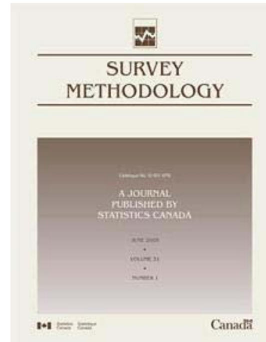
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Rescaled bootstrap for stratified multistage sampling



by John Preston

December 2009

Survey Methodology, December 2009
Vol. 35, No. 2, pp. 227-234
Statistics Canada, Catalogue No. 12-001-X

Rescaled bootstrap for stratified multistage sampling

John Preston [1]

Abstract

In large scaled sample surveys it is common practice to employ stratified multistage designs where units are selected using simple random sampling without replacement at each stage. Variance estimation for these types of designs can be quite cumbersome to implement, particularly for non-linear estimators. Various bootstrap methods for variance estimation have been proposed, but most of these are restricted to single-stage designs or two-stage cluster designs. An extension of the rescaled bootstrap method (Rao and Wu 1988) to stratified multistage designs is proposed which can easily be extended to any number of stages. The proposed method is suitable for a wide range of reweighting techniques, including the general class of calibration estimators. A Monte Carlo simulation study was conducted to examine the performance of the proposed multistage rescaled bootstrap variance estimator.

Key Words: Bootstrap; Calibration; Multistage sampling; Stratification; Variance estimation.

1. Introduction

Stratified multistage sampling designs are especially suited to large scaled sampled surveys because of the advantage of clustering collection effort. Various methods exist for variance estimation for these complex survey designs. The most commonly used methods are the linearization (or Taylor) method, and resampling methods, such as the jackknife, balance repeated replication and the bootstrap. The linearization method can be quite cumbersome to implement for complex survey designs as it requires the derivation of separate variance formulae for each non-linear estimator. Some approximations are normally required for the variance of non-linear functions, such as ratios and correlation and regression coefficients, and functionals, such as quantiles.

On the other hand, the various resampling methods employ a single variance formulae for all estimators. The replication methods can reflect the effects of a wide range of reweighting techniques, including calibration, and adjustments due to provider non-response and population undercoverage. The jackknife and balance repeated replication methods are only applicable to stratified multistage designs where the clusters are sampled with replacement or the first-stage sampling fractions are negligible. A number of different bootstrap methods for finite population sampling have been proposed in the literature, including the with-replacement bootstrap (McCarthy and Snowden 1985), the rescaled bootstrap (Rao and Wu 1988), the mirror match bootstrap (Sitter 1992a), and the without-replacement bootstrap (Gross 1980; Bickel and Freedman 1984; Sitter 1992b). A summary of these bootstrap methods can be found in Shao and Tu (1995).

Most of these bootstrap methods are restricted to single-stage designs or multistage designs where the first-stage sampling units are selected with replacement or the first-stage sampling fractions are small in most strata. However, in many large scaled sample surveys it is common practice to employ highly stratified multistage designs where units are selected using simple random sampling without replacement at each stage. Some typical examples of these types of surveys are employer-employee surveys, such as the Survey of Employee Earnings and Hours (ABS 2008), and school-student surveys, such as the National Survey on the Use of Tobacco by Australian Secondary School Students (White and Hayman 2006).

McCarthy and Snowden (1985) proposed an extension of their with-replacement bootstrap to two-stage sampling in the special case of equal cluster sizes and equal within cluster sample sizes, while Rao and Wu (1988) and Sitter (1992a) have given extensions of their rescaled bootstrap and mirror match bootstrap methods to two-stage cluster sampling. More recently, Funaoka, Saigo, Sitter and Toida (2006) proposed two Bernoulli-type bootstrap methods, the general Bernoulli bootstrap and the short cut Bernoulli bootstrap, which can easily handle multistage stratified designs where units are selected using simple random sampling without replacement at each stage. The general Bernoulli bootstrap has the advantage that it can handle any combination of sample sizes, but it requires a much larger number of random number generations than the short cut Bernoulli bootstrap.

In this paper, an extension of the rescaled bootstrap procedure to stratified multistage sampling where units are selected using simple random sampling without replacement at each stage is proposed. In Section 2, the notation for stratified multistage sampling is introduced. In Section 3, the extension of the rescaled bootstrap estimator to multistage sampling is described. The main findings of a simulation study are reported in Section 4. Some concluding remarks are provided in Section 5.

1. John Preston, Australian Bureau of Statistics, 639 Wickham Street, Fortitude Valley QLD 4006, Australia. E-mail: john.preston@abs.gov.au.

2. Stratified multistage sampling

For simplicity, the case of stratified three-stage sampling is presented. Consider a finite population U divided into H nonoverlapping strata $U = \{U_1, ..., U_H\}$, where U_h is comprised of N_{1h} primary sampling units (PSU's). At the first-stage, a simple random sample without replacement (SRSWOR) of n_{1h} PSU's are selected with selection probabilities $\pi_{1hi} = n_{1h} / N_{1h}$ within each stratum h. Suppose selected PSU i in stratum h is comprised of N_{2hi} secondary sampling units (SSU's). At the second-stage, a SRSWOR of size n_{2hi} SSU's are selected with selection probabilities $\pi_{2hij} = n_{2hi} / N_{2hi}$ within each selected PSU. Suppose selected SSU j in selected PSU i in stratum h is comprised of N_{3hij} ultimate sampling units (USU's). At the third-stage, a SRSWOR of size n_{3hij} USU's are selected with selection probabilities $\pi_{3hijk} = n_{3hij} / N_{3hij}$ within each selected SSU.

