Generalized Variance Estimation for Business Surveys

James Chipperfield and John Preston
james.chipperfield@abs.gov.au, john.preston@abs.gov.au
Australian Bureau of Statistics

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 are highly stratified, 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), a variation of Rao and Wu (1998)'s With Replacement Scaled Bootstrap (WSB), that was implemented. 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 method for point-in-time and movement variance estimates. Simulation results obtained as part of the evaluation process show that the WOSB was more efficient than the WSB.

1. Introduction

In 2000, the Australian Bureau of Statistics (ABS) first obtained a register of business 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 Jacknife 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 typically equal probably 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 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 (1998)'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 and mentions further methodological developments to ABSEST.

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 $Y_h=\sum_{i\in U_h} y_{ih}$ and h=1,…H. Within stratum h, the sample s_h of n_h units is selected 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 $\underset{\sim}{x}_i=(x_{1i},...,x_{ki},...,x_{Ki})^T$ is available for $i\in s$ and the corresponding vector of population totals $\underline{X}=\sum_{i\in U}\underset{\sim}{x}_i$ are known. The GREG estimator (Särndal, Swenssen, and Wretman, 1992 pp 227) is given by $\hat{Y}_{reg}=\sum_{i\in s}\tilde{w}_i y_i=\sum_{i\in s} w_i y_i+\left(\underset{\sim}{X}-\hat{\underset{\sim}{X}}_\pi\right)^T\hat{\underset{\sim}{B}}$, where $\tilde{w}_i=w_i g_i$, $w_i=N_h/n_h$, $\hat{\underset{\sim}{B}}=\underset{\sim}{T}^{-1}\underset{\sim}{t}$ with $\underset{\sim}{T}^{-1}$ being the generalised inverse of $\underset{\sim}{T}$, $\hat{\underset{\sim}{X}}_\pi=\sum_{i\in s} w_i\underset{\sim}{x}_i$ and $g_i=\left(1+\underset{\sim}{x}_i^T\underset{\sim}{T}^{-1}\left(\underset{\sim}{X}-\hat{\underset{\sim}{X}}_\pi\right)^T\right)$, $\underset{\sim}{T}=\left(\sum_{i\in s} w_i\underset{\sim}{x}_i\underset{\sim}{x}_i^T\sigma_i^{-1}\right)$, $\underset{\sim}{t}=\left(\sum_{i\in s} w_i\underset{\sim}{x}_i y_i\sigma_i^{-1}\right)$, σ_i is a constant motivated by the superpopulation model $y_i=\underset{\sim}{x}_i^T\underset{\sim}{\beta}+\varepsilon_i$ such that ε_i is independently and identically distributed with mean 0 and variance σ_i^2, and $E\left(\hat{\underset{\sim}{B}}\right)=\underset{\sim}{\beta}$. It is well known that $\hat{Y}_{reg}$ 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 $\hat{Y}_{reg}$ can be adapted to a range of practical situations (see Estevao, Hidiroglou, Sarndal 1996). For example, when $\underset{\sim}{x}_i=1$, $\hat{Y}_{reg}$ becomes the Horvitz-Thompson estimator given by $\hat{Y}_\pi=\sum_h w_h\sum_{i\in s_h} y_{hi}$ with estimated variance $V\hat{a}r(\hat{Y}_\pi)=\sum_h N_h^2/n_h(1-f_h)\hat{s}_h^2$, where $\hat{s}_h^2=1/(n_h-1)\sum_{i\in s_h}(y_{hi}-\hat{\bar{y}}_h)^2$, $\hat{\bar{y}}_h=\sum_{i\in s_h} y_{hi}/n_h$ 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 type of estimator and sample design. Also, strong consideration was given to minimise the computational costs.

