

## Article

# A Bootstrap Variance Estimation Method for Multistage Sampling and Two-Phase Sampling When Poisson Sampling Is Used at the Second Phase

Jean-François Beaumont \* and Nelson Émond

Statistics Canada, Ottawa, ON K1A 0T6, Canada; nelson.emond@statcan.gc.ca

\* Correspondence: jean-francois.beaumont@statcan.gc.ca

**Abstract:** The bootstrap method is often used for variance estimation in sample surveys with a stratified multistage sampling design. It is typically implemented by producing a set of bootstrap weights that is made available to users and that accounts for the complexity of the sampling design. The Rao–Wu–Yue method is often used to produce the required bootstrap weights. It is valid under stratified with-replacement sampling at the first stage or fixed-size without-replacement sampling provided the first-stage sampling fractions are negligible. Some surveys use designs that do not satisfy these conditions. We propose a simple and unified bootstrap method that addresses this limitation of the Rao–Wu–Yue bootstrap weights. This method is applicable to any multistage sampling design as long as valid bootstrap weights can be produced for each distinct stage of sampling. Our method is also applicable to two-phase sampling designs provided that Poisson sampling is used at the second phase. We use this design to model survey nonresponse and derive bootstrap weights that account for nonresponse weighting. The properties of our bootstrap method are evaluated in three limited simulation studies.



**Citation:** 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. <https://doi.org/10.3390/stats5020019>

Academic Editor: Fulvia Mecatti

Received: 29 January 2022

Accepted: 16 March 2022

Published: 22 March 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/>).

**Keywords:** bootstrap weights; two-stage sampling; multistage sampling; non-negligible sampling fraction; two-phase sampling; nonresponse

## 1. Introduction

The bootstrap method is often used for estimating the variance of survey estimates in Statistics Canada's social surveys and sometimes in its economic surveys. It is implemented by producing a set of bootstrap weights that is made available to users along with the survey data. This implementation of the bootstrap method is convenient for users as they can typically easily replicate their statistics using the bootstrap weights instead of the original survey weights.

Most of Statistics Canada's surveys that implement the bootstrap method have a complex stratified two-stage or three-stage sampling design. The Rao–Wu–Yue bootstrap weights [1] are often computed in those surveys. The Rao–Wu–Yue bootstrap weights are applicable when the first-stage sample is drawn with replacement within strata. With-replacement sampling is never applied in practice; however, the Rao–Wu–Yue method provides a reasonable approximation for fixed-size without-replacement sampling when the first-stage sampling fractions are small.

In some multistage surveys, the first-stage sampling fractions may be large in some strata, and the Rao–Wu–Yue bootstrap weights may lead to substantial overestimation of the variance of survey estimates. The literature on this topic is mostly limited to the case of simple random sampling at each stage (e.g., [2–4]). Chaudhuri and Saha [5] considered two-stage sampling with the Rao–Hartley–Cochran (RHC) method [6] at both stages. Inspired by [7], Osiewicz and Pérez-Duarte [8] provided bootstrap weights for the case of RHC sampling at the first stage followed by simple random sampling without replacement at the second stage.

In addition to RHC and simple random sampling, Poisson sampling and sampling without replacement with probability proportional to size are sometimes used in one of the stages. Beaumont and Patak [9] proposed a general bootstrap method that can be applied to any design, including multistage sampling designs, as long as joint selection probabilities can be computed or at least accurately approximated. However, the method is somewhat complex to implement, as it requires a spectral decomposition, and negative bootstrap weights may occur. A recent review of bootstrap methods for finite population sampling is provided in [10].

To our knowledge, there does not appear to exist a simple and general method for constructing non-negative bootstrap weights in multistage sampling with different sampling methods at each stage. We thus developed a unified bootstrap procedure that is straightforward to implement to fill this methodological gap. Our procedure can be applied to any number of stages as long as valid bootstrap weights can be produced for each distinct stage of sampling.

Our bootstrap procedure is also directly applicable to two-phase sampling designs when Poisson sampling is used at the second phase. This specific two-phase design can be viewed as a special case of two-stage sampling where first-stage units contain only one second-stage unit. It is also a useful design to model survey nonresponse. Our bootstrap procedure can thus be applied to estimate the variance of nonresponse-weighted estimators. Replication variance estimation for this specific two-phase design was also considered by [11].

In Section 2, we introduce notation, provide textbook variance estimators for the double expansion estimator of the population total under two-stage sampling and describe basic bootstrap concepts for finite populations. A few commonly-used methods of deriving bootstrap weights for single-stage cluster sampling are briefly reviewed in Section 3 for different sampling designs. For probability proportional to size without replacement sampling, we propose a simple heuristic extension of the Rao–Wu–Yue bootstrap weights.

Our bootstrap procedure for two-stage sampling is presented in Section 4. Section 5 considers the special case of two-phase sampling with Poisson sampling at the second phase, and the application to nonresponse is discussed in Section 6. Section 7 describes how to apply our bootstrap procedure for multistage sampling using the example of three-stage sampling. Bootstrap theory for nonlinear estimators is briefly discussed in Section 8. Section 9 provides the results of three limited simulation studies to evaluate the performance of our proposed bootstrap methods. Finally, a few concluding remarks are given in Section 10.

## 2. Background and Notation

Let us first consider the estimation of the population total

$$\theta = \sum_{k \in U} \sum_{i \in U_k} y_{ki} = \sum_{k \in U} Y_k,$$

where  $U$  is the population of clusters,  $U_k$  is the population of units within cluster  $k$ ,  $y_{ki}$  is the value of variable of interest  $y$  for unit  $i$  in cluster  $k$  and  $Y_k = \sum_{i \in U_k} y_{ki}$  is the total of variable  $y$  within cluster  $k$ .

### 2.1. Single-Stage Cluster Sampling

At the first stage, a sample of clusters,  $s$ , is selected from the population  $U$  according to a probability sampling design. The probability of selecting a given cluster  $k \in U$  at the first stage is  $\pi_{1k}$ , and the first-stage design weight of cluster  $k$  is defined as  $w_{1k} = 1/\pi_{1k}$ . The joint probability of selecting clusters  $k \in U$  and  $l \in U$  is  $\pi_{1kl}$ . We assume that both  $\pi_{1k}$  and  $\pi_{1kl}$  are strictly greater than 0 for  $k \in U$  and  $l \in U$ . The basic estimator of  $\theta$  under single-stage cluster sampling is the simple expansion estimator,

$$\tilde{\theta} = \sum_{k \in s} w_{1k} Y_k.$$

It is design-unbiased, i.e.,  $E_p(\tilde{\theta}) = \theta$ , where the subscript  $p$  indicates that the expectation is taken with respect to the sampling design. The design variance of  $\tilde{\theta}$  is given by

$$V_1 = \text{var}_p(\tilde{\theta}) = \sum_{k \in U} \sum_{l \in U} (\pi_{1kl} - \pi_{1k}\pi_{1l})(w_{1k}Y_k)(w_{1l}Y_l). \tag{1}$$

We consider estimators of  $V_1$  that takes the following quadratic form:

$$\tilde{V}_1 = \sum_{k \in s} \sum_{l \in s} \Delta_{1kl}(w_{1k}Y_k)(w_{1l}Y_l), \tag{2}$$

where  $\Delta_{1kl}$  is determined so as to ensure that  $\tilde{V}_1$  is design-unbiased for  $V_1$ , i.e.,  $E_p(\tilde{V}_1) = V_1$ . The typical choice is

$$\Delta_{1kl} = \frac{\pi_{1kl} - \pi_{1k}\pi_{1l}}{\pi_{1kl}}. \tag{3}$$

Under regularity conditions, this also ensures that  $\tilde{V}_1$  is design-consistent for  $V_1$ .

### 2.2. Two-Stage Sampling

At the second stage, a sample  $s_k$  is drawn from  $U_k$ , independently from one selected cluster  $k \in s$  to another, using some probability sampling design. The probability of selecting unit  $i$  in sample cluster  $k$  is  $\pi_{2ki}$ , and the second-stage design weight of unit  $i$  in sample cluster  $k$  is defined as  $w_{2ki} = 1/\pi_{2ki}$ . The joint probability of selecting units  $i$  in the sample cluster  $k$  and unit  $j$  in the sample cluster  $l$  is  $\pi_{2ki,lj}$ .

Again, we assume that  $\pi_{2ki}$  and  $\pi_{2ki,lj}$  are all strictly greater than 0. Due to the above independence assumption,  $\pi_{2ki,lj} = \pi_{2ki}\pi_{2lj}$  for two units  $i$  and  $j$  in different clusters. This assumption is key for the validity of our bootstrap method, but we do not require the typical invariance assumption in two-stage sampling (see [12], Chapter 4, and [13]).