The objective is to estimate the population total $Y = \sum_{h=1}^{H} \sum_{i=1}^{N_{1h}} \sum_{j=1}^{N_{2hi}} \sum_{k=1}^{N_{3hij}} y_{hijk}$, where y_{hijk} is the value for the variable of interest y for USU k in SSU j in PSU i in stratum h. An unbiased estimate of Y is given by:

$$\hat{Y} = \sum_{h=1}^{H} \hat{Y}_h = \sum_{h=1}^{H} \frac{N_{1h}}{n_{1h}} \sum_{i=1}^{n_{1h}} \frac{N_{2hi}}{n_{2hi}} \sum_{j=1}^{n_{2hi}} \frac{N_{3hij}}{n_{3hij}} \sum_{k=1}^{n_{3hij}} y_{hijk}$$

where $\hat{Y}_h = (N_{1h} / n_{1h}) \sum_{i=1}^{n_{1h}} \hat{Y}_{hi}$, $\hat{Y}_{hi} = (N_{2hi} / n_{2hi}) \sum_{j=1}^{n_{2hi}} \hat{Y}_{hij}$ and $\hat{Y}_{hij} = (N_{3hij} / n_{3hij}) \sum_{k=1}^{n_{3hij}} y_{hijk}$. This estimator can also be written as $\hat{Y} = \sum_{h=1}^{H} \sum_{i=1}^{n_{1h}} \sum_{j=1}^{n_{2hi}} \sum_{k=1}^{n_{3hij}} w_{hijk} y_{hijk}$, where $w_{hijk} = w_{1hi} w_{2hij} w_{3hijk} = (N_{1h} / n_{1h})(N_{2hi} / n_{2hi})(N_{3hij} / n_{3hij})$ is the sampling weight for USU k in SSU j in PSU i in stratum h.

An unbiased estimate of $\operatorname{Var}(\hat{Y})$ is given by (Särndal, Swensson and Wretman 1992):

$$\begin{aligned} \hat{\operatorname{V}}\text{ar}(\hat{Y}) &= \sum_{h=1}^{H} \frac{N_{1h}^2}{n_{1h}} (1 - f_{1h}) s_{1h}^2 \\ &+ \sum_{h=1}^{H} \frac{N_{1h}}{n_{1h}} \sum_{i=1}^{n_{1h}} \frac{N_{2hi}^2}{n_{2hi}} (1 - f_{2hi}) s_{2hi}^2 \\ &+ \sum_{h=1}^{H} \frac{N_{1h}}{n_{1h}} \sum_{i=1}^{n_{1h}} \frac{N_{2hi}}{n_{2hi}} \sum_{j=1}^{n_{2hi}} \frac{N_{3hij}^2}{n_{3hij}} (1 - f_{3hij}) s_{3hij}^2 \end{aligned} \quad (2.1)$$

where $f_{1h} = (n_{1h}/N_{1h})$, $f_{2hi} = (n_{2hi}/N_{2hi})$, $f_{3hij} = (n_{3hij}/N_{3hij})$, $\bar{\hat{Y}}_h = \sum_{i=1}^{n_{1h}} \hat{Y}_{hi}/n_{1h}$, $\bar{\hat{Y}}_{hi} = \sum_{j=1}^{n_{2hi}} \hat{Y}_{hij}/n_{2hi}$, $\bar{y}_{hij} = \sum_{k=1}^{n_{3hij}} y_{hijk}/n_{3hij}$, $s_{1h}^2 = \sum_{i=1}^{n_{1h}} (\hat{Y}_{hi} - \bar{\hat{Y}}_h)^2/(n_{1h} - 1)$, $s_{2hi}^2 = \sum_{j=1}^{n_{2hi}} (\hat{Y}_{hij} - \bar{\hat{Y}}_{hi})^2/(n_{2hi} - 1)$ and $s_{3hij}^2 = \sum_{k=1}^{n_{3hij}} (y_{hijk} - \bar{y}_{hij})^2/(n_{3hij} - 1)$.

3. Rescaled bootstrap for stratified multistage sampling

Rao and Wu (1988) proposed a rescaling of the standard bootstrap method for various sampling designs including stratified sampling. Since the rescaling factors are applied to the survey data values, this method is only applicable to smooth statistics. Rao, Wu and Yue (1992) presented a modification to this rescaled bootstrap method where the rescaling factors are applied to the survey weights, rather than the survey data values. This modified rescaled bootstrap method is equivalent to the original rescaled bootstrap method, but has the added advantage that it is applicable to non-smooth statistics as well as smooth statistics. Kovar, Rao and Wu (1988) showed that when using a bootstrap sample size of $n_h^* = n_h - 1$ the rescaled bootstrap estimator performed well for smooth statistics.

Although bootstrap samples are usually selected with replacement, Chipperfield and Preston (2007) modified the rescaled bootstrap method to the situation where the bootstrap samples are selected without replacement. Under this without replacement rescaled bootstrap method it can be shown that the choice of either $n_h^* = \lfloor n_h / 2 \rfloor$ or $n_h^* = \lceil n_h / 2 \rceil$ is optimal, where the operators $\lfloor x \rfloor$ and $\lceil x \rceil$ round the argument x down and up respectively to the nearest integer. The choice of $n_h^* = \lfloor n_h / 2 \rfloor$ has the desirable property that the bootstrap weights will never be negative.