We considered the relative merits of a number of variance estimators for implementation in ABSEST. The TS method was not suitable as its variance expression for complex estimands could involve many terms specific to the estimand, making it difficult to adapt into a generalized system. For the same reason other linearized variance estimators (described in Estevao, Hidiroglou, Sarndal 1996 and evaluated in Yung and Wu (1996)) were rejected, despite good theoretical properties, good empirical results and being computationally efficient. We also considered the Bootstrap, Jacknife and Balanced Repeated Replication (BRR) methods (Shao and Tu 1995, Rao and Wu 1988, and Wolter, 1985).

Consider estimating the variance of a function $g(\hat{\underline{\theta}})$, where $\hat{\underline{\theta}}$ is a P vector of estimates and g 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 to get the replicate samples s^r, where r =1, ..., R; (ii) computing $w_i^r = b_i^r w_i$, where b_i^r depends upon the number of times unit i is selected in s^r; (iii) calculate $g(\hat{\underline{\theta}}^r)$, where $\hat{\underline{\theta}}^r = \sum_{i \in s} w_i^r \underline{\theta}_i$ and is an estimate of $g(\hat{\underline{\theta}})$ based on s and $\underline{\theta}_i$ is the response vector from unit i; (iv) estimate variance of $g(\hat{\underline{\theta}})$ by $\hat{Var}_{rep}(g(\hat{\underline{\theta}})) = (R-1)^{-1} \sum_{r=1}^{R} (g(\hat{\underline{\theta}}^r) - g(\hat{\underline{\theta}}))^2$. Note: the expression for replicate weights, $w_i^r = b_i^r w_i$, includes the Jacknife, Bootstrap and Balanced Repeated Replication as special cases; see reference list for the appropriate choice of b_i^r for each of these methods.

From section 2, it is possible to express $\hat{Y}_{reg}^r = g(\hat{\underline{\theta}})$, where $\hat{\underline{\theta}} = (\mathbf{t}_\pi, \mathbf{T}_\pi, X_\pi, Y_\pi)$. In the case of estimating the variance of $\hat{Y}_{reg}^r$, steps (iii) and (iv) above respectively become: (iii) calculate $\hat{Y}_{reg}^r = \sum_{i \in s^r} \tilde{w}_i^r y_i$, where $\tilde{w}_i^r = w_i^r g_i^r$ and g_i^r has the same form as g_i but is calculated using the weights w_i^r instead of the weights w_i; (iv) estimating variance by $\hat{Var}_{rep}(\hat{Y}_{reg}) = (R-1)^{-1} \sum_{r=1}^{R} (\hat{Y}_{reg}^r - \hat{Y}_{reg})^2$.

The attractive feature of these replication methods is that only the selection of the replicate samples and the value b_i^r needs to be varied to calculate unbiased variance estimates for many commonly used sample designs and for estimators that have good first order Taylor Series approximations. For example, under SRSWOR the Jacknife forms replicate samples s^r by dropping the r th unit from s for r =1, .., R and using the values $b_i^r = n/(n-1)\sqrt{(n-1)/n(1-n/N)}$ if $i \in s^r$ and 0 otherwise, and. If the replicate weights, $\tilde{w}_i^r$ are stored the variance estimates of $\hat{Y}_{reg}^r$ 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.

We now consider the replication methods mentioned above. The drop-one Jacknife forms replicate samples, s^r, 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 Jacknife, 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 pp. 251), the Jacknife 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^r for each r = 1, …, R and 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. Rao and Shao (1995) show in their simulation study that BRR, "is very unstable". Further, 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 Jacknife 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.

