Université de Montréal

Méthodes de rééchantillonnage en méthodologie d'enquête

par

Zeinab Mashreghi

Département de mathématiques et de statistique

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de

Philosophiæ Doctor (Ph.D.)
en Statistique

octobre 2014

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée

Méthodes de rééchantillonnage en méthodologie d'enquête

présentée par

Zeinab Mashreghi

a été évaluée par un jury composé des personnes suivantes :

Pierre Duchesne
(président-rapporteur)

Christian Léger
(directeur de recherche)

David Haziza
(co-directeur)

Jean-François Angers
(membre du jury)

John Michael Brick
(examinateur externe)

Silvia Goncalves
(représentant du doyen de la FES)

Thèse acceptée le:
15 septembre 2014

SOMMAIRE

Le sujet principal de cette thèse porte sur l'étude de l'estimation de la variance d'une statistique basée sur des données d'enquête imputées via le bootstrap (ou la méthode de Cyrano). L'application d'une méthode bootstrap conçue pour des données d'enquête complètes (en absence de non-réponse) en présence de valeurs imputées et faire comme si celles-ci étaient de vraies observations peut conduire à une sous-estimation de la variance. Dans ce contexte, Shao et Sitter (1996) ont introduit une procédure bootstrap dans laquelle la variable étudiée et l'indicateur de réponse sont rééchantillonnés ensemble et les non-répondants bootstrap sont imputés de la même manière qu'est traité l'échantillon original. L'estimation bootstrap de la variance obtenue est valide lorsque la fraction de sondage est faible. Dans le chapitre 1, nous commençons par faire une revue des méthodes bootstrap existantes pour les données d'enquête (complètes et imputées) et les présentons dans un cadre unifié pour la première fois dans la littérature. Dans le chapitre 2, nous introduisons une nouvelle procédure bootstrap pour estimer la variance sous l'approche du modèle de non-réponse lorsque le mécanisme de non-réponse uniforme est présumé. En utilisant seulement les informations sur le taux de réponse, contrairement à Shao et Sitter (1996) qui nécessite l'indicateur de réponse individuelle, l'indicateur de réponse bootstrap est généré pour chaque échantillon bootstrap menant à un estimateur bootstrap de la variance valide même pour les fractions de sondage non-négligeables. Dans le chapitre 3, nous étudions les approches bootstrap par pseudo-population et nous considérons une classe plus générale de mécanismes de non-réponse. Nous développons deux procédures bootstrap par pseudo-population pour estimer la variance d'un estimateur imputé par rapport à l'approche du modèle de non-réponse et à celle du modèle d'imputation. Ces procédures sont également valides même pour des fractions de sondage non-négligeables.

Mots-clés: bootstrap, poids bootstrap, estimation doublement robuste, imputation, modèle d'imputation, non-réponse partielle, modèle de non-résponse, bootstrap par pseudo-population, estimation de la variance.

SUMMARY

The aim of this thesis is to study the bootstrap variance estimators of a statistic based on imputed survey data. Applying a bootstrap method designed for complete survey data (full response) in the presence of imputed values and treating them as true observations may lead to underestimation of the variance. In this context, Shao and Sitter (1996) introduced a bootstrap procedure in which the variable under study and the response status are bootstrapped together and bootstrap non-respondents are imputed using the imputation method applied on the original sample. The resulting bootstrap variance estimator is valid when the sampling fraction is small. In Chapter 1, we begin by doing a survey of the existing bootstrap methods for (complete and imputed) survey data and, for the first time in the literature, present them in a unified framework. In Chapter 2, we introduce a new bootstrap procedure to estimate the variance under the non-response model approach when the uniform non-response mechanism is assumed. Using only information about the response rate, unlike Shao and Sitter (1996) which requires the individual response status, the bootstrap response status is generated for each selected bootstrap sample leading to a valid bootstrap variance estimator even for non-negligible sampling fractions. In Chapter 3, we investigate pseudo-population bootstrap approaches and we consider a more general class of non-response mechanisms. We develop two pseudo-population bootstrap procedures to estimate the variance of an imputed estimator with respect to the non-response model and the imputation model approaches. These procedures are also valid even for non-negligible sampling fractions.

