbol{y}_i$ is the response vector from unit i with elements y_{pi}, and θ is a smooth function. Estimating the variance using a replication method involves the following steps:

(i) independently sub-sampling from the set s a total of R times;

(ii) for each of the R sub-samples computing $w_i^* = b_i^* w_i$, where b_i^* depends upon the number of times unit i is selected in the sub-sample;

(iii) calculate $\hat{\theta}^* = \hat{\theta}(\hat{\boldsymbol{Y}}^*)$, where $\hat{\boldsymbol{Y}}^* = \sum_{i \in s} w_i^* \boldsymbol{y}_i$;

(iv) estimate variance of θ by $\widehat{\text{Var}}_{\text{rep}}(\hat{\theta}) = (R-1)^{-1} \sum_{r=1}^{R} (\hat{\theta}^{*(r)} - \hat{\theta})^2$, where $\hat{\theta}^{*(r)}$ is the estimate of θ based on the rth replicate sample. Note: the expression for replicate weights, $w_i^* = b_i^* w_i$, includes the Jackknife, Bootstrap and Balanced Repeated Replication as special cases.

As we can express $\hat{Y}_{\text{reg}}$ by a function $\hat{\theta}$, the variance of $\hat{Y}_{\text{reg}}^*$ can be calculated by the above steps where specifically steps (iii) and (iv) respectively become: (iii) calculate $\hat{Y}_{\text{reg}}^* = \sum_{i \in s} \tilde{w}_i^* y_i$, where $\tilde{w}_i^* = w_i^* g_i^*$ and g_i^* has the same form as g_i but is calculated using the weights w_i^* instead of the weights w_i for $i \in s$; (iv) estimating variance by $\widehat{\text{Var}}_{\text{rep}}(\hat{Y}_{\text{reg}}) = (R-1)^{-1} \sum_{r=1}^{R} (\hat{Y}_{\text{reg}}^{*(r)} - \hat{Y}_{\text{reg}})^2$.

The attractive feature of these replication methods is that only the selection of the replicate samples and the value b_i^* is required to calculate unbiased variance estimates for many commonly used sample designs and for estimators that have good first order Taylor Series approximations. Also if the replicate weights, $\tilde{w}_i^*$ are stored the variance estimates of $\hat{Y}_{\text{reg}}^*$ require simple calculations that can be completed in a short time; this approach of storing replicate weights has been applied successfully by the ABS' generalized estimation system for household surveys. Once replicate weights are available, calculation of variance for a variety of analysis, such as linear regression, involves simple calculations that require little time and does not require the analyst to have knowledge about the sample design. Next we consider the relative merits of some replicate variance estimators for implementation in ABSEST.

The drop-one Jackknife forms replicate samples, s^*, by dropping one unit at a time. This implies that $R = n$. For large-scale surveys this storage requirement is excessive. The delete-a-group Jackknife, while reducing R by dropping a group of units within a stratum at a time, would still have at least $R = 2H$ replicates- a minimum of two groups per stratum is required to calculate variance. Despite performing well in an empirical study where $n_h = 2$ (see Shao and Tu 1995, page 251), the Jackknife was rejected on the basis of its excessive storage requirement.

For stratified designs the scaled Balanced Repeated Replication (BRR) requires approximately $R = H$ replicate weights. Firstly, the replicate samples are formed by randomly splitting the stratum sample s_h into two groups then allocating one of these groups to s_h^* for each

$h = 1, ..., H$. The allocation of groups to replicates, defined by a Hadamard matrix, is done in such a way to eliminate between stratum covariance in the replicate samples. The Grouped BRR (GBRR) (see Shao and Tu 1995) can arbitrarily reduce R at the cost of introducing between stratum covariance in the replicate samples. Preston and Chipperfield (2002) showed in an empirical evaluation for a typical ABS business survey that BRR (and GBRR) was significantly more unstable than the Bootstrap.

In their summary of the literature, Kovar, Rao and Wu (1988) found that the scaled Bootstrap tended to have a larger bias compared with the Jackknife or TS when estimating the variance of GREG estimates. As the relative assessment of these methods varied according to the underlying simulation model and the stratum sample size it was important to make an assessment that was based on a model and sample design that were typical of ABS business surveys. Section 6 shows these properties to be acceptable. Unlike the other replication methods, the value of R for the Bootstrap may be chosen arbitrarily and so meet storage and computation restrictions. Further, the selection of the Bootstrap replicate samples is more easily specified in a computer system compared with selection of the BRR replicate samples.