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

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 $w_{hi}^r = w_{hi}(1-\gamma_h + \gamma_h \frac{n_h}{m_h}\delta_{hi}^r)$ for $i \in s_h$, where $\gamma_h = \sqrt{(1-f_h)m_h/(n_h - m_h)}$, δ_{ri} is 1 if $i \in s_h^r$ and 0 otherwise; and (c) calculating $\tilde{w}_{hi}^r = w_{hi}^r g_{hi}^r$ for $i \in s$; and (d) calculating the r th Bootstrap estimate of Y, $\hat{Y}_{reg}^r = \sum_{i\in s} \tilde{w}_{hi}^r y_i$. The Bootstrap variance estimator is given by the Monte Carlo approximation, $Var_B\left(\hat{Y}_{reg}\right) = 1/(R-1)\sum_{b=1}^{R}\left(\hat{Y}_{reg}^r - \hat{Y}_{reg}\right)^2$. 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 $Var(g(\hat{\underline{\theta}}))$. Briefly dropping the stratum subscript, the TS approximate variance is given by $V\hat{a}r(g(\hat{\underline{\theta}})) = \nabla g(\hat{\underline{\theta}})'\hat{V}(\hat{\underline{\theta}})\nabla g(\hat{\underline{\theta}})$, where $\hat{V}(\hat{\underline{\theta}})$ is P x P matrix with elements $C\hat{o}v(\hat{\theta}_{\pi,p}, \hat{\theta}_{\pi,p'}) = N^2(1-f)\hat{s}_{p,p'}^2/n$, where $\hat{s}_{p,p'}^2 = 1/(n-1)\sum_{i\in s}(\theta_{pi} - \hat{\bar{\theta}}_p)(\theta_{p'i} - \hat{\bar{\theta}}_{p'})$, $\hat{\bar{\theta}}_p = 1/n\sum_{i\in s}\theta_{pi}$, $\hat{\theta}_{\pi,p} = \sum_{i\in s} w_i\theta_{pi}$, $p, p' = 1,...P$, and $\nabla = (\partial/\partial\theta_1, ...\partial/\partial\theta_P)|_{\hat{\underline{\theta}}}$. It is easy to see that $E^r\left(V\hat{a}r(g(\hat{\underline{\theta}}^r))\right) = \nabla g(\hat{\underline{\theta}})'E^r\left(\hat{V}(\hat{\underline{\theta}}^r)\right)\nabla g(\hat{\underline{\theta}})$ $= \nabla g(\hat{\underline{\theta}})'V(\hat{\underline{\theta}})\nabla g(\hat{\underline{\theta}})$, by noting that $E^r\left[C\hat{o}v(\hat{\theta}_{\pi,p}^r, \hat{\theta}_{\pi,p'}^r)\right] = C\hat{o}v(\hat{\theta}_{\pi,p}, \hat{\theta}_{\pi,p'})$ where E^r 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 appropriate finite population correction factor is reflected in $Cov(\hat{\theta}_{\pi,p}, \hat{\theta}_{\pi,p'})$. It therefore follows that the Monte Carlo approximation to the variance, $Var_B\left(g(\hat{\underline{\theta}})\right) = 1/(R-1)\sum_{b=1}^{R}\left(g(\hat{\underline{\theta}}^r) - g(\hat{\underline{\theta}}^r)\right)^2$, is unbiased for $Var(g(\hat{\underline{\theta}}))$.

5. Movement Variance between Single Phase Estimates

A key output requirements of many business surveys is the estimate of change between two time points. Denote the finite population at time t by $U^{(t)} = \left\{U_1^{(t)}, U_2^{(t)} ..., U_H^{(t)}\right\}$, 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. We also 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 $U_h^{(1)} \cap U_h^{(2)}$, $s_h^{(c)} = s_h^{(1)} \cap s_h^{(2)}$, $s_{h\bar{c}}^{(1)} = s_h^{(1)} - s_{hc}$, and $s_{h\bar{c}}^{(2)} = s_h^{(2)} - s_{hc}$ respectively. When sampling on more than one occasion, ABS business surveys designs are analogous to sampling plan A of Tam (1985) where the size of overlapping sample, n_{hc}, is controlled by the Permanent Random Number method (see Brewer, Gross and Lee 1999).