Keywords: Bootstrap, Bootstrap weights approach, Doubly robust estimation, Imputation, Imputation model approach, Item non-response, Non-response model approach, Pseudo-population bootstrap approach, Variance estimation.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENT

This research project would not have been possible without the support of many people. I wish to express my deepest gratitude and appreciation to my supervisor Professor Christian Léger and my co-supervisor Professor David Haziza for their support, invaluable assistance and great patience throughout these years.

I would also like to extend my appreciation to the members of the supervisory committee, Professor John Michael Brick, Professor Pierre Duchesne and Professor Jean-François Angers.

I would also like to convey thanks to the Faculty, especially my supervisor and co-supervisor, for providing the financial means for my studies. In addition, I must thank the Department for the privilege of serving as a lecturer and teaching assistant. I also thank *Le Fonds de recherche du Québec Nature et technologies (FQRNT)* for the honor of receiving the doctoral research scholarship, as well as *MITACS* and *Institut des sciences mathématiques (ISM)* for generous contributions.

Above all, my most heartfelt gratitude goes to my family: my parents for their unconditional love and support, my husband (Mostafa) and my brother (Javad) for their understanding, endless love and advice through the duration of my studies and my sister (Fatemeh) and my sister-in-law (Shahzad) who have been sources of joy and happiness. Were it not for their support and encouragement, this work would never have been completed. This thesis is dedicated to them.

INTRODUCTION

Official statistical agencies have long collected data of interest to governments to inform the development of public policy through surveys. Aggregate indicators were usually sought to describe the overall situation. Research in social and health sciences has demonstrated the need for more focussed surveys, as well as the need for information at various levels, such as families, neighborhoods, schools, etc. While statistical agencies understood this need many years ago by providing such complex surveys, confidentiality issues were such that only aggregated data were available to researchers outside the statistical agencies. Unfortunately, they need access to the micro-level data to assess the role of persons, or families, or schools on various social issues through adequate sophisticated modeling.

To fulfill this long-felt need, Research Data Centers were opened by statistical agencies, such as the Census Bureau, the National Center for Health Statistics, and Statistics Canada among others. In these centers, academic researchers become, in the case of Statistics Canada, "deemed employees" of the organization thereby allowing them access to detailed micro-level data while preserving confidentiality. The files provided to the researchers have a matrix form. Each row corresponds to an ultimate unit in the survey with columns corresponding to the different variables under study, plus other columns for survey weights. While the availability of complex survey data sets to social and health researchers will not cause them much difficulty to compute point estimates of various quantities, often through Horvitz-Thompson-type estimators, variance estimation for estimators other than the mean or total is more complicated.

The quality and the volume of literature published about variance estimation bear witness to the theoretical and practical interests that this issue produces.

All existing methods have been obtained so far through a linearization method or one of the resampling methods: the balanced repeated replication, the jackknife and the bootstrap methods.

The linearization method is available when the parameter of interest can be written as a differentiable function of totals. The variance estimate is based on a Taylor series expansion of the estimator. To apply this method, a separate formula is required for each nonlinear statistic. This is not very convenient nor easy to apply for researchers who are not familiar with the mathematical tools. In addition, the linearization method cannot be implemented when the parameter of interest is not a differentiable function of totals, such as the median. It is to overcome these difficulties that researchers have given a lot of attention to resampling methods.

In this thesis, we concentrate on the bootstrap. The bootstrap method was first proposed by Efron (1979) in the context of classical statistics, where data are independently and identically distributed (i.i.d.) from an unknown distribution. This method consists of first estimating the unknown distribution by the empirical distribution function and then generating the i.i.d. bootstrap samples from the estimated distribution. This is equivalent to taking simple random samples with replacement from the original sample. The bootstrap variance estimator can then be approximated by the Monte Carlo variance of the bootstrap statistics computed on the resulting bootstrap samples. However, in a sampling design context, the data are usually not i.i.d. Therefore, to have a valid variance estimator, the bootstrap procedure must be modified to reflect the variability under the survey design. This thesis is a compilation of three independent research papers about bootstrap methods for survey data in different contexts. Each of these papers is presented in a single chapter. In Chapter 1 of this thesis, all important existing bootstrap methods for survey data are studied in a survey of the field. Afterwards, in Chapters 2 and 3, some new bootstrap procedures are proposed for imputed survey data when the problem of item non-response arises.

