

Article

A Comparison of Existing Bootstrap Algorithms for Multi-Stage Sampling Designs

Sixia Chen ¹, David Haziza ^{2,*} and Zeinab Mashreghi ³

¹ Department of Biostatistics and Epidemiology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA; sixia-chen@ouhsc.edu

² Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON K1N 6N5, Canada

³ Department of Mathematics and Statistics, University of Winnipeg, Winnipeg, MB R3B 2E9, Canada; z.mashreghi@uwinnipeg.ca

* Correspondence: dhaziza@uottawa.ca

Abstract: Multi-stage sampling designs are often used in household surveys because a sampling frame of elements may not be available or for cost considerations when data collection involves face-to-face interviews. In this context, variance estimation is a complex task as it relies on the availability of second-order inclusion probabilities at each stage. To cope with this issue, several bootstrap algorithms have been proposed in the literature in the context of a two-stage sampling design. In this paper, we describe some of these algorithms and compare them empirically in terms of bias, stability, and coverage probability.

Keywords: bootstrap algorithms; multi-stage sampling; Taylor linearization; variance estimation



Citation: Chen, S.; Haziza, D.; Mashreghi, Z. A Comparison of Existing Bootstrap Algorithms for Multi-Stage Sampling Designs. *Stats* **2022**, *5*, 521–537. <https://doi.org/10.3390/stats5020031>

Academic Editor: Fulvia Mecatti

Received: 20 April 2022

Accepted: 29 May 2022

Published: 6 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Many surveys conducted by national statistical offices use stratified multi-stage sampling designs for selecting a sample. Reasons for using multi-stage sampling rather than direct element sampling include the lack of element-level sampling frames and cost considerations when data collection involves face-to-face interviews. Stratified multi-stage sampling designs include some form of stratification, selection of primary sampling units (psu), and subsampling within selected psus. This is especially common in social and health surveys. For general multi-stage sampling designs, unbiased variance estimation is a complex task as it relies on the availability of the second-order inclusion probabilities at each stage. If the first-stage sampling fraction is small, a common variance estimation strategy is to pretend that the psus were selected with replacement and use the customary with replacement variance estimator. The resulting estimator is generally conservative in the sense that it may suffer from a small positive bias.

Another approach to variance estimation for survey data is bootstrap variance estimation originally proposed by Efron [1] in the context of independent and identically distributed observations. In a finite population sampling, bootstrap procedures can be classified into two broad groups. In the first, bootstrap samples are selected from the original sample; e.g., [2,3] among others. Rao and Wu [2] applied a scale adjustment directly to the survey data values so as to recover the usual variance formulae. Rao et al. [4] presented a modification of the method of Rao and Wu [2], where the scale adjustment is applied to the survey weights rather than to the data values. The second group of procedures consists of first creating a pseudo-population from the original sample. Bootstrap samples are then selected from the pseudo-population using the same sampling design utilized to select the original samples; see [5–8], among others. Many of the aforementioned bootstrap procedures may be implemented by randomly generating bootstrap weights so that the first two (or more) design moments of the sampling error are tracked by the corresponding bootstrap moments; see [9,10]. These procedures are often referred to as bootstrap weight procedures.

For a comprehensive review of bootstrap procedures for survey data, the reader is referred to Mashreghi et al. [11].

The goal of this paper is to empirically compare several existing bootstrap algorithms that have been proposed in the literature for two-stage sampling designs. The bootstrap procedures are compared with respect to bias, stability, and coverage probability of confidence intervals. In Section 2 we present the basic setup and discuss some classical variance estimation procedures for two-stage sampling designs. In Section 3, we present some bootstrap algorithms proposed in the case of simple random sampling without replacement in both stages. Bootstrap algorithms for unequal probability sampling designs are described in Section 4. In Section 5, we present the results from a simulation study. We make some final remarks in Section 6.

2. The Setup

Consider a finite population U consisting of N primary sampling units (psu), U_1, \dots, U_N , of size M_1, \dots, M_N , such that $U = \cup_{i \in U} U_i$ and $U_i \cap U_j = \emptyset$ if $i \neq j$. Let $M_0 = \sum_{i \in U} M_i$ be the total number of elements in the population. We are interested in estimating a population total t_y of a survey variable y :

$$t_y = \sum_{i \in U} t_i,$$

where $t_i = \sum_{k \in U_i} y_{ik}$ denotes the i th psu total, $i = 1, \dots, N$, and y_{ik} denotes the y -value for the k th element in the i th psu. To that end, we select a sample according to a two-stage sampling design:

- (i) A sample S_{1st} of psus, of size n , is selected according to a given sampling design $P(S_{1st})$ with first-order inclusion probabilities, $\pi_i = P(i \in S_{1st})$, and with second-order inclusion probabilities, $\pi_{ij} = P(i \in S_{1st} \ \& \ j \in S_{1st})$. Finally, let $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$.
- (ii) In the i th psu sampled at the first stage, $i \in S_{1st}$, a subsample of the elements, S_i , of size m_i is selected according to a given sampling design $P(S_i | S_{1st})$ with first-order inclusion probabilities $\pi_{k|i} = P(k \in S_i | i \in S_{1st})$ and second-order inclusion probabilities $\pi_{k\ell|i} = P(k \in S_i \ \& \ \ell \in S_i | i \in S_{1st})$. Subsampling in a given psu is carried out independently of subsampling in any other psu.

A design-unbiased estimator of t_y is the Horvitz–Thompson estimator given by

$$\hat{t}_y = \sum_{i \in S_{1st}} \pi_i^{-1} \hat{t}_i, \tag{1}$$

where $\hat{t}_i = \sum_{k \in S_i} y_{ik} / \pi_{k|i}$. The estimator (1) can be written as $\hat{t}_y = \sum_{k \in \tilde{S}} w_k y_k$, where $w_k = \pi_k^{-1}$ with $\pi_k = \pi_i \times \pi_{k|i}$, and \tilde{S} denotes the sample of elements of size $\sum_{i \in S_{1st}} m_i$.

The design variance of \hat{t}_y , denoted by $\mathbb{V}_p(\hat{t}_y)$, can be unbiasedly estimated by

$$\hat{V}(\hat{t}_y) = \sum_{i \in S_{1st}} \sum_{j \in S_{1st}} \frac{\Delta_{ij}}{\pi_{ij}} \frac{\hat{t}_i}{\pi_i} \frac{\hat{t}_j}{\pi_j} + \sum_{i \in S_{1st}} \pi_i^{-1} \hat{V}_i, \tag{2}$$

where

$$\hat{V}_i = \sum_{k \in S_i} \sum_{\ell \in S_i} \frac{\Delta_{k\ell|i}}{\pi_{k\ell|i}} \frac{y_{ik}}{\pi_{k|i}} \frac{y_{i\ell}}{\pi_{\ell|i}}$$

and $\Delta_{k\ell|i} = \pi_{k\ell|i} - \pi_{k|i} \pi_{\ell|i}$. That is, $\mathbb{E}_1 \mathbb{E}_2 \left\{ \hat{V}(\hat{t}_y) \mid S_{1st} \right\} = \mathbb{V}_p(\hat{t}_y)$, where $\mathbb{E}_1(\cdot)$ denotes the expectation with respect to the first-stage sampling design, and $\mathbb{E}_2(\cdot \mid S_{1st})$ denotes the expectation with respect to the second-stage sampling design conditionally on S_{1st} . In the case of simple random sampling without replacement at both stages, the estimator (2) reduces to

$$\hat{V}(\hat{t}_y, \pi) = N^2 \left(1 - \frac{n}{N} \right) \frac{s_f^2}{n} + \frac{N}{n} \sum_{i \in S} M_i^2 \left(1 - \frac{m_i}{M_i} \right) \frac{s_{yi}^2}{m_i}, \tag{3}$$

where

$$s_t^2 = \frac{1}{n-1} \sum_{i \in S_{1st}} \left(\hat{t}_i - \frac{\sum_{i \in S} \hat{t}_i}{n} \right)^2$$

and

$$s_{y_i}^2 = \frac{1}{m_i-1} \sum_{k \in S_i} (y_{ik} - \bar{y}_i)^2$$

with $\bar{y}_i = m_i^{-1} \sum_{k \in S_i} y_{ik}$.