We considered the relative merits of a number of other variance estimators for implementation in ABSEST. The TS method was not suitable as its variance expression for complex estimands involves many terms specific to the estimand, making it difficult to adapt into a generalized system. This problem was addressed by Nordberg (2000) who described a method for variance calculation that automatically generates the Taylor Series expansions. They implemented the method in a computer system, called CLAN. CLAN can handle any function of means under Probability Proportional to Size (PPS) and cluster sampling. A limitation of CLAN is that it does not produce replicate weights which support complex analysis, such as regression analysis, either within or outside a statistical agency; it requires knowledge of the sample design; and it would be a relatively complex system to specify and maintain in a generalized system (in comparison to the Bootstrap). For the same reason other linearized variance estimators (described in Estevao, Hidiroglou and Särndal 1995 and evaluated in Yung and Rao 1996) were rejected, despite good theoretical properties, good empirical results and being computationally efficient.

On the above considerations, the preferred variance estimation method for ABSEST was the Bootstrap. In the next section we describe the WOSB and WSB, where only the former is implemented in ABSEST.

4. Without replacement scaled bootstrap (WOSB) for point in time estimates

4.1 Method

For point-in-time GREG estimates, the Without Replacement Scaled Bootstrap (WOSB) variance estimator involves repeating the following R times:

(a) forming the set s^r by selecting m_h units by SRSWOR from s_h independently within each stratum $h = 1, ..., H$, where $m_h = [n_h / 2]$ and the operator [.] rounds down its argument down to the nearest integer;

(b) calculating $w^*_{hi} = w_{hi}(1 - \gamma_h + \gamma_h\, n_h / m_h\, \delta^*_{hi})$ for $i \in s_h$, where $\gamma_h = \sqrt{(1 - f_h) m_h / (n_h - m_h)}$, δ^*_{hi} is 1 if $i \in s^r_h$ and 0 otherwise; and

(c) calculating $\tilde{w}^*_{hi} = w^*_{hi}\, g^*_{hi}$ for $i \in s$; and

(d) calculating the rth Bootstrap estimate of Y, $\hat{Y}^*_{\text{reg}} = \sum_{i \in s} \tilde{w}^*_{hi}\, y_i$. The justification $m_h = [n_h / 2]$ is given in section 4.2. The Bootstrap variance estimator is given by the Monte Carlo approximation, $\widehat{\text{Var}}_B(\hat{Y}_{\text{reg}}) = (R-1)^{-1} \sum_{r=1}^{R} (\hat{Y}^{*(r)}_{\text{reg}} - \hat{Y}_{\text{reg}})^2$. The WSB method is the same as WOSB except that the replicate samples are selected by SRSWR and the scaling factor is instead $\gamma_h = \sqrt{(1-f_h) m_h / (n_h - 1)}$, where m_h is often set to $n_h - 1$ in the literature. Preston and Chipperfield (2002) found that WOSB was found to have significantly less replication error than the WSB- the error due to replicate sampling and conditional on the sample set.

It is easy to see that the WOSB and WSB estimators are unbiased estimators of $\text{Var}(\hat{\theta})$. The TS approximate variance is given by $\widehat{\text{Var}}(\hat{\theta}) = \nabla' \hat{\theta}\, \hat{V}(\hat{\boldsymbol{Y}}) \nabla \hat{\theta}$, where $\hat{V}(\hat{\boldsymbol{Y}})$ is a $P \times P$ matrix with elements

$$\widehat{\text{Cov}}(\hat{Y}_p, \hat{Y}_{p'}) = \frac{N^2(1 - f)}{n}\, \hat{s}_{p,\, p'}$$

where

$$\hat{s}_{p,\, p'} = \frac{1}{n-1} \sum_{i \in s} (y_{pi} - \hat{\bar{y}}_p)(y_{p'i} - \hat{\bar{y}}_{p'});$$

