Journal of Official Statistics, Vol. 26, No. 1, 2010, pp. 193–207

Comparing Four Bootstrap Methods for Stratified Three-Stage Sampling

Hiroshi Saigo[1]

For a stratified three-stage sampling design with simple random sampling without replacement at each stage, only the Bernoulli bootstrap is currently available as a bootstrap for design-based inference under arbitrary sampling fractions. This article extends three other methods (the mirror-match bootstrap, the rescaling bootstrap, and the without-replacement bootstrap) to the design and conducts simulation study that estimates variances and constructs coverage intervals for a population total and selected quantiles. The without-replacement bootstrap proves the least biased of the four methods when estimating the variances of quantiles. Otherwise, the methods are comparable.

Key words: Multistage sampling; high sampling fractions; resampling methods; quantile estimation.

1. Introduction

For most stratified multi-stage sampling designs, unbiased variance estimators of statistics expressed by linear functions of the observations are available. However, for nonlinear statistics and functionals, closed-form variance formulas are often unavailable. Consequently, bootstrap methods can serve for variance estimation of such statistics under stratified multi-stage sampling designs.

When the sampling fractions at the first stage are small, bootstrap methods for stratified multi-stage sampling designs are simplified because without-replacement sampling can be approximated by with-replacement sampling (Shao and Tu 1995, p. 235). But when the sampling fractions at the first stage are not negligible, bootstrap methods for consistent variance estimation become complicated and few have been developed. For instance, for a stratified three-stage with simple random sampling without replacement at each stage ($\mathrm{ST-SI}^3$) with arbitrary sampling fractions, no bootstrap procedure is available except the Bernoulli Bootstrap (BBE) proposed by Funaoka, Saigo, Sitter, and Toida (2006) for the 1997 Japanese National Survey of Prices (NSP).

A resampling method for quantile estimation is particularly important for the NSP because to analyze price formations for major consumers' goods, comprehensive quantile estimates are presented in the NSP report. In 1997, $\mathrm{ST-SI}^3$ was conducted in the NSP.

---

[1] Faculty of Political Science and Economics, Waseda University, Tokyo 169-8050, Japan. Email: saigo@waseda.jp

Acknowledgments: This work is supported by a grant from the Japan Society for the Promotion of Science. The author thanks Dr. Randy R. Sitter for suggesting the theme of this article. The author also thanks an associate editor and anonymous referees for valuable comments that substantially improved the article.

First, 3,233 municipalities (the primary sampling units) were stratified into 537 strata according to prefectures, economic spheres, and population sizes. At the first stage, the simple random sampling without replacement was conducted. Since price formations were locally correlated, large first-stage sampling fractions were adopted according to population size: 1/1, 2/3, 1/3, 1/5, and 1/15. At the second stage, all large-scale outlets were enumerated, while for small scale outlets, sampled municipalities were divided into survey areas (the secondary sampling units) each consisting of about 100 outlets. Systematic sampling was used to choose survey areas. The second-stage sampling fractions were between 0.1 and 1.0. Finally, in each selected survey area, about 40 small outlets were selected via ordered systematic sampling. Funaoka et al. (2006) regarded the systematic sampling at the second and third stages as simple random sampling without replacement and applied the BBE for quantile estimation.

However, Funaoka et al. (2006) studied only the BBE. A comparative study of several resampling methods is necessary. Although thorough simulation studies have been conducted about the bootstrap for sampling from a finite population (e.g., Kovar, Rao, and Wu 1988; Sitter 1992a, 1992b), none of them has compared the methods under $\mathrm{ST-SI}^3$. The study is particularly important not only for the NSP but also the Family Income Expenditure Survey in Japan, where the $\mathrm{ST-SI}^3$ design is employed.

In this article, we extend the following bootstrap approaches to $\mathrm{ST-SI}^3$ and compare them with the BBE for variance estimation as well as interval estimation through a simulation study using pseudo-populations: the mirror-match bootstrap (BMM) by Sitter (1992a) which covers $\mathrm{ST-SI}^2$; the rescaling bootstrap (BRS) originally proposed by Rao and Wu (1988) for $\mathrm{ST-SI}^2$ by rescaling the study variable and then modified by Rao, Wu, and Yue (1992) for $\mathrm{ST-SI}$ with replacement by rescaling the sampling weights to handle quantile estimation; and the without-replacement bootstrap (BWO) originally argued by Gross (1980) for $\mathrm{ST-SI}$ and then extended by Sitter (1992b) to $\mathrm{ST-SI}^2$. The with-replacement bootstrap (BWR) is included as a special case of the BMM (Sitter 1992a). A theoretical comparison of the four methods is beyond the scope of this article.

This article is organized as follows. Section 2 describes $\mathrm{ST-SI}^3$. Section 3 presents the BBE and an extension of the BMM, the BRS, and the BWO to $\mathrm{ST-SI}^3$. In Section 4, we conduct a simulation study to compare the four bootstrap methods under $\mathrm{ST-SI}^3$. Concluding remarks are made in Section 5.

2. Stratified Three-Stage Design

In this section, we describe $\mathrm{ST-SI}^3$, a stratified three-stage design with simple random sampling without replacement at each stage (see Särndal, Swensson, and Wretman 1992, pp. 146-150).

Suppose that a population is stratified into H strata, labeled as h ($h = 1, 2, \ldots, H$). Stratum h has N_h primary sampling units (PSUs) in it, labeled as $i \in \mathcal{P}_{h1} = \{1, 2, \ldots, N_h\}$. Primary sampling unit $i \in \mathcal{P}_{h1}$ has M_{hi} secondary sampling units (SSUs) in it, labeled as $j \in \mathcal{P}_{h2i} = \{1, 2, \ldots, M_{hi}\}$. Secondary sampling unit $j \in \mathcal{P}_{h2i}$ has L_{hij} ultimate sampling units (USUs) in it, labeled as $k \in \mathcal{P}_{h3ij} = \{1, 2, \ldots, L_{hij}\}$. Ultimate sampling unit $k \in \mathcal{P}_{h3ij}$ has the characteristic(s) of interest y_{hijk}. The population total is given by $Y_{....} = \sum_{h=1}^{H} \sum_{i \in \mathcal{P}_{h1}} \sum_{j \in \mathcal{P}_{h2i}} \sum_{k \in \mathcal{P}_{h3ij}} y_{hijk}$.

Under ST−SI^3, sampling is carried out independently in different strata. In stratum h, we take a simple random sample without replacement (SI) of size n_h from $\mathcal{P}_{h1}$ and denote the set of the sampled labels by $\mathcal{S}_{h1}$. Then, for $i \in \mathcal{S}_{h1}$, we take an SI of size m_{hi} from $\mathcal{P}_{h2i}$ and denote the set of the sampled labels by $\mathcal{S}_{h2i}$. Finally, for $j \in \mathcal{S}_{h2i}$, we take an SI of size l_{hij} from $\mathcal{P}_{h3ij}$ and denote the set of the sampled labels by $\mathcal{S}_{h3ij}$.