For general two-stage sampling designs, the computation of (2) is cumbersome as it requires the availability of the second-order inclusion probabilities at each stage. A simplified variance estimator is given by

$$\hat{V}_{sim}(\hat{t}_y) = \sum_{i \in S_{1st}} \sum_{j \in S_{1st}} \frac{\Delta_{ij}}{\pi_{ij}} \frac{\hat{t}_i}{\pi_i} \frac{\hat{t}_j}{\pi_j}. \tag{4}$$

That is, only the first term of (2) is kept. The bias of $\hat{V}_{sim}(\hat{t}_y)$, which is always negative, is expected to be small provided that the first-stage sampling fraction, $f_1 = n/N$, is small; see [12,13].

An alternative simplified variance estimator can be obtained by pretending that the psus are selected with replacement. It is given by

$$\hat{V}(\hat{t}_{y,wr}) = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{\hat{t}_i}{p_i} - \hat{t}_{y,wr} \right)^2, \tag{5}$$

where $\hat{t}_{y,wr} = \sum_{i \in S} \frac{\hat{t}_i}{np_i}$, with p_i denoting the probability of selection of the i th psu at any given draw. If the first-stage sampling fraction, f_1 , is small, we expect (5) to suffer from a small positive bias. Unlike (4), the estimator does not require the availability of the second-order inclusion probabilities π_{ij} .

So far, we have considered the case of a population total t_y . In practice, it may be of interest to estimate more complex parameters such as distribution functions and quantiles. Let θ_N be defined as the solution of the following census estimating equation:

$$U_N(\theta) = \frac{1}{M_0} \sum_{i \in U} \sum_{k \in U_i} u(y_{ik}; \theta) = 0, \tag{6}$$

where $u(y_{ik}; \theta)$ can be either a smooth (i.e., a function that is differentiable and whose derivatives are continuous) or a non-smooth function of θ . When $u(\cdot)$ is smooth, the solution of (6) is called a smooth parameter; otherwise, it is called a non-smooth parameter. Common parameters include: (i) the population mean obtained with $u(y_{ik}; \theta) = y_{ik} - \theta$; (ii) the finite population distribution function obtained with $u(y_{ik}; \theta) = I(y_{ik} \leq t) - \theta, t \in \mathbb{R}$; (iii) the τ -th population percentile obtained with $u(y_{ik}; \theta) = I(y_{ik} \leq \theta) - \tau$. The population mean is an example of a smooth parameter, whereas distribution functions and quantiles are examples of non-smooth parameters.

An estimator $\hat{\theta}$ of θ_N can be obtained by solving the following sample estimating equation:

$$\hat{U}_S(\theta) = \frac{1}{M_0} \sum_{k \in \hat{S}} w_k u(y_k; \theta) = 0. \tag{7}$$

The variance of $\hat{\theta}$ may be obtained using a first-order Taylor expansion or by using a resampling method such as balanced repeated replication, jackknife, and bootstrap; see [14] for a discussion of resampling methods. In the remainder of this paper, we confine to bootstrap.

3. Bootstrap Procedures for Simple Random Sampling without Replacement at Both Stages

In this section, we describe several bootstrap algorithms for a two-stage sampling design with simple random sampling without replacement at both stages.

3.1. The Rescaling Bootstrap Algorithm

Rao and Wu [2] proposed a rescaled bootstrap algorithm for both uni-stage and two-stage sampling designs. Because the rescaling factor is applied to the y -values, this method is applicable to smooth statistics but not to the case of non-smooth statistics such as quantiles. The algorithm can be described as follows:

- Step 1. Draw a sample of size n psus from S_{1st} , according to simple random sampling with replacement.
- Step 2. From each psu selected in Step 1, select a sample of elements, of size m_i according to simple random sampling with replacement. For a psu selected more than once in Step 1, perform independent subsampling.
- Step 3. Let y_{ik}^* be the y -value of the k th bootstrap element in the i th bootstrap psu and m_i^* be the m_i -value of the i th bootstrap psu and M_i^* is defined similarly. Let

$$\tilde{y}_{ik} = \hat{Y} + \left\{ \frac{n(1-f_1)}{n-1} \right\}^{1/2} \left(\frac{\hat{t}_i^*}{\bar{M}_0} - \hat{Y} \right) + \left\{ \frac{m_i^* f_1 (1-f_{2i}^*)}{m_i^* - 1} \right\}^{1/2} \left(\frac{M_i^* y_{ik}^*}{\bar{M}_0} - \frac{\hat{t}_i^*}{\bar{M}_0} \right),$$

where $f_{2i}^* = m_i^* / M_i^*$, $\hat{t}_i^* = \frac{M_i^*}{m_i^*} \sum_{k=1}^{m_i^*} y_{ik}^*$, $\bar{M}_0 = M_0 / N$ and $\hat{Y} = \frac{1}{M_0} \frac{N}{n} \sum_{i \in S_{1st}} \frac{M_i}{m_i} \sum_{k \in S_i} y_{ik}$.

- Step 4. Compute $\hat{\theta}^*$ using the same formulae that were used to obtain the original point estimator.
- Step 5. Repeat Steps 1–4 a large number of times, B , to obtain $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$.
- Step 6. The bootstrap variance estimator is $var^*(\hat{\theta}^*)$. In practice, the Monte Carlo approximation of $var^*(\hat{\theta}^*)$ is applied

$$\widehat{var}^* = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\theta}^*)^2,$$

where $\bar{\theta}^* = B^{-1} \sum_{b=1}^B \hat{\theta}_b^*$.

Rao and Wu [2] showed that in the case of a population total, the above algorithm matches the standard variance estimator (3). Rao et al. [4] proposed a weighted version of the Rao–Wu method, whereby the rescaling is applied to the sampling weights rather than the y -values; see also [10]. The method of Rao et al. [4] is described in Section 4.