Chapter 1 is based on the paper Mashreghi, Haziza, and Léger (2014b) entitled *A survey of bootstrap methods in finite population sampling*. There we discuss the existing bootstrap methods where, for the first time in the literature, these procedures are unified. This contribution will greatly help researchers compare the existing bootstrap methods and assess their advantages and disadvantages.

We classified the bootstrap methods for complete (full response) survey data into three main groups: the pseudo-population bootstrap, the direct bootstrap and the bootstrap weights methods.

In the pseudo-population bootstrap methods, a pseudo-population is created by repeating the elements of the original sample and bootstrap samples are selected from the original sampling scheme; see Gross (1980), Booth et al. (1994) and Chauvet (2007), among others. In fact, the nature of this group is similar to the case of classical statistics where the unknown distribution function is first estimated by the empirical distribution function and then an i.i.d. bootstrap sample is generated from the estimated distribution function. Here, the unknown is the population that is first estimated by constructing a pseudo-population. This pseudo-population is built by repeating the observations in the original sample using the original sampling design. Then, again using the original sampling design, the bootstrap sample is drawn from the resulting pseudo-population.

In the direct bootstrap group, bootstrap samples are obtained through i.i.d. resampling from the observations or vectors of observations from the original sample or a rescaled version of it; see Rao and Wu (1988), McCarthy and Snowden (1985) and Sitter (1992b). Such a with replacement sampling design is of course usually different from the original sampling design. However, to have a correct bootstrap estimator which will reflect the variability under the sampling design, some modifications have to be done either on the data set or the way bootstrap samples are taken.

In the third group, the bootstrap weights methods, a set of bootstrap survey weights are generated and applied to the original sample instead of generating bootstrap samples; see Rao et al. (1992) and Beaumont and Patak (2012), for instance. These bootstrap weights are the result of making adjustments on the original survey weights. In most cases, these adjustments are made so that the first two bootstrap moments match the sample moments of the distribution of the estimator in the case of the population total.

These methods are very easy for users of public data files prepared by agencies such as Statistics Canada. Very often, these users are not familiar with complex statistical methods. With these methods, using the resulting bootstrap weights with the original data set to compute many bootstrap estimators easily leads to a bootstrap variance estimator.

Unfortunately, life is rarely that simple and one of the important practical problems in statistical surveys is the presence of non-respondents in most data files. There are two types of non-response: complete non-response and item non-response. Complete non-response is not too difficult to handle and is usually dealt with by reweighting the respondents. But item non-response produces empty cells in the data files which is not easy to deal with particularly for researchers who are not familiar with complex statistical concepts. Item non-response is usually compensated using single imputation which fills the holes in the data set. A well-known fact is that treating the imputed values as if they were observed values may lead to serious underestimation of the variance of point estimators since bootstrap methods for complete survey data only account for the sampling variability in the observations, and not the added variability due to item non-response and imputation. These underestimations can be significant as we will illustrate in a study based on a real-life example in Section 2.7 that I have done as a MITACS trainee at Statistics Canada. Therefore, the bootstrap procedures have to be modified by taking into account the non-respondents and imputation method.

Working with item non-response, two inferential approaches can be used in order to assess the properties of point and variance estimators: the non-response model approach that requires explicit assumptions on the unknown non-response mechanism and the imputation model approach that requires the specification of a model describing the distribution of the variable under study in need of imputation. In Chapter 1, a broad study is also done on the existing bootstrap methods for this context. The most famous method is the one proposed by Shao and Sitter (1996). In this method, they utilize any direct bootstrap method to draw a bootstrap sample from the set of pairs made of the imputed data and the corresponding response status, followed by reimputation of the bootstrap sample of non-respondents using the same imputation method that was used on the original data. The estimator is computed based on the imputed bootstrap data and the process is repeated a large number of times, leading to a bootstrap variance estimate.