An unbiased estimator of the population total is given by

$$\hat{Y}.... = \sum_{h=1}^{H} \sum_{i \in \mathcal{S}_{h1}} \sum_{j \in \mathcal{S}_{h2i}} \sum_{k \in \mathcal{S}_{h3ij}} w_{hijk} y_{hijk} \tag{1}$$

where $w_{hijk} = f_{h1}^{-1} f_{h2i}^{-1} f_{h3ij}^{-1}$ with $f_{h1} = n_h / N_h$, $f_{h2i} = m_{hi}/M_{hi}$, and $f_{h3ij} = l_{hij}/L_{hij}$. An unbiased variance estimator for $\hat{Y}....$ is given by

$$v(\hat{Y}....) = \sum_{h=1}^{H} \left\{ \frac{N_h^2 (1 - f_{h1})}{n_h} s_{h1}^2 + \sum_{i=1}^{n_h} \frac{N_h M_{hi}^2 (1 - f_{h2i})}{n_h m_{hi}} s_{h2i}^2 \right.$$
$$\left. + \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \frac{N_h M_{hi} L_{hij}^2 (1 - f_{h3ij})}{n_h \, m_{hi} \, l_{hij}} s_{h3ij}^2 \right\}$$

where $s_{h1}^2 = (n_h - 1)^{-1} \sum_{i \in \mathcal{S}_{h1}} \left(\hat{Y}_{hi}.. - \bar{\hat{Y}}_{h}... \right)^2$, $s_{h2i}^2 = (m_{hi} - 1)^{-1} \sum_{j \in \mathcal{S}_{h2ij}} \left(\hat{Y}_{hij}. - \bar{\hat{Y}}_{hi}.. \right)^2$, and $s_{h3ij}^2 = (l_{hji} - 1)^{-1} \sum_{i \in \mathcal{S}_{h3ijk}} \left(y_{hijk} - \bar{y}_{hij}. \right)^2$ with $\hat{Y}_{hi}. = \sum_{j \in \mathcal{S}_{h2i}} \sum_{k \in \mathcal{S}_{h3ij}} f_{h2i}^{-1} f_{h3ij}^{-1} y_{hijk}$, $\bar{\hat{Y}}_{h}... = n_h^{-1} \sum_{j \in \mathcal{S}_{h1}} \hat{Y}_{hi}..$, $\hat{Y}_{hij}. = \sum_{k \in \mathcal{S}_{h3ij}} f_{h3ij}^{-1} y_{hijk}$, $\bar{\hat{Y}}_{hi}.. = m_{hi}^{-1} \sum_{j \in \mathcal{S}_{h2i}} \hat{Y}_{hij}.$, and $\bar{y}_{hij}. = l_{ijk}^{-1} \sum_{k \in \mathcal{S}_{h3ij}} y_{hijk}$.

To estimate the distribution function $F(x) = \sum_{h=1}^{H} \sum_{i \in \mathcal{P}_{h1}} \sum_{j \in \mathcal{P}_{h2i}} \sum_{k \in \mathcal{P}_{h3ij}} I(y_{hijk} \le x) / \sum_{h=1}^{H} \sum_{i \in \mathcal{P}_{h1}} \sum_{j \in \mathcal{P}_{h2i}} L_{hij}$, where $I(\cdot)$ is the indicator function, an unbiased point estimator is given by

$$\hat{F}(x) = \sum_{h=1}^{H} \sum_{i \in \mathcal{S}_{h1}} \sum_{j \in \mathcal{S}_{h2i}} \sum_{k \in \mathcal{S}_{h3ij}} w_{hijk} I(y_{hijk} \le x) \Big/ \sum_{h=1}^{H} \sum_{i \in \mathcal{S}_{h1}} \sum_{j \in \mathcal{S}_{h2i}} \sum_{k \in \mathcal{S}_{h3ij}} w_{hijk}$$

A closed-form variance formula is provided by replacing y_{hijk} with $I(y_{hijk} \le x)$ in $v(\hat{Y}....)$.

For estimating quantile $F^{-1}(p) = \inf\{x : F(x) \ge p\}$ for $p \in (0, 1)$, the direct inversion estimator $\hat{F}^{-1}(p) = \inf\{x : \hat{F}(x) \ge p\}$ is available. However, no closed-form variance formula is available for $\hat{F}^{-1}(p)$. Although the Woodruff method can handle variance estimation (Woodruff 1952; see also Shao and Tu 1995, p. 238), the bootstrap is a reasonable choice because it can accommodate nonsmoothed statistics other than quantiles as well.

3. Bootstrap Methods

The bootstrap methods considered here can be described as follows. Suppose an estimator of the parameter θ can be written as $\hat{\theta} = t(w_{hijk}, y_{hijk}; h = 1, 2, \ldots, H, i \in \mathcal{S}_{1h}, j \in \mathcal{S}_{h2i}, k \in \mathcal{S}_{h3ij})$, where the sample weight w_{hijk} is given in (1). Then, through the bootstrap method employed, obtain a bootstrap sample $\left\{ \mathcal{S}_{h1}^*, h = 1, 2, \ldots, H; \mathcal{S}_{h2i}^*, i \in \mathcal{S}_{h1}^*; \mathcal{S}_{h3ij}^*, j \in \mathcal{S}_{h2i}^* \right\}$ and calculate $\hat{\theta}^* = t(w_{hijk}^*, y_{hijk}; h = 1, 2, \ldots, H, i \in \mathcal{S}_{1h}^*, j \in \mathcal{S}_{h2i}^*, k \in \mathcal{S}_{h3ij}^*)$. The methods below are different in values of w_{hijk}^* and creation of a bootstrap sample $\{\mathcal{S}_{h1}^*, h = 1, 2, \ldots, H; \mathcal{S}_{h2i}^*, i \in \mathcal{S}_{h1}^*; \mathcal{S}_{h3ij}^*, j \in \mathcal{S}_{h2i}^*\}$. But all of them satisfy the condition that for $\hat{Y}_{\cdots\cdot}^* = \sum_{h=1}^{H} \sum_{i \in \mathcal{S}_{h1}^*} \sum_{j \in \mathcal{S}_{h2i}^*} \sum_{k \in \mathcal{S}_{h3ij}^*} w_{hijk}^* y_{hijk}$, $E_*(\hat{Y}_{\cdots\cdot}^*) = \hat{Y}_{\cdots\cdot}$ and $V_*(\hat{Y}_{\cdots\cdot}^*) = v(\hat{Y}_{\cdots\cdot})$, where $E_*(\cdot)$ and $V_*(\cdot)$ are the expectation and variance under repeated bootstrap resampling, respectively. The proofs of the results for the four methods are given in a separate appendix available from the author upon request.

For variance estimation, one can perform a Monte Carlo simulation, i.e., repeat the above resampling and calculation of $\hat{\theta}^*$ a large number of times B and obtain an estimate by $v_{boot}(\hat{\theta}) = (B-1)\sum_{b=1}^{B}\left(\hat{\theta}_b^* - \hat{\theta}_{\cdot}^*\right)^2$, where $\hat{\theta}_b^*$ is the value of $\hat{\theta}^*$ in the bth bootstrap sample and $\hat{\theta}_{\cdot}^* = B^{-1}\sum_{b=1}^{B}\hat{\theta}_b^*$.

3.1. Bernoulli Bootstrap

In the BBE proposed by Funaoka et al. (2006), a bootstrap sample is constructed through random replacement of the sampled units. The procedure is performed independently for $h = 1, 2, \ldots, H$.

Step 1. Choose $(n_h - 1)$ labels by simple random sampling with replacement from $\mathcal{S}_{h1}$. Denote the candidate set by $\mathcal{C}_{h1}^*$. For each $i \in \mathcal{S}_{h1}$, we: (a) keep it in the bootstrap sample with probability $p_{h1} = 1 - (1/2)\left(1 - n_h^{-1}\right)^{-1}(1 - f_{1h})$; or (b) replace it with one randomly selected from $\mathcal{C}_{h1}^*$. If (a) is the case, go to Step 2.

Step 2. For i kept at Step 1, choose $(m_{hi} - 1)$ labels by simple random sampling with replacement from $\mathcal{S}_{h2i}$. Denote the candidate set by $\mathcal{C}_{h2i}^*$. For each $j \in \mathcal{S}_{h2i}$, we: (c) keep it in the bootstrap sample with probability $p_{h2i} = 1 - (1/2)p_h^{-1} f_{1h}(1 - m_{hi}^{-1})^{-1}(1 - f_{2hi})$; or (d) replace it with one randomly selected from $\mathcal{C}_{h2i}^*$. If (c) is the case, go to Step 3.

Step 3. For j kept at Step 2, choose $(l_{hij} - 1)$ labels by simple random sampling with replacement from $\mathcal{S}_{h3ij}$. Denote the candidate set by $\mathcal{C}_{h3ij}^*$. For each $k \in \mathcal{S}_{h3ij}$, we: (e) keep it in the bootstrap sample with probability $p_{h3ij} = 1 - (1/2)p_h^{-1} f_{1h} p_{h2i}^{-1} f_{2hi}(1 - l_{hij}^{-1})^{-1}(1 - f_{3hij})$; or (f) replace it with one randomly selected from $\mathcal{C}_{h3ij}^*$.