3.2. The Mirror-Match Bootstrap Algorithm

Sitter [3] proposed an extension of his mirror-match bootstrap to the case of a two-stage sampling design. In [3], the algorithm assumed that the number of repetitions k_1 and k_{2i} (see below) are integers. It can be described as follows:

- Step 1. Choose $1 \leq n' < n$ and draw a sample of size n' psus from S_{1st} , according to simple random sampling without replacement.
- Step 2. Repeat Step 1 $k_1 = n(1-f_1^*) / \{n'(1-f_1)\}$ times independently to obtain a bootstrap sample of psus of size $n^* = k_1 n'$, where $f_1^* = n' / n$.
- Step 3. Choose $1 \leq m_i' < m_i$ and draw according to simple random sampling without replacement m_i' units within the i th psu obtained in Steps 1 and 2.
- Step 4. Repeat Step 3 $k_{2i} = N m_i (1-f_{2i}^*) / \{n^* m_i' (1-f_{2i})\}$ times independently to obtain a bootstrap sample of size $m_i^* = k_{2i} m_i'$ from the i th psu drawn in Step 3, where $f_{2i}^* = m_i' / m_i$.

Step 5. Compute $\hat{\theta}^*$ using the same formulae that were used to obtain the original point estimator.

Step 6. Repeat Steps 1–5 a large number of times, B , to obtain $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$.

Step 7. The bootstrap variance estimator is $var^*(\hat{\theta}^*)$. In practice, the Monte Carlo approximation of $var^*(\hat{\theta}^*)$ is applied

$$\widehat{var}^* = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\theta}^*)^2,$$

where $\bar{\theta}^* = B^{-1} \sum_{b=1}^B \hat{\theta}_b^*$.

Sitter [3] showed that in the case of a population total, the above algorithm matches the standard variance estimator (3). If k_1 and k_{2i} are not integers, a randomization between bracketing integer values is proposed in [15].

3.3. The Without-Replacement Bootstrap Algorithm

Sitter [16] proposed a pseudo-population bootstrap algorithm, referred to as the without-replacement bootstrap (BWO) method, in the case of uni-stage and two-stage sampling designs. We focus on the latter. In [16], the algorithm assumed that the quantities k_1, n', k_{2i} , and m'_i (see below) are integers. It can be described as follows:

Step 1: Create a pseudo-population by replicating each psu in S_{1st} k_1 times and each unit within the i th psu k_{2i} times. Let U' be the resulting pseudo-population consisting of $N' = nk_1$ psus, $U'_1, \dots, U'_{N'}$, of size $M'_1, \dots, M'_{N'}$, where there exists $j \in S_{1st}$ such that $M'_i = m_j k_{2j}$. Let $M'_0 = \sum_{i \in U'} M'_i$ be the total number of elements in the pseudo-population.

Step 2: From the pseudo-population U' , select a sample of psus, S_{1st}^* , of size n' , according to simple random sampling without replacement. In each selected psu, select a sample, S_i^* , of size m'_i according to simple random sampling without replacement.

Step 3: Compute $\hat{\theta}^*$ using the formulae that were used to obtain the original point estimator.

Step 4: Repeat Steps 2 and 3 a large number of times, B , to obtain $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$.

Step 5: The bootstrap variance estimator is $var^*(\hat{\theta}^*)$. In practice, the Monte Carlo approximation of $var^*(\hat{\theta}^*)$ is applied

$$\widehat{var}^* = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\theta}^*)^2,$$

where $\bar{\theta}^* = B^{-1} \sum_{b=1}^B \hat{\theta}_b^*$.

In the case of the population total (or the population mean), Sitter [16] showed that the bootstrap variance estimator reduces to the standard variance estimator provided that n' and k_1 satisfy

$$f'_1 = f_1 \text{ and } \frac{k_1(n-1)}{n'(k_1n-1)} = \frac{1}{n'} \tag{8}$$

and m'_i and k_{2i} satisfy

$$f'_{2i} = f_{2i} \text{ and } \frac{k_{2i}(m_i-1)}{n'(k_{2i}m_i-1)} = \frac{f_1}{nm_i}, \tag{9}$$

where $f'_1 = n'/N'$ and $f'_{2i} = m'_i/M'_i$, for each i . If k_1, n', k_{2i} , and m'_i are not integers, a randomization between bracketing integer values was proposed in [15]. In Appendix A, we show that, if we define k_1, n', k_{2i} , and m'_i as in (8) and (9), the bootstrap variance estimator does not reduce to the standard variance estimator in (3). We suggest a modification to (8) and (9) so that the bootstrap variance estimators reduces to the standard variance estimator (3). In the simulation study (see Section 5), we show that the modified version of Sitter [16] works well in terms of bias and coverage probability of confidence intervals.

3.4. The Bernoulli Bootstrap Algorithm

Funaoka et al. [17] proposed two bootstrap procedures, referred to as Bernoulli bootstrap, for stratified three-stage sampling. Here, we consider the special case of two-stage sampling. Funaoka et al. [17] proposed a short-cut algorithm and a general algorithm. The general algorithm can handle any combination of sample sizes but is more computationally intensive. The general algorithm can be described as follows:

Step 1. Draw a sample, \tilde{S}_1 , of size $n' = n - 1$, from the original sample of clusters, S_{1st} , according to simple random sampling with replacement. Generate n Bernoulli random variables, I_{1i} , with probability

$$p_1 = 1 - \frac{1 - f_1}{2(1 - n^{-1})}.$$

For each $i \in S_{1st}$, keep the i th cluster in the bootstrap sample S^* and go to Step 2, if $I_{1i} = 1$, and replace the i th cluster with one randomly selected cluster from \tilde{S}_{1st} , if $I_{1i} = 0$.

Step 2. For cluster i kept in Step 1, draw a sample, \tilde{S}_{2i} , of size $m'_i = m_i - 1$, from the original sample S_{2i} according to simple random sampling with replacement. Generate m_i Bernoulli random variable, I_{2ij} , with probability

$$p_{2i} = 1 - \frac{f_1(1 - f_{2i})}{2p_1^{-1}(1 - m_i^{-1})}.$$

For each $(ij) \in S_{2i}$, keep the (ij) th element in the bootstrap sample, S^* , if $I_{2ij} = 1$, and replace it with one randomly selected element from \tilde{S}_{2i} , if $I_{2ij} = 0$.

Step 3. Compute $\hat{\theta}^*$ using the formulae that were used to obtain the original point estimator.

Step 4. Repeat Steps 1–3 a large number of times, B , to obtain $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$.

Step 5. The bootstrap variance estimator is $var^*(\hat{\theta}^*)$. In practice, the Monte Carlo approximation of $var^*(\hat{\theta}^*)$ is applied

$$\widehat{var}^* = \frac{1}{B - 1} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2,$$

where $\bar{\hat{\theta}}^* = B^{-1} \sum_{b=1}^B \hat{\theta}_b^*$.

Funaoka et al. [17] argued that the resulting bootstrap variance estimator is consistent for both smooth and non-smooth parameters.

3.5. The Preston Bootstrap Weights Algorithm

Preston [18] proposed a bootstrap weights approach for stratified three-stage sampling. Here, we focus on the special case of two-stage sampling. The algorithm can be described as follows:

Step 1. Draw a sample of size n' psus from S_{1st} , according to simple random sampling without replacement. Let $\delta_i = 1$ if the i th psu is selected and $\delta_i = 0$, otherwise.