However, two problems may arise in the application of the Shao and Sitter (1996) method. The first one is the requirement of the presence of an imputation flag for each item under study. These indicators are usually not present in the files of research data centres. Therefore, the Shao and Sitter (1996) method is often unapplicable in practice, at least by researchers in research data centres. The second one is that their variance estimate is consistent only when the sampling fraction, the ratio of the sample size to the population size, is negligible. This result is proven in our second paper in Chapter 2 through a detailed analysis of their method using the reverse framework of Fay (1991) and Shao and Steel (1999). An example in Section 2.7 shows that this condition does not always hold in practice which in turn implies that the Shao and Sitter (1996) method may not work sometimes, even if the response status was available.

In Chapter 2, which is based on the paper Mashreghi, Léger, and Haziza (2014) entitled *Bootstrap methods for imputed data from regression, ratio and hot deck imputation*, published in The Canadian Journal of Statistics, the two drawbacks of the Shao and Sitter (1996) method are addressed by introducing a new bootstrap method, called the independent bootstrap method. Our theory is applicable to stratified simple random sample without replacement with uniform non-response in each stratum. Using the estimated response rate of the item under study in each stratum rather than the response status for each sample unit, our proposed bootstrap variance estimator is asymptotically consistent under the non-response model approach when the parameter of interest can be written as a function of means. The procedure is applied independently across strata. It consists of first selecting a bootstrap sample of observations using one of the direct bootstrap methods. Then, independently, bootstrap response indicators are regenerated mimicking the initial non-response mechanism. Unlike the Shao and Sitter (1996) method, the bootstrap sample of observations and the bootstrap response status are generated independently. This is why this method is called the independent bootstrap method. Since the sampling mechanism used in most direct bootstrap methods differs from simple (or stratified) random sampling, they all involve a constant which contains the sampling fraction and guarantees that when they are applied to the estimator of the total, they consistently estimate the variance of the estimator. These constants do not take into account the non-response mechanism and the method of imputation, so they need to be modified in the independent bootstrap, whereas Shao and Sitter (1996) use the original constants.

In Chapter 3, which is based on the paper Mashreghi, Haziza, and Léger (2014a) entitled *Pseudo-population bootstrap methods for imputed survey data*, two different bootstrap methods under the pseudo-population bootstrap approach are presented in order to estimate the variance of an imputed estimator under the non-response model and the imputation model approaches. In this paper, the class of doubly robust linear regression imputation is considered. These imputation methods, which are built using both the non-response and the imputation models, lead to doubly robust imputed estimators. That is, it remains asymptotically unbiased and consistent for the true parameter if either model (non-response or imputation) is true; e.g., Haziza and Rao (2006) and Kim and Haziza (2014).

Assuming the data are Missing At Random (MAR) (Rubin, 1976), the proposed pseudo-population bootstrap procedures are valid even for large sampling fractions unlike the Shao and Sitter (1996) procedure. The first bootstrap method is the non-response model approach that requires assumptions about the non-response mechanism and leads to an approximately unbiased variance estimator with respect to the non-response model approach. The second one is the imputation model approach that requires assumption about the distribution of the variable being imputed and leads to an approximately unbiased variance estimator with respect to the imputation model approach. In addition, combining the first two procedures, a doubly robust bootstrap variance estimator results. That is, the resulting bootstrap variance estimator is approximately unbiased for the true variance if one model or the other is correctly specified.

It should be noted that the first paper was written after the other two papers which is why the methods of Chapters 2 and 3 are surveyed in Chapter 1.