For simplicity, the case of stratified three-stage sampling is presented, but the proposed procedure can easily be extended to any number of stages. The without replacement rescaled bootstrap procedure for stratified three-stage sampling is as follows:

(a) Draw a simple random sample of n_{1h}^* PSU's without replacement from the n_{1h} PSU's in the sample. Let δ_{1hi} be equal to 1 if PSU i in stratum h is selected and 0 otherwise. Calculate the PSU bootstrap weights:

$$w_{1hi}^* = w_{1hi} \left(1 - \lambda_{1h} + \lambda_{1h} \frac{n_{1h}}{n_{1h}^*} \delta_{1hi} \right)$$

where $\lambda_{1h} = \sqrt{n_{1h}^* (1 - f_{1h}) / (n_{1h} - n_{1h}^*)}$.

(b) Within each of the PSU's in the sample, draw a simple random sample of n_{2hi}^* SSU's without replacement from the n_{2hi} SSU's in the sample. Let δ_{2hi} be equal to 1 if SSU j in PSU i in stratum h is selected and 0 otherwise. Calculate the conditional SSU bootstrap weights:

$$w_{2hij}^* = $$
$$w_{2hij}\frac{w_{1hi}}{w_{1hi}^*}\left(1-\lambda_{1h}+\lambda_{1h}\frac{n_{1h}}{n_{1h}^*}\delta_{1hi}\right.$$
$$\left.-\lambda_{2hi}\sqrt{\frac{n_{1h}}{n_{1h}^*}}\delta_{1hi}+\lambda_{2hi}\sqrt{\frac{n_{1h}}{n_{1h}^*}}\delta_{1hi}\frac{n_{2hi}}{n_{2hi}^*}\delta_{2hij}\right)$$

where $\lambda_{2hi} = \sqrt{n_{2hi}^* f_{1h}(1-f_{2hi})/(n_{2hi}-n_{2hi}^*)}$.

(c) Within each of the SSU's in the sample, draw a simple random sample of n_{3hij}^* USU's without replacement from the n_{3hij} USU's in the sample. Let δ_{3hijk} be equal to 1 if USU k in SSU j in PSU i in stratum h is selected and 0 otherwise. Calculate the conditional USU bootstrap weights:

$$w_{3hijk}^* = $$
$$w_{3hijk}\frac{w_{1hi}}{w_{1hi}^*}\frac{w_{2hij}}{w_{2hij}^*}\left(1-\lambda_{1h}+\lambda_{1h}\frac{n_{1h}}{n_{1h}^*}\delta_{1hi}\right.$$
$$-\lambda_{2hi}\sqrt{\frac{n_{1h}}{n_{1h}^*}}\delta_{1hi}+\lambda_{2hi}\sqrt{\frac{n_{1h}}{n_{1h}^*}}\delta_{1hi}\frac{n_{2hi}}{n_{2hi}^*}\delta_{2hij}$$
$$-\lambda_{3hij}\sqrt{\frac{n_{1h}}{n_{1h}^*}}\delta_{1hi}\sqrt{\frac{n_{2hi}}{n_{2hi}^*}}\delta_{2hij}$$
$$\left.+\lambda_{3hij}\sqrt{\frac{n_{1h}}{n_{1h}^*}}\delta_{1hi}\sqrt{\frac{n_{2hi}}{n_{2hi}^*}}\delta_{2hij}\frac{n_{3hij}}{n_{3hij}^*}\delta_{3hijk}\right)$$

where $\lambda_{3hij} = \sqrt{n_{3hij}^* f_{1h} f_{2hi}(1-f_{3hij})/(n_{3hij}-n_{3hij}^*)}$.

(d) Calculate the bootstrap estimates:

$$\hat{Y}^* = \sum_{h=1}^{H}\sum_{i=1}^{n_{1h}}\sum_{j=1}^{n_{2hi}}\sum_{k=1}^{n_{3hij}} w_{hijk}^* y_{hijk}, \quad \hat{\theta} = g(\hat{Y}^*)$$

where $w_{hijk}^* = w_{1hi}^* w_{2hij}^* w_{3hijk}^*$.

(e) Independently repeat steps (a) to (d) a large number of times, B, and calculate the bootstrap estimates, $\hat{\theta}^{(1)}$, $\hat{\theta}^{(2)}$, ..., $\hat{\theta}^{(B)}$.

(f) The bootstrap variance estimator of $\hat{\theta}$ is given by:

$$\operatorname{Var}(\hat{\theta}) = E_*(\hat{\theta}-E_*(\hat{\theta}))^2 \tag{3.1}$$

or the Monte Carlo approximation:

$$\operatorname{Var}(\hat{\theta}) = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^{(b)}-\overline{\hat{\theta}})^2$$

where $\overline{\hat{\theta}} = \sum_{b=1}^{B}\hat{\theta}^{(b)}/B$.

It is shown in the Appendix that the multistage rescaled bootstrap variance estimator for stratified three-stage sampling as defined by (3.1) reduces to the standard unbiased three-stage variance estimator (2.1) in the case of $\hat{\theta}$ being a linear estimator. The choice of $n_{1h}^* = \lfloor n_{1h}/2\rfloor$, $n_{2hi}^* = \lfloor n_{2hi}/2\rfloor$ and $n_{3hij}^* = \lfloor n_{3hij}/2\rfloor$ will be optimal and will have the desirable property that the bootstrap weights will never be negative.

The proposed procedure can easily be extended to any number of stages by adding terms of the form $-\lambda_R(\prod_{r=1}^{R-1}\sqrt{(n_r/n_r^*)}\delta_r)+\lambda_R(\prod_{r=1}^{R-1}\sqrt{(n_r/n_r^*)}\delta_r)(n_R/n_R^*)\delta_R$ at each stage, R, to the bootstrap weight adjustments, where $\lambda_R = \sqrt{n_R^*(\prod_{r=1}^{R-1} f_r)(1-f_R)/(n_R-n_R^*)}$.