Step 2. Define the psu bootstrap weights:

$$w_i^* = \left\{ 1 + \lambda_1 \left(\frac{n}{n'} \delta_i - 1 \right) \right\} \pi_i^{-1},$$

where $\lambda_1 = \{n'(1 - f_1)/(n - n')\}^{1/2}$.

Step 3. Within each of the sample of psus selected in Step 1, draw a simple random sample without replacement, of size m'_i . Let $\delta_{ik} = 1$ if the k th element in the i th psu is

selected and $\delta_{ik} = 0$, otherwise. We define the conditional element bootstrap weights:

$$w_{ik}^* = \left\{ 1 + \lambda_1 \left(\frac{n}{n'} \delta_i - 1 \right) + \lambda_{2i} \left(\frac{n}{n'} \right)^{1/2} \delta_i \left(\frac{m_i}{m'_i} \delta_{ik} - 1 \right) \right\} \frac{\pi_i^{-1}}{w_i^*} \pi_{k|i}^{-1},$$

where $\lambda_{2i} = \{m'_i f_1(1 - f_{2i}) / (m_i - m'_i)\}^{1/2}$.

Step 4. Compute $\hat{\theta}^*$ using the formulae that were used to obtain the original point estimator with the original weights replaced by the bootstrap weights w_{ik}^* .

Step 5. Repeat Steps 1–4 a large number of times, B , to obtain $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$.

Step 6. The bootstrap variance estimator is $var^*(\hat{\theta}^*)$. In practice, the Monte Carlo approximation of $var^*(\hat{\theta}^*)$ is applied

$$\widehat{var}^* = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\theta}^*)^2,$$

where $\bar{\theta}^* = B^{-1} \sum_{b=1}^B \hat{\theta}_b^*$.

In the case of the population total, Preston [18] showed that the bootstrap variance estimator reduces to the textbook variance estimator (3). He suggested that the choice of $n' = \lfloor n/2 \rfloor$ and $m'_i = \lfloor m_i/2 \rfloor$ will be optimal and lead to non-negative bootstrap weights, where $\lfloor \cdot \rfloor$ denotes the integer part.

4. Bootstrap Procedures for Unequal Probability Sampling Designs

4.1. The Rao-Wu-Yue Bootstrap Weights Algorithm

Rao et al. [4] proposed a bootstrap weights approach for stratified multi-stage sampling designs. Unlike the method of Rao and Wu [2], it can be applied to estimate the variance of smooth and non-smooth parameters (e.g., quantiles).

Step 1. Select n' psus according to simple random sampling with replacement from S_{1st} .

Step 2. Define the bootstrap weight as

$$w_{ik}^* = \left\{ 1 + \left(\frac{n'}{n-1} \right)^{1/2} \left(\frac{nn_i^*}{n'} - 1 \right) \right\} \pi_i^{-1} \pi_{k|i}^{-1},$$

where n_i^* denotes the number of times the i th psu is selected in the bootstrap sample.

Step 3. Compute $\hat{\theta}^*$ using the formulae that were used to obtain the original point estimator with the original weights replaced by the bootstrap weights w_{ik}^* .

Step 4. Repeat Steps 1–3 B times to obtain $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$.

Step 5. The bootstrap variance estimator is $var^*(\hat{\theta}^*)$. In practice, the Monte Carlo approximation of $var^*(\hat{\theta}^*)$ is applied

$$\widehat{var}^* = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\theta}^*)^2,$$

where $\bar{\theta}^* = B^{-1} \sum_{b=1}^B \hat{\theta}_b^*$.

The algorithm of Rao et al. [4] leads to consistent variance estimators provided that the first-stage sampling fraction is negligible. The choice $0 < n' \leq n - 1$ leads to non-negative bootstrap weights.

4.2. The Pseudo-population Bootstrap Algorithm

Chauvet [8] proposed a pseudo-population bootstrap approach (PPB) in the case of unequal two-stage sampling designs. It can be described as follows:

- Step 1. Each unit $k \in S_i$ is duplicated $\lceil \pi_{k|i}^{-1} \rceil$ times to create a second-stage pseudo-population denoted by U_i^* , $i \in S_{1st}$, where $\lceil \cdot \rceil$ denotes the closet integer.
- Step 2. Each pair (S_i, U_i^*) is duplicated $\lceil \pi_i^{-1} \rceil$ times. The population of pairs is completed by selecting a sample in the set $\{(S_i, U_i^*); i \in S_{1st}\}$ by means of sampling design with first-order inclusion probabilities $\pi_i^{-1} - \lceil \pi_i^{-1} \rceil$. This leads to the pseudo-population U^* .
- Step 3. Select a first-stage bootstrap sample S_{1st}^* from U^* using the original first-stage sampling design with first-order inclusion probabilities π_i .
- Step 4. Select a second-stage bootstrap sample S_i^{**} from U_i^* using the original second-stage sampling design. We set $S_i^* = S_i^{**}$ with probability π_i and $S_i^* = S_i$ with probability $1 - \pi_i$. This procedure is applied to each pair $(S_i, U_i^*) \in S_{1st}^*$. The union of the S_i^* 's leads to the bootstrap sample S^* .
- Step 5. Compute $\hat{\theta}^*$ using the formulae that were used to obtain the original point estimator.
- Step 6. Steps 3–5 are repeated B_{PPB} times to obtain the bootstrap statistics $\hat{\theta}_1^*, \dots, \hat{\theta}_{B_{PPB}}^*$. Let

$$\hat{v}^* = \frac{1}{B_{PPB} - 1} \sum_{b=1}^{B_{PPB}} (\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2,$$

where $\bar{\hat{\theta}}^* = B_{PPB}^{-1} \sum_{b=1}^{B_{PPB}} \hat{\theta}_b^*$.

- Step 7. Steps 2–6 are repeated A_{PPB} times to obtain $\hat{v}_1^*, \dots, \hat{v}_{A_{PPB}}^*$. The variance of $\hat{\theta}$ is estimated by

$$\widehat{var}^* = \frac{1}{A_{PPB}} \sum_{a=1}^{A_{PPB}} \hat{v}_a^*.$$

Chauvet [8] showed that in the case of high entropy sampling design (e.g., [19–22]) at both stages, the above algorithm leads to a consistent estimator in the context of a population total. In the case of a fixed-size sampling design, Chauvet [8] suggested completing the pseudo-population in Step 2, by applying Poisson sampling design with inclusion probabilities $\pi_i^{-1} - \lceil \pi_i^{-1} \rceil$.

5. Simulation Study

We conducted a simulation study to assess the performance of the bootstrap procedures described in Sections 3 and 4 in terms of bias, stability, and coverage rate of confidence intervals based on the t -distribution with $n - 1$ degrees of freedom. The simulation study was carried out using the R software. We created two finite populations consisting of $N = 200$ primary sampling units. The cluster sizes M_i were generated according to a Poisson distribution with a mean equal to 50. In each population, we generated a survey variable y according to

$$y_{ij} = 10 + x_i + \varepsilon_{ij},$$

where

$$x_i \sim N\left(0, \sqrt{\frac{\rho}{1 - \rho}}\right) \quad \text{and} \quad \varepsilon_{ij} \sim N(M_i, 1). \tag{10}$$

The parameter ρ in (10) was set to 0.1 for Population 1 and 0.3 for Population 2. We were interested in estimating the population total of the y -variable, t_y , as well as the finite population median.