Chapter 1

A SURVEY OF BOOTSTRAP METHODS IN FINITE POPULATION SAMPLING

1.1. Introduction

Statistical agencies, such as the Census Bureau and Statistics Canada, provide researchers with access to detailed micro-level data while preserving confidentiality. Each table of data contains ultimate sample units in its rows and the different variables under study in its columns, plus other columns for survey weights. Parameters of interest can be easily estimated based on these values. However, a crucial step is to use the data to estimate some accuracy measures of a given statistic, such as the variance, something which is not always easy to obtain through analytical methods. For this purpose, many statistical agencies apply bootstrap resampling methods. Data files prepared by these agencies contain also a large number of columns for bootstrap survey weights. Each column of bootstrap survey weights with sample units is used to compute the bootstrap version of the given statistic. The Monte Carlo variance estimator of the resulting bootstrap statistics is used to estimate the variance under study. Since the bootstrap methods are readily applicable for many estimators, these methods are attractive from a practical point of view.

The bootstrap was first introduced by Efron (1979) in the context of classical statistics where data are independently and identically distributed (i.i.d.) from an unknown distribution. Since survey data are not necessarily i.i.d., many bootstrap resampling methods have been proposed in the context of survey sampling over the past thirty years. These methods are obtained after making some modifications on the classical i.i.d. bootstrap in order to adapt it for survey data.

A full study of the various bootstrap methods in the context of survey sampling has never been done in the literature. In this paper, we classify the methods in different groups according to their features and we present them in a unified way that shows the similarities and the differences among the methods in a given group. This comprehensive survey should be useful to researchers who need to use or better understand existing bootstrap methods in survey sampling. It provides sufficient details to help researchers apply the methods or develop new ones.

We classify the bootstrap methods for complete (full response) survey data in three groups. The first one is the class of the pseudo-population bootstrap methods in which a pseudo-population is first created by repeating the units of the original sample and bootstrap samples are then selected from the resulting pseudo-population, e.g. Gross (1980), Booth et al. (1994) and Chauvet (2007). The second one, called the direct bootstrap methods, consists of directly selecting bootstrap samples from the original sample or a rescaled version of it, e.g. Rao and Wu (1988) and Sitter (1992b). In the third group, called the bootstrap weights methods, an appropriate adjustment is made on the original survey weights to obtain a new set of weights called the bootstrap weights, e.g. Rao et al. (1992) and Beaumont and Patak (2012). Users of public data files prepared by agencies such as Statistics Canada, who are usually not familiar with complex statistical methods, can easily use the generated bootstrap weights. They only need to replace the original weights by the resulting bootstrap weights in the estimator of the parameter of interest to define the bootstrap statistics.

The paper is organized as follows. Basic concepts concerning sampling designs, parameter estimation, and estimation of its variance that will be used in the sequel are introduced in Section 1.2. The jackknife and the balanced repeated replication, which are resampling methods introduced before the bootstrap, are briefly discussed in Section 1.3. After introducing the i.i.d. bootstrap in Section 1.4, a detailed presentation of the three classes of bootstrap methods is the topic of Section 1.5. Note that the preceding methods are designed for finite population parameters where the population under study is treated as fixed. The bootstrap methods introduced in Section 1.6 are applicable when the study variables in the finite population are seen as a realization of a statistical model and the goal is to estimate the variance of the estimator of the parameter of that statistical model.

In practice, we often must be able to deal with imputed data which are used to compensate item non-response. Treating imputed data as true observations may lead to an underestimation of the variance. Therefore, some bootstrap methods that account for the added variability due to item non-response and imputation have been proposed and are studied in Section 1.7.

1.2. Preliminaries

Let U be a finite population consisting of N distinct units. Let $y_1, \ldots, y_J$ be J study variables and $\boldsymbol{y}_i = (y_{1i}, \ldots, y_{Ji})^\top$ denote the vector of study variables associated with the i-th unit, $i = 1, \ldots, N$. We are interested in estimating a finite population parameter, denoted by θ, which is a function of the N values, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$. A simple but important parameter, in the case where $J = 1$, is the population total of a study variable y defined as $\theta \equiv t = \sum_{i \in U} y_i$. Many parameters encountered in practice can be expressed as a function of population totals:

$$\theta = g(t_1, \ldots, t_J) \quad \text{with } t_j = \sum_{i \in U} y_{ji} \text{ for } j = 1, \ldots, J. \tag{1.2.1}$$