$$\hat{\bar{y}}_p = \frac{1}{n} \sum_{i \in s} y_{pi};$$

$$\hat{Y}_p = \sum_{i \in s} w_i\, y_{pi}$$

for $p, p' = 1, ..., P$, and $\nabla' = (\partial / \partial Y_1, ..., \partial / \partial Y_P)|_{\hat{\boldsymbol{Y}}}$. It is easy to see that

$$E_*(\widehat{\text{Var}}(\hat{\theta}^*)) = \nabla \hat{\theta}' E_*(\hat{V}(\hat{\boldsymbol{Y}}^*)) \nabla \hat{\theta} = \nabla' \hat{\theta}\, \hat{V}(\hat{\boldsymbol{Y}}) \nabla \hat{\theta},$$

by noting that

$$E_*[\widehat{\text{Cov}}(\hat{Y}^*_p, \hat{Y}^*_{p'})] = \widehat{\text{Cov}}(\hat{Y}_p, \hat{Y}_{p'})$$

where E_* denotes the expectation with respect to re-sampling. Note the scaling constants applied to w_{hi} to calculate the replicate weights are chosen so that the correct finite population correction factor is obtained. It therefore follows that the Monte Carlo approximation to the variance, $\widehat{\mathrm{Var}}_B(\hat{\theta}) = (R-1)^{-1}\sum_{r=1}^{R}(\hat{\theta}^{*(r)} - \hat{\theta})^2$, is unbiased for $\mathrm{Var}\,(\hat{\theta})$.

4.2 A note on the relative efficiency of WSB and WOSB sampling

To simplify notation, let $\hat{v}_{\text{boot}} = \widehat{\mathrm{Var}}_B(\hat{\theta})$. The variance of the Bootstrap variance estimator can be written as

$$\mathrm{Var}\,(\hat{v}_{\text{boot}}) = \mathrm{Var}_s(E_*[\hat{v}_{\text{boot}} \mid s]) + E_s[\mathrm{Var}_*(\hat{v}_{\text{boot}} \mid s)],$$

where s denotes the expectation with respect to the sample design. If $\hat{v}_{\text{boot}}$ is unbiased (*i.e.*, $E_*[\hat{v}_{\text{boot}} \mid s] = \mathrm{Var}(\hat{\theta})$) then $\mathrm{Var}\,(\hat{v}_{\text{boot}})$ does not depend upon how the replicate samples are selected. The term $\mathrm{Var}_*(\hat{v}_{\text{boot}} \mid s)$ is the replication error conditional on the sample and is inversely proportional to R. The value of R is chosen to be sufficiently large such that $\mathrm{Var}_s(E_*[\hat{v}_{\text{boot}} \mid s])$ is small relative to $\hat{v}_{\text{boot}}$, the estimated sample variance. The efficiency of two Bootstrap estimators can be compared by the size of $\mathrm{Var}_s(E_*[\hat{v}_{\text{boot}} \mid s])$ when both estimators have the same value of R. Next we summarise empirical results based on actual data that show the WOSB can be significantly more efficient than WSB. The benefits of efficiency are either reduced computation time and/or more accurate variance estimates.

Preston and Chipperfield (2002) compared the efficiency of WOSB with $m_h = [n_h/2]$ and WSB with $m_h = n_h - 1$ (see Rao and Wu 1984) for the Australian Quarterly Economic Activity Survey in March 2000. This survey has a stratum level sample size that varies from 4 and into the 100s. The results (derived from Preston and Chipperfield 2002, Table 1) show at the national level the size of $\mathrm{Var}_s(E_*[\hat{v}_{\text{boot}} \mid s])$ was 54% smaller for WOSB compared with WSB sampling when $R = 100$ (See Preston and Chipperfield 2002 for more empirical estimates of $\mathrm{Var}_s(E_*[\hat{v}_{\text{boot}} \mid s])$ for WSB and WSOB). In other words, WOSB required about half the number of replicates to achieve the same replication error as WSB. This represents a significant efficiency gain. Another benefit of WOSB over WSB is that the computational time in selecting the replicate samples is considerably less.

From empirical investigations, the choice of $m_h = [n_h/2]$ for WOSB minimized $\mathrm{Var}_s(E_*[\hat{v}_{\text{boot}} \mid s])$. As n increases, we suspect that the difference between WOSB and WSB will reduce to approximately zero. More work needs to be done to establish these properties.