From each population, we selected $K = 3000$ samples according to a two-stage sampling design:

- (i) At the first stage, we selected n psus according to two sampling designs: simple random sampling without replacement and inclusion probability-proportional-to-size randomized systematic sampling. The value of n was set to $n = 10$ which corresponds to a first-stage sampling fraction, $f_1 = 5\%$, and $n = 40$, which corresponds to $f_1 = 20\%$.
- (ii) At the second stage, $m_i = 5$ elements within each psu selected at the first stage were selected according to simple random sampling without replacement.

In each sample, we computed the estimator \hat{t}_y given by (1). Its variance was estimated using the variance estimation procedures listed in Table 1. Except of the procedure of Chauvet [8], we used $B = 500$ bootstrap samples for all the other bootstrap procedures. For the procedure of Chauvet [8], we used $A_{PPB} = 10$ and $B_{PPB} = 50$ ($B = A_{PPB} \times B_{PPB}$).

Table 1. Variance estimation procedures for each population parameter/sampling design.

Sampling Design	Population Total	Population Median
SRSWOR/SRSWOR	Textbook variance estimator	Linearization variance estimator
	Rao and Wu [2]	Rao and Wu [2]
	Rao et al. [4]	Rao et al. [4]
	Modified Sitter	Modified Sitter
	Funaoka et al. [17]	Funaoka et al. [17]
	Chauvet [8]	Chauvet [8]
	Preston [18]	Preston [18]
IPPSWOR/SRSWOR	Textbook variance estimator	Linearization variance estimator
	Rao et al. [4]	Rao et al. [4]
	Chauvet [8]	Chauvet [8]

As a measure of bias of a variance estimator \widehat{var} , we computed its Monte Carlo percent relative bias

$$RB_{MC}(\widehat{var}) = 100 \times \frac{\mathbb{E}_{MC}(\widehat{var}) - \mathbb{V}_{MC}(\hat{t}_y)}{\mathbb{V}_{MC}(\hat{t}_y)},$$

where $\mathbb{E}_{MC}(\widehat{var})$ denotes the Monte Carlo expectation of \widehat{var} and $\mathbb{V}_{MC}(\hat{t}_y)$ denotes the Monte Carlo variance estimator of \hat{t}_y . As a measure of stability of a variance estimator \widehat{var} , we computed its Monte Carlo percent coefficient of variation given by

$$CV_{MC} = 100 \times \frac{\sqrt{\mathbb{V}_{MC}(\widehat{var})}}{\mathbb{E}_{MC}(\widehat{var})},$$

where $\mathbb{V}_{MC}(\widehat{var})$ denotes the Monte Carlo variance estimator of \widehat{var} . Finally, we computed the Monte Carlo coverage rate of t -based confidence intervals and their corresponding Monte Carlo average length.

Tables 2–5 show the results for the bootstrap methods in the case of SRSWOR/SRSWOR. Table 2 shows the Monte Carlo percent relative bias of the bootstrap variance estimators. For the population total, all the procedures led to small biases for $f_1 = 5\%$ with the value of absolute relative bias ranging from 1.08% to 8.66%. For the population median, except for the method of Rao and Wu [2], the other procedures led to good results with an absolute relative bias ranging from 0.02% to 16.58%. As expected, the method of Rao and Wu [2] led to substantial bias because it cannot be applied to non-smooth statistics. For $f_1 = 20\%$, the absolute relative bias varied between 2.49% and 10.00% for all bootstrap methods except for the method of Rao et al. [4] who suffered from a significant positive bias. This can be explained by the fact that the sampling fraction was not small. For $f_1 = 20\%$ the absolute relative bias varied between 2.60% and 13.20% for all bootstrap methods except for the methods of Rao and Wu [2] and Rao et al. [4].

Table 3 shows the percent CV. All the bootstrap methods led to similar Monte Carlo coefficients of variation (CV). For $f_1 = 5\%$, the CV varied between 41.2% and 46.0% for the population total, and between 59.0% and 64.3% (except for the method of Rao and Wu [2] that led to a CV of 69.1% for $\rho = 0.3$) for the population median. For $f_1 = 20\%$, the CV varied between 18.9% and 21.0% for the population total, and between 36.4% and 42.3% for the population median.

Tables 4 and 5 show the coverage probability and the average length of 95% confidence intervals based on the t -distribution, respectively. All the bootstrap methods led to good coverage and similar average length except the method of Rao and Wu [2] for the population

median. The coverage rate in all cases, except the method of Rao and Wu [2] for the population median, varied between 93.17% and 96.73%.

Tables 6–9 show the results for the bootstrap methods in the case of randomized IPPS systematic/SRSWOR. Table 6 shows the percent relative bias of the bootstrap variance estimators. The method of Chauvet [8] worked generally better than the method of Rao et al. [4] in terms of relative bias, especially in the case of the population median. The percent CVs presented in Table 7 were very similar for both methods.

Tables 8 and 9 respectively show the coverage probability and the average length of the 95 percent confidence intervals based on the *t*-distribution for both methods. Both bootstrap methods led to good coverage and similar average length. The coverage rate in all cases varied from 93.23% to 96.43%.

Table 2. Monte Carlo percent RB of the bootstrap variance estimators to estimate the variance of the point estimator based on 3000 samples selected according to SRSWOR/SRSWOR.

		$\rho = 0.1$		$\rho = 0.3$	
$\hat{\theta}$	Bootstrap Method	$f_1 = 5\%$ ($n = 10$)	$f_1 = 20\%$ ($n = 40$)	$f_1 = 5\%$ ($n = 10$)	$f_1 = 20\%$ ($n = 40$)
\hat{t}_y	Textbook	1.99	5.10	1.28	5.05
	Chauvet [8]	−7.79	2.49	−8.39	2.54
	Rao et al. [4]	6.99	29.07	6.27	29.38
	Rao and Wu [2]	8.66	10.00	6.96	9.20
	Preston [18]	1.88	5.20	1.13	5.11
	Modified Sitter	6.97	9.93	5.84	9.49
	Funaoka et al. [17]	1.80	4.00	1.08	4.15
$\hat{\theta}_{1/2}$	Textbook	11.71	7.42	19.31	10.96
	Chauvet [8]	−0.02	7.56	0.03	8.67
	Rao et al. [4]	11.19	20.39	13.05	27.92
	Rao and Wu [2]	930.28	729.81	502.50	367.14
	Preston [18]	10.97	7.58	16.58	11.00
	Modified Sitter	11.81	12.98	10.74	13.20
	Funaoka et al. [17]	7.38	2.60	9.21	6.76

Table 3. Monte Carlo percent CV based on 3000 samples selected according to SRSWOR/SRSWOR.