The double expansion estimator of the population total  $\theta$  is

$$\hat{\theta} = \sum_{k \in s} w_{1k} \sum_{i \in s_k} w_{2ki}y_{ki} = \sum_{k \in s} w_{1k} \hat{Y}_k, \tag{4}$$

where  $\hat{Y}_k = \sum_{i \in s_k} w_{2ki}y_{ki}$ . It follows from  $E_p(\hat{Y}_k | s) = Y_k, k \in s$ , that  $\hat{\theta}$  is design-unbiased for  $\theta$ . The design variance of  $\hat{\theta}$  is given by

$$\text{var}_p(\hat{\theta}) = \text{var}_p(\tilde{\theta}) + E_p\left(\sum_{k \in s} w_{1k}^2 V_{2k}\right), \tag{5}$$

where

$$V_{2k} = \text{var}_p(\hat{Y}_k | s), k \in s.$$

The first term on the right-hand side of (5),  $V_1 = \text{var}_p(\tilde{\theta})$ , is the variance under single-stage cluster sampling given in (1). The second term,  $V_2 = E_p(\sum_{k \in s} w_{1k}^2 V_{2k})$ , reflects the increase in variance due to the second stage of sampling.

The estimator  $\tilde{V}_1$  of  $V_1$ , given in (2), cannot be used in two-stage sampling because the cluster totals  $Y_k, k \in s$ , are unknown. Instead, we consider estimators of  $V_1$  that take the form:

$$\hat{V}_1 = \sum_{k \in s} \sum_{l \in s} \Delta_{1kl}(w_{1k}\hat{Y}_k)(w_{1l}\hat{Y}_l) - \sum_{k \in s} \Delta_{1kk}w_{1k}^2 \hat{V}_{2k}, \tag{6}$$

where

$$\hat{V}_{2k} = \sum_{i \in s_k} \sum_{j \in s_k} \Delta_{2ki,kj}(w_{2ki}y_{ki})(w_{2kj}y_{kj}). \tag{7}$$

The quantity  $\Delta_{2ki,kj}$  in (7) is determined so as to ensure that  $\hat{V}_{2k}$ ,  $k \in s$ , is conditionally design-unbiased for  $V_{2k} = \text{var}_p(\hat{Y}_k|s)$ , i.e.,  $E_p(\hat{V}_{2k}|s) = V_{2k}$ . The typical choice is

$$\Delta_{2ki,kj} = \frac{\pi_{2ki,kj} - \pi_{2ki}\pi_{2kj}}{\pi_{2ki,kj}}. \tag{8}$$

The independence assumption at the second stage implies that  $E_p(\hat{Y}_k\hat{Y}_l|s) = Y_kY_l$  for two different sample clusters  $k$  and  $l$ . As a result, it is straightforward to show that  $\hat{V}_1$  in (6) is design-unbiased for  $V_1$ .

A design-unbiased estimator of  $V_2 = E_p(\sum_{k \in s} w_{1k}^2 V_{2k})$  is simply obtained as

$$\hat{V}_2 = \sum_{k \in s} w_{1k}^2 \hat{V}_{2k}.$$

The overall design variance in (5) can thus be estimated by  $\hat{V}_{12} = \hat{V}_1 + \hat{V}_2$ . Under regularity conditions,  $\hat{V}_{12}$  is design-consistent for the design variance (5). An alternative decomposition for  $\hat{V}_{12}$  is  $\hat{V}_{12} = \hat{V}_1^{alt} + \hat{V}_2^{alt}$ , where

$$\hat{V}_1^{alt} = \sum_{k \in s} \sum_{l \in s} \Delta_{1kl} (w_{1k} \hat{Y}_k) (w_{1l} \hat{Y}_l) \tag{9}$$

and

$$\hat{V}_2^{alt} = \sum_{k \in s} (1 - \Delta_{1kk}) w_{1k}^2 \hat{V}_{2k}. \tag{10}$$

This alternative decomposition is used to develop our bootstrap methodology in Section 4. For stratified sampling at the first stage, when all the first-stage sampling fractions are small, the term  $\hat{V}_2^{alt}$  is typically negligible. In that scenario, it is common practice to estimate the overall design variance (5) by  $\hat{V}_1^{alt}$  only and assume with-replacement sampling at the first stage. The Rao–Wu–Yue bootstrap method is based on this simplification. Although it is valid in many surveys, there are cases where some first-stage sampling fractions are not small, and  $\hat{V}_2^{alt}$  cannot be neglected.

Using (7), (9) and (10), the overall variance estimator  $\hat{V}_{12}$  can be rewritten as

$$\hat{V}_{12} = \sum_{k \in s} \sum_{i \in s_k} \sum_{l \in s} \sum_{j \in s_l} \Delta_{12ki,lj} (w_{12ki} y_{ki}) (w_{12lj} y_{lj}), \tag{11}$$

where  $w_{12ki} = w_{1k} w_{2ki}$  and

$$\Delta_{12ki,lj} = \Delta_{1kl} + (1 - \Delta_{1kl}) \Delta_{2ki,lj}, \tag{12}$$

with  $\Delta_{2ki,lj} = 0$ , for  $k \neq l$ , due to the independence assumption at the second stage. Assuming (3) and (8) are used, Equation (12) can be expressed as

$$\Delta_{12ki,lj} = \frac{\pi_{12ki,lj} - \pi_{12ki}\pi_{12lj}}{\pi_{12ki,lj}}, \tag{13}$$

where  $\pi_{12ki} = \pi_{1k}\pi_{2ki}$  and  $\pi_{12ki,lj} = \pi_{1kl}\pi_{2ki,lj}$ . Expression (11) has a form similar to (2) for single-stage cluster sampling and is useful for the extension to three-stage sampling discussed in Section 7.

### 2.3. Bootstrap Weights and Bootstrap Variance

Let us define the first-stage bootstrap weight for a sample cluster as  $w_{1k}^* = w_{1k} a_{1k}$ ,  $k \in s$ , where  $a_{1k}$  is a random bootstrap weight adjustment for cluster  $k$  that is intended to account for the first-stage sampling variability. Section 3 discusses how to generate  $a_{1k}$  with a few examples for some common sampling designs. The second-stage bootstrap weight for a sample unit  $i$  in cluster  $k$  is defined as  $w_{2ki}^* = w_{2ki} a_{2ki}$ ,  $i \in s_k$  and  $k \in s$ , where  $a_{2ki}$  is a random bootstrap weight adjustment for unit  $i$  in cluster  $k$  that is intended to account for

the second-stage sampling variability. A procedure for generating  $a_{2ki}$  will be discussed in Section 4. The overall bootstrap weight for a sample unit  $i$  in cluster  $k$  is defined as  $w_{12ki}^* = w_{12ki}a_{1k}a_{2ki} = w_{1k}^*w_{2ki}^*$ .

Using the bootstrap weights  $w_{1k}^*$  and  $w_{2ki}^*$ , we obtain bootstrap versions of the simple expansion estimator,  $\tilde{\theta} = \sum_{k \in s} w_{1k} Y_k$ , and the double expansion estimator  $\hat{\theta}$ , given in (4), as  $\tilde{\theta}^* = \sum_{k \in s} w_{1k}^* Y_k$  and

$$\hat{\theta}^* = \sum_{k \in s} w_{1k}^* \sum_{i \in s_k} w_{2ki}^* y_{ki} = \sum_{k \in s} w_{1k}^* \hat{Y}_k^* \tag{14}$$

respectively, where  $\hat{Y}_k^* = \sum_{i \in s_k} w_{2ki}^* y_{ki}$ . The bootstrap weights are constructed so that the bootstrap expectation and variance of  $\tilde{\theta}^*$  and  $\hat{\theta}^*$  are design-unbiased and design-consistent for the design expectation and variance of  $\tilde{\theta}$  and  $\hat{\theta}$ , respectively. Greater detail is given in Sections 3 and 4.

Suppose now that  $\hat{P}$  is a nonlinear design-weighted estimator of a certain population parameter  $P$ , not necessarily a total, such that, for large samples,  $E_p(\hat{P} - P) \approx 0$ . The bootstrap version  $\hat{P}^*$  of  $\hat{P}$  is obtained by replacing the first-stage and second-stage design weights by their corresponding bootstrap weights. Section 8 provides an example where  $\hat{P}$  is a smooth function of design-weighted estimators of totals. A valid bootstrap method ensures that, for large samples,  $E_*(\hat{P}^* - \hat{P}) \approx 0$  and  $\text{var}_*(\hat{P}^*) \approx E_*(\hat{P}^* - \hat{P})^2$  is design-consistent for  $\text{var}_p(\hat{P}) \approx E_p(\hat{P} - P)^2$ , where the subscript  $*$  indicates the moments are taken with respect to the bootstrap mechanism, conditionally on the selected two-stage sample.

Except for linear estimators, such as  $\tilde{\theta}^*$  and  $\hat{\theta}^*$ , discussed in Sections 3 and 4, respectively, a closed-form expression for the bootstrap variance  $\text{var}_*(\hat{P}^*)$  does not exist. It is generally approximated through Monte Carlo simulations by generating bootstrap adjustments independently a large number of times, say  $B$ . We denote these  $B$  replicate versions of  $a_{1k}$  and  $a_{2ki}$  by  $a_{1k}^{(b)}$  and  $a_{2ki}^{(b)}$ , respectively,  $b = 1, \dots, B$ . From these  $B$  bootstrap adjustments, we can compute  $B$  replicate versions  $\hat{P}^{*(b)}$  of  $\hat{P}^*$ ,  $b = 1, \dots, B$ . The bootstrap variance,  $\text{var}_*(\hat{P}^*)$ , can then be approximated as

$$\text{var}_*(\hat{P}^*) \approx \frac{1}{B} \sum_{b=1}^B (\hat{P}^{*(b)} - \hat{P})^2 .$$

In Sections 3–7, we focus on the estimation of the population total  $\theta = \sum_{k \in U} \sum_{i \in U_k} y_{ki}$ . Other population parameters are considered in Section 8 and in the simulation study in Section 9.