5. Movement variance between single phase estimates

A key output requirement of many business surveys is the estimate of change between two time points. Denote the finite population at time t by $U^{(t)} = \{U_1^{(t)}, U_2^{(t)}, ..., U_H^{(t)}\}$, where $U_h^{(t)}$ is the stratum h population at time t that is made up of $N_h^{(t)}$ units. The population total at time t is $Y^{(t)} = \sum_h \sum_{i \in U_h^t} y_{hi}^{(t)}$. Estimating the variance of $\Delta^{(t)} = \hat{Y}^{(t)} - \hat{Y}^{(t-1)}$, the difference between two time periods, is the focus of this section. The terms corresponding to n_h, f_h and s_h^2 at time t are denoted by $n_h^{(t)}$, $f_h^{(t)}$ and $s_h^{(t)2}$ respectively. When sampling on two occasions define N_c, n_{hc}, $n_{\bar{c}h}^{(1)}$, and $n_{\bar{c}h}^{(2)}$ to be the number of units in the following sets $U_h^{(1)} \cap U_h^{(2)}$, $s_h^{(c)} = s_h^{(1)} \cap s_h^{(2)}$, $s_{h\bar{c}}^{(1)} = s_h^{(1)} - s_h^{(c)}$, and $s_{h\bar{c}}^{(2)} = s_h^{(2)} - s_h^{(c)}$ respectively. In ABS business surveys the time 1 sample of size $n_h^{(1)}$ is an SRSWOR from $U_h^{(1)}$. The time 2 sample is the union of the following two samples: an SRSWOR of n_{hc} units from $s_h^{(c)}$ and an SRSWOR of $n_{\bar{c}h}^{(2)}$ units from $U_h^{(2)} - (U_h^{(1)} \cap U_h^{(2)})$. The time 2 sample is effectively an SRSWOR from $U_h^{(2)}$. At the ABS, the size of the overlapping sample, n_{hc}, is controlled by the Permanent Random Number method (see Brewer, Gross and Lee 1999).

The estimator of $\mathrm{Var}\,(\hat{\Delta})$ can be expressed as

$$\mathrm{Var}\,(\hat{\Delta}) = \mathrm{Var}\,(\hat{Y}^{(1)}) + \mathrm{Var}\,(\hat{Y}^{(2)}) - 2\mathrm{Cov}\,(\hat{Y}_1^{(1)}, \hat{Y}_2^{(2)}).$$

Consider the Horvitz-Thompson estimator $\hat{\Delta} = \hat{Y}^{(2)} - \hat{Y}^{(1)}$, where $t = 1, 2$ and $\hat{Y}^t$ is defined analogously to $\hat{Y}$. Tam (1985) show that when $U_h^{(1)} = U_h^{(2)}$, an unbiased estimator of $\mathrm{Var}\,(\hat{\Delta})$ under the above sampling scheme is

$$\widehat{\mathrm{Var}}\,(\hat{\Delta}) = \widehat{\mathrm{Var}}\,(\hat{Y}^{(1)}) + \widehat{\mathrm{Var}}\,(\hat{Y}^{(2)}) - 2\widehat{\mathrm{Cov}}\,(\hat{Y}^{(1)}, \hat{Y}^{(2)}),$$

where

$$\widehat{\mathrm{Var}}\,(\hat{Y}^{(t)}) = \sum_h N_h^2 (1 - f_t) s_h^{(t)2} / n_h^{(t)},$$

$$\widehat{\mathrm{Cov}}\,(Y^{(1)}, Y^{(2)}) = \sum_h N_h^2 (1 - f_{12,h})\, s_h^{(12)} n_{hc} / (n_h^{(1)} n_h^{(2)}),$$

$$s_h^{(12)} = (n_{hc} - 1)^{-1} \sum_{i \in s_h^{(c)}} (y_{1i} - \hat{\bar{y}}_1)(y_{2i} - \hat{\bar{y}}_2),$$

$$\hat{\bar{y}}_t = n_{hc}^{-1} \sum_{i \in s_h^{(c)}} y_{ti}$$

for $t = 1, 2$ and $f_{12,h} = n_h^{(1)} n_h^{(2)} / n_{hc} N_h$.