The estimator of $\operatorname{Var}\left(\hat{\Delta}\right)$ can be expressed as $\operatorname{Var}(\hat{\Delta}\ ) = \operatorname{Var}(\hat{Y}_1) + \operatorname{Var}(\hat{Y}_2) - 2\operatorname{Cov}(\hat{Y}_1, \hat{Y}_2)$. Consider the Horvitz –Thompson estimator $\hat{\Delta}_\pi = \hat{Y}_\pi^{(2)} - \hat{Y}_\pi^{(1)}$, where t = 1, 2 and $\hat{Y}_\pi^t$ is defined analogously to $\hat{Y}_\pi$. Using a result from Tam (1985) when $U_h^{(1)} = U_h^{(2)}$, an unbiased estimator of $\operatorname{Var}\left(\hat{\Delta}_\pi\right)$ under sampling plan A is $\widehat{\operatorname{Var}}(\hat{\Delta}_\pi) = \widehat{\operatorname{Var}}(\hat{Y}_\pi^{(1)}) + \widehat{\operatorname{Var}}(\hat{Y}_\pi^{(2)}) - 2\widehat{\operatorname{Cov}}(\hat{Y}_\pi^{(1)}, \hat{Y}_\pi^{(2)})$, where $\widehat{\operatorname{Var}}(\hat{Y}_\pi^{(t)}) = \sum_h N_h^2(1-f_t)s_h^{(t)2} / n_h^{(t)}$, $\widehat{\operatorname{Cov}}(Y_\pi^{(1)}, Y_\pi^{(2)}) = \sum_h N_h^2(1-f_{12})S_h^{(12)}n_c / (n_h^{(1)}n_h^{(2)})$ and $f_{12,h} = \dfrac{n_h^{(1)}n_h^{(2)}}{n_{hc}N_h}$.

When $U_h^{(1)} \neq U_h^{(2)}$, a more general form of Tam's estimator is given by $\widehat{\operatorname{Var}}\left(\hat{\Delta}_\pi\right)$, except that $\widehat{\operatorname{Var}}(\hat{Y}_\pi^{(t)}) = \sum_h N_h^{(t)2}(1-f_t)s_h^{(t)2} / n_h^{(t)}$, $\operatorname{Cov}(Y_\pi^{(1)}, Y_\pi^{(1)}) = \sum_h N_h^{(1)}N_h^{(2)} / (n_h^{(1)}n_h^{(2)})n_c(1-f_{12,h})S_h^{(12)}$ and $f_{12,h} = \dfrac{n_h^{(1)}n_h^{(2)}N_{hc}}{n_{hc}N_h^{(1)}N_h^{(2)}}$. For the reminder of this section we assume that $\widehat{\operatorname{Var}}(\hat{\Delta}_\pi)$ is unbiased for $\operatorname{Var}(\hat{\Delta}_\pi)$ when $U_h^{(1)} \neq U_h^{(2)}$.

Estimating the variance of $\hat{\Delta}_{reg} = \hat{Y}_{reg}^{(1)} - \hat{Y}_{reg}^{(2)}$, the movement between GREG estimates at times 1 and 2, using WOSB involves repeating the following R times: (a) forming the set s^r by independently selecting $m_{ch} = \left[n_{ch}/2\right]$, $m_{\bar{c}h}^{(1)} = \left[n_{\bar{c}h}^{(1)}/2\right]$ and $m_{\bar{c}h}^{(2)} = \left[n_{\bar{c}h}^{(2)}/2\right]$ 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

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

for $i \in s_{hc}$,

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

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

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

$\gamma_{ch} = \sqrt{\left(1 - f_{12,h}\right)m_{ch} / \left(n_{ch} - m_{ch}\right)}$ and $\delta_{hi}^{r(t)}$ equals 1 if unit i is selected in replicate group r at time point t and zero otherwise; (c) calculating weights defined analogously for $i \in s_h^{(2)}$; (d) calculating $\tilde{w}_{hi}^{r(t)} = w_{hi}^{r(t)}g_i^{(t)r}$ for $i \in s_h^{(1)}, s_h^{(2)}$, where $g_i^{(t)r}$ has the same form as g_i but is calculated using the weights $w_{hi}^{r(t)}$ instead of $w_{hi}^{(t)}$; (e) calculating $\hat{\Delta}_{reg}^r = \hat{Y}_{reg}^{(2)r} - \hat{Y}_{reg}^{(1)r}$, where $\hat{Y}_{reg}^{(t)r} = \sum_{i \in s^{(1)}} \tilde{w}_{hi}^{(t)r}y_i$. The WOSB variance estimator is given by