Denote the resultant bootstrap sample by $\left\{\mathcal{S}_{h1}^*, h = 1, 2, \ldots, H; \mathcal{S}_{h2i}^*, i \in \mathcal{S}_{h1}^*; \mathcal{S}_{h3ij}^*, j \in \mathcal{S}_{h2i}^*\right\}$ and let $w_{hijk}^* = w_{hijk}$.

Creating the candidate sets is necessary to make the procedure feasible for any n_h, m_{hi}, $l_{hij} \geq 2$ (Funaoka et al. 2006).

Obviously, the BBE retains the original sample sizes and the original sample weights. This is desirable in dealing with randomly imputed survey data (Saigo, Shao, and Sitter 2001).

3.2. Mirror-Match Bootstrap

The BMM proposed by Sitter (1992a) can be extended to ST–SI^3 as follows. The procedure is performed independently for $h = 1, 2, \ldots, H$.

Step 0. Choose n_h', m_h', and l_{hij}' such that $1 \leq n_h' \leq n_h'/(2 - f_{1h})$, $1 \leq m_{hi}' \leq m_{hi}/\{1 + (1 - f_{h2i})(n_h - n_h')/(N_h - n_h)\}$, and $1 \leq l_{hij}' \leq l_{hij}/\{1 + (1 - f_{h3ij})(m_{hi} - m_{hi}')/(M_{hi} - m_{hi})\}$. Let $f_{h1}^* = n_h'/n_h$, $f_{h2i}^* = m_{hi}'/m_{hi}$, $f_{h3ij}^* = l_{hij}'/l_{hij}$, $k_{h1} = \{n_h(1 - f_{h1}^*)\}/\{n_h'(1 - f_{h1})\}$, $k_{h2i} = \{N_h(1 - f_{h1})\}/\{n_h(1 - f_{h1}^*)\} \cdot \{m_{hi}(1 - f_{h2i}^*)\}/$

$\{m'_h(1-f_{h2i})\}$, and $k_{h3ij}=\{M_{hi}(1-f_{h2i})\}/\{m_{hi}(1-f^*_{h2i})\}\cdot\{l_{hij}(1-f^*_{h3ij})\}/\{l'_{hij}(1-f_{h3ij})\}$. If desirable, we may randomize k_{h1}, k_{h2i}, and k_{h3ij} to realize $E_*(f^*_{h1})=f_{h1}$, $E_*(f^*_{h2i})=f_{h2i}$, and $E_*(f^*_{h3ij})=f_{h3ij}$. See the comment made in the second paragraph below Step 3.

Step 1. If k_{h1} is an integer, $\tilde{k}_{h1}=k_{h1}$. If not, let $\tilde{k}_{h1}=\lfloor k_{h1}\rfloor$ with probability $p_{h1}=\left(k_{h1}^{-1}-\lceil k_{h1}\rceil^{-1}\right)/\left(\lfloor k_{h1}\rfloor^{-1}-\lceil k_{h1}\rceil^{-1}\right)$ or $\tilde{k}_{h1}=\lceil k_{h1}\rceil$ otherwise. Repeat independently $\tilde{k}_{h1}$ times simple random sampling without replacement of size n'_h from $\mathcal{S}_{h1}$. Denote the subsampled $\tilde{n}^*_h=n'_h\tilde{k}_{h1}$ labels by $\mathcal{S}^*_{h1}$.

Step 2. For each $i\in\mathcal{S}^*_{h1}$: if K_{h2i} is an integer, $\tilde{k}_{h2i}=k_{h2i}$; if not, let $\tilde{k}_{h2i}=\lfloor k_{h2i}\rfloor$ with probability $p_{h2i}=\left(k_{h2i}^{-1}-\lceil k_{h2i}\rceil^{-1}\right)/\left(\lfloor k_{h2i}\rfloor^{-1}-\lceil k_{h2i}\rceil^{-1}\right)$ or $\tilde{k}_{h2i}=\lceil k_{h2i}\rceil$ otherwise. Repeat independently $\tilde{k}_{h2i}$ times simple random sampling without replacement of size m'_{hi} from $\mathcal{S}_{h2i}$. Denote the subsampled $\tilde{m}^*_{hi}=m'_{hi}\tilde{k}_{h2i}$ labels by $\mathcal{S}^*_{h2i}$.

Step 3. For each $j\in\mathcal{S}^*_{h2i}$: if k_{h3ij} is an integer, $\tilde{k}_{h3ij}=k_{h3ij}$; if not, let $\tilde{k}_{h3ij}=\lfloor k_{h3ij}\rfloor$ with probability $p_{h3ij}=\left(k_{h3ij}^{-1}-\lceil k_{h3ij}\rceil^{-1}\right)/\left(\lfloor k_{h3ij}\rfloor^{-1}-\lceil k_{h3ij}\rceil^{-1}\right)$ or $\tilde{k}_{h3ij}=\lceil k_{h3ij}\rceil$ otherwise. Repeat independently $\tilde{k}_{h3ij}$ times simple random sampling without replacement of size l'_{hij} from $\mathcal{S}_{h3ij}$. Denote the subsampled $\tilde{l}^*_{hij}=l'_{hij}\tilde{k}_{h3ij}$ labels by $\mathcal{S}^*_{h3ij}$.

The bootstrap sample weight is given by $w^*_{hijk}=\left(N_h/\tilde{n}^*_h\right)\left(M_{hi}/\tilde{m}^*_{hi}\right)\left(L_{hij}/\tilde{l}^*_{hij}\right)$. To conduct a Monte Carlo simulation, repeat Steps 1–3 a large number of times B. In the separate appendix, it is shown that $n_h/(2-f_h)\geq 1$, $m_{hi}/\{1+(1-f_{h2i})(n_h-n'_h)/(N_h-n_h)\}\geq 1$ and $l_{hij}/\{1+(1-f_{h3ij})(m_{hi}-m'_{hi})/(M_{hi}-m_{hi})\}\geq 1$ for $2\leq n_h<N_h$, $2\leq m_{hi}<M_{hi}$, and $2\leq l_{hij}<L_{hij}$. Thus, the method is always feasible. Note that by letting $n'_h=m'_{hi}=l'_{hij}=1$, the with-replacement bootstrap (BWR) follows from the BMM.

If $1\leq n_hf_{h1}$, $1\leq m_{hi}f_{h2i}$, and $1\leq l_{hij}f_{h3ij}$, randomizing n'_h, m'_{hi}, and l'_{hij} may yield $E_*\left(f^*_{h1}\right)=f_{h1}$, $E_*\left(f^*_{h2i}\right)=f_{h2i}$, and $E_*\left(f^*_{h3ij}\right)=f_{h3ij}$. Specifically, replace the first sentence in step 0 by "Let $\tilde{n}'_h=\lfloor n_hf_{h1}\rfloor$ with probability $\lceil n_hf_{h1}\rceil-n_hf_{h1}$ or $\tilde{n}'_h=\lceil n_hf_{h1}\rceil$ otherwise; let $\tilde{m}'_{hi}=\lfloor m_{hi}f_{h2i}\rfloor$ with probability $\lceil m_{hi}f_{h2i}\rceil-m_{hi}f_{h2i}$ or $\tilde{m}'_{hi}=\lceil m_{hi}f_{h2i}\rceil$ otherwise; and let $\tilde{l}'_h=\lfloor l_{hij}f_{h3ij}\rfloor$ with probability $\lceil l_{hij}f_{h3ij}\rceil-l_{hij}f_{h3ij}$ or $\tilde{l}'_{hij}=\lceil l_{hij}f_{h3ij}\rceil$ otherwise." Then replace n'_h, m'_{hi}, and l'_{hij} with $\tilde{n}'_h, \tilde{m}'_{hi}$, and $\tilde{l}'_{hij}$ in the second sentence in Step 0.

Under ST–SI, the conditions $E_*\left(f^*_{h1}\right)=f_{h1}(h=1,2,\ldots,H)$ ensure third-order moment matching (Sitter 1992a). Although no theoretical explanation is available for a multistage design, we may pursue the conditions $E_*\left(f^*_{h1}\right)=f_{h1}$, $E_*\left(f^*_{h2i}\right)=f_{h2i}$, and $E_*\left(f^*_{h3ij}\right)=f_{h3ij}$ that possibly improve the BMM's performance.