When $U_h^{(1)} \neq U_h^{(2)}$, a more general form of Tam's estimator is given by $\widehat{\mathrm{Var}}\,(\hat{\Delta})$, except that

$$\widehat{\mathrm{Var}}\,(\hat{Y}^{(t)}) = \sum_h N_h^{(t)2} (1 - f_t) s_h^{(t)2} / n_h^{(t)},$$

$$\widehat{\mathrm{Cov}}(\hat{Y}^{(1)}, \hat{Y}^{(1)}) = \sum_h N_h^{(1)} N_h^{(2)} / (n_h^{(1)} n_h^{(2)}) n_c (1 - f_{12,h}) s_h^{(12)}$$

and

$$f_{12,h} = \frac{n_h^{(1)} n_h^{(2)} N_{hc}}{n_{hc} N_h^{(1)} N_h^{(2)}}.$$

For the reminder of this section we assume that $\widehat{\mathrm{Var}}(\hat{\Delta})$ is unbiased for $\mathrm{Var}(\hat{\Delta})$ when $U_h^{(1)} \neq U_h^{(2)}$. (It is worthwhile noting that $\widehat{\mathrm{Var}}(\hat{\Delta})$ can take negative values when $U_h^{(1)} \neq U_h^{(2)}$. Nordberg (2000) gives an unbiased estimator of $\mathrm{Var}(\hat{\Delta})$ for the regression estimator when $U_h^{(1)} \neq U_h^{(2)}$, but there is no obvious way in which it can be used with the Bootstrap as described in this paper.)

Estimating the variance of $\hat{\Delta}_{\mathrm{reg}} = \hat{Y}_{\mathrm{reg}}^{(1)} - \hat{Y}_{\mathrm{reg}}^{(2)}$, the movement between GREG estimates at times 1 and 2, using WOSB involves repeating the following R times:

(a) forming the set s^* by independently selecting $m_{ch} = [n_{ch}/2]$, $m_{\bar{c}h}^{(1)} = [n_{\bar{c}h}^{(1)}/2]$ and $m_{\bar{c}h}^{(2)} = [n_{\bar{c}h}^{(2)}/2]$ units by SRSWOR from the sets s_{hc}, $s_{h\bar{c}}^{(1)}$, and $s_{h\bar{c}}^{(2)}$ respectively;

(b) for $i \in s_h^{(1)}$ calculate the replicate weights

$$w_{hi}^{*(1)} = N / n_h^{(1)} \left[1 - \gamma_{\mathrm{ch}} \frac{n_{ch}}{n_h^{(1)}} - \gamma_{1\bar{c}\mathrm{h}} \frac{n_{h\bar{c}}^{(1)}}{n_h^{(1)}} + \gamma_{\mathrm{ch}} \frac{n_{ch}}{m_{ch}} \delta_{hi}^{r(t)} \right]$$

for $i \in s_{hc}$,

$$w_{hi}^{*(1)} = \left[1 - \gamma_{\mathrm{ch}} \frac{n_{ch}}{n_{1h}} - \gamma_{1\bar{c}\mathrm{h}} \frac{n_{1\bar{c}h}}{n_{1h}} + \gamma_{1\bar{c}\mathrm{h}} \frac{n_{1\bar{c}h}}{m_{1\bar{c}h}} \delta_{hi}^{r(1)} \right]$$

for $i \in s_{h\bar{c}}^{(1)}$, where

$$\gamma_{1\bar{c}\mathrm{h}} = \sqrt{\frac{[n_{1h}(1 - f_h) - n_{ch}(1 - f_{12,h})]\, m_{1\bar{c}h}}{\{n_{1\bar{c}h}(n_{1\bar{c}h} - m_{1\bar{c}h})\}}},$$

$$\gamma_{\mathrm{ch}} = \sqrt{(1 - f_{12,h})\, m_{ch} / (n_{ch} - m_{ch})}$$

and $\delta_{hi}^{*(t)}$ equals 1 if unit i is selected in the replicate group at time point t and zero otherwise;

(c) calculate weights defined analogously for $i \in s_h^{(2)}$;

(d) calculate $\tilde{w}_{hi}^{*(t)} = w_{hi}^{*(t)} g_i^{(t)*}$ for $i \in s_h^{(1)}, s_h^{(2)}$, where $g_i^{(t)*}$ has the same form as g_i but is calculated using the weights $w_{hi}^{*(t)}$ instead of $w_{hi}^{(t)}$;