$$\widehat{\operatorname{Var}}_B\left(\hat{\Delta}_{reg}\right) = 1/(R-1)\sum_{r=1}^{R}\left(\hat{\Delta}_{reg}^{(t)r} - \hat{\Delta}_{reg}\right)^2,$$

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

The proof that $\widehat{\operatorname{Var}}_B\left(\hat{\Delta}_{reg}\right)$ is unbiased is straight-forward and is similar to the proof that $\operatorname{Var}_B\left(g(\underline{\hat{\theta}})\right)$ is unbiased (see section 4).

The approach described above requires a separate set of replicate weights for movement and level variance estimates. Roberts, Kovacevic, Mantel, and Phillips (2001) consider approximate Bootstrap variance estimators of movement that only use the level replicate weights.

6. Simulation Study

This section summarizes the simulation study for point-in-time and movement estimates carried out to empirically measure the bias and variability of the WOSB and WSB over repeated sampling. A population of size 5000 was generated from the following models, $y_{1i} = \operatorname{abs}(x_{1i} + \psi_{1i})$ and $y_{2i} = \operatorname{abs}(x_{1i} + x_{2i} + \psi_{2i})$, $x_{1i} = \operatorname{abs}(\xi_{1i})$, $x_{2i} = \operatorname{abs}(\xi_{2i})$, $\xi_{1i} \sim N(0,25)$, $\xi_{2i} \sim N(0,25)$, $\psi_{1i} \sim N\left(0, \sqrt{x_{1i}}\right)$ and $\psi_{2i} \sim N\left(0, \sqrt{x_{1i} + x_{2i}}\right)$. Each population unit was assigned to one of 10 strata on the basis of z_i, where $z_i = \operatorname{abs}(x_{1i} + \zeta_i)$ and $\zeta_i \sim N\left(0, \sqrt{x_{1i}}\right)$; the stratum boundaries were

$z_i = 5,10,15,20,25,30,40,50,75$ resulting in stratum population sizes that ranged from 20 to 900. A total of 2500 simulated stratified simple random samples without replacement were taken from the population, where $n_h = 10$ for all h.

From the j th simulation sample we define $\hat{S}_k^j = S\hat{e}_B\left(\hat{Y}_{k,reg}^j\right) = \sqrt{Var_B\left(\hat{Y}_{k,reg}^j\right)}$, where $\hat{Y}_{k,reg}^j$ is the GREG population estimate of Y_k based on the j th simulation sample, $\hat{Y}_{1,reg}$ and $\hat{Y}_{2,reg}$ are defined as in section 2 with $\underset{\sim}{x}_i = x_{1i}$ and $\underset{\sim}{x}_i = (x_{1i}, x_{2i})$ respectively, and $\sigma_i = 1$. The true standard error of the Bootstrap's estimated standard error

Next we consider the WOSB for movement estimates. The population at time point 1 is described above. For every population unit at time point 1, the corresponding characteristic of interest at time point 2 was generated using the model $y_{ki}^{(2)} = abs\left(1.1 \times y_{ki}^{(1)} + \psi_{2i}\right)$. From this population 2500 samples at time 1 and 2 were taken using Tam's sampling plan A to control overlap, where $n_h^{(t)} = 10$ and $n_{hc} = 6$ for all h. The results in the table show that the bias and the RRMSE of the WOSB were acceptably small.

was calculated by $S_k = \sqrt{\frac{1}{2500}\sum_{j=1}^{2500}\left(\hat{Y}_{k,reg}^j - Y_k\right)^2}$. We define the Relative Bias (RB) of the Bootstrap's estimated standard error by $Bias(\hat{S}_k) = \frac{1}{2500 S_k}\sum_{j=1}^{2500}\left(\hat{S}_k^j - S_k\right)$ and its Relative Root Mean Squared Error (RRMSE) by $RRMS(\hat{S}_k) = \frac{1}{\bar{S}_k}\sqrt{\frac{1}{2500}\sum_{j=1}^{2500}\left(\hat{S}_k^j - S_k\right)^2}$. The results in the table show that the bias and the RRMSE of the WOSB and WSB are both acceptably small.