### 3. Bootstrap for Single-Stage Cluster Sampling

We consider distributions for generating the bootstrap adjustments  $a_{1k}$ ,  $k \in s$ , that satisfy the following two conditions:

$$E_*(a_{1k}) = 1 \tag{15}$$

and

$$\text{cov}_*(a_{1k}, a_{1l}) = \Delta_{1kl} , \tag{16}$$

for every pair of sample clusters  $k$  and  $l$ . Conditions (15) and (16) ensure that

$$E_*(\tilde{\theta}^* - \tilde{\theta}) = 0 \tag{17}$$

and

$$E_*(\tilde{\theta}^* - \tilde{\theta})^2 = \text{var}_*(\tilde{\theta}^*) = \tilde{V}_1 . \tag{18}$$

Conditions (17) and (18) are key for the validity of any bootstrap method. They ensure that the bootstrap expectation and variance of  $\tilde{\theta}^*$  are design-unbiased and design-consistent for the design expectation and variance of  $\tilde{\theta}$ , respectively.

Any distribution can be used to generate bootstrap adjustments  $a_{1k}$  as long as conditions (15) and (16) are satisfied. It may also be desirable to ensure that the bootstrap adjustments are all non-negative leading to non-negative bootstrap weights  $w_{1k}^*$ . We provide below some common expressions of bootstrap weight adjustments that achieve these objectives for different sampling designs.

### 3.1. Poisson Sampling

Under Poisson sampling, clusters are selected independently of one another. Assuming (3) is used, we have  $\Delta_{1kl} = 0, k \neq l$ , and  $\Delta_{1kk} = 1 - \pi_{1k}$ . Beaumont and Patak [9] studied several distributions that satisfy conditions (15) and (16), including some that ensured non-negative bootstrap adjustments  $a_{1k}$ . A main conclusion of their simulation study is that the choice of the distribution does not have a large impact on the properties of the bootstrap variance estimator as long as heavily-skewed distributions are avoided. We thus propose to generate  $a_{1k}, k \in s$ , independently of one another using the gamma distribution with  $E_*(a_{1k}) = 1$  and  $\text{var}_*(a_{1k}) = 1 - \pi_{1k}$ . This distribution satisfies conditions (15) and (16), ensures non-negative values of  $a_{1k}$  and is easy to implement.

### 3.2. Sampling with Replacement with Unequal Probabilities

Suppose that a sample  $s$  of size  $n_1$  is drawn with replacement from the population  $U$  of size  $N$ . At each draw, the population cluster  $k$  is chosen with probability  $p_{1k}$  and we have  $\sum_{k \in U} p_{1k} = 1$ . Note that some population clusters may appear multiple times in the sample  $s$ . In other words, each element of  $s$  is one of the  $n_1$  draws. Here, we define  $\pi_{1k} = n_1 p_{1k}$  and  $w_{1k} = \pi_{1k}^{-1}$ . (This is a slight abuse of notation as  $\pi_{1k}$  is not the selection probability of population cluster  $k$  in with-replacement sampling.) The standard variance estimator for  $\tilde{\theta} = \sum_{k \in s} w_{1k} Y_k$ , proposed by [14], can be written in the form (2) with  $\Delta_{1kl} = -(n_1 - 1)^{-1}, k \neq l$ , and  $\Delta_{1kk} = 1$ . The Rao–Wu–Yue bootstrap weight adjustment can be used for this design. It is given by

$$a_{1k} = 1 - \sqrt{\frac{m_1}{n_1 - 1}} + \sqrt{\frac{m_1}{n_1 - 1} \frac{n_1}{m_1} m_{1k}^*}, \tag{19}$$

for  $k \in s$ . The so-called multiplicity  $m_{1k}^*$  is the number of times cluster  $k \in s$  is chosen out of  $m_1$  independent random trials, where each  $k \in s$  has a constant probability of being chosen at each trial equal to  $n_1^{-1}$ . By observing that the multiplicities  $m_{1k}^*$  follow a multinomial distribution, it is straightforward to show that the bootstrap weight adjustments (19) satisfy conditions (15) and (16). They are non-negative provided that  $m_1 \leq n_1 - 1$ . The typical choice is  $m_1 = n_1 - 1$ . Using this choice, the bootstrap adjustment (19) reduces to  $a_{1k} = \frac{n_1}{n_1 - 1} m_{1k}^*$ .

If  $n_1$  is large and  $m_1 = n_1$ , the bootstrap weight adjustment (19) can be approximated as  $a_{1k} \approx m_{1k}^*$ . Using  $a_{1k} = m_{1k}^*$  is equivalent to applying the original bootstrap method of [15] for independently and identically distributed observations.

### 3.3. Simple Random Sampling without Replacement

Suppose now that a simple random sample  $s$  of size  $n_1$  is drawn without replacement from the population  $U$  of size  $N$ . Under this design, it can be easily shown that (3) reduces to  $\Delta_{1kl} = -(1 - f_1)/(n_1 - 1), k \neq l$ , and  $\Delta_{1kk} = 1 - f_1$ , where  $f_1 = n_1 N^{-1}$  is the sampling fraction. Rao, Wu and Yue [1] did not provide bootstrap weights for this design. However, [9] showed that the bootstrap method of [7] can be implemented by using the following bootstrap weight adjustment:

$$a_{1k} = 1 - \sqrt{\frac{m_1(1 - f_1)}{n_1 - 1}} + \sqrt{\frac{m_1(1 - f_1)}{n_1 - 1} \frac{n_1}{m_1} m_{1k}^*}, \tag{20}$$

for  $k \in s$ . Expression (20) is similar to (19) but it incorporates the finite population correction  $1 - f_1$ . Again, it is straightforward to show that the bootstrap weight adjustments (20) satisfy

conditions (15) and (16). They are non-negative provided that  $m_1 \leq (n_1 - 1)/(1 - f_1)$ . The choice  $m_1 = n_1 - 1$  dominates in practice.

### 3.4. Sampling without Replacement with Unequal Probabilities

The case of sampling  $n_1$  clusters without replacement with unequal probabilities is, in general, more difficult to handle as a simple expression, like (19) or (20), does not exist, and the joint selection probabilities  $\pi_{1kl}$  are often unknown. For the RHC sampling design, a design-unbiased variance estimator that takes the form (2) is available, and it is thus possible to construct valid bootstrap weights. The RHC sampling design was first considered by [7], and an expression for the bootstrap weights is given in [8]. Chaudhuri and Saha [5] also developed a bootstrap method for the RHC sampling design.

Beaumont and Patak [9] considered a method for generating bootstrap weight adjustments that satisfy (15) and (16) for general sampling designs. Their method requires the computation of joint selection probabilities or an accurate approximation of these probabilities. It is somewhat complex to implement and may produce negative bootstrap weights. As a simple alternative, we suggest below a heuristic extension of the bootstrap adjustment (20):

$$a_{1k} = 1 - \sqrt{\frac{m_1(1 - \pi_{1k})}{n_1 - 1}} + \sqrt{\frac{m_1(1 - \pi_{1k})}{n_1 - 1} \frac{n_1}{m_1} m_{1k}^*}. \tag{21}$$

The bootstrap adjustment (21) satisfies condition (15) and, if (3) is used, also satisfies  $\text{var}_*(a_{1k}) = \Delta_{1kk} = 1 - \pi_{1k}$ . However, it does not satisfy (16) entirely since

$$\text{cov}_*(a_{1k}, a_{1l}) = \tilde{\Delta}_{1kl} = -\frac{\sqrt{(1 - \pi_{1k})(1 - \pi_{1l})}}{n_1 - 1}, \tag{22}$$

for  $k \neq l$ ; the right-hand side of (22) can only be viewed as an approximation of  $\Delta_{1kl}$  given in (3). As a result, replacing  $\Delta_{1kl}$  in (2) with  $\tilde{\Delta}_{1kl}$  does not lead to an exactly design-unbiased estimator of  $V_1$ . By equating the right-hand side of (3) to the right-hand side of (22), we obtain the following approximation of the joint selection probability  $\pi_{1kl}, k \neq l$ :

$$\pi_{1kl} \approx \pi_{1k}\pi_{1l} \frac{(n_1 - 1)}{(n_1 - 1) + \sqrt{(1 - \pi_{1k})(1 - \pi_{1l})}}. \tag{23}$$

We note that, if all the selection probabilities  $\pi_{1k}$  are equal to  $n_1 N^{-1}$ , then the right-hand side of (23) reduces to the joint selection probability under simple random sampling without replacement and the bootstrap adjustment (21) reduces to (20). The bootstrap adjustment (21) is also approximately equal to (19) if all the selection probabilities  $\pi_{1k}$  are small. Note also that the bootstrap adjustment (21) is exactly equal to 1, as expected, for clusters selected with certainty in the sample.