To implement a Monte Carlo simulation, repeat Steps 0–3 a large number of times B.

3.3. *Rescaling Bootstrap*

Rao and Wu (1988) proposed a BRS method that rescales residuals to provide consistent variance estimation for a parameter defined as a smooth function of the population means. This approach, however, cannot handle variance estimation for sample quantiles. Rao, Wu, and Yue (1992) studied a modified BRS method that rescales sample weights to accommodate quantile estimation under stratified multistage sampling where the first stage sampling fractions are negligible. Here, we present a weight-rescaling BRS for ST–SI^3 with any sampling fractions. The procedure is conducted independently for $h=1,2,\ldots,H$.

Step 0. Choose positive integers n^*_h, m^*_{hi}, and l^*_{hij} for $i \in \mathcal{S}_{h1}, j \in \mathcal{S}_{h2i}$. To avoid negative weights, choose $n^*_h \le (1 - f_{h1})^{-1}(n_h - 1)$, $m^*_{hi} \le n_h(n_h - 1)^{-1}(1 - f_{h1})f^{-1}_{h1}(m_{hi} - 1)$ $(1 - f_{h2i})^{-1}$, and $l^*_{hij} \le m_{hi}(m_{hi} - 1)^{-1}(1 - f_{h2i})f^{-1}_{h2i}(l_{hij} - 1)(1 - f_{h3ij})^{-1}$ if possible. For example, if $f_{h1} \le 1/2$ and $f_{h2i} \le 1/2$ for all h and i, we may choose $n^*_h = n_h - 1, m^*_{hi} = m_{hi} - 1$, and $l^*_{hij} = l_{hij} - 1$.

Step 1. Choose n^*_h labels randomly with replacement from $\mathcal{S}_{h1}$. Let $\tilde{n}^*_{hi}$ be the number of times label $i \in \mathcal{S}_{h1}$ is selected. We may equivalently carry out this step by letting $\left(\tilde{n}^*_{h1}, \tilde{m}^*_{h2}, \ldots, \tilde{n}^*_{hn_h}\right) \sim \mathrm{MN}\left(n^*_h; 1/n_h, 1/n_h, \ldots, 1/n_h\right)$, where MN stands for the multinomial distribution.

Step 2. If $\tilde{n}^*_{hi} \ge 1$ for $i \in \mathcal{S}_{h1}$, choose $\tilde{n}^*_h m^*_{hi}$ labels randomly with replacement from $\mathcal{S}_{h2i}$. Let $\tilde{m}^*_{hij}$ be the number of times label $j \in \mathcal{S}_{h2i}$ is selected. Equivalently, let $\left(\tilde{m}^*_{hi1}, \tilde{m}^*_{hi2}, \ldots, \tilde{m}^*_{him_{hi}}\right) \sim \mathrm{MN}\left(n^*_{hi}m^*_{hi}; 1/m_{hi}, 1/m_{hi}, \ldots, 1/m_{hi}\right)$, If $\tilde{n}^*_{hi} = 0$, let $\tilde{m}^*_{hij} = 0$ for $j \in \mathcal{S}_{h2i}$.

Step 3. If $\tilde{m}^*_{hij} \ge 1$, choose $\tilde{m}^*_{hij}\tilde{l}^*_{hij}$ labels randomly with replacement from $\mathcal{S}_{h3ij}$. Let $\tilde{l}^*_{hijk}$ be the number of times label $k \in \mathcal{S}_{h3ij}$ is selected. Equivalently, let $\left(\tilde{l}^*_{hij1}, \tilde{l}^*_{hij2}, \ldots, \tilde{l}^*_{hijl_{hij}}\right) \sim \mathrm{MN}\left(\tilde{m}^*_{hi}l^*_{hij}; 1/l_{hij}, 1/l_{hij}, \ldots, 1/l_{hij}\right)$. If $\tilde{m}^*_{hij} = 0$, let $\tilde{l}^*_{hijk} = 0$ for $k \in \mathcal{S}_{h3ij}$.

The bootstrap sample is the same as the original sample, but the bootstrap sample weights are given by

$$w^*_{hijk} = w_{hijk}\left\{1 + \alpha_{hi}\left(\tilde{n}^*_{hi} - n^*_h/n_h\right) + \beta_{hij}\left(\tilde{m}^*_{hij} - \tilde{n}^*_{hi}m^*_{hi}/m_{hi}\right) + \gamma_{hijk}\left(\tilde{l}^*_{hijk} - \tilde{m}^*_{hij}\tilde{l}^*_{hij}/l_{hij}\right)\right\},$$

where

$$\alpha_{hi} = n_h\sqrt{1 - f_{h1}}\Big/\sqrt{n^*_h(n_h - 1)},$$

$$\beta_{hij} = \sqrt{n_h f_{h1}/n^*_h}\, m_{hi}\sqrt{1 - f_{h2i}}\Big/\sqrt{m^*_{hi}(m_{hi} - 1)},$$

$$\gamma_{hijk} = \sqrt{n_h f_{h1}/n^*_h}\sqrt{m_{hi} f_{h2i}/m^*_{hi}}\, l_{hij}\sqrt{1 - f_{h3ij}}\Big/\sqrt{l^*_{hij}(l_{hij} - 1)}.$$

Rao and Wu (1988) studied the choice of the resample sizes for ST–SI with replacement when rescaling the study variable y to match the third order moments and to capture the second term of Edgeworth expansions with known strata variances. Note, however, that no theory has been developed for multistage designs.

For variance estimation, repeat Steps 1–3 for a large number of times B and perform a Monte Carlo simulation.

3.4. *Without-Replacement Bootstrap*

An extended BWO method for a stratified two-stage design (Sitter 1992b) can be extended further to ST–SI^3 as follows. The procedure is conducted independently in $h = 1, 2, \ldots, H$.

Step 0. We need the following integer random variables to create a pseudo-population by copying the original sampled units.

For PSUs Compute the following constants:

$$v_h = n_h\left(1 - n_h^{-1} + N_h^{-1}\right), \bar{v}_h = \lceil v_h \rceil, \text{ and } \underline{v}_h = \lfloor v_h \rfloor;$$

$$k_{h1} = N_h n_h^{-2} v_h, \bar{k}_{h1} = \lceil \kappa_{h1} \rceil, \text{ and } \underline{\kappa}_{h1} = \lfloor \kappa_{h1} \rfloor; \text{ and}$$

$$\bar{a}_{h1} = \bar{k}_{h1}\{1 - \underline{v}_h/(n_h \bar{k}_{h1})\}/\{\underline{v}_h(n_h \bar{k}_{h1} - 1)\}, \text{ and}$$

$$\underline{a}_{h1} = \underline{k}_{h1}\{1 - \bar{v}_h/(n_h \underline{k}_{h1})\}/\{\bar{v}_h(n_h \underline{k}_{h1} - 1)\}.$$

Define $(\tilde{n}_h^*, \tilde{k}_{h1})$ as :

$$\Pr\left\{\left(\tilde{n}_h^*, \tilde{k}_{h1}\right) = \left(\underline{v}_h, \bar{k}_{h1}\right)\right\} = p_{h1} \text{ and } \Pr\left\{\left(\tilde{n}_h^*, \tilde{k}_{h1}\right) = \left(\bar{v}_h, \underline{k}_{h1}\right)\right\} = 1 - p_{h1},$$

where $p_{h1} = [(1 - f_{h1})/\{n_h(n_h - 1)\} - \underline{a}_{h1}]/\{\bar{a}_{h1} - \underline{a}_{h1})$