Yeo, Mantel and Liu (1999) presented an enhancement to the rescaled bootstrap which accounted for adjustments made to the design weights, such as post-stratification. For example, consider a simple case of non-integrated calibration using auxiliary information for two-stage stratified sampling (Estevao and Särndal 2006), which has the dual objectives of producing estimates for both a first-stage variable of interest $Y_1 = \sum_{(hi)\in U} y_{1hi}$ as well as a second-stage variable of interest, $Y_2 = \sum_{(hij)\in U} y_{2hij}$. Assume there exists:

(i) a set of p first-stage auxiliary variables $\boldsymbol{x}_{1hi}$ for which the population totals $\boldsymbol{X}_1 = \sum_{(hi)\in U}\boldsymbol{x}_{1hi}$ are known, and where the population totals are generated from a list frame of PSU's for which the $\boldsymbol{x}_{1hi}$ are known for every PSU in the population; and

(ii) a set of q second-stage auxiliary variables $\boldsymbol{x}_{2hij}$ for which the population totals $\boldsymbol{X}_2 = \sum_{(hij)\in U}\boldsymbol{x}_{2hij}$ are known, where the population totals are acquired from an external source.

The auxiliary variables can be used to form the first-stage and second-stage calibration estimators:

$$\hat{Y}_{\mathrm{CAL1}} = \sum_{(hi)\in s_1}\tilde{w}_{1hi}y_{1hi}$$

$$\hat{Y}_{\mathrm{CAL2}} = \sum_{(hij)\in s_2}\tilde{w}_{12hij}y_{2hij}$$

where the first-stage calibration weights, $\tilde{w}_{1hi}$, and the combined first-stage and second-stage calibration weights, $\tilde{w}_{12hij}$, are given by:

$$\tilde{w}_{1hi} = w_{1hi}\left(1+\left(\boldsymbol{X}_1-\sum_{(hi)\in s_1}w_{1hi}\boldsymbol{x}_{1hi}\right)^T\left(\sum_{(hi)\in s_1}w_{1hi}\boldsymbol{x}_{1hi}\boldsymbol{x}_{1hi}^T\right)^{-1}\boldsymbol{x}_{1hi}\right)$$

$$\tilde{w}_{12hij} = w_{1hi}w_{2hij}\left(1+\left(\boldsymbol{X}_2-\sum_{(hij)\in s_2}w_{1hi}w_{2hij}\boldsymbol{x}_{2hij}\right)^T\left(\sum_{(hij)\in s_2}w_{1hi}w_{2hij}\boldsymbol{x}_{2hij}\boldsymbol{x}_{2hij}^T\right)^{-1}\boldsymbol{x}_{2hij}\right).$$

Then the multistage rescaled bootstrap method can easily be modified in a similar manner to handle these calibration estimators by replacing step (d) in the procedure as follows:

(d) Calculate the first-stage and second-stage calibrated bootstrap weights in the same manner as the first-stage and second-stage calibrated weights:

$$\tilde{w}^*_{1hi} = w^*_{1hi}\left(1 + \left(\boldsymbol{X}_1 - \sum_{(hi)\in s_1} w^*_{1hi}\boldsymbol{x}_{1hi}\right)^T \left(\sum_{(hi)\in s_1} w^*_{1hi}\boldsymbol{x}_{1hi}\boldsymbol{x}^T_{1hi}\right)^{-1} \boldsymbol{x}_{1hi}\right)$$

$$\tilde{w}^*_{12hij} = w^*_{1hi}w^*_{2hij}\left(1 + \left(\boldsymbol{X}_2 - \sum_{(hij)\in s_2} w^*_{1hi}w^*_{2hij}\boldsymbol{x}_{2hij}\right)^T \left(\sum_{(hij)\in s_2} w^*_{1hi}w^*_{2hij}\boldsymbol{x}_{2hij}\boldsymbol{x}^T_{2hij}\right)^{-1} \boldsymbol{x}_{2hij}\right).$$

The first-stage and second-stage calibrated bootstrap estimates are calculated as:

$$\hat{Y}^*_{\text{CAL1}} = \sum_{(hi)\in s_1} \tilde{w}^*_{1hi} y_{1hi}$$

$$\hat{Y}^*_{\text{CAL2}} = \sum_{(hij)\in s_2} \tilde{w}^*_{12hij} y_{2hij}.$$

This procedure can easily be modified to any type of calibration and extended to any number of stages. This modification of the rescaled bootstrap takes into account adjustments made to the design weights due to calibration. Ideally all adjustments made to the design weights, including adjustments due to provider non- response and population under-coverage should also be made to the bootstrap weights.

4. Simulation study

A Monte Carlo simulation study was conducted to examine the performance of the multistage rescaled bootstrap variance estimator. The study was restricted to stratified two-stage sampling. The simulation study was based on ten artificial populations, each of which was stratified into $H = 5$ strata, with $N_{1h} = 50$ first-stage units within each stratum, and $N_{2hi} = 40$ second-stage units within each first-stage unit.