A drawback of the bootstrap adjustment (21) is that the bootstrap version of  $n_1$ ,  $n_1^* = \sum_{k \in s} a_{1k}$ , is random and not necessarily equal to the fixed sample size  $n_1$ . This can be dealt with by calibrating the bootstrap adjustment (21) so as to ensure that the calibrated bootstrap adjustments  $a_{1k}^{cal}$  satisfy  $\sum_{k \in s} a_{1k}^{cal} = n_1$ . An alternative to (21) is thus

$$a_{1k}^{cal} = a_{1k} \frac{n_1}{\sum_{l \in s} a_{1l}}, \tag{24}$$

with  $a_{1k}$  given in (21). Note that clusters selected with certainty, if any, should be excluded from this calibration and keep their bootstrap adjustment of 1. Both (21) and (24) yield non-negative bootstrap adjustments when  $m_1 \leq n_1 - 1$  and are easy to implement. This simplicity comes at the expense of losing the strict design-unbiasedness property of the resulting variance estimator.

It is beyond the scope of this paper to rigorously assess the quality of the approximation (23) and the bootstrap adjustments (21) and (24). However, these bootstrap adjustments are evaluated in a limited simulation study in Section 9.1. Our simulation study shows

that the bootstrap adjustment (21) alone leads to substantial overestimation of the design variance but yields small biases if it is modified by the calibration adjustment (24). In [16], and recently in [17], bootstrap methods were proposed for without-replacement sampling with probability proportional to size that could be used as an alternative to (24).

### 3.5. Stratified Sampling

All the above sampling designs are typically applied within strata, and sampling is done independently from one stratum to another. The extension of the above bootstrap methods to stratified sampling is straightforward. It only suffices to generate the bootstrap weight adjustments independently from one stratum to another using the appropriate method.

## 4. Bootstrap for Two-Stage Sampling

First, bootstrap adjustments  $a_{1k}$ ,  $k \in s$ , that satisfy conditions (15) and (16) given in Section 3, are generated. The objective is then to generate  $a_{2ki}$  for all sample units so as to ensure that

$$E_*(\hat{\theta}^* - \hat{\theta}) = 0 \quad (25)$$

and

$$E_*(\hat{\theta}^* - \hat{\theta})^2 = \text{var}_*(\hat{\theta}^*) = \hat{V}_{12} = \hat{V}_1^{alt} + \hat{V}_2^{alt}, \quad (26)$$

where  $\hat{\theta}^*$  is given in (14). We assume that the adjustments  $a_{2ki}$  are generated independently from  $a_{1k}$ ,  $k \in s$  and that  $E_*(a_{2ki}) = 1$ . This ensures that condition (25) is satisfied. If we further assume that  $a_{2ki}$ ,  $i \in s_k$ , are generated independently from one cluster  $k \in s$  to another, then it is straightforward to show that

$$\text{var}_*(\hat{\theta}^*) = \hat{V}_1^{alt} + \sum_{k \in s} w_{1k}^2 (1 + \Delta_{1kk}) \sum_{i \in s_k} \sum_{j \in s_k} \text{cov}_*(a_{2ki}, a_{2kj}) (w_{2ki} y_{ki}) (w_{2kj} y_{kj}).$$

As a result, conditions (25) and (26) are satisfied if

$$E_*(a_{2ki}) = 1 \quad (27)$$

and

$$\text{cov}_*(a_{2ki}, a_{2kj}) = \frac{1 - \Delta_{1kk}}{1 + \Delta_{1kk}} \Delta_{2ki,kj}. \quad (28)$$

There are many possible ways of generating  $a_{2ki}$ ,  $i \in s_k$ , that satisfy conditions (27) and (28) for each  $k \in s$ . We propose generating the second-stage bootstrap adjustments as

$$a_{2ki} = 1 - \sqrt{\frac{1 - \Delta_{1kk}}{1 + \Delta_{1kk}}} + \sqrt{\frac{1 - \Delta_{1kk}}{1 + \Delta_{1kk}}} \tilde{a}_{2ki}, \quad (29)$$

where  $\tilde{a}_{2ki}$ ,  $i \in s_k$ , are preliminary second-stage bootstrap adjustments, generated independently from one cluster  $k \in s$  to another, that satisfy

$$E_*(\tilde{a}_{2ki}) = 1 \quad (30)$$

and

$$\text{cov}_*(\tilde{a}_{2ki}, \tilde{a}_{2kj}) = \Delta_{2ki,kj}. \quad (31)$$

The methods described in Section 3.1, Section 3.2, Section 3.3, Section 3.4, which satisfy conditions (15) and (16), can be adapted in a straightforward manner to generate  $\tilde{a}_{2ki}$ ,  $i \in s_k$ , that satisfy conditions (30) and (31). For instance, if Poisson sampling is used at the second stage along with Equation (8), then  $\tilde{a}_{2ki}$ ,  $i \in s_k$ , can be generated independently of one another using the gamma distribution with  $E_*(\tilde{a}_{2ki}) = 1$  and  $\text{var}_*(\tilde{a}_{2ki}) = 1 - \pi_{2ki}$ . In Section 9.2, we provide an expression for  $\tilde{a}_{2ki}$  under simple random sampling without replacement at the second stage.

Three remarks are in order: (i) the bootstrap adjustment  $a_{2ki}$  in (29) is computable only if  $-1 < \Delta_{1kk} \leq 1$ ; (ii) the bootstrap adjustment  $a_{2ki}$  is non-negative when  $\tilde{a}_{2ki}$  is non-negative and  $\Delta_{1kk} \geq 0$ ; and (iii) if (3) is used,  $0 \leq \Delta_{1kk} = 1 - \pi_{1k} < 1$ , and the bootstrap adjustment (29) can be rewritten as

$$a_{2ki} = 1 - \sqrt{\frac{\pi_{1k}}{2 - \pi_{1k}}} + \sqrt{\frac{\pi_{1k}}{2 - \pi_{1k}}} \tilde{a}_{2ki}. \quad (32)$$

It is non-negative provided that  $\tilde{a}_{2ki} \geq 0$ . The bootstrap adjustments (29) or (32) satisfy conditions (27) and (28).

We can also show that

$$\text{cov}_* \left( \sum_{k \in s} w_{1k}^* X_k, \hat{\theta}^* \right) = \sum_{k \in s} \sum_{l \in s} \Delta_{1kl} (w_{1k} X_k) (w_{1l} \hat{Y}_l), \quad (33)$$

where  $X_k = \sum_{i \in U_k} x_{ki}$  is the total of a certain variable  $x$  within cluster  $k$ . The right-hand side of (33) is a design-unbiased and design-consistent estimator of  $\text{cov}_p(\sum_{k \in s} w_{1k} X_k, \hat{\theta})$ . The property (33) is required for estimators that involve sums over first-stage and second-stage sample units (see Section 8). The nonresponse-adjusted estimator discussed in Section 6 is a special case that involves sums over first-phase and second-phase sample units.

Our bootstrap method can be summarized using the following steps:

- Step 1: Generate first-stage bootstrap adjustments  $a_{1k}$ ,  $k \in s$ , that satisfy conditions (15) and (16).
- Step 2: Generate preliminary second-stage bootstrap adjustments  $\tilde{a}_{2ki}$ ,  $i \in s_k$ , independently from one cluster  $k \in s$  to another, that satisfy conditions (30) and (31).
- Step 3: Compute the second-stage bootstrap adjustments  $a_{2ki}$ ,  $i \in s_k$ , using (29) or (32) for each  $k \in s$ .
- Step 4: Compute the first-stage bootstrap weights  $w_{1k}^* = w_{1k} a_{1k}$ ,  $k \in s$  and, for each  $k \in s$ , the overall bootstrap weights  $w_{12ki}^* = w_{12ki} a_{1k} a_{2ki}$ ,  $i \in s_k$ .

The above steps are repeated independently  $B$  times. Bootstrap variance estimates can then be computed, as described in Section 2.3, using the bootstrap weights obtained in step 4. Note that standard bootstrap methods for single-stage sampling can be used to obtain bootstrap adjustments in steps 1 and 2. This is discussed in Section 3 for the first-stage bootstrap adjustments in step 1. Preliminary second-stage bootstrap adjustments in step 2 can be obtained similarly.