For SSUs Compute the following constants:

$$E_{*(\tilde{n}_h^*, \tilde{k}_{h1})}\left(\tilde{n}_h^{*-1}\right) = p_{h1}\underline{v}_h^{-1} + (1 - p_{h1})\bar{v}_h^{-1};$$

$$\mu_{hi} = m_{hi}\left\{N_h E_{*(\tilde{n}_h^*, \tilde{k}_{h1})}\left(\tilde{n}_h^{*-1}\right)\left(1 - m_{hi}^{-1}\right) + M_{hi}^{-1}\right\}, \bar{\mu}_{hi} = \lceil \mu_{hi} \rceil, \text{ and } \underline{\mu}_{hi} = \lfloor \mu_{hi} \rfloor;$$

$$k_{h2i} = M_{hi} m_{hi}^{-2} \mu_{hi}, \bar{k}_{h2i} = \lceil k_{h2i} \rceil, \text{ and } \underline{k}_{h2i} = \lfloor k_{h2i} \rfloor; \text{ and}$$

$$\bar{a}_{h2i} = \bar{k}_{h2i}\left\{1 - \underline{\mu}_{hi}/(m_{hi}\bar{k}_{h2i})\right\}\Big/\left\{\underline{\mu}_{hi}(m_{hi}\bar{k}_{h2i} - 1)\right\}, \text{ and}$$

$$\underline{a}_{h2i} = \underline{k}_{h2i}\left\{1 - \bar{\mu}_{hi}/(m_{hi}\underline{k}_{h2i})\right\}\Big/\left\{\bar{\mu}_{hi}(m_{hi}\underline{k}_{h2i} - 1)\right\}$$

Define $\left\{\left(\tilde{m}_{hi}^*, \tilde{k}_{h2i}\right)\right\}$ as:

$$\Pr\left\{\left(\tilde{m}_{hi}^*, \tilde{k}_{h2i}\right) = (\underline{\mu}_{hi}, \bar{k}_{h2i})\right\} = p_{h2i} \text{ and } \Pr\left\{\left(\tilde{m}_{hi}^*, \tilde{k}_{h2i}\right) = (\bar{\mu}_{hi}, \underline{k}_{h2i})\right\}$$

$= 1 - p_{h2i}$, where

$$p_{h2i} = \left[(1 - f_{h2i})/\left\{N_h E_{*\left(\tilde{n}_h^*, \tilde{k}_{hi}\right)}\left(\tilde{n}_h^{*-1}\right) m_{hi}(m_{hi} - 1)\right\} - \underline{a}_{h2i}\right]/(\bar{a}_{h2i} - \underline{a}_{h2i})$$

For USUs Compute the following constants:

$$E_{*(\tilde{m}_{hi}^*, \tilde{k}_{h2i})}\left(\tilde{m}_h^{*-1}\right) = p_{h2i}\underline{\mu}_{hi}^{-1} + (1 - p_{h2i})\bar{\mu}_{hi}^{-1};$$

$$\lambda_{hij} = l_{hij}\left(N_h E_{*(\tilde{n}_h^*,\tilde{k}_{h1})}\left(\tilde{n}_h^{*-1}\right) M_h E_{*(\tilde{m}_{hi}^*,\tilde{k}_{h2ij})}\left(\tilde{m}_{hi}^{*-1}\right)\left(1 - l_{hij}^{-1}\right) + L_{hij}^{-1}\right),$$

$\bar{\lambda}_{hij} = \lceil\lambda_{hij}\rceil$, and $\underline{\lambda}_{hij} = \lfloor\lambda_{hij}\rfloor$;

$k_{h3ij} = L_{hij} l_{hij}^{-2}\lambda_{hij}, \bar{k}_{h3ij} = \lceil k_{h3ij}\rceil$, and $\underline{k}_{h3ij} = \lfloor k_{h3ij}\rfloor$; and

$\bar{a}_{h3ij} = \bar{k}_{h3ij}\{1 - \underline{\lambda}_{hij}/l_{hij}\bar{k}_{h3ij})\}/\{\underline{\lambda}_{hij}(l_{hij}\bar{k}_{h3ij} - 1)\}$, and

$$\underline{a}_{h3ij} = \underline{k}_{h3ij}\{1 - \bar{\lambda}_{hij}/l_{hij}\underline{k}_{h3ij})\} / \left\{\bar{\lambda}_{hij}\left(l_{hij}\underline{k}_{h3ij} - 1\right)\right\}$$

Define $\left(\tilde{l}_{hij}^*, \tilde{\kappa}_{h3ij}\right)$ as :

$\Pr\left\{\left(\tilde{l}_{hij}^*, \bar{k}_{h3ij}\right) = \left(\underline{\lambda}_{hij}, \bar{k}_{h3ij}\right)\right\} = p_{h3ij}$ and $\Pr\left\{\left(\tilde{l}_{hij}^*, \tilde{k}_{h3ij}\right) = \left(\bar{\lambda}_{hij}, \underline{\kappa}_{h3ij}\right)\right\}$

$= 1 - p_{h3ij}$, where

$$p_{h3ij} = \left[(1 - f_{h3ij})/\left\{N_h E_{*\left(\tilde{n}_h^*,\tilde{k}_{h1}\right)}\left(n_h^{*-1}\right) M_{hi} E_{*\left(\tilde{m}_{hi}^*,\tilde{k}_{h2i}\right)}\left(\tilde{m}_{hi}^{*-1}\right) l_{hij}(l_{hij} - 1)\right\}\right.$$
$$\left. - \underline{a}_{h3ij}\right]/(\bar{a}_{h3ij} - \underline{a}_{h3ij}).$$

Step 1. Generate $\left(\tilde{n}_h^*, \tilde{k}_{h1}\right)$. Copy $S_{h1}\tilde{\kappa}_{h1}$ times to create $\mathcal{P}_{h1}^*$ of size $n_h\tilde{\kappa}_{h1}$.
Step 2. For each $i \in \mathcal{P}_{h1}^*$, generate $(\tilde{m}_{hi}^*, \tilde{\kappa}_{h2i})$. Copy $\mathcal{S}_{h2i}(i \in \mathcal{P}_{h1}^*)\tilde{\kappa}_{h2i}$ times to create $\mathcal{P}_{h2i}^*$ of size $m_{hi}\tilde{k}_{h2i}$.
Step 3. For each $j \in \mathcal{P}_{h2i}^*(i \in \mathcal{P}_{h1}^*)$, generate $\left(\tilde{l}_{hij}^*, \tilde{k}_{h3ij}\right)$. Copy $\mathcal{S}_{h3ij}\left(j \in \mathcal{P}_{h2i}^*, i \in \mathcal{P}_{h1}^*\right)\bar{k}_{h3ij}$ times to create $\mathcal{P}_{h3ij}^*$ of size $l_{hij}\bar{k}_{h3ij}$.
Step 4. Conduct ST–SI^3 from the pseudo-population to obtain a bootstrap sample: first, take an SI of size $\tilde{n}_h^*$ from $\mathcal{P}_{h1}^*$ to get $\mathcal{S}_{h1}^*$; then for $i \in \mathcal{S}_{h1}^*$, take an SI of size $\tilde{m}_{hi}^*$ from $\mathcal{P}_{h2i}^*$ to get $\mathcal{S}_{h2i}^*$ and finally for $j \in \mathcal{S}_{h2i}^*\left(i \in \mathcal{S}_{h1}^*\right)$, take an SI of size $\tilde{l}_{hij}^*$ from $\mathcal{P}_{h2ij}^*$ to get $\mathcal{S}_{h3ij}^*$.

The bootstrap sampling weights are given by $w_{hijk}^* = \left(N_h/\tilde{n}_h^*\right)\left(M_{hi}/\tilde{m}_{hi}^*\right)\left(L_{hij}/\tilde{l}_{hij}^*\right)$. It is shown in the separate appendix that $p_{h1}, p_{h2i}, p_{h3ij} \in [0, 1]$ and that v_h, k_{h1}, μ_{hi}, k_{h2i}, λ_{hij}, and k_{h3ij} are all positive integers for n_h, m_{hi}, $l_{hij} \geq 2$. For variance estimation, repeat Steps 1–4 for a large number of times B.