Table: Bootstrap Estimate of the standard error for movement and level estimates

Estimator	Replicate sampling scheme	Variable	Point in time		Movement	
			RRMSE (%)	**RB (%)**	**RRMSE (%)**	**RB (%)**
HT	SRSWOR	y_1	11.8	-1.5	16.0	1.4
HT	SRSWR	y_1	11.9	-1.5	15.8	1.5
GREG	SRSWOR	y_1	11.9	3.3	14.5	1.5
GREG	SRSWR	y_1	11.8	2.5	14.7	2.2
HT	SRSWOR	y_2	13.0	1.9	15.7	-2.3
HT	SRSWR	y_2	13.2	1.9	15.7	-2.3
GREG	SRSWOR	y_2	12.4	1.9	12.6	0.5
GREG	SRSWR	y_2	12.8	2.8	12.7	-0.3

7. Summary

From the simulation results, both the WOSB and WSB were considered to be reliably accurate. The mean squared error of the Bootstrap standard error estimator is made up of: (i) the variance due to repeated sampling from the population, (ii) the variance due to replicate sampling conditional on the sample, and (iii) the bias squared. We know from standard theory that (i) is independent of the Bootstrap method and so clouds any comparison of the WOSB and WSB. Therefore, when comparing the accuracy of the WOSB and WSB only (ii) and (iii) are relevant. We know from the table of results that (iii) for both the WSB and WOSB is negligible. Using the ABS' Quarterly Economy Wide Survey, where a stratum's sample size is often as low as 6, Preston and Chipperfield (2002) found that (iii) for the WOSB was 40% smaller than the WSB: this meant that WOSB required 40% fewer replicates (R) than WSB. As a result, the WOSB was considered to be more efficient and subsequently implemented in ABSEST.

ABSEST has adopted the paradigm of storing replicate weights for variance calculation. A limitation of this approach is that, with the exception of mean imputation, accounting for the variance in estimation due to item non-response could potentially require RM replicate weights per sample unit, where M is the number of data items requiring imputation. However, estimating the variance due to imputation could be implemented in ABSEST, but would require additional calculations not reflected in the replicate weights (see Shao and Sitter (1996)).

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.

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 pp. 25-44

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. (1988) *Resampling inference with complex survey data*, Journal of the American Statistical Association, March, Vol 83, No.401 pp231-241.

Rao, J. N. K. and Sitter, R. R. (1997) *Variance estimation Under Stratified Two-Phase Sampling with Applications to Measurement Bias* . Survey Measurement and Process Quality, John Wiley & Sons, New York pp. 759-762

Roberts, G., Kovacevic, M., Mantel, H. and Phillips, O. (2001), *Cross sectional inference based on longitudinal surveys: some experience with Statistics Canada Surveys*, Statistics Canada.

Rust, K. F. and Rao, J. N. K. (1996) *Variance estimation for complex surveys using replication techniques,* Statistical Methods in Medical Research 283- 310

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

Shao, J.and Sitter, R.R.(1996) *Bootstrap for Imputed Data*, Journal of the American Statistical Association, 91, pp 1278- 1288.

Sing, A.C. & Mohl, C.A. (1996) *Calibration estimators in survey sampling under benchmark constraints and range restrictions*, Survey Methodology

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

Yung, W. and Rao, J. N. K. (1996) *Jacknife linearization variance estimators under stratified multi-stage design* Survey Methodology pp. 56-66.