## 5. Bootstrap for Two-Phase Sampling with Poisson Sampling at the Second Phase

In the next two sections, we consider, in greater detail, the case where all the population clusters are of size 1. Due to the independence assumption at the second stage, this special case of two-stage sampling is usually called two-phase sampling with Poisson sampling at the second phase. For the single unit  $i$  in population cluster  $k$ , we thus have  $y_{ki} = Y_k \equiv y_k$ . The selection of a unit  $i$  at the second stage (or phase) is equivalent to the selection of its cluster  $k$ , and we can write  $\pi_{2ki} = \pi_{2k}$ ,  $w_{2ki} = w_{2k}$  and  $w_{12ki} = w_{12k} = w_{1k} w_{2k}$ . Assuming (8) is used, we can also write  $\Delta_{2ki,ki} = 1 - \pi_{2ki} = 1 - \pi_{2k}$ . The estimated cluster total  $\hat{Y}_k$  can be written as  $\hat{Y}_k = I_{2k} w_{2k} y_k$ , where  $I_{2k} = 1$  if the unit in cluster  $k$  is selected at the second phase and  $I_{2k} = 0$ , otherwise. We denote by  $s_2$ , the set of clusters (or units) selected at the second phase, i.e., the set of units  $k \in s$  such that  $I_{2k} = 1$ . The double expansion estimator (4) reduces to  $\hat{\theta} = \sum_{k \in s_2} \frac{w_{1k}}{\pi_{2k}} y_k$ . The estimated variance (7) becomes

$$\hat{V}_{2k} = I_{2k} (1 - \pi_{2k}) (w_{2k} y_k)^2.$$

The alternative variance estimators  $\hat{V}_1^{alt}$  and  $\hat{V}_2^{alt}$ , given in (9) and (10), become

$$\hat{V}_1^{alt} = \sum_{k \in s_2} \sum_{l \in s_2} \Delta_{1kl} (w_{12k} y_k) (w_{12l} y_l) \quad (34)$$

and

$$\hat{V}_2^{alt} = \sum_{k \in s_2} (1 - \Delta_{1kk})(1 - \pi_{2k})(w_{12k}y_k)^2, \tag{35}$$

respectively. The bootstrap procedure in Section 4 can still be applied. With the modified notation in this section, it becomes:

- Step 1: Generate first-stage (first-phase) bootstrap adjustments  $a_{1k}$ ,  $k \in s$ , that satisfy conditions (15) and (16).
- Step 2: Generate preliminary second-stage (second-phase) bootstrap adjustments  $\tilde{a}_{2k}$ ,  $k \in s_2$ , independently of one another so as to satisfy  $E_*(\tilde{a}_{2k}) = 1$  and  $\text{var}_*(\tilde{a}_{2k}) = 1 - \pi_{2k}$ . We suggest using the gamma distribution.
- Step 3: Compute the second-stage (second-phase) bootstrap adjustments  $a_{2k}$ ,  $k \in s_2$ , using (29) or (32) with  $\tilde{a}_{2ki}$  and  $a_{2ki}$  replaced with  $\tilde{a}_{2k}$  and  $a_{2k}$ , respectively.
- Step 4: Compute the first-stage (first-phase) bootstrap weights  $w_{1k}^* = w_{1k}a_{1k}$ ,  $k \in s$  and the overall bootstrap weights  $w_{12k}^* = w_{12k}a_{1k}a_{2k}$ ,  $k \in s_2$ .

The application of these four steps ensures that conditions (25) and (26) are satisfied with  $\hat{V}_1^{alt}$  and  $\hat{V}_2^{alt}$  given in (34) and (35), respectively. The bootstrap estimator  $\hat{\theta}^*$  involved in these conditions can be expressed as  $\hat{\theta}^* = \sum_{k \in s_2} w_{1k}^* \frac{a_{2k}}{\pi_{2k}} y_k$ .

### 6. Application to Nonresponse

Nonresponse is often viewed as a second phase of sampling, and it is typically assumed that sample units respond independently of one another. Poisson sampling is thus a useful model to handle nonresponse. However, the response probability  $\pi_{2k}$ ,  $k \in s_2$ , is unknown. We assume that a consistent estimator  $\hat{\pi}_{2k}$  of  $\pi_{2k}$  can be obtained under a specified nonresponse model. The bootstrap procedure of Section 5 can still be used but  $\text{var}_*(\tilde{a}_{2k}) = 1 - \pi_{2k}$  in step 2 must be replaced with  $\text{var}_*(\tilde{a}_{2k}) = 1 - \hat{\pi}_{2k}$ . Condition (25) remains satisfied, and we can show that

$$\text{var}_*(\hat{\theta}^*) = \hat{V}_1^{alt} + \hat{V}_2^{alt} - \sum_{k \in s_2} (1 - \Delta_{1kk})(\hat{\pi}_{2k} - \pi_{2k})(w_{12k}y_k)^2,$$

where  $\hat{V}_1^{alt}$  and  $\hat{V}_2^{alt}$  are given in (34) and (35), respectively. Condition (26) is thus asymptotically satisfied provided that  $\hat{\pi}_{2k} - \pi_{2k} = o_p(1)$ .

The linear double expansion estimator  $\hat{\theta} = \sum_{k \in s_2} \frac{w_{1k}}{\pi_{2k}} y_k$  of the population total  $\theta$  cannot be used in the context of nonresponse because  $\pi_{2k}$  is unknown. It can be replaced with  $\hat{\pi}_{2k}$  to obtain the nonresponse-adjusted estimator  $\hat{\theta}_{NA} = \sum_{k \in s_2} \frac{w_{1k}}{\hat{\pi}_{2k}} y_k$ . The bootstrap version of the nonresponse-adjusted estimator  $\hat{\theta}_{NA}^*$  is  $\hat{\theta}_{NA}^* = \sum_{k \in s_2} w_{1k}^* \frac{a_{2k}}{\hat{\pi}_{2k}^*} y_k = \sum_{k \in s_2} w_{NA,k}^* y_k$ , where  $w_{NA,k}^* = w_{1k}^* \frac{a_{2k}}{\hat{\pi}_{2k}^*}$  is the bootstrap nonresponse-adjusted weight for unit  $k \in s_2$  and  $\hat{\pi}_{2k}^*$  is the bootstrap version of  $\hat{\pi}_{2k}$ .

In Section 6.1, we provide two examples of response probability models with their associated estimators  $\hat{\pi}_{2k}$  and  $\hat{\pi}_{2k}^*$ . An alternative replication method was proposed by [11]. This is discussed in Section 6.2.

#### 6.1. Examples

##### Uniform model

In the uniform model, the unknown response probability  $\pi_{2k}$  is assumed to be constant, i.e.,  $\pi_{2k} = \pi_2$ ,  $k \in s$ . It can be estimated by the design-weighted response rate

$$\hat{\pi}_2 = \frac{\sum_{k \in s_2} w_{1k}}{\sum_{k \in s} w_{1k}}, \tag{36}$$

provided that the design weight  $w_{1k}$  does not explain the response indicator  $I_{2k}$ . The bootstrap version of  $\hat{\pi}_2$  in (36) is

$$\hat{\pi}_2^* = \frac{\sum_{k \in s_2} w_{1k} a_{1k} a_{2k}}{\sum_{k \in s} w_{1k} a_{1k}}. \tag{37}$$

Unweighted versions of (36) and (37), obtained by omitting the design weight  $w_{1k}$  in the above two equations, are also used in practice.

The uniform model is usually assumed to hold within response homogeneous groups so that  $\pi_{2k} = \pi_{2h}$  for all sample units  $k \in s$  that fall into group  $h$ ,  $h = 1, \dots, H$ , where  $H$  is the number of groups. The response rates (36) and (37) are then computed separately for each group so as to obtain  $\hat{\pi}_{2h}$  and  $\hat{\pi}_{2h}^*$ ,  $h = 1, \dots, H$ .

*Logistic model*

Suppose now that the unknown response probability  $\pi_{2k}$  can be modelled using the logistic function  $\pi_{2k}(\alpha) = [1 + \exp(-x_k^T \alpha)]^{-1}$ , where  $x_k$  is a vector of auxiliary variables available for all  $k \in s$  and  $\alpha$  is a vector of unknown model parameters. The estimator  $\hat{\alpha}$  of  $\alpha$  can be obtained by solving for  $\alpha$  in the design-weighted (or pseudo) maximum likelihood estimating equation

$$\sum_{k \in s} w_{1k} [I_{2k} - \pi_{2k}(\alpha)] x_k = \mathbf{0}, \tag{38}$$

provided that the design weight  $w_{1k}$  does not explain  $I_{2k}$  after conditioning on  $x_k$ . The response probability can then be estimated as  $\hat{\pi}_{2k} = \pi_{2k}(\hat{\alpha})$ . The bootstrap version of (38) is

$$\sum_{k \in s} w_{1k}^* [a_{2k} I_{2k} - \pi_{2k}(\alpha)] x_k = \mathbf{0}. \tag{39}$$

Solving (39) for  $\alpha$  yields  $\hat{\alpha}^*$  and the bootstrap version of  $\hat{\pi}_{2k}$  is  $\hat{\pi}_{2k}^* = \pi_{2k}(\hat{\alpha}^*)$ . When  $x_k = 1, k \in s$ , the logistic model reduces to the uniform model.

6.2. *The Kim-Yu Method*

Kim and Yu [11] proposed a replication variance estimation method for two-phase sampling with Poisson sampling at the second phase. Their method is not applicable to the double expansion estimator but can be used with reweighted estimators, such as the nonresponse-adjusted estimator discussed above under the uniform model. Using the notation of our paper, the second-phase bootstrap adjustments proposed by [11] are given by

$$a_{2k} = 1 + (\delta_k - p) b_k, \quad k \in s_2, \tag{40}$$

where  $\delta_k$  follows a Bernoulli distribution with probability  $p$  and

$$b_k = \frac{\sqrt{\pi_{1k}(1 - \hat{\pi}_{2k})}}{\sqrt{p(1 - p)}}.$$

The adjustments (40) proposed by [11] can be rewritten as

$$a_{2k} = 1 - \sqrt{\pi_{1k}} + \sqrt{\pi_{1k}} \tilde{a}_{2k}, \tag{41}$$

where

$$\tilde{a}_{2k} = 1 + \sqrt{1 - \hat{\pi}_{2k}} \frac{(\delta_k - p)}{\sqrt{p(1 - p)}}$$

has a mean equal to 1 and a variance equal to  $1 - \hat{\pi}_{2k}$ .