For efficient computations, we may avoid unnecessary random number generation by replacing Steps 1–4 with the following steps:

Step 1′. Generate $\left(\tilde{n}_h^*, \tilde{k}_{h1}\right)$. Copy $\mathcal{S}_{h1}, \tilde{k}_{h1}^*$ times to create $\mathcal{P}_{h1}^*$ of size $n_h\tilde{k}_{h1}$. Sample $\mathcal{S}_{h1}^*$ of size $\tilde{n}_h^*$ from $\mathcal{P}_{h1}^*$ without replacement.
Step 2′. For each $i \in \mathcal{S}_{h1}^*$, generate $\left(\tilde{m}_{hi}^*, \bar{k}_{h2i}\right)$. Copy $\mathcal{S}_{h2i}^*\left(i \in \mathcal{P}_{h1}^*\right)\tilde{k}_{h2i}$ times to create $\mathcal{P}_{h2i}^*$ of $m_{hi}\tilde{k}_{h2i}$. Sample $\mathcal{S}_{h2i}^*$ of size $\tilde{m}_{hi}^*$ from $\mathcal{P}_{h2i}^*$ without replacement.
Step 3′. For each $j \in \mathcal{S}_{h2i}^*\left(i \in \mathcal{S}_{h1}^*\right)$, generate $\left(\tilde{l}_{hij}^*, \bar{k}_{h3ij}\right)$. Copy $\mathcal{S}_{h3ij}^*\left(j \in \mathcal{S}_{h2i}^*, i \in \mathcal{S}_{h1}^*\right)\tilde{k}_{h3ij}$ times to create $\mathcal{P}_{h3ij}^*$ of size $l_{hij}\bar{k}_{h3ij}$. Sample $\mathcal{S}_{h3ij}^*$ of size $\tilde{l}_{hij}^*$ from $\mathcal{P}_{h3ij}^*$.

This procedure is similar to the BMM.

3.5. Numerical Illustration

To illustrate, let us consider a single-stratum ($H = 1$) population which has $N = 10$ PSUs each composed of $M_i = 20$ SSUs each with $L_{ij} = 30$ USUs in it. The sample sizes used are $n = 4, m_i = 8$, and $l_{ij} = 3$. Since $H = 1$, we suppress subscript h in what follows. Creation of a bootstrap sample in the four methods is carried out as follows.

In the BBE:

Step 1. Choose 3 PSUs as a candidate set through simple random sampling with replacement from the 4 PSUs in the sample. For each of the 4 PSUs, keep it with probability 0.60 or replace it with one randomly selected from the 3 PSUs in the candidate set. For the PSUs kept in this step, go to the next step.
Step 2. Choose 7 SSUs as a candidate set through simple random sampling with replacement from the 8 SSUs in the sample. For each of the 8 SSUs, keep it with probability 0.771 or replace it with one randomly selected from the 7 PSUs in the candidate set. For the SSUs kept in this step, go to the next step.
Step 3. Choose 2 USUs as a candidate set through simple random sampling with replacement from the 3 USUs in the sample. For each of the 3 USUs, keep it with probability 0.767 or replace it with one randomly selected from the 2 USUs in the candidate set.

In the BMM:

Step 0. Let $n' = 2, m'_i = 3$, and $l'_{ij} = 1$, say.
Step 1. Let $\tilde{k}_1 = 1$ with probability 0.20 or $= 2$ otherwise. Generate $\tilde{k}_1$ and repeat $\tilde{k}_1$ times simple random sampling without replacement of size 2 from the 4 PSUs.
Step 2. Let $\tilde{k}_{2i} = 8$ with probability 0.64 or $= 9$ otherwise. For each of the PSUs taken in Step 1, generate $\tilde{k}_{2i}$ and repeat $\tilde{k}_{2i}$ times simple random sampling without replacement of size 3 from the 8 SSUs.
Step 3. Let $\tilde{k}_{3ij} = 5$ with probability 0.63 or $= 6$ otherwise. For each of the SSUs taken in Step 2, generate $\tilde{k}_{3ij}$ and repeat $\tilde{k}_{3ij}$ times simple random sampling without replacement of size 1 from the 3 USUs (This is equivalent to simple random sampling with replacement of size $\tilde{k}_{3ij}$).

In the BRS:

Step 0. Let $n^* = 3$, $m^*_i = 7$, and $l^*_{ij} = 2$, say.
Step 1. Choose 3 PSUs from 4 PSUs via simple random sampling with replacement. Let $\tilde{n}^*_i$ be the number of times that PSU i is selected.
Step 2. For i such that $\tilde{n}^*_i \geq 1$, choose $7\tilde{n}^*_i$ SSUs from 8 SSUs in PSU i via simple random sampling with replacement. Let $\tilde{m}^*_{ij}$ be the number of times that SSU j in PSU i is selected. For i such that $\tilde{n}^*_i = 0$, let $\tilde{m}^*_{ij} = 0$;
Step 3. For ij such that $\tilde{m}^*_{ij} \geq 1$, choose $2\tilde{m}^*_{ij}$ USUs from 2 USUs in SSU j in PSU i via simple random sampling with replacement. Let $\tilde{l}^*_{ijk}$ be the number of times that USU k in SSU j in PSU i is selected. For ij such that $\tilde{m}^*_{ij} = 0$, let $\tilde{l}^*_{ijk} = 0$. Use $\alpha_i = 1.03$, $\beta_{ij} = 0.65$, and $\gamma_{ijk} = 0.70$ for weight rescaling.

In the BWO:

Step 0. Define the following integer random variables: $(\tilde{n}^*, \tilde{k}_1) = (3, 3)$ with probability 0.44 or $= (4{,}2)$ otherwise; $(\tilde{m}_i^*, \tilde{k}_{2i}) = (20, 7)$ with probability 0.47 or $= (21,\ 6)$ otherwise; and $(\tilde{l}_{ij}^*, \tilde{k}_{3ij}) = (5, 19)$ with probability 0.29 or $= (6,\ 18)$ otherwise.
Step 1. Generate $(\tilde{n}^*, \tilde{k}_1)$ and copy 4 PSUs $\tilde{k}_1$ times.
Step 2. For each i in the copied PSUs in Step 1, generate $\left(\tilde{m}_i^*, \tilde{k}_{2i}\right)$ and copy 6 SUSs $\tilde{k}_{2i}$ times.
Step 3. For each j in the copied SSUs in PSU i in Step 2, generate $\left(\tilde{l}_{ij}^*, \tilde{k}_{3ij}\right)$ and copy 2 USUs $\tilde{k}_{3ij}$ times.
Step 4. Mimic the original sampling under the created pseudo-population with realized sample sizes $\tilde{n}^*, \tilde{m}_i^*$, and $\tilde{l}_{ij}^*$.

4. Simulation Study

We compared the four bootstrap methods through a limited simulation study. To focus on design-based properties, we generated a population and fixed it under repeated sampling. We created two populations. In Population I, stratification was less effective and the intra-cluster correlation was low. On the other hand, in Population II, stratification was more effective and the intra-cluster correlation was high. Both had $H = 4$ strata. Each stratum had $N_h = 10$ primary sampling units, each of which had $M_{hi} = 20$ secondary sampling units, each having $L_{hij} = 30$ ultimate sampling units. We chose the small number of strata to obtain a clear picture of the pseudo-populations under study. The study variables y_{hijk} were generated as follows. First, we generated $\mu_{hi} = \mu_h + \sigma_h u_{hi}$, where μ_h and σ_h are listed in Table 1 and $u_{hi} \sim \text{iidN}(0, 1)$. Second, we generated $\mu_{hij} = \mu_{hi} + \sqrt{(1 - \rho_2)/\rho_2}\sigma_h u_{hij}$, where $u_{hij} \sim \text{iidN}(0, 1)$. Finally, we generated $y_{hijk} = \mu_{hij} + \sqrt{(1 - \rho_3)/\rho_3}\sigma_h u_{hijk}$, where $u_{hijk} \sim \text{iidN}(0, 1)$. We set $(\rho_2, \rho_3) = (0.2, 0.3)$ for Population I (low intra-cluster correlation) and $(\rho_2, \rho_3) = (0.5, 0.5)$ for Population II (high intra-cluster correlation).

Figure 1 shows the histograms of y_{hijk} in Stratum $h\,(h = 1, 2, 3, 4)$ in Population I, with five vertical lines showing the population quantiles for $p = 0.10, 0.25, 0.50, 0.75, 0.90$. Since stratification is weak in Population I, the characteristic y_{hijk} in a stratum overlaps those in the other strata.

On the other hand, stratification in Population II is effective. In particular, the first quartile $F^{-1}(0.25)$, the median $F^{-1}(0.5)$, and the third quartile $F^{-1}(0.75)$ are located at the stratum boundary between $h = 1$ and $h = 2$, $h = 2$ and $h = 3$, and $h = 3$ and $h = 4$, respectively (see Figure 2).