Firstly, the first-stage auxiliary variable x_{1hi} for each first-stage unit i in stratum h was generated from the normal distribution $N(\mu_{x1h}, (1-\rho_{x1b})\,\sigma^2_{x1b}/\rho_{x1b})$. Secondly, the second-stage auxiliary variable, x_{2hij}, and the second-stage target variables, y_{2hij} and z_{2hij}, for each second-stage unit j within first-stage unit i in stratum h were then generated from the multivariate normal distribution $N_3(\boldsymbol{\mu}_{2hi}, \boldsymbol{\Sigma}_{2hi})$ where $\boldsymbol{\mu}_{2hi}$ is the mean vector:

$$\boldsymbol{\mu}_{2hi} = \begin{bmatrix} \mu_{x2hi} \\ \mu_{y2hi} \\ \mu_{z2hi} \end{bmatrix}$$

with $\mu_{x2hi} = \mu_{y2hi} = \mu_{z2hi} = x_{1hi}$, and $\boldsymbol{\Sigma}_{2hi}$ is the variance-covariance matrix:

$$\boldsymbol{\Sigma}_{2hi} = \begin{bmatrix} \sigma^2_{x2hi} & \rho_{xy2hi}\sigma_{x2hi}\sigma_{y2hi} & \rho_{xz2hi}\sigma_{x2hi}\sigma_{z2hi} \\ \rho_{xy2hi}\sigma_{x2hi}\sigma_{y2hi} & \sigma^2_{y2hi} & \rho_{yz2hi}\sigma_{y2hi}\sigma_{z2hi} \\ \rho_{xz2hi}\sigma_{x2hi}\sigma_{z2hi} & \rho_{yz2hi}\sigma_{y2hi}\sigma_{z2hi} & \sigma^2_{z2hi} \end{bmatrix}$$

with $\sigma^2_{x2hi} = \sigma^2_{y2hi} = \sigma^2_{z2hi} = (1-\rho_{w2hi})\sigma^2_{w2hi}/\rho_{w2hi}$.

The parameter values that were kept stable across all ten populations were $\mu_{x1h} = 25 \times (h+1)$, $\sigma^2_{b1h} = 10$, $\sigma^2_{w2hi} = 100$, $\rho_{xy2hi} = \rho_{xz2hi} = 0.75$ and $\rho_{yz2hi} = 0.50$. The parameter values that were varied across the ten populations were f_{1h}, the first-stage sampling fractions, f_{2hi}, the second-stage sampling fractions, ρ_{b1h} and ρ_{w2hi}. These parameter values are presented in Table 1.

Table 1
Characteristics of simulation populations

	f_{1h}	f_{2hi}	ρ_b	ρ_w
Pop I	0.1	0.1	0.75	0.75
Pop II	0.1	0.1	0.25	0.75
Pop III	0.1	0.5	0.75	0.75
Pop IV	0.1	0.5	0.25	0.75
Pop V	0.1	0.5	0.25	0.25
Pop VI	0.5	0.1	0.75	0.75
Pop VII	0.5	0.1	0.75	0.25
Pop VIII	0.5	0.1	0.25	0.25
Pop IX	0.3	0.3	0.75	0.25
Pop X	0.3	0.3	0.25	0.25

The parameters of interest used in the simulation study were the population mean, μ_y, the population ratio, $R_{yz} = \mu_y/\mu_z$, the population correlation coefficient, $\rho_{yz} = \sigma_{yz}/\sigma_y\sigma_z$, the population regression coefficient, $\beta_{yz} = \sigma_{yz}/\sigma^2_y$, and the population median, M_y.

In order to estimate these parameters of interest using the multistage bootstrap variance estimators, a total of S = 20,000 independent two-stage simple random samples were selected without replacement from each of the ten artificial populations. In addition, a grand total of $T = 100{,}000$ independent two-stage simple random samples were selected without replacement from each of the ten artificial populations in order to estimate the true population variances for the parameters of interest. The multistage bootstrap variance estimators were calculated using $B = 100$ bootstrap samples.

The accuracy of the multistage bootstrap variance estimators were compared using the relative biases (RB) and the relative root mean square error (RRMSE). These measures were calculated as:

$$\text{RB} = \frac{1}{\hat{\text{V}}\text{ar}(\hat{Y})}\left[\frac{1}{S}\sum_{s=1}^{S}(\text{Var}_*(\hat{Y}_s) - \hat{\text{V}}\text{ar}(\hat{Y}))\right]$$

$$\text{RRMSE} = \frac{1}{\hat{\text{V}}\text{ar}(\hat{Y})}\sqrt{\frac{1}{S}\sum_{s=1}^{S}(\text{Var}_*(\hat{Y}_s) - \hat{\text{V}}\text{ar}(\hat{Y}))^2}$$

where $\hat{\text{V}}\text{ar}(\hat{Y}) = T^{-1}\sum_{t=1}^{T}(\hat{Y}_t - Y)^2$ is the estimated true population variance, and $\text{Var}_*(\hat{Y}_s)$ are the multistage bootstrap variance estimators for the s^{th} simulation sample.

The multistage rescaled bootstrap variance estimator (MRBE) was compared against the single-stage rescaled bootstrap variance estimator (SRBE) and the multistage general Bernoulli bootstrap variance estimator (BBE) proposed by Funaoka *et al.* (2006), with bootstrap samples using the non-calibration estimation weights, $w_{hij} = w_{1hi}w_{2hij}$. The relative biases and relative root mean square errors of MRBE, SRBE and BBE using the non-calibration estimation weights for the ten artificial populations are given in Tables 2 and 3.