The second-phase adjustment (41) is similar but not identical to our proposed adjustments. It would be identical to the adjustment (32) if  $\sqrt{\pi_{1k}}$  in (41) were replaced with  $\sqrt{\pi_{1k}/(2 - \pi_{1k})}$ . The adjustment (41) is expected to lead to overestimation of the variance. However, our simulation study in Section 9.3 shows that the overestimation remains moderate.

### 7. Bootstrap for Multistage Sampling

Designs with more than two stages of sampling are sometimes used in practice. For these multistage sampling designs, our bootstrap procedure in Section 4 can be applied repeatedly to generate bootstrap adjustments for each additional stage. First, bootstrap weight adjustments are generated for the first two stages of sampling. Then, the first two stages are viewed as a single stage, bootstrap weight adjustments for the third stage are generated and so on.

We provide below greater detail on our bootstrap procedure for three-stage sampling. To simplify the notation in this section, the first two stages are combined into a single stage so that the subscript  $k$  refers to second-stage units and the subscript  $i$  refers to third-stage units. As a result,  $s$  denotes the sample of second-stage units, and  $s_k$  denotes the sample of third-stage units within the selected second-stage unit  $k$ . Using this revised notation,  $\pi_{12ki}$  and the weight  $w_{12ki}$ , defined in Section 2.2, and  $\Delta_{12ki,lj}$ , given in (12), are now written  $\pi_{12k}$ ,  $w_{12k}$  and  $\Delta_{12kl}$ , respectively.

For three-stage sampling, our bootstrap method can be summarized using the following steps:

- Step 1: Generate combined first-stage and second-stage bootstrap adjustments  $a_{12k}$ ,  $k \in s$ , that satisfy  $E_*(a_{12k}) = 1$  and  $\text{cov}_*(a_{12k}, a_{12l}) = \Delta_{12kl}$ . The bootstrap procedure in Section 4 can be used to generate  $a_{12k}$ ,  $k \in s$ , as the product of the first-stage and second-stage bootstrap adjustments.
- Step 2: Generate preliminary third-stage bootstrap adjustments  $\tilde{a}_{3ki}$ ,  $i \in s_k$ , independently from one second-stage unit  $k \in s$  to another, that satisfy  $E_*(\tilde{a}_{3ki}) = 1$  and  $\text{cov}_*(\tilde{a}_{3ki}, \tilde{a}_{3kj}) = \Delta_{3ki,kj}$ , where  $\Delta_{3ki,kj}$  is determined similarly to  $\Delta_{2ki,kj}$  in Section 2.2.
- Step 3: Compute the third-stage bootstrap adjustments  $a_{3ki}$ ,  $i \in s_k$ , similarly to (29), using

$$a_{3ki} = 1 - \sqrt{\frac{1 - \Delta_{12kk}}{1 + \Delta_{12kk}}} + \sqrt{\frac{1 - \Delta_{12kk}}{1 + \Delta_{12kk}}} \tilde{a}_{3ki},$$

for each second-stage unit  $k \in s$ . Note that  $\Delta_{12kk} = 1 - \pi_{12k}$  if  $\Delta_{12kl}$  can be written in the form (13).

- Step 4: Compute the overall bootstrap weights  $w_{123ki}^* = w_{123ki} a_{12k} a_{3ki}$ , for each  $i \in s_k$  and  $k \in s$ , where  $w_{123ki} = w_{12k} w_{3ki}$ ,  $w_{3ki} = \pi_{3ki}^{-1}$  and  $\pi_{3ki}$  is the probability of selecting unit  $i \in s_k$  at the third stage.

### 8. Bootstrap for Nonlinear Estimators

In Section 4, we developed the theory for the linear double expansion estimator (4) and proposed random bootstrap weights that satisfy conditions (25) and (26). In practice, nonlinear estimators are often used. The nonresponse-adjusted estimator in Section 6 is an example. The bootstrap continues to be valid for nonlinear estimators under certain regularity conditions.

We provide below, through linearization techniques, an informal justification of the validity of the bootstrap method for nonlinear estimators that can be written as smooth functions of totals. A more formal justification is given in [18] for simple random sampling without replacement at the first stage with negligible first-stage sampling fractions. Justification of the bootstrap method for nonlinear estimators that are implicitly defined through estimating equations is discussed in [9,19]. The latter provides a rigorous treatment of this topic. Greater detail on the theory of bootstrap can be found in [20].

Suppose that we are interested in estimating the design variance of the nonlinear estimator  $\hat{P} = g(\tilde{T}_x, \hat{T}_y)$ , where the function  $g$  is a smooth function,  $\tilde{T}_x = \sum_{k \in s} w_{1k} X_k$ ,  $X_k = \sum_{i \in U_k} x_{ki}$ ,

$$\hat{T}_y = \sum_{k \in s} w_{1k} \sum_{i \in s_k} w_{2ki} y_{ki},$$

and  $x_{ki}$  and  $y_{ki}$  are vectors for unit  $i$  in cluster  $k$ . Under regularity conditions, the estimator  $\hat{P}$  is design-consistent for the population parameter  $P = g(T_x, T_y)$ , where  $T_x = \sum_{k \in U} X_k$ ,  $T_y = \sum_{k \in U} Y_k$  and  $Y_k = \sum_{i \in U_k} y_{ki}$ . The bootstrap version of  $\hat{P}$  is  $\hat{P}^* = g(\tilde{T}_x^*, \hat{T}_y^*)$ , where  $\tilde{T}_x^* = \sum_{k \in s} w_{1k}^* X_k$  and

$$\hat{T}_y^* = \sum_{k \in s} w_{1k}^* \sum_{i \in s_k} w_{2ki}^* y_{ki}.$$

Using a first-order Taylor linearization, the sampling error  $\hat{P} - P$  can be approximated as

$$\hat{P} - P \approx [h_x^T(T_x, T_y)](\tilde{T}_x - T_x) + [h_y^T(T_x, T_y)](\hat{T}_y - T_y), \tag{42}$$

where  $h_x(c, d) = \partial g(c, d) / \partial c$  and  $h_y(c, d) = \partial g(c, d) / \partial d$ . Using (42), the design variance of  $\hat{P}$  is approximated as

$$\begin{aligned} \text{var}_p(\hat{P}) &\approx h_x^T(T_x, T_y) [\text{var}_p(\tilde{T}_x)] h_x(T_x, T_y) \\ &+ h_y^T(T_x, T_y) [\text{var}_p(\hat{T}_y)] h_y(T_x, T_y) \\ &+ 2h_x^T(T_x, T_y) [\text{cov}_p(\tilde{T}_x, \hat{T}_y)] h_y(T_x, T_y). \end{aligned} \tag{43}$$

Similarly, we can again use a first-order Taylor linearization to approximate the bootstrap error  $\hat{P}^* - \hat{P}$  as

$$\hat{P}^* - \hat{P} \approx [h_x^T(\tilde{T}_x, \hat{T}_y)](\tilde{T}_x^* - \tilde{T}_x) + [h_y^T(\tilde{T}_x, \hat{T}_y)](\hat{T}_y^* - \hat{T}_y). \tag{44}$$

From (44), the bootstrap variance of  $\hat{P}^*$  is approximated as

$$\begin{aligned} \text{var}_*(\hat{P}^*) &\approx h_x^T(\tilde{T}_x, \hat{T}_y) [\text{var}_*(\tilde{T}_x^*)] h_x(\tilde{T}_x, \hat{T}_y) \\ &+ h_y^T(\tilde{T}_x, \hat{T}_y) [\text{var}_*(\hat{T}_y^*)] h_y(\tilde{T}_x, \hat{T}_y) \\ &+ 2h_x^T(\tilde{T}_x, \hat{T}_y) [\text{cov}_*(\tilde{T}_x^*, \hat{T}_y^*)] h_y(\tilde{T}_x, \hat{T}_y). \end{aligned} \tag{45}$$

The right-hand side of (45) is design-consistent for the right-hand side of (43) provided that  $\tilde{T}_x, \hat{T}_y, \text{var}_*(\tilde{T}_x^*), \text{var}_*(\hat{T}_y^*)$  and  $\text{cov}_*(\tilde{T}_x^*, \hat{T}_y^*)$  are design-consistent for  $T_x, T_y, \text{var}_p(\tilde{T}_x), \text{var}_p(\hat{T}_y)$  and  $\text{cov}_p(\tilde{T}_x, \hat{T}_y)$ , respectively. Conditions (15), (16), (27) and (28) are key to obtain a valid bootstrap method. They ensure that  $\text{var}_*(\tilde{T}_x^*), \text{var}_*(\hat{T}_y^*)$  and  $\text{cov}_*(\tilde{T}_x^*, \hat{T}_y^*)$  are identical to standard textbook estimators of  $\text{var}_p(\tilde{T}_x), \text{var}_p(\hat{T}_y)$  and  $\text{cov}_p(\tilde{T}_x, \hat{T}_y)$ , respectively.