Table 1. Parameter Values to Create a Population

	Population I		Population II	
h	μ_h	σ_h	μ_h	σ_h
1	200	20.0	200	10.0
2	150	15.0	150	7.5
3	120	12.0	120	6.0
4	100	10.0	100	5.0

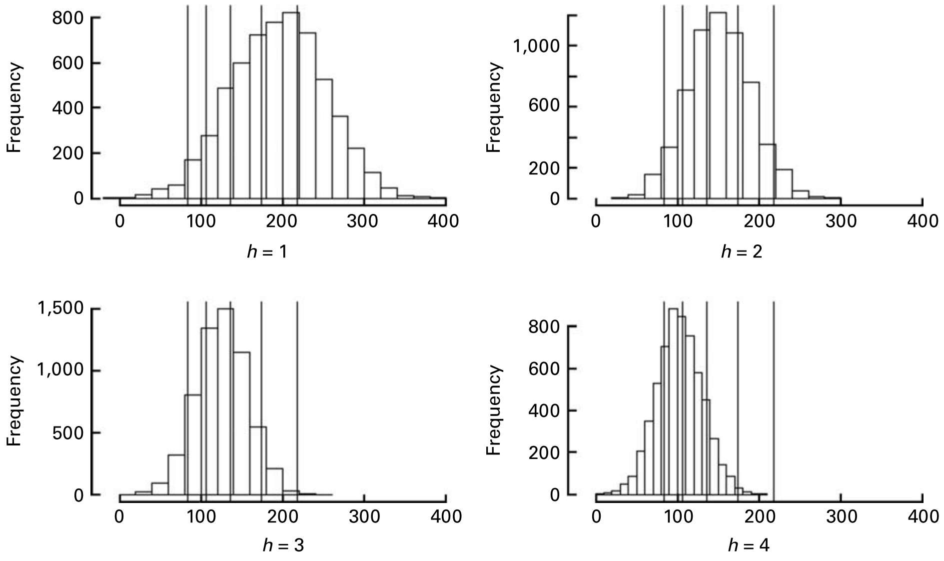

Fig. 1. Histograms of y_{hijk} in Strata h ($h = 1, 2, 3, 4$) in Population I. Note: The five vertical lines show $F^{-1}(p)$ with $p = 0.1, 0.25, 0.50, 0.75, 0.9$

Since high sampling fractions were the concern, all the first stage sampling fractions were nonnegligible. Table 2 shows the first- and second-stage sampling fractions in the simulation. The third-stage sampling fractions were set to be $f_{h3ij} = 0.1$.

In the BMM, n'_h, m'_h, and l'_{hij} were randomized to get $E_*\left(f^*_{h1}\right) = f_{h1}$, $E_*\left(f^*_{h2i}\right) = f_{h2i}$, and $E_*\left(f^*_{h3ij}\right) = f_{h3ij}$ when possible (see Section 3.2) or $n'_h = 1, m'_{hi} = 1$ and $l'_{hij} = 1$

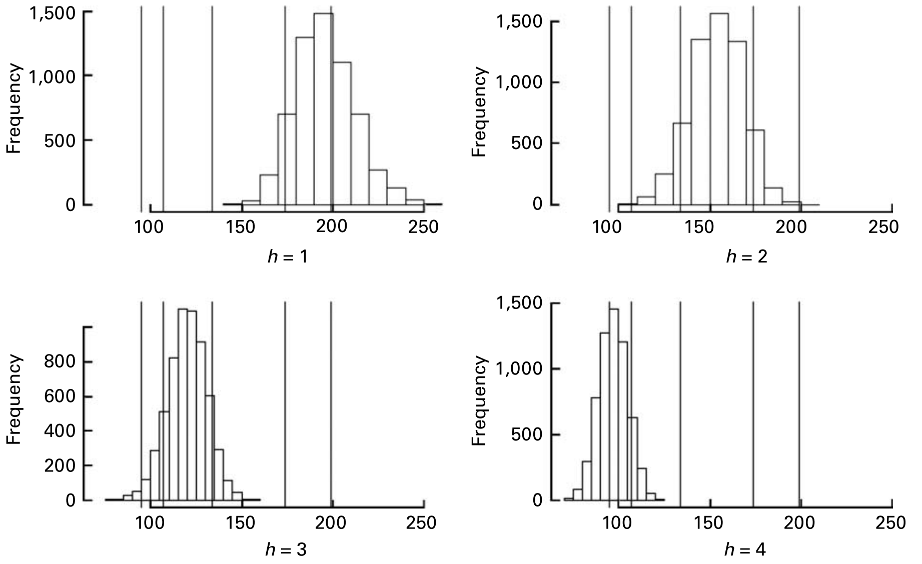

Fig. 2. Histograms of y_{hijk} in Strata h ($h = 1, 2, 3, 4$) in Population II. Note: The five vertical lines show $F^{-1}(p)$ with $p = 0.1, 0.25, 0.50, 0.75, 0.9$

Table 2. The sampling Fractions in the Simulation Study

h	f_{h1}	f_{h2i}	f_{h3ij}
1	0.5	0.4	0.1
2	0.4	0.4	0.1
3	0.2	0.2	0.1
4	0.2	0.2	0.1

otherwise. In the BRS, $n_h^* = n_h - 1, m_{hi}^* = m_{hi} - 1$, and $l_{hij}^* = l_{hij} - 1$, which assures that $w_{hijk}^* > 0$.

The parameters of interest θ were the population total $Y....$ and quantiles $F^{-1}(p)\,(p = 0.1, 0.25, 0.5, 0.75, 0.9)$, and their point estimators were given by $\hat{Y}....$ and $\hat{F}^{-1}(p)$ in Section 2. In each simulation run, we generated $B = 1,000$ bootstrap samples to estimate variance and made a bootstrap histogram. Variance was estimated by $v_{boot}(\hat{\theta})$ in Section 3. To evaluate the empirical coverage, the 0.1 and 0.9 points of the bootstrap histogram were employed to calculate the lower- and upper-tail errors defined below (see Shao and Tu 1995, pp. 132–133, for using the bootstrap histogram to construct a confidence interval). The relative bias (%Bias) and the instability (the coefficient of variation, %Instb) in variance estimation, and the lower- and upper-tail errors (%L and %U) were evaluated through $S = 1,000$ simulation runs while the variance $V(\hat{\theta})$ was estimated by 10,000 iterations:

$$\%\text{Bias} = \left\{ S^{-1} \sum_{s=1}^{S} v_{boot}^{(s)}(\hat{\theta}) - V(\hat{\theta}) \right\} \Big/ V(\hat{\theta}) \times 100,$$

$$\%\text{Instb} = \left\{ S^{-1} \sum_{s=1}^{S} \left(v_{boot}^{(s)}(\hat{\theta}) - V(\hat{\theta}) \right)^2 \right\}^{1/2} \Big/ V(\hat{\theta}) \times 100,$$

$$\%\text{L} = S^{-1} \# \{ 0 < \text{the 0.1 point of bootstrap histgram of } \hat{\theta}_b^* \} \times 100,$$

$$\%\text{U} = S^{-1} \# \{ 0 > \text{the 0.9 point of bootstrap histgram of } \hat{\theta}_b^* \} \times 100,$$

The empirical coverage rate was computed as (100 − %L − %U) percent.

Tables 3 and 4 show the results for the four bootstrap methods for Populations I and II, respectively. We observe from Tables 3 and 4 the following points.

1. In estimating variances for the estimated population total in both Populations I and II, the performance measures for the four methods are almost identical. This was expected because the four methods satisfy $E_*\left(\hat{Y}_{....}^*\right) = \hat{Y}...$ and $V_*\left(\hat{Y}_{....}^*\right) = v\left(\hat{Y}...\right)$, so the differences in %Bias reflect only Monte Carlo errors.
2. In estimating variances for the five estimated quantiles in Population I, the four methods perform similarly although the BWO is the least biased and the most stable.
3. In estimation for Population II, the four methods show remarkable differences: the bias in variance estimation by the BRS can be serious; the instability of the BBE tends to be greater than that of the BMM and the BWO; and the BMM and the BWO perform similarly, although the latter is marginally less biased and more stable.