In the case of linear functions, such as means, and non-linear functions, such as ratios, correlation coefficients and regression coefficients, the MRBE performed better than the SRBE and BBE with respect to relative bias and relative root mean square error. While the MRBE performed consistently well across all ten artificial populations, the SRBE only performed well for artificial populations III, IV and V, where the first-stage sampling fractions were small $(f_{1h} = 0.1)$ and the second-stage sampling fractions were large $(f_{2hi} = 0.5)$, and the BBE only performed well for artificial populations VI, VII and VIII, where the first-stage sampling fractions were large $(f_{1h} = 0.5)$ and the second-stage sampling fractions were small $(f_{2hi} = 0.1)$. These sampling fractions were similar to the first-stage and second-stage sampling fractions used in the simulation study presented in Funaoka *et al.* (2006). The different levels of correlation between the first-stage units, and between the second-stage units within the first-stage units, controlled by varying the parameters ρ_b and ρ_w, had little impact on the performance of the variance estimators.

Table 2
Relative bias (%) of variance estimators

	Mean (μ_y)			Mean (μ_z)			Ratio (R_{yz})		
	MRBE	SRBE	BBE	MRBE	SRBE	BBE	MRBE	SRBE	BBE
Pop I	-0.28	-6.73	27.10	0.42	-6.63	27.32	0.00	-9.07	36.22
Pop II	-0.05	-2.21	11.83	0.59	-1.64	12.54	-0.43	-9.26	36.40
Pop III	-0.79	-2.63	3.62	-0.93	-2.66	3.40	-0.17	-5.30	5.19
Pop IV	-0.23	-0.52	3.60	-0.18	-0.46	3.61	0.53	-4.65	5.98
Pop V	0.15	-1.60	4.55	0.15	-1.64	4.54	0.52	-4.85	6.41
Pop VI	0.70	-39.18	-0.34	0.65	-39.36	-0.28	1.57	-46.40	1.30
Pop VII	0.19	-46.19	-0.26	-0.06	-46.48	-0.57	-0.27	-48.19	-0.73
Pop VIII	0.37	-38.62	-0.41	0.23	-39.36	-0.46	-0.26	-47.93	-0.62
Pop IX	0.42	-20.85	-7.76	-0.51	-20.03	-8.41	0.13	-23.13	-8.87
Pop X	-0.56	-12.35	-6.08	0.70	-10.87	-6.93	-0.72	-23.70	-9.51
	Correlation Coefficient (ρ_{yz})			Regression Coefficient (β_{yz})			Median (M_y)		
	MRBE	SRBE	BBE	MRBE	SRBE	BBE	MRBE	SRBE	BBE
Pop I	-2.31	-10.23	32.17	-0.08	-9.05	36.41	19.04	-19.86	33.21
Pop II	-1.51	-8.41	29.65	0.05	-8.74	36.41	19.29	2.42	40.85
Pop III	0.36	-4.37	5.69	0.05	-5.12	5.42	7.50	4.28	9.72
Pop IV	2.18	-0.60	7.17	0.28	-5.05	5.70	17.40	16.17	34.37
Pop V	0.79	-2.71	5.95	0.26	-5.40	6.34	8.29	4.78	11.49
Pop VI	0.32	-46.67	0.14	0.89	-46.59	0.69	13.57	-33.56	9.15
Pop VII	-0.07	-46.78	-0.39	-0.21	-47.85	-0.60	14.68	-38.16	11.86
Pop VIII	0.31	-44.25	-0.27	-0.09	-47.54	-0.55	2.09	-38.90	-0.64
Pop IX	-0.93	-23.02	-9.30	-0.20	-23.48	-9.20	8.08	-17.23	-1.97
Pop X	-0.82	-19.35	-8.24	-1.02	-23.89	-9.75	2.10	-13.84	-5.46

Note: The largest simulation error on the relative biases was less than 0.7%.

In the case of non-smooth statistics, such as medians, both the MRBE and the BBE tended to overestimate the true population variances, while the SRBE tended to underestimate the true population variances. Furthermore, the relative root mean square errors for medians were up to 3 times larger than the relative root mean square errors for means. The MRBE performed better than the BBE for the artificial populations I to V where the first-stage sampling fractions were smaller ($f_{1h} = 0.1$), while the BBE performed slightly better than the MRBE for the artificial populations VI to X where the first-stage sampling fractions were larger ($f_{1h} = 0.3$ or 0.5).

This overestimation of the multistage rescaled bootstrap for medians was similar to the findings shown in the studies by Kovar *et al.* (1988) and Rao *et al.* (1992) for the single-stage rescaled bootstrap. It should be noted that the original rescaled bootstrap introduced by Rao and Wu (1988) was developed only for smooth statistics, such as means, ratios, and correlation and regression coefficients.

The MRBE was examined using the calibration estimation weights, $\tilde{w}_{hij} = w_{1hi}\tilde{w}_{2hij}$, which satisfy the calibration constraint $\sum_{(hij)\in s_2} w_{1hi}\tilde{w}_{2hij}x_{2hij} = X_2$, where $X_2 = \sum_{(hij)\in U} x_{2hij}$ is the population total for the second-stage auxiliary variable. The relative biases and relative root mean square errors of the MRBE using the calibration estimation weights for the four artificial populations II, IV, VII and IX are given in Table 4.