Special cases of (1.2.1) include the ratio of two population totals, $\theta = t_1/t_2$, and the finite population distribution function

$$F_N(z) = \frac{1}{N} \sum_{i \in U} I(y_i < z), \tag{1.2.2}$$

where $I(A)$ is the indicator function of the event A taking the value 1 when A occurs and 0 otherwise, and z is a real number. Note that $F_N(z)$ represents the proportion of units in the population with a y-value smaller than z. A parameter closely related to the distribution function is the finite population median, which is the value separating the higher half of data from the lower half. More formally, the population median m is defined as

$$m = F_N^{-1}(0.5),$$

where $F_N^{-1}(\cdot)$, the inverse function of $F_N(\cdot)$, is defined as

$$F_N^{-1}(b) = \inf\{y_i |\ F_N(y_i) \geq b;\ i \in U\} \tag{1.2.3}$$

with $0 \leq b \leq 1$.

A sample $s \subseteq U$ of (expected) size n, is randomly selected according to a given sampling design $p(s)$ with first-order inclusion probabilities $\pi_i = Prob(i \in s)$. Common sampling designs include simple random sampling without replacement and stratified simple random sampling, which are both fixed size sampling designs. Fixed size sampling designs are those for which the sample size is fixed prior to sampling. While simple random sampling without replacement is seldom used in practice, stratified simple random sampling is widely applied, especially in business surveys. Under this design, the population U is first divided into L non-overlapping strata $U_1, \ldots, U_L$ with N_h units in the h-th stratum, $h = 1, \ldots, L$. Then, a sample s_h of size n_h is selected from U_h according to simple random sampling without replacement, independently across strata. The first-order inclusion probability of unit i in stratum h is n_h/N_h, $h = 1, \ldots, L$. Except in the case of proportional allocation, stratified simple random sampling is an example of an unequal probability sampling design as units in different strata have different inclusion probabilities. Another unequal probability sampling design is Poisson sampling, which consists of performing N independent *Bernoulli* trials with probability π_i for unit i and selecting a unit in the sample when the trial is a “success”. Unlike simple random sampling without replacement and stratified simple random sampling, Poisson sampling is a random size sampling design.

Estimators of finite population parameters are constructed on the basis of the sample values and, possibly, auxiliary information, which is a set of variables collected for the sample units and for which the corresponding total in the population is known. We start by examining the case of a population total t and consider a general linear estimator of the form

$$\hat{t} = \sum_{i \in s} w_i(s) y_i, \tag{1.2.4}$$

where $w_i(s)$ is a survey weight associated with the i-th unit. The Horvitz-Thompson estimator $\hat{t}_{HT}$ (Horvitz and Thompson, 1952), is an important special case of (1.2.4) with

$$w_i(s) = w_i = \pi_i^{-1}. \tag{1.2.5}$$

Suppose that a l-vector of auxiliary variables $\boldsymbol{x}_i = (x_{1i}, \dots, x_{li})^\top$ is available for all the sample units and that the vector of population totals, $t_{\boldsymbol{x}} = \sum_{i \in U} \boldsymbol{x}_i$, is known. Another linear estimator of t is the so-called Generalized REGression (GREG) estimator, $\hat{t}_G$, given by (1.2.4) with

$$w_i(s) = \pi_i^{-1} \left\{ 1 + (t_{\boldsymbol{x}} - \hat{t}_{\boldsymbol{x}HT})^\top \hat{\boldsymbol{T}}^{-1} c_i^{-1} \boldsymbol{x}_i \right\}, \tag{1.2.6}$$

where $\hat{t}_{\boldsymbol{x}HT} = \sum_{i \in s} \pi_i^{-1} \boldsymbol{x}_i$, $\hat{\boldsymbol{T}} = \sum_{i \in s} \pi_i^{-1} \boldsymbol{x}_i c_i^{-1} \boldsymbol{x}_i^\top$ and c_i is a known positive constant attached to unit i. Note that the GREG estimator can also be viewed as a function of estimated totals since it can be expressed as

$$\hat{t}_G = \hat{t}_{HT} + \left(t_{\boldsymbol{x}} - \hat{t}_{\boldsymbol{x}HT} \right)^\top \hat{\boldsymbol{\beta}}, \tag{1.2.7}$$

where

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i \in s} \pi_i^{-1} \boldsymbol{x}_i c_i^{-1} \boldsymbol{x}_i^\top \right)^{-1} \sum_{i \in s} \pi_i^{-1} \boldsymbol{x}_i c_i^{-1} y_i.$$