		$\rho = 0.1$		$\rho = 0.3$	
$\hat{\theta}$	Bootstrap Method	$f_1 = 5\%$ ($n = 10$)	$f_1 = 20\%$ ($n = 40$)	$f_1 = 5\%$ ($n = 10$)	$f_1 = 20\%$ ($n = 40$)
\hat{t}_y	Textbook	43.4	18.9	45.3	19.8
	Chauvet [8]	43.9	19.9	45.9	20.9
	Rao et al. [4]	44.0	20.1	45.8	21.0
	Rao and Wu [2]	41.2	19.3	43.3	20.2
	Preston [18]	43.8	20.2	45.6	21.0
	Modified Sitter	43.6	19.9	45.0	20.6
	Funaoka et al. [17]	44.1	20.2	46.0	21.0
$\hat{\theta}_{1/2}$	Textbook	62.0	37.2	62.4	36.4
	Chauvet [8]	61.0	41.1	60.5	38.9
	Rao et al. [4]	59.9	41.1	59.3	38.0
	Rao and Wu [2]	64.3	37.4	69.1	39.7
	Preston [18]	63.8	38.0	64.0	37.3
	Modified Sitter	59.4	40.4	59.0	38.3
	Funaoka et al. [17]	59.8	42.3	59.3	39.5

Table 4. Coverage rate (CR) of the 95% *t*-distribution based confidence intervals constructed using bootstrap standard error estimators based on 3000 samples selected according to SRSWOR/SRSWOR.

		$\rho = 0.1$		$\rho = 0.3$	
$\hat{\theta}$	Bootstrap Method	$f_1 = 5\%$ ($n = 10$)	$f_1 = 20\%$ ($n = 40$)	$f_1 = 5\%$ ($n = 10$)	$f_1 = 20\%$ ($n = 40$)
\hat{t}_y	Textbook	95.33	95.07	95.40	94.73
	Chauvet [8]	94.50	94.87	94.70	94.67
	Rao et al. [4]	95.50	96.63	95.77	96.73
	Rao and Wu [2]	96.10	95.33	96.00	95.47
	Preston [18]	95.37	95.07	95.30	94.77
	Modified Sitter	95.80	95.47	95.73	95.30
	Funaoka et al. [17]	95.27	94.97	95.33	94.73

Table 4. Cont.

		$\rho = 0.1$		$\rho = 0.3$	
$\hat{\theta}_{1/2}$	Textbook	94.63	95.23	94.27	95.20
	Chauvet [8]	93.90	94.67	93.17	94.57
	Rao et al. [4]	95.10	95.97	94.70	96.03
	Rao and Wu [2]	100.00	99.97	99.97	99.93
	Preston [18]	94.50	95.23	93.70	94.80
	Modified Sitter	95.07	95.13	94.67	94.93
	Funaoka et al. [17]	94.73	94.27	94.20	94.13

Table 5. Average length (AL) of the 95% *t*-distribution based confidence intervals constructed using bootstrap standard error estimators based on 3000 samples selected according to SRSWOR/SRSWOR.

		$\rho = 0.1$		$\rho = 0.3$	
$\hat{\theta}$	Bootstrap Method	$f_1 = 5\%$	$f_1 = 20\%$	$f_1 = 5\%$	$f_1 = 20\%$
		($n = 10$)	($n = 40$)	($n = 10$)	($n = 40$)
\hat{t}_y	Textbook	21,470.7	8725.5	23,157.5	9721.9
	Chauvet [8]	20,404.4	8612.6	22,012.1	9599.7
	Rao et al. [4]	21,978.5	9663.6	23,707.2	10,782.5
	Rao and Wu [2]	22,218.5	8925.4	23,857.5	9910.3
	Preston [18]	21,450.9	8724.4	23,134.4	9718.9
	Modified Sitter	21,988.4	8919.4	23,684.7	9921.4
	Funaoka et al. [17]	21,435.4	8674.5	23,119.5	9674.2
$\hat{\theta}_{1/2}$	Textbook	0.9796	0.4081	1.3847	0.5705
	Chauvet [8]	0.9284	0.4071	1.2717	0.5634
	Rao et al. [4]	0.9798	0.4306	1.3537	0.6117
	Rao and Wu [2]	2.9770	1.1343	3.0965	1.1681
	Preston [18]	0.9736	0.4081	1.3654	0.5701
	Modified Sitter	0.9795	0.4175	1.3407	0.5753
	Funaoka et al. [17]	0.9634	0.3972	1.3306	0.5582

Table 6. Monte Carlo percent RB of the bootstrap variance estimators to estimate the variance of the point estimator based on 3000 samples selected according to IPPS/SRSWOR.

		$\rho = 0.1$		$\rho = 0.3$	
		$f_1 = 5\%$	$f_1 = 20\%$	$f_1 = 5\%$	$f_1 = 20\%$
$\hat{\theta}$	Bootstrap Method	($n = 10$)	($n = 40$)	($n = 10$)	($n = 40$)
\hat{t}_y	Textbook	−1.26	−0.36	−1.98	−0.05
	Chauvet [8]	−13.65	−5.74	−11.86	−3.83
	Rao et al. [4]	1.24	9.93	2.06	18.03
$\hat{\theta}_{1/2}$	Textbook	18.37	3.92	22.58	5.14
	Chauvet [8]	0.16	−3.24	2.65	2.09
	Rao et al. [4]	16.45	15.71	17.72	21.56

Table 7. Monte Carlo percent CV based on 3000 samples selected according to IPPS/SRSWOR.

		$\rho = 0.1$		$\rho = 0.3$	
		$f_1 = 5\%$	$f_1 = 20\%$	$f_1 = 5\%$	$f_1 = 20\%$
$\hat{\theta}$	Bootstrap Method	($n = 10$)	($n = 40$)	($n = 10$)	($n = 40$)
\hat{t}_y	Textbook	44.9	18.9	44.6	19.0
	Chauvet [8]	46.2	21.0	46.7	21.3
	Rao et al. [4]	46.9	22.4	46.1	21.3
$\hat{\theta}_{1/2}$	Textbook	62.0	37.7	59.3	36.1
	Chauvet [8]	61.3	40.5	61.0	40.1
	Rao et al. [4]	60.3	41.2	58.4	37.5

Table 8. Coverage rate (CR) of the 95% *t*-distribution based confidence intervals constructed using bootstrap standard error estimators based on 3000 samples selected according to IPPS/SRSWOR.

		$\rho = 0.1$		$\rho = 0.3$	
		$f_1 = 5\%$	$f_1 = 20\%$	$f_1 = 5\%$	$f_1 = 20\%$
$\hat{\theta}$	Bootstrap Method	($n = 10$)	($n = 40$)	($n = 10$)	($n = 40$)
\hat{t}_y	Textbook	95.27	95.50	94.73	94.83
	Chauvet [8]	93.23	94.83	93.63	94.70
	Rao et al. [4]	95.40	96.07	95.03	96.43
$\hat{\theta}_{1/2}$	Textbook	94.53	94.63	95.70	93.63
	Chauvet [8]	93.77	93.80	93.27	93.93
	Rao et al. [4]	95.17	94.93	95.27	95.43

Table 9. Average length (AL) of the 95% t -distribution based confidence intervals constructed using bootstrap standard error estimators based on 3000 samples selected according to IPPS/SRSWOR.