Table 3
Relative root mean square error (%) of variance estimators

	Mean (μ_y)			Mean (μ_z)			Ratio (R_{yz})		
	MRBE	**SRBE**	**BBE**	**MRBE**	**SRBE**	**BBE**	**MRBE**	**SRBE**	**BBE**
Pop I	31.9	32.1	44.6	31.7	31.8	44.4	31.8	32.3	51.4
Pop II	33.9	33.8	38.2	33.4	33.3	38.1	32.2	32.9	51.7
Pop III	33.8	33.8	35.9	33.0	33.0	35.0	33.0	33.1	35.1
Pop IV	35.3	35.3	37.4	35.2	35.2	37.3	32.8	32.8	35.0
Pop V	32.0	31.9	34.2	34.3	34.2	36.5	33.0	33.1	35.6
Pop VI	16.4	40.7	16.5	16.4	40.9	16.5	16.5	47.5	16.5
Pop VII	16.1	47.4	16.4	16.1	47.8	16.4	16.1	49.0	16.1
Pop VIII	16.3	40.3	16.5	16.7	40.9	16.3	16.2	48.8	16.1
Pop IX	19.2	26.7	20.0	19.3	26.3	20.0	19.2	28.6	20.2
Pop X	19.8	22.4	20.2	19.9	21.6	20.3	19.1	29.0	20.6
	Correlation Coefficient (ρ_{yz})			**Regression Coefficient (β_{yz})**			**Median (M_y)**		
	MRBE	**SRBE**	**BBE**	**MRBE**	**SRBE**	**BBE**	**MRBE**	**SRBE**	**BBE**
Pop I	47.8	46.3	68.7	36.6	37.2	55.3	88.7	80.1	89.8
Pop II	48.4	47.1	66.6	37.4	37.9	55.6	93.4	91.0	115.9
Pop III	35.9	35.6	38.4	37.5	37.6	39.9	80.4	80.3	81.1
Pop IV	42.6	42.2	45.4	38.0	38.0	40.3	97.5	96.6	127.3
Pop V	40.3	40.0	43.3	37.3	37.5	40.1	31.5	30.7	63.3
Pop VI	21.6	48.4	21.7	16.9	47.8	17.0	55.3	51.4	52.0
Pop VII	21.4	48.4	21.3	16.9	49.0	16.8	53.5	51.4	51.4
Pop VIII	21.6	46.3	21.5	17.0	48.6	16.9	41.8	49.7	40.3
Pop IX	21.5	29.4	22.5	20.5	29.9	21.6	46.1	42.7	42.7
Pop X	22.7	27.8	23.4	20.6	30.2	21.9	39.7	38.9	37.9

Table 4
Relative bias (%) and relative root mean square error (%) of rescaled bootstrap variance estimator

	μ_y	R_{yz}	ρ_{yz}	β_{yz}	M_y
		Relative Bias (%)			
Pop II	-0.42	-0.29	-1.51	-0.08	20.98
Pop IV	0.40	0.49	1.83	0.08	18.28
Pop VII	-0.22	-0.24	-0.03	-0.28	12.24
Pop IX	0.62	0.19	-1.00	-0.20	7.24
		Relative Root Mean Square Error (%)			
Pop II	32.6	32.4	48.4	37.3	97.8
Pop IV	32.8	32.8	44.6	37.9	99.4
Pop VII	16.2	16.1	21.5	16.9	50.0
Pop IX	19.1	19.2	21.6	20.5	43.8

Note: The largest simulation error on the relative biases was less than 0.6%.

The relative biases and relative root mean square errors of the MRBE using the calibration estimation weights were similar to those using the non-calibration estimation weights.

5. Conclusion

This paper extends the rescaled bootstrap procedure to multistage sampling where units are selected using simple random sampling without replacement at each stage. Under the proposed multistage rescaled bootstrap method, the bootstrap samples are selected without replacement and rescaling factors are applied to the survey weights. This proposed method is relatively simple to implement and requires considerably less random number generations than the multistage general Bernoulli bootstrap method. The proposed method is also suitable for a wide range of reweighting techniques, including calibration, and adjustments due to provider non-response and population under-coverage. Furthermore, the results of the Monte Carlo simulation study indicate that the multistage rescaled bootstrap performs much better than the single-stage rescaled bootstrap and the multistage Bernoulli bootstrap for smooth statistics, such as means, ratios, and correlation and regression coefficients.

Appendix

In this Appendix it is shown that the multistage rescaled bootstrap variance estimator for stratified three-stage sampling reduces to the standard unbiased three-stage variance estimator (2.1) in the case of $\hat{\theta}$ being the linear estimator, $\hat{Y}^* = \sum_{h=1}^{H}\sum_{i=1}^{n_{1h}}\sum_{j=1}^{n_{2hi}}\sum_{k=1}^{n_{3hij}} w^*_{hijk} y_{hijk}$. The bootstrap variance estimator of $\hat{Y}^*$ is given by:

$$\mathrm{Var}(\hat{Y}^*) = \mathrm{Var}_{1*}(E_{2*}(E_{3*}(\hat{Y}^*))) + E_{1*}(\mathrm{Var}_{2*}(E_{3*}(\hat{Y}^*))) + E_{1*}(E_{2*}(\mathrm{Var}_{3*}(\hat{Y}^*))).$$