### 9. Simulation Studies

We conducted three simulation studies to evaluate our proposed bootstrap methods under three different sampling designs. First, our population for the simulation studies was taken from 50 clusters of the Canadian Health Measures Survey (CHMS). We considered two variables of interest from the CHMS: Body Mass Index (BMI) and Height. The CHMS aims to collect important health information through a household interview and direct physical measures and is based on a stratified three-stage sampling design. Greater detail on the CHMS can be found at: <https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5071> (accessed on 28 January 2022).

From this population, we generated  $R = 10,000$  samples using one of the three sampling designs described in Section 9.1, Section 9.2, Section 9.3. Suppose that  $\hat{P}_r$  is

a weighted estimate of a certain population parameter  $P$  for the  $r^{\text{th}}$  generated sample. The design variance of the estimator  $\hat{P}$ ,  $\text{var}_p(\hat{P})$ , is approximated by the Monte Carlo variance

$$V_{MC} = \frac{1}{R} \sum_{r=1}^R (\hat{P}_r - P)^2.$$

Then, for each of the  $R$  samples, 1000 sets of bootstrap weights were generated using one of the methods discussed in Section 3, Section 4, Section 5, Section 6, and the bootstrap variance was obtained as described at the end of Section 2.3. Let us denote by  $v_r^{\text{boot}}$ , the bootstrap variance estimate for the  $r^{\text{th}}$  sample for any given sampling design and bootstrap method. The percent Monte Carlo Relative Bias (RB) of the bootstrap variance estimator is defined as

$$RB = \frac{1}{R} \sum_{r=1}^R e_r,$$

and its percent Monte Carlo Relative Root Mean Square Error (RRMSE) is defined as

$$RRMSE = \sqrt{\frac{1}{R} \sum_{r=1}^R e_r^2},$$

where  $e_r = \frac{(v_r^{\text{boot}} - V_{MC})}{V_{MC}} \times 100\%$  is the percent relative error. We evaluated bootstrap methods for the three sampling designs described in Section 9.1, Section 9.2, Section 9.3 by comparing their RBs and RRMSEs. The SAS code used to obtain our simulation results is available upon request from the authors.

### 9.1. Single-Stage Cluster Sampling without Replacement with Unequal Probabilities

The objective of this simulation study is to evaluate the accuracy of the bootstrap adjustments (21) and (24) under single-stage cluster sampling without replacement with probability proportional to size.

From our population of  $N = 50$  clusters, we generated  $R = 10,000$  samples using sequential Poisson sampling [21], where the size measure is the cluster size. We considered two sample sizes:  $n_1 = 10$  and  $n_1 = 30$ . The sampling fraction is 0.20 for  $n_1 = 10$  and 0.60 for  $n_1 = 30$ . We compared the RB and RRMSE of different bootstrap methods of estimating the design variance of the expansion estimator  $\hat{\theta} = \sum_{k \in s} w_{1k} Y_k$  for two variables of interest: BMI and Height. Table 1 describes the five methods we considered in this simulation study to obtain bootstrap weight adjustments  $a_{1k}$ ,  $k \in s$ .

**Table 1.** Bootstrap methods for single-stage cluster sampling.

Method	Description
RWY	Bootstrap adjustment (19) obtained under with-replacement sampling with $m_1 = n_1 - 1$ as proposed in [1].
SRS-RWY	Bootstrap adjustment (20) obtained under simple random sampling without replacement.
PPS-RWY	Proposed bootstrap adjustment (21) for sampling without replacement with probability proportional to size.
PPS-RWY-CAL	Proposed bootstrap adjustment (21) for sampling without replacement with probability proportional to size modified by the calibration adjustment (24).
POI-CAL	Bootstrap adjustment under Poisson sampling, using the gamma distribution discussed in Section 3.1, modified by the calibration adjustment (24).

Relative biases are provided in Table 2. As expected, the RWY method shows large positive biases, especially for the largest sampling fraction, as it is valid under with-replacement sampling. The SRS-RWY method is surprisingly not overly poor in terms of bias, except for the variable BMI and  $n_1 = 30$ , where the RB is not small (34.6%). The PPS-RWY method shows very large biases, whereas PPS-RWY-CAL is the best method in terms of bias.

**Table 2.** The percent relative biases of the bootstrap variance estimators.

Method	BMI		Height	
	$n_1=10$	$n_1=30$	$n_1=10$	$n_1=30$
RWY	31.2	236.4	25.1	145.8
SRS-RWY	4.9	34.6	0.1	−1.7
PPS-RWY	127.7	4288.6	1847.8	68,959.1
PPS-RWY-CAL	1.0	1.5	−0.2	−1.9
POI-CAL	−15.8	−2.1	−17.3	−6.3

In other words, it appears that modifying the bootstrap adjustment (21) by the calibration adjustment (24) is effective for taking into account the unknown joint selection probabilities and the fixed sample size. The POI-CAL method is also effective, although it shows negative RBs, especially for the smallest sample size. The POI-CAL method is sometimes used in practice for its simplicity.

Table 3 gives the RRMSEs. Both RWY and PPS-RWY show large RRMSEs due to their large biases. The other three methods are comparable, perhaps with a small advantage for POI-CAL.

**Table 3.** The percent relative root mean square errors of the bootstrap variance estimators.

Method	BMI		Height	
	$n_1=10$	$n_1=30$	$n_1=10$	$n_1=30$
RWY	64.0	243.4	56.0	152.3
SRS-RWY	45.0	41.6	40.1	17.7
PPS-RWY	161.4	4335.6	2021.1	69,700.3
PPS-RWY-CAL	44.2	30.0	41.8	24.9
POI-CAL	39.9	28.8	38.5	24.8

Overall, the proposed PPS-RWY-CAL method offers a good compromise in terms of both bias and mean square error. The POI-CAL method is an interesting alternative, especially for larger sample sizes.

*9.2. Two-Stage Sampling with Poisson Sampling at the First Stage and Simple Random Sampling at the Second Stage*

The objective of this second simulation study is to evaluate the accuracy of the bootstrap adjustment (32) under a two-stage sampling design. At the first stage, Poisson sampling was used to select a sample of clusters, with probability proportional to cluster size, among the  $N = 50$  population clusters. We considered two values of the expected sample size:  $E_p(n_1) = 10$  and  $E_p(n_1) = 30$ . For each of the selected clusters at the first stage, a second-stage sample of 10 units was selected using simple random sampling without replacement.

This process was repeated  $R = 10,000$  times. For each of the 10,000 samples, we computed double expansion estimates  $\hat{\theta} = \sum_{k \in s} \sum_{i \in s_k} w_{1k} w_{2ki} y_{ki}$  of the population total for the variables BMI and Height. We also computed weighted estimates of the population

median for these variables using the methodology implemented in the procedure Univariate of SAS for weighted percentiles.

We considered two bootstrap methods for estimating design variances: (i) the RWY method described in Table 1 and (ii) the proposed method described in Section 4. In our proposed method, we generated first-stage bootstrap adjustments using the gamma distribution described in Section 3.1. Then, we generated preliminary second-stage bootstrap adjustments within each selected cluster  $k \in s$  by applying the method in Section 3.3 for simple random sampling without replacement, i.e., we used

$$\tilde{a}_{2ki} = 1 - \sqrt{\frac{m_{2k}(1 - f_{2k})}{n_{2k} - 1}} + \sqrt{\frac{m_{2k}(1 - f_{2k})}{n_{2k} - 1} \frac{n_{2k}}{m_{2k}} m_{2ki}^*},$$

where  $n_{2k}$  is the number of units selected in cluster  $k$ ,  $f_{2k}$  is the sampling fraction in cluster  $k$ ,  $m_{2ki}^*$  is the number of times sample unit  $i$  in cluster  $k$  is chosen out of  $m_{2k} = n_{2k} - 1$  independent random trials, where each sample unit  $i$  in cluster  $k$  has a constant probability of being chosen at each trial equal to  $n_{2k}^{-1}$ . Finally, we computed second-stage bootstrap adjustments using Equation (32).

Table 4 provides RBs for both the RWY bootstrap method and the proposed method, whereas the RRMSEs are given in Table 5. The RWY method leads to significant negative biases for the estimation of population totals because it does not account for the random sample size at the first stage due to Poisson sampling. As expected, the proposed method shows small biases and smaller RRMSEs compared with the RWY method.

**Table 4.** The percent relative biases of the bootstrap variance estimators.

Method	Population Parameter	BMI		Height	
		$E_p(n_1)=10$	$E_p(n_1)=30$	$E_p(n_1)=10$	$E_p(n_1)=30$
RWY	Total	−98.8	−97.2	−99.7	−99.3
Proposed	Total	−1.8	1.3	0.71	4.3
RWY	Median	14.6	24.7	21.5	12.8
Proposed	Median	−2.9	10.0	3.8	6.8

**Table 5.** The percent relative root mean square errors of the bootstrap variance estimators.

Method	Population Parameter	BMI		Height	
		$E_p(n_1)=10$	$E_p(n_1)=30$	$E_p(n_1)=10$	$E_p(n_1)=30$
RWY	Total	98.8	97.2	99.7	99.3
Proposed	Total	29.1	16.6	28.9	17.5
RWY	Median	88.3	58.6	126.4	50.6
Proposed	Median	65.9	42.9	87.6	41.1

Estimators of population means or medians are typically less affected by random sample sizes than estimators of population totals. As a result, the RWY method shows moderate positive biases for the estimation of population medians due to the non-negligible first-stage sampling fractions. Again, the proposed method shows smaller biases and RRMSEs than the RWY method.