		$\rho = 0.1$		$\rho = 0.3$	
		$f_1 = 5\%$	$f_1 = 20\%$	$f_1 = 5\%$	$f_1 = 20\%$
$\hat{\theta}$	Bootstrap Method	($n = 10$)	($n = 40$)	($n = 10$)	($n = 40$)
\hat{t}_y	Textbook	7694.4	3401.9	11,021.3	4746.8
	Chauvet [8]	7317.2	3314.3	10,508.3	4650.7
	Rao et al. [4]	7772.1	3566.9	11228.1	5152.5
$\hat{\theta}_{1/2}$	Textbook	0.9797	0.4184	1.3709	0.5597
	Chauvet [8]	0.9247	0.4110	1.2627	0.5472
	Rao et al. [4]	0.9736	0.4403	1.3461	0.6011

6. Final Remarks

The results of the simulation studies suggest that most bootstrap procedures work well in terms of bias and coverage rate of confidence intervals for estimating smooth or non-smooth parameters. An exception is the method of Rao and Wu [2] for quantiles and the method of Rao et al. [4] for appreciable first-stage sampling fractions. In terms of stability of the variance estimators, there is little difference between the bootstrap procedures. Our results are aligned with those of Saigo [23] who empirically compared four bootstrap procedures in the context of stratified three-stage sampling with simple random sampling without replacement at each stage: the procedure of Funaoka et al. [17], the mirror match bootstrap of Sitter [3], the method of Rao et al. [4], and the BWO method of Sitter [16].

In this paper, we have compared several bootstrap algorithms in the context of a two-stage sampling design. The algorithms were described in an ideal setup. In practice, the design weights π_k^{-1} undergo a weighting process that involves a nonresponse adjustment followed by some form of calibration, whose goal is to ensure consistency between survey estimates and known population quantities. When the first-stage sampling fraction is small, the method of Rao et al. [4] is typically used in surveys conducted by national statistical offices. To account for unit nonresponse and calibration, the bootstrap weights need to undergo the same weighting process (non-response adjustment and calibration) that was used in the original sample were.

Bootstrap may be used to estimate the variance of imputed estimators in the context of imputation for item nonresponse. If the first-stage sampling fraction is small, one can re-impute the missing values in each bootstrap sample using the same imputation procedure that was utilized in the original sample, see [24]. The case of non-negligible first-stage sampling fractions requires further research.

We end this paper by pointing out a very recent paper by Beaumont and Émond [25], who proposed a bootstrap weight approach for multi-stage sampling designs. It would be interesting to compare this method to the procedures considered in this paper.

Author Contributions: Conceptualization, S.C., D.H. and Z.M.; methodology, S.C., D.H. and Z.M.; software, S.C. and Z.M.; validation, S.C., D.H. and Z.M.; formal analysis, NA; investigation, S.C., D.H. and Z.M.; resources, S.C., D.H. and Z.M.; data curation, NA; writing—original draft preparation, S.C., D.H. and Z.M.; writing—review and editing, S.C., D.H. and Z.M.; visualization, NA; supervision, NA; project administration, S.C., D.H. and Z.M.; funding acquisition, NA. All authors have read and agreed to the published version of the manuscript.

Funding: Sixia Chen is partly supported by the National Institute on Minority Health and Health Disparities at the National Institutes of Health (1R21MD014658-01A1) and the Oklahoma Shared

Clinical and Translational Resources (U54GM104938) with an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The work of David Haziza and Z.M. is funded by grants of the Natural Sciences and Engineering Research Council of Canada.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank the Associate Editor and three reviewers for their comments that helped improve the overall quality of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this section, we show that, obtaining k_1 , n' , k_{2i} , and m'_i from Equations (8) and (9) in the Sitter [16] algorithm in Section 3.3, the resulting bootstrap variance estimator does not reduce to the standard variance estimator in the case of the population total $\theta = t_y$. In the case of $\theta = t_y$, the bootstrap estimator is

$$\hat{\theta}^* = \frac{N'}{n'} \sum_{i \in S_{1st}^*} \frac{M'_i}{m'_i} \sum_{k \in S_i^*} y_{ik}^*$$

where y_{ik}^* is the y -value for the k th selected element in S_i^* in Step 2. The bootstrap variance of $\hat{\theta}^*$ is

$$var^*(\hat{\theta}^*) = \mathbb{V}_{1*} \mathbb{E}_{2*}(\hat{\theta}^* | S_{1st}^*) + \mathbb{E}_{1*} \mathbb{V}_{2*}(\hat{\theta}^* | S_{1st}^*), \quad (\text{A1})$$

where the subscripts 1* and 2* respectively denote the expectation and the variance with respect to the first-stage and second-stage resampling design in Step 2. Let $t'_i = \sum_{k \in U'_i} y_{ik}$ be the total of y -values in U'_i in Step 1. The first and the second component of the bootstrap variance estimator $var^*(\hat{\theta}^*)$ in (A1) respectively are

$$\begin{aligned} \mathbb{V}_{1*} \mathbb{E}_{2*}(\hat{\theta}^* | S_{1st}^*) &= \mathbb{V}_{1*} \left(\frac{N'}{n'} \sum_{i \in S_{1st}^*} t'_i \right) \\ &= N'^2 \frac{1-f'_1}{n'} \frac{1}{N'-1} \sum_{i \in U'} \left(t'_i - \frac{1}{N'} \sum_{j \in U'} t'_j \right)^2 \\ &= N'^2 \frac{1-f'_1}{n'} \frac{k_1}{k_1 n - 1} \sum_{i \in S_{1st}} \left(k_{2i} m_i \bar{y}_i - \frac{1}{n} \sum_{j \in S_{1st}} k_{2j} m_j \bar{y}_j \right)^2 \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}_{1*} \mathbb{V}_{2*}(\widehat{\theta}^* | S_{1st}^*) &= \mathbb{E}_{1*} \left\{ \frac{N'^2}{n'^2} \sum_{i \in S_{1st}^*} M_i'^2 \frac{1-f_{2i}'}{m_i'} \frac{1}{M_i'-1} \sum_{k \in U_i'} \left(y_{ik}^* - \frac{1}{M_i'} \sum_{l \in U_i'} y_{il}^* \right)^2 \right\} \\ &= \frac{N'}{n'} \sum_{i \in U'} M_i'^2 \frac{1-f_{2i}'}{m_i'} \frac{1}{M_i'-1} \sum_{k \in U_i'} \left(y_{ik}^* - \frac{1}{M_i'} \sum_{l \in U_i'} y_{il}^* \right)^2 \\ &= N'^2 \frac{1}{nn'} \sum_{i \in S_{1st}} (k_{2i} m_i)^2 \frac{1-f_{2i}'}{m_i'} \frac{k_{2i}}{k_{2i} m_i - 1} \sum_{k \in S_i} (y_{ik} - \bar{y}_i)^2 \\ &= \sum_{i \in S_{1st}} (N' k_{2i} m_i)^2 \frac{1-f_{2i}'}{nn' m_i'} \frac{k_{2i} (m_i - 1)}{k_{2i} m_i - 1} s_{yi}^2, \end{aligned}$$

where $\bar{y}_i = \sum_{i \in S_i} / m_i$. In [16], it is claimed that the first and the second components of the bootstrap variance estimator respectively are

$$N^2 \frac{k_1(n-1)}{k_1 n - 1} \frac{1-f_1'}{n'} s_i^2,$$

and

$$N^2 \sum_{i \in S_{1st}} M_i^2 \frac{1-f_{2i}'}{nn' m_i'} \frac{k_{2i} (m_i - 1)}{k_{2i} m_i - 1} s_{yi}^2,$$

see Equation (3.5) in Section 3.2 in [16]. This is true only when

$$N' = k_1 n = N, \quad \text{and} \quad k_{2i} m_i = M_i \quad \text{for all } i \in S_{1st}. \tag{A2}$$

In other words, we have to define $k_1 = N/n$ and $k_{2i} = M_i/m_i$ which is contradictory to the way Sitter [16] defined k_1, n', k_{2i} , and m_i' using Equations (8) and (9).