(e) calculate $\hat{\Delta}_{\mathrm{reg}}^* = \hat{Y}_{\mathrm{reg}}^{*(2)} - \hat{Y}_{\mathrm{reg}}^{*(1)}$, where $\hat{Y}_{\mathrm{reg}}^{*(t)} = \sum_{i \in s^{(1)}} \tilde{w}_{hi}^{*(t)} y_i$. The WOSB variance estimator is given by

$$\widehat{\mathrm{Var}}_B(\hat{\Delta}_{\mathrm{reg}}) = (R-1)^{-1} \sum_{r=1}^{R} (\hat{\Delta}_{\mathrm{reg}}^* - \hat{\Delta}_{\mathrm{reg}})^2,$$

where $\hat{\Delta}_{\mathrm{reg}} = \hat{Y}_{\mathrm{reg}}^{(1)} - \hat{Y}_{\mathrm{reg}}^{(2)}$ and $\hat{Y}_{\mathrm{reg}}^{(t)} = \sum_{i \in s^{(t)}} \tilde{w}_{hi}^{(t)} y_i$.

The proof that $\widehat{\mathrm{Var}}_B(\hat{\Delta}_{\mathrm{reg}})$ is unbiased is straightforward and is similar to the proof that $\mathrm{Var}_B(\hat{\theta})$ is unbiased (see section 4).

The approach described above requires a separate set of replicate weights for movement and level variance estimates. Roberts, Kovačević, Mantel and Phillips (2001) consider approximate Bootstrap variance estimators of movement that only use the level replicate weights, hence reducing computational costs and simplifying the method and its implementation in a computing system.

6. Simulation study

This section summarizes a simulation study for point-in-time and movement estimates carried out to empirically measure the bias and variability of WOSB and WSB over repeated sampling when $R = 100$. A population was generated at time points 1 and 2 from the following models, $y_i^{(1)} = (0.75x_{1i} + 0.25x_{2i})\, W(0, 2.5, 1)$ and $y_i^{(2)} = 1.5y_i^{(1)}\, W(0, 5, 1)$, where the auxiliary variables are given by $x_{1i} = 0.25x_{2i} + 0.75\,[100L(0, 1, 1)]$ and $x_{1i} = 100L(0, 1, 1)$ where $W(\mu, \gamma, \alpha)$ and $L(\mu, \gamma, \alpha)$ are the Weibull and Log-normal distributions with location, shape and scale parameters given by μ, γ and α. These distributions reflect the long tails that are typical of economic survey data. The times 1 and 2 populations were of size 3,000, with 2,500 population units common to both time points. Each population unit, i, was assigned to one of 5 strata at both time points using z_i, where $z_i = x_{1i}\, W(0, 2.5, 1)$ and the stratum boundaries were $z_i = 50, 100, 150, 250$. This resulted in stratum population sizes that ranged from 400 to 1,000.

A total of 3,000 simulated stratified SRSWOR were taken from the population at times 1 and 2, where $n_h^{(t)} = 12$, $n_{\bar{c}h}^{(1)} = n_{\bar{c}h}^{(2)} = 4$ and $n_{ch} = 8$ for all h and $t = 1, 2$. For WOSB the replicate sample sizes are given in sections 4 and 5. For WSB the replicate sample sizes for movements were $m_{ch} = [n_{ch} - 1]$, $m_{\bar{c}h}^{(1)} = [n_{\bar{c}h}^{(1)} - 1]$ and $m_{\bar{c}h}^{(2)} = [n_{\bar{c}h}^{(2)} - 1]$ and for levels were $m_h^{(t)} = n_h^{(t)} - 1$. The WSB estimator for movements has the same form as WOSB but has a slightly different scaling factors and takes replicate samples with replacement.