We now turn to the case of parameters that can be expressed as functions of totals, $\theta = g(t_1, \dots, t_J)$. In this case, we use the plug-in principle that consists of replacing each unknown population total by its corresponding estimator; see Cassel et al. (1976). This leads to the so-called plug-in estimator

$$\hat{\theta} = g(\hat{t}_1, \dots, \hat{t}_J),$$

where $\hat{t}_j = \sum_{i \in s} w_i(s) y_{ji}$ is a linear estimator of t_j; e.g., the Horvitz-Thompson estimator, for $j = 1, \dots, J$. For example, the ratio of two totals $\theta = t_1/t_2$ may be estimated by $\hat{\theta} = \hat{t}_{1HT}/\hat{t}_{2HT}$.

Similarly, an estimator of the distribution function (1.2.2) is given by

$$\tilde{F}_n(z) = \frac{1}{\sum_{i \in s} w_i(s)} \sum_{i \in s} w_i(s) I(y_i < z)$$

noting that the population size N in the definition of $F_N(t)$ can be expressed as $N = \sum_{i \in U} 1$. It follows that an estimator of the population median, m, is given by

$$\hat{m} = \tilde{F}_n^{-1}(0.5),$$

where $\tilde{F}_n^{-1}(\cdot)$, the inverse function of $\tilde{F}_n(\cdot)$, is defined as in (1.2.3).

The above discussion suggests that an estimator of a finite population parameter θ can be viewed as a function of the sample units in s and the survey weights; i.e., $\hat{\theta} = \hat{\theta}\left(s; w_1(s), \ldots, w_n(s)\right)$. This will prove useful when studying the bootstrap weights methods described in Section 1.5.3.

In this paper, the properties of estimators (e.g., bias and variance) are studied with respect to the design-based approach. In this approach, the population U is held fixed and the properties of estimators are evaluated with respect to repeated sampling.

The expectation and the variance with respect to the design-based approach are defined as

$$E_p\left(\hat{\theta}\right) = \sum_{s \subset U} \hat{\theta}(s) p(s) \quad \text{and} \quad V_p\left(\hat{\theta}\right) = E_p\left\{\left[\hat{\theta} - E_p\left(\hat{\theta}\right)\right]^2\right\},$$

where the subscript p denotes the sampling design. An estimator is design-unbiased if $E_p\left(\hat{\theta}\right) = \theta$. While the Horvitz-Thompson estimator, $\hat{t}_{HT}$, is design-unbiased for t, the GREG estimator, $\hat{t}_G$, is only asymptotically design-unbiased for t; see, e.g., Isaki and Fuller (1982).

We now turn to the variance of point estimators and variance estimation. We start by examining the case of the Horvitz-Thompson estimator. The design-variance of $\hat{t}_{HT}$ is given by

$$V_p\left(\hat{t}_{HT}\right) = \sum_{i \in U} \sum_{j \in U} \Delta_{ij} y_i y_j, \tag{1.2.8}$$

where

$$\Delta_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}$$

with $\pi_{ij} = Prob(i \in s \ \& \ j \in s)$ denoting the second-order inclusion probability of units i and j in the sample. The variance (1.2.8) can be estimated unbiasedly by

$$\hat{V}\left(\hat{t}_{HT}\right) = \sum_{i\in s}\sum_{j\in s}\frac{\Delta_{ij}}{\pi_{ij}}y_i y_j. \tag{1.2.9}$$