Table 3. Variance Estimation and Tail Errors for Population I

	%Bias	%Instb	%L	%U	%Bias	%Instb	%L	%U
	$\hat{Y}\ldots$				$\hat{F}^{-1}(0.10)$			
BBE	−1.7	60.4	12.2	13.0	9.4	101.6	21.4	9.4
BMM	−1.7	60.4	11.9	13.1	9.3	98.2	19.6	8.7
BRS	−1.7	60.5	12.2	13.2	10.0	101.9	20.1	8.6
BWO	−3.1	59.1	12.1	13.4	6.7	95.5	20.1	8.8
	$\hat{F}^{-1}(0.25)$				$\hat{F}^{-1}(0.50)$			
BBE	4.9	91.4	15.4	17.2	9.1	96.6	8.0	19.0
BMM	4.2	89.6	15.2	17.2	9.1	95.7	8.5	18.3
BRS	4.3	91.1	15.4	16.8	9.0	97.7	9.1	18.3
BWO	2.1	84.4	16.0	17.0	6.8	93.4	9.2	18.4
	$\hat{F}^{-1}(0.75)$				$\hat{F}^{-1}(0.90)$			
BBE	4.6	47.6	12.2	10.1	6.4	54.7	10.7	13.2
BMM	3.9	47.8	12.4	9.8	7.1	56.3	10.7	12.5
BRS	3.6	47.8	12.1	9.6	9.4	58.2	10.1	12.1
BWO	4.2	47.5	13.0	9.7	6.7	54.8	10.8	12.6

%Bias $= \left\{S^{-1}\sum_{s=1}^{S} v_{boot}^{(s)}(\hat{\theta}) - V(\hat{\theta})\right\} \Big/ V(\hat{\theta}) \times 100;$

%Instb $= \left\{S^{-1}\sum_{s=1}^{S} \left(v_{boot}^{(s)}(\hat{\theta}) - V(\hat{\theta})\right)^2\right\}^{1/2} / V(\hat{\theta}) \times 100;$

%L $= S^{-1}\#\{\theta <$ the 0.10 lower trail of $\hat{\theta}_b^*\} \times 100;$

%U $= S^{-1}\#\{\theta <$ the 0.10 upper trail of $\hat{\theta}_b^*\} \times 100.$

Table 4. Variance Estimation and Tail Errors for Population II

	%Bias	%Instb	%L	%U	%Bias	%Instb	%L	%U
	$\hat{Y}\ldots$				$\hat{F}^{-1}(0.10)$			
BBE	−2.0	58.6	12.0	13.5	5.1	109.5	20.7	18.1
BMM	−2.1	58.7	11.5	13.4	5.2	106.6	20.8	15.5
BRS	−2.0	58.8	11.8	13.2	11.0	118.0	20.7	16.1
BWO	−3.6	57.4	12.2	13.5	2.8	104.3	20.9	15.5
	$\hat{F}^{-1}(0.25)$				$\hat{F}^{-1}(0.50)$			
BBE	−5.8	88.7	10.6	21.2	1.4	119.4	3.0	52.9
BMM	9.7	99.5	12.5	14.6	−2.4	83.1	3.5	31.3
BRS	−20.3	77.6	16.3	18.9	−62.4	74.7	7.3	50.6
BWO	9.5	105.6	13.4	15.3	4.6	100.9	3.1	48.2
	$\hat{F}^{-1}(0.75)$				$\hat{F}^{-1}(0.90)$			
BBE	22.6	85.2	10.0	17.9	3.6	49.5	10.8	12.6
BMM	21.5	83.4	11.3	13.4	2.1	51.0	11.3	10.8
BRS	−22.2	58.4	14.7	18.4	6.8	52.9	10.6	10.3
BWO	25.3	85.4	11.2	13.6	1.7	48.6	11.5	12.1

%Bias $= \left\{S^{-1}\sum_{s=1}^{S} v_{boot}^{(s)}(\hat{\theta}) - V(\hat{\theta})\right\} \Big/ V(\hat{\theta}) \times 100;$

%Instb $= \left\{S^{-1}\sum_{s=1}^{S} \left(v_{boot}^{(s)}(\hat{\theta}) - V(\hat{\theta})\right)^2\right\}^{1/2} \Big/ V(\hat{\theta}) \times 100;$

%L $= S^{-1}\#\{\theta <$ the 0.10 lower trail of $\hat{\theta}_b^*\} \times 100;$

%U $= S^{-1}\#\{\theta >$ the 0.10 upper trail of $\hat{\theta}_b^*\} \times 100.$

4. Variance of sample quantiles is usually overestimated by the bootstrap methods. However, the BRS seriously underestimates variance of $\hat{F}^{-1}(0.25)$, $\hat{F}^{-1}(0.50)$, and $\hat{F}^{-1}(0.75)$ in Population II. A possible explanation for this is as follows. To fix the idea, define $\tilde{W}^*_{hijk}$ such that $\hat{Y}^*_{....} = \sum_{h=1}^{H} \sum_{i\in\mathcal{S}_{h1}} \sum_{j\in\mathcal{S}_{h2i}} \sum_{k\in\mathcal{S}_{h3ij}} \tilde{W}^*_{hijk} y_{hijk}$. That is, $\tilde{W}^*_{hijk}$ are the sum of bootstrap weights associated with USU k in SSU j in PSU i in Stratum h in the original sample. In the BRS, $\hat{W}^*_{hijk} = w^*_{hijk} > 0$ for all the units in the original sample. Namely, the order statistics in bootstrap samples do not change at all. By contrast, in the BBE, the BMM, and the BWO, $\tilde{W}_{hijk}$ can be zero. In Population II, y_{hijk} around $F^{-1}(p)$, where $p = 0.25, 0.5,$ and 0.75, are scarce, and so are the sampled y_{hijk} around the points. Possibly, the bootstrap pseudo-estimates, $\hat{F}^{-1*}(p)$, where $p = 0.25, 0.5,$ and 0.75, in the BRS vary too smoothly because of the fixed order statistics while those in the BBE, the BMM, and the BWO fluctuate largely since some $\tilde{W}_{hijk}$ are zero.
5. Overall poor performances of interval estimation are probably due to a small population size with a small sample size. This is particularly true for the median in Population II, where y values around the quantile are scarce. To test this point, we conducted a simulation using the parameter values for Population II with doubled $N_h = 20$ and $M_{hi} = 40$ to find the actual tail error rates closer to the nominal rates. The result showed that the lower and the upper-tail error rates were, respectively, about 8% and 24%. We can, perhaps, improve the coverage rates through a more sophisticated bootstrap confidence interval. But that is beyond the scope of this article and we do not intend to pursue it here.

5. Conclusion

In this article, we first extended the three bootstrap methods to a stratified three-stage design with simple random sampling without replacement at each stage: the mirror-match bootstrap (BMM), the rescaling bootstrap (BRS), and the without-replacement bootstrap (BWO). Then, we conducted a simulation study to examine the three methods as well as the Bernoulli bootstrap (BBE). The simulation showed that (1) the four methods perform similarly for estimating the variance of the estimated population total; (2) the four methods perform differently for quantile estimation when stratification is effective; and (3) overall, the BWO was the least biased while bias in variance estimation by the BRS was sometimes remarkably large. The last observation supports the intuition that methods which better mimic the original sampling will perform better as the estimators and the sampling design become more complex (Sitter 1992b, p. 153). Theoretical research is beyond the scope of this article and will be a future topic.

6. References

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 American Statistical Association, 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. The Canadian Journal of Statistics, 16, 25–45.

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.

Saigo, H., Shao, J., and Sitter, R.R. (2001). A Repeated Half-Sample Bootstrap and Balanced Repeated Replications for Randomly Imputed Data. Survey Methodology, 27, 189–196.

Särndal, C.-E., Swensson, 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. The Canadian Journal of Statistics, 20, 135–154.

Woodruff, R.S. (1952). Confidence Intervals for Median and Other Position Measures. Journal of the American Statistical Association, 47, 635–646.

Received April 2008
Revised September 2009