Using standard results on the expectation and variance with respect to the SRSWOR bootstrap sampling and some tedious but straightforward algebra, the components of bootstrap variance estimator are given below. The conditional expectation of $\hat{Y}^*$ given s_3 is

$$E_{3*}(\hat{Y}^*) = \sum_{h=1}^{H}\sum_{i=1}^{n_{1h}}\sum_{j=1}^{n_{2hi}} w_{1i} w_{2ij}\left(1 - \lambda_{1h} + \lambda_{1h}\frac{n_{1h}}{n^*_{1h}}\delta_{1hi} - \lambda_{2hi}\sqrt{\frac{n_{1h}}{n^*_{1h}}\delta_{1hi}} + \lambda_{2hi}\sqrt{\frac{n_{1h}}{n^*_{1h}}\delta_{1hi}}\,\frac{n_{2hi}}{n^*_{2hi}}\delta_{2hij}\right)\hat{Y}_{ij}$$

and the conditional variance of $\hat{Y}^*$ given s_3 is

$$\mathrm{Var}_{3*}(\hat{Y}^*) = \sum_{h=1}^{H}\frac{N_{1h}}{n^*_{1h}}\sum_{i=1}^{n_{1h}}\frac{N_{2hi}}{n^*_{2hi}}\sum_{j=1}^{n_{2hi}}\delta_{1hi}\,\delta_{2hij}\frac{N^2_{3hij}}{n_{3hij}}(1 - f_{3hij})\,s^2_{3hij}.$$

The conditional expectation of $E_{3*}(\hat{Y}^*)$ and $\mathrm{Var}_{3*}(\hat{Y}^*)$ given s_2 are

$$E_{2*}(E_{3*}(\hat{Y}^*)) = \sum_{h=1}^{H}\sum_{i=1}^{n_{1h}} w_{1i}(1 - \lambda_{1h} + \lambda_{1h}\frac{n_{1h}}{n^*_{1h}}\delta_{1hi})\,\hat{Y}_{hi}$$

$$E_{2*}(\mathrm{Var}_{3*}(\hat{Y}^*)) = \sum_{h=1}^{H}\frac{N_{1h}}{n^*_{1h}}\sum_{i=1}^{n_{1h}}\frac{N_{2hi}}{m_{2hi}}\sum_{j=1}^{n_{2hi}}\delta_{1hi}\frac{N^2_{3hij}}{n_{3hij}}(1 - f_{3hij})\,s^2_{3hij}$$

and the conditional variance of $E_{3*}(\hat{Y}^*)$ given s_2 is

$$\mathrm{Var}_{2*}(E_{3*}(\hat{Y}^*)) = \sum_{h=1}^{H}\frac{N_{1h}}{n^*_{1h}}\sum_{i=1}^{n_{1h}}\delta_{1hi}\frac{N^2_{2hi}}{n_{2hi}}(1 - f_{2hi})\,s^2_{2hi}.$$

Finally, the conditional expectation of $E_{2*}(\mathrm{Var}_{3*}(\hat{Y}^*))$ and $\mathrm{Var}_{2*}(E_{3*}(\hat{Y}^*))$ given s_1 are

$$E_{1*}(E_{2*}(\mathrm{Var}_{3*}(\hat{Y}^*))) = \sum_{h=1}^{H}\frac{N_{1h}}{n_{1h}}\sum_{i=1}^{n_{1h}}\frac{N_{2hi}}{n_{2hi}}\sum_{j=1}^{n_{2hi}}\frac{N^2_{3hij}}{n_{3hij}}(1 - f_{3hij})\,s^2_{3hij}$$

$$E_{1*}(\mathrm{Var}_{2*}(E_{3*}(\hat{Y}^*))) = \sum_{h=1}^{H}\frac{N_{1h}}{n_{1h}}\sum_{i=1}^{n_{1h}}\frac{N^2_{2hi}}{n_{2hi}}(1 - f_{2hi})\,s^2_{2hi}$$

which are equal to the third and second terms of (2.1) respectively, and the conditional variance of $E_{2*}(E_{3*}(\hat{Y}^*))$ given s_1 is

$$\mathrm{Var}_{1*}(E_{2*}(E_{3*}(\hat{Y}^*))) = \sum_{h=1}^{H}\frac{N^2_{1h}}{n_{1h}}(1 - f_{1h})\,s^2_{1h}$$

which is equal to the first term of (2.1).

References

Australian Bureau of Statistics (ABS) (2008). Employee Earnings and Hours, Catalogue Number 6306.0.

Bickel, P.J., and Freedman, D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Annals of Statistics*, 12, 470-482.

Chipperfield, J., and Preston, J. (2007). Efficient bootstrap for business surveys. *Survey Methodology*, 33, 167-172.

Estevao, V., and Särndal, C.-E. (2006). Survey estimates by calibration on complex auxiliary information. *International Statistical Review*, 74, 127-147.

Funaoka, F., Saigo, H., Sitter, R.R. and Toida, T. (2006). Bernoulli bootstrap for stratified multistage sampling. *Survey Methodology*, 32, 151-156.

Gross, S. (1980). Median estimation in sample surveys. *Proceedings of the Section on Survey Research Methods*, 181-184.

Kovar, J.G., Rao, J.N.K. and Wu, C.F.J. (1988). Bootstrap and other methods to measure errors in survey estimates. *Canadian Journal of Statistics*, 16, 25-45.

McCarthy, P.J., and Snowden, C.B. (1985). The bootstrap and finite population sampling. Vital and Health Statistics (Series 2 No 95), Public Health Service Publication 85-1369, Washington, DC: U.S. Government Printing Office.

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

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18, 209-217.

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

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

Sitter, R.R. (1992a). A resampling procedure for complex survey data. *Journal of the American Statistical Association*, 87, 755-765.

Sitter, R.R. (1992b). Comparing three bootstrap methods for survey data. *Canadian Journal of Statistics*, 20, 135-154.

White, V., and Hayman J. (2006). Smoking behaviours of Australian secondary students in 2005**.** National Drug Strategy Monograph Series No. 59. Canberra: Australian Government Department of Health and Ageing.

Yeo, D., Mantel, H. and Liu T.-P. (1999). Bootstrap variance estimation for the National Population Health Survey. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 778-783.