In the following, we suggest how the method of Sitter [16] can be modified. We first define

$$k_1 = \frac{N}{n} \quad \text{and} \quad k_{2i} = \frac{M_i}{m_i}. \tag{A3}$$

To have the equality between the first and the second component of the bootstrap variance estimator and the first and the second component of the standard variance estimator in (3), respectively, we need to have

$$\begin{aligned} \frac{k_1(n-1)}{k_1 n - 1} \frac{(1-f_1')}{n'} &= \frac{1-f_1}{n} \\ \implies \frac{k_1(n-1)}{k_1 n - 1} \frac{1-\frac{n'}{N}}{n'} &= \frac{1-f_1}{n} \\ \implies n' &= \frac{N}{1 + \frac{1-f_1}{f_1} \frac{k_1 n - 1}{k_1 (n-1)}} \end{aligned} \tag{A4}$$

and

$$\begin{aligned} \frac{k_{2i}(m_i-1)}{k_{2i} m_i - 1} \frac{1-\frac{m_i'}{M_i'}}{n' m_i'} &= \frac{f_1}{n m_i} (1-f_{2i}) \\ \implies \frac{1-\frac{m_i'}{M_i'}}{m_i'} &= \frac{f_1(1-f_{2i})}{n m_i} \frac{n'(k_{2i} m_i - 1)}{k_{2i}(m_i - 1)} \\ \implies m_i' &= \frac{M_i}{1 + \frac{f_1(1-f_{2i})}{n f_{2i}} \frac{n'(k_{2i} m_i - 1)}{k_{2i}(m_i - 1)}} \end{aligned} \tag{A5}$$

Defining k_1 , n' , k_{2i} , and m'_i as in Equations (A3)–(A5), the bootstrap variance estimator reduces to the usual variance estimator in (3). In the simulation study, the method "Modified Sitter" refers to the method of Sitter [16] while applying the modified k_1 , n' , k_{2i} , and m'_i defined in Equations (A3)–(A5). When k_1 , n' , k_{2i} , or m'_i is not integer, we simply rounded it to the closest integer value.

References

1. Efron, B. Bootstrap methods: Another look at the jackknife. *Ann. Stat.* **1979**, *7*, 1–26. [[CrossRef](#)]
2. Rao, J.N.K.; Wu, C.F.J. Resampling inference with complex survey data. *J. Am. Stat. Assoc.* **1988**, *83*, 231–241. [[CrossRef](#)]
3. Sitter, R.R. A resampling procedure for complex survey data. *J. Am. Stat. Assoc.* **1992**, *87*, 755–765. [[CrossRef](#)]
4. Rao, J.N.K.; Wu, C.F.J.; Yue, K. Some recent work on resampling methods for complex surveys. *Surv. Methodol.* **1992**, *18*, 209–217.
5. Gross, S. Median estimation in sample surveys. *Proc. Sect. Surv. Res. Methods* **1980**, 1814184.
6. Bickel, P.J.; Freedman, D.A. Asymptotic normality and the bootstrap in stratified sampling. *Ann. Stat.* **1984**, *12*, 470–482. [[CrossRef](#)]
7. Booth, J.G.; Butler, R.W.; Hall, P. Bootstrap methods for finite populations. *J. Am. Stat. Assoc.* **1994**, *89*, 1282–1289. [[CrossRef](#)]
8. Chauvet, G. Méthodes de Bootstrap en Population Finie. Ph.D. Thesis, École Nationale de Statistique et Analyse de l'Information, Bruz, France, 2007.
9. Antal, E.; Tillé, Y. A direct bootstrap method for complex sampling designs from a finite population. *J. Am. Stat. Assoc.* **2011**, *106*, 534–543. [[CrossRef](#)]
10. Beaumont, J.F.; Patak, Z. On the generalized bootstrap for sample surveys with special attention to Poisson sampling. *Int. Stat. Rev.* **2012**, *80*, 127–148. [[CrossRef](#)]
11. Mashreghi, Z.; Haziza, D.; Léger, C. A survey of bootstrap methods in finite population sampling. *Stat. Surv.* **2016**, *10*, 1–52. [[CrossRef](#)]
12. Särndal, C.E.; Swensson, B.; Wretman, J. *Model-Assisted Survey Sampling*; Springer: New York, NY, USA, 1992.
13. Beaumont, J.F.; Béliveau, A.; Haziza, D. Clarifying some aspects of variance estimation in two-phase sampling. *J. Surv. Stat. Methodol.* **2015**, *3*, 524–542. [[CrossRef](#)]
14. Wolter, K.M. *Introduction to Variance Estimation*; Springer Series in Statistics: New York, NY, USA, 2007.
15. Sitter, R.R. Resampling Procedures for Complex Survey Data. Ph.D. Thesis, University of Waterloo, Waterloo, ON, Canada, 1989.
16. Sitter, R.R. Comparing three bootstrap methods for survey data. *Can. J. Stat.* **1992**, *20*, 135–154. [[CrossRef](#)]
17. Funaoka, F.; Saigo, H.; Sitter, R.R.; Toida, T. Bernoulli bootstrap for stratified multistage sampling. *Surv. Methodol.* **2006**, *32*, 151–156.
18. Preston, J. Rescaled bootstrap for stratified multistage sampling. *Surv. Methodol.* **2009**, *35*, 227–234.
19. Hájek, J. Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Stat.* **1964**, *35*, 1491–1523. [[CrossRef](#)]
20. Berger, Y.G. Rate of convergence for asymptotic variance of the Horvitz–Thompson estimator. *J. Stat. Plan. Inference* **1998**, *74*, 149–168. [[CrossRef](#)]
21. Matei, A.; Tillé, Y. Evaluation of variance approximations and estimators in maximum entropy sampling with unequal probability and fixed sample size. *J. Off. Stat.* **2005**, *21*, 543–570.
22. Haziza, D.; Mecatti, F.; Rao, J.N.K. Evaluation of some approximate variance estimators under the Rao–Sampford unequal probability sampling design. *Metron* **2008**, *66*, 91–108.
23. Saigo, H. Comparing four bootstrap methods for stratified three-stage sampling. *J. Off. Stat.* **2010**, *26*, 193–207.
24. Shao, J.; Sitter, R.R. Bootstrap for imputed survey data. *J. Am. Stat. Assoc.* **1996**, *91*, 1278–1288. [[CrossRef](#)]
25. Beaumont, J.F.; Émond, N. A bootstrap variance estimation method for multistage sampling and two-phase sampling when poisson sampling is used at the second phase. *Stats* **2022**, *5*, 339–357. [[CrossRef](#)]