From each of the 3,000 simulated samples $\hat{Y}_{\mathrm{reg}}^j$ is calculated, where $\hat{Y}_{\mathrm{reg}}^j$ is given by $\hat{Y}_{\mathrm{reg}}$ with $\boldsymbol{x}_i = (x_{1i}, x_{2i})$, $\sigma_i = 1$ and $j = 1, 2, \ldots, 3{,}000$. The true standard error of $\hat{Y}_{\mathrm{reg}}$ is calculated by

$$S = \sqrt{\frac{1}{3{,}000} \sum_{j=1}^{3{,}000} (\hat{Y}_{\mathrm{reg}}^j - Y)^2}.$$

The Bootstrap's estimated standard error of $\hat{Y}_{\text{reg}}$ from the jth sample is

$$\hat{S}^j = \sqrt{\frac{1}{100}\sum_{r=1}^{100}(\hat{Y}_{\text{reg}}^{j*(r)} - Y)^2},$$

where $\hat{Y}_{\text{reg}}^{j*(r)}$ is defined analogously to $\hat{Y}_{\text{reg}}^{*(r)}$. The Relative Bias (RB) of the Bootstrap's standard error is

$$\text{RB}(\hat{S}) = \frac{1}{3{,}000S}\sum_{j=1}^{3{,}000}(\hat{S}^j - S).$$

The Relative Root Mean Squared Error (RRMSE) of the Bootstrap's estimated standard error is

$$\text{RRMS}(\hat{S}) = \frac{1}{S}\sqrt{\frac{1}{3{,}000}\sum_{j=1}^{3{,}000}(\hat{S}^j - S)^2}.$$

Similar definitions for RRMSE and bias are used when estimating the movement variance. The 95% coverage probabilities, the percentage of 95% confidence intervals containing the true population total, of WOSB and WSB for levels and movement are also compared.

The results in Table 1 show that the RB and the RRMSE of the WOSB and WSB are both acceptably small. The bias of WSB's time point 1 estimates are slightly higher than WOSB resulting in slightly worse coverage probabilities.

Table 1
Bootstrap estimate of the standard error for movements and point-in-time estimates

	Time point 1			Movement		
Method	RB	RRMSE	C95%	RB	RRMSE	C95(%)
WOSB	0.7	17.3	94.7	2.1	20.7	95.3
WSB	-3.1	15.8	93.7	-1.3	19.6	94.6

7. Summary

From the simulation results, both the WOSB and WSB were considered to be reliably accurate over repeated sampling. Conditional on the sample, the WOSB was found to be significantly more efficient (up to 50%) than WSB for stratified sampling when the stratum sample size is sometimes small. As a result, the WOSB was implemented in ABSEST.

References

Brewer, K.R.W., Gross, W.F. and Lee, G.F. (1999). PRN sampling: The Australian experience. *Proceedings of the International Associations of Survey Statisticians*, Helsinki.

Estevao, V., Hidiroglou, M.A. and Särndal, C.-E. (1995). Methodological principles for a generalised estimation system at Statistics Canada, *Journal of Official Statistics*, 11, 2, 181-204.

Kovar, J., Rao, J.N.K. and Wu, C.F.J. (1988). Bootstrap and other methods to measure errors in survey estimates. *The Canadian Journal of Statistics*, 25-44.

Nordberg, L. (2000). On variance estimation for measures of change when samples are coordinated by the use of permanent random numbers. *Journal of Official Statistics*, 16, 4, 363-378.

Preston, J., and Chipperfield, J.O. (2002). Using a generalised estimation methodology for ABS business surveys. Methodology Advisory Committee (available at www.abs.gov.au).

Rao, J.N.K., and Wu, C.F.J. (1984). Bootstrap inference for sample surveys. *Proceeding Section on Survey Methods Research, Journal of the American Statistical Association*, 106-112.

Rao, J.N.K., and Wu, C.F.J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, March, 83, 401, 231-241.

Roberts, G., Kovačević, M., Mantel, H. et Phillips, O. (2001). Cross sectional inference based on longitudinal surveys: Some experience with Statistics Canada Surveys. Presented at the 2001 FCSM Research Conference, Arlington, VA, November 14-16, proceedings available at www.fcsm.gov.

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

Shao, J., and Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer-Verlag.

Singh, A.C., and Mohl, C.A. (1996). Understanding calibration estimators in survey sampling. *Survey Methodology*, 22, 107-115.

Tam, S.M. (1985). On covariance in finite population sampling. *The Statistician*, 34, 429-433.

Yung, W., and Rao, J.N.K. (1996). Jackknife linearization variance estimators under stratified multi-stage sampling. *Survey Methodology*, 22, 23-31.