These results are not surprising since the RWY method is not intended for Poisson sampling and/or non-negligible first-stage sampling fractions. It remains a useful method in many practical scenarios. This simulation study simply illustrates that the sampling design must be properly taken into account when choosing a bootstrap method.

9.3. Two-Phase Sampling with Simple Random Sampling at the First Phase and Nonresponse at the Second Phase

The objective of this last simulation study is to evaluate our bootstrap procedure in Sections 5 and 6 under two-phase sampling with nonresponse at the second phase. Our population for the variable BMI is a random subset of all the units in the 50 population clusters that were used in Sections 9.1 and 9.2. This gave a total of 6227 population units. The population for the variable Height was constructed similarly by selecting a random subset of all the units in the same 50 clusters, which resulted in 6439 units.

For both populations, simple random sampling without replacement was used to select a first-phase sample with a sampling fraction of 0.5. From the first-phase sample units, nonresponse was then generated using Bernoulli sampling with probability 0.05 (uniform nonresponse). This process was again repeated  $R = 10,000$  times. For each of the 10,000 samples, we computed nonresponse-adjusted estimates  $\hat{\theta}_{NA} = \sum_{k \in s_2} \frac{w_{1k}}{\hat{\pi}_2} y_k$  of the population total for the variables BMI and Height, where the estimated probability  $\hat{\pi}_2$  is given in (36).

We considered two bootstrap methods for estimating design variances: (i) the Kim–Yu method described in Section 6.2 with  $p = 0.5$  and (ii) the proposed method described in Sections 5 and 6. In our proposed method, we generated first-phase bootstrap adjustments using Equation (20) in Section 3.3. Then, we generated preliminary second-phase bootstrap adjustments using the gamma distribution with mean 1 and variance  $1 - \hat{\pi}_2$  and computed the second-phase bootstrap adjustments using (32) as described in step (3) of Section 5. In both methods, the bootstrap version of  $\hat{\theta}_{NA}$  is  $\hat{\theta}_{NA}^* = \sum_{k \in s_2} w_{1k}^* \frac{a_{2k}}{\hat{\pi}_2^*} y_k$ , where  $\hat{\pi}_2^*$  is given in (37).

Table 6 provides RBs and RRMSEs for both bootstrap methods. The Kim–Yu method shows a moderate positive bias for both variables whereas the proposed method has a negligible bias, albeit negative. The larger RRMSEs of the Kim–Yu method can be explained by its larger bias.

**Table 6.** The percent RBs and percent RRMSEs of the bootstrap variance estimators.

Method	RB		RRMSE	
	BMI	Height	BMI	Height
Kim–Yu	22.8	23.0	30.6	29.2
Proposed	−2.6	−2.4	16.4	14.4

10. Concluding Remarks

We developed a simple and unified method for producing non-negative bootstrap weights that can be applied to any multistage sampling design provided that valid bootstrap weights can be produced for each distinct stage of sampling. This is not a major limitation in practice as most bootstrap methods for single-stage sampling can be implemented through the computation of bootstrap weights, including the pseudo-population method (see [9,10,16,22]).

We provided a few expressions for bootstrap weights under different single-stage sampling designs, including probability proportional to size without replacement sampling. For this design, we proposed a simple heuristic extension of the Rao–Wu–Yue bootstrap weights that performed well in our empirical investigations.

We assumed that the selection and joint selection probabilities were strictly greater than 0 at each stage, which excludes systematic sampling in one of the stages. For systematic sampling, the usual practical solution is to approximate this design with stratified simple random sampling without replacement, where strata are formed based on the order of the population units in the sampling frame.

Our bootstrap method for multistage sampling can be used even when first-stage sampling fractions are not small—unlike the commonly used Rao–Wu–Yue bootstrap weights. The results of simulation studies indicate that our method has good properties in terms of bias and mean square error.

As indicated in the introduction, there are alternatives bootstrap methods for some specific multistage sampling designs, such as simple random sampling without replacement at each stage. Although we did not investigate these methods, we do not expect our bootstrap method to have necessarily better (or worse) statistical properties. They are all valid methods that correctly track the first two design moments of estimators. The main appeal of our method is its simplicity and generality; it can be applied to most multistage sampling designs used in practice through a simple formula that combines the bootstrap weights at each distinct stage of sampling.

Our bootstrap method is also directly applicable to two-phase sampling designs when Poisson sampling is used at the second phase; it can thus handle nonresponse weighting. More research is required for general two-phase or multi-phase sampling designs.

**Author Contributions:** Methodology, J.-F.B.; software, N.É.; validation, J.-F.B. and N.É.; writing—original draft preparation, J.-F.B.; writing—review and editing, J.-F.B. and N.É. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data are not publicly available due to confidentiality restrictions.

**Acknowledgments:** We would like to thank Fulvia Mecatti for the kind invitation to contribute to this special issue and all three reviewers for their comments, which helped improve the clarity of our manuscript.

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

**Disclaimer:** The content of this paper represents the authors' opinions and not necessarily those of Statistics Canada.

## References

1. 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.
2. Funaoka, F.; Saigo, H.; Sitter, R.R.; Toida, T. Bernoulli bootstrap for stratified multistage sampling. *Surv. Methodol.* **2006**, *32*, 151–156.
3. Preston, J. Rescaled bootstrap for stratified multistage sampling. *Surv. Methodol.* **2009**, *35*, 227–234.
4. Saigo, H. Comparing four bootstrap methods for stratified three-stage sampling. *J. Off. Stat.* **2010**, *26*, 193–207.
5. Chaudhuri, A.; Saha, A. Extending Sitter's mirror-match bootstrap to cover Rao-Hartley-Cochran sampling in two-stages with simulated illustrations. *Sankhyā Indian J. Stat.* **2004**, *66*, 791–802.
6. Rao, J.N.K.; Hartley, H.O.; Cochran, W.G. On a simple procedure of unequal probability sampling without replacement. *J. R. Stat. Soc. Ser. B* **1962**, *24*, 482–491. [[CrossRef](#)]
7. Rao, J.N.K.; Wu, C.F.J. Resampling inference with complex survey data. *J. Am. Stat. Assoc.* **1988**, *83*, 231–241. [[CrossRef](#)]
8. Osiewicz, M.; Pérez-Duarte, S. Flexible variance estimation in complex sample surveys: Rescaled bootstrap in multistage, pps surveys. In Proceedings of the New Techniques and Technologies for Statistics, Brüssel, Belgium, 5–7 March 2013.
9. 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](#)]
10. Mashreghi, Z.; Haziza, D.; Léger, C. A survey of bootstrap methods in finite population sampling. *Stat. Surv.* **2016**, *10*, 1–52. [[CrossRef](#)]
11. Kim, J.K.; Yu, C.L. Replication variance estimation under two-phase sampling. *Surv. Methodol.* **2011**, *37*, 67–74.
12. Särndal, C.-E.; Swensson, B.; Wretman, J. *Model Assisted Survey Sampling*; Springer: New York, NY, USA, 1992.
13. Beaumont, J.-F.; Haziza, D. A note on the concept of invariance in two-phase sampling designs. *Surv. Methodol.* **2016**, *42*, 319–323.
14. Hansen, M.H.; Hurwitz, W.N. On the theory of sampling from finite populations. *Ann. Math. Stat.* **1943**, *14*, 333–362. [[CrossRef](#)]
15. Efron, B. Bootstrap methods: Another look at the jackknife. *Ann. Stat.* **1979**, *7*, 1–26. [[CrossRef](#)]
16. Ranalli, M.G.; Mecatti, F. Comparing recent approaches for bootstrapping sample survey data: A first step towards a unified approach. In *JSM Proceedings*; Survey Research Methods Section; American Statistical Association: Alexandria, VA, USA, 2012; pp. 4088–4099.
17. Conti, P.L.; Mecatti, F.; Nicolussi, F. Efficient unequal probability resampling from finite populations. *Comput. Stat. Data Anal.* **2022**, *167*, 107366. [[CrossRef](#)]
18. Chauvet, G. Coupling methods for multistage sampling. *Ann. Stat.* **2015**, *43*, 2484–2506. [[CrossRef](#)]
19. Kim, J.K.; Rao, J.N.K.; Wang, Z. Hypotheses testing from complex survey data using bootstrap weights: A Unified Approach. *arXiv* **2019**, arXiv:1902.08944.

- 
20. Shao, J.; Tu, D. *The Jackknife and Bootstrap*; Springer Science & Business Media: New York, NY, USA, 2012.
  21. Ohlsson, E. Sequential poisson sampling. *J. Off. Stat.* **1998**, *14*, 149–162.
  22. Conti, P.L.; Mecatti, F. Resampling under complex sampling designs: Roots, development and the way forward. *Stats* **2022**, *5*, 258–269. [[CrossRef](#)]