That is, $E_p\left\{\hat{V}\left(\hat{t}_{HT}\right)\right\} = V_p\left(\hat{t}_{HT}\right)$. For example, under simple random sampling without replacement, (1.2.9) reduces to the textbook variance estimator of $\hat{t}_{HT}$:

$$\hat{V}\left(\hat{t}_{HT}\right) = N^2(1-f)\frac{s^2}{n}, \tag{1.2.10}$$

where $f = n/N$ is the sampling fraction and

$$s^2 = \frac{1}{n-1}\sum_{i\in s}(y_i - \bar{y})^2$$

with $\bar{y} = n^{-1}\sum_{i\in s} y_i$. For Poisson sampling, noting that $\pi_{ij} = \pi_i \pi_j$ for $i \neq j$, (1.2.9) reduces to

$$\hat{V}\left(\hat{t}_{HT}\right) = \sum_{i\in s}\frac{1-\pi_i}{\pi_i^2}y_i^2. \tag{1.2.11}$$

In contrast, the variance of the GREG estimator is virtually untractable, the latter being a complex function of estimated totals. The same is true for parameters that are expressed as functions of totals such as the ratio of two population totals. To overcome this difficulty, we settle for an approximate expression of the design-variance, which is obtained through a first-order Taylor expansion. Suppose that $\hat{\theta}$ is expressed as a function of estimated totals, $\hat{\theta} = g(\hat{t}_{1HT}, \ldots, \hat{t}_{JHT})$, where $g(\cdot)$ is a differentiable function. Under mild regularity conditions, a first-order Taylor expansion of $\hat{\theta}$ leads to

$$\hat{\theta} - \theta = \sum_{i\in s}\pi_i^{-1}z_i - \sum_{i\in U} z_i + O_p\left(n^{-1}\right), \tag{1.2.12}$$

where

$$z_i = \sum_{j=1}^{J} y_{ji}\left.\frac{\partial g(\hat{t}_{1HT}, \ldots, \hat{t}_{JHT})}{\partial \hat{t}_{jHT}}\right|_{\hat{t}_{1HT}=t_1,\ldots,\hat{t}_{JHT}=t_J} \tag{1.2.13}$$

is the so-called linearized variable. For instance, in the case of a ratio, $\theta = t_1/t_2$, the linearized variable is $z_i = (y_{1i} - \theta y_{2i})/t_2$. Ignoring the higher-order terms in (1.2.12), the design-variance of $\hat{\theta}$ can be approximated by (1.2.8), where y_i is replaced with z_i. That is, the approximate variance of $\hat{\theta}$ is given by

$$AV_p\left(\hat{\theta}\right) = \sum_{i\in U}\sum_{j\in U}\Delta_{ij} z_i z_j. \tag{1.2.14}$$

As mentioned above, the GREG estimator, $\hat{t}_G$, can also be viewed as a function of estimated totals. In this case, the linearized variable (1.2.13) reduces to

$$z_i = y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_U \tag{1.2.15}$$

with

$$\boldsymbol{\beta}_U = \left(\sum_{i \in U} \boldsymbol{x}_i c_i^{-1} \boldsymbol{x}_i^\top \right)^{-1} \sum_{i \in U} \boldsymbol{x}_i c_i^{-1} y_i.$$

The approximate variance of $\hat{t}_G$ is thus given by (1.2.14) with z_i given by (1.2.15). The approximate variance (1.2.14) is unknown as the linearized variable z depends on unknown quantities. To estimate (1.2.14), we start by estimating z by $\hat{z}$. For example, in the case of an estimated ratio, $\hat{\theta} = \hat{t}_{1HT}/\hat{t}_{2HT}$, we have $\hat{z}_i = (y_{1i} - \hat{\theta} y_{2i})/\hat{t}_{2HT}$. An estimator of the approximate variance is obtained from (1.2.9) by replacing y_i with $\hat{z}_i$, which leads to

$$\hat{V}\left(\hat{\theta}\right) = \sum_{i \in s} \sum_{j \in s} \frac{\Delta_{ij}}{\pi_{ij}} \hat{z}_i \hat{z}_j. \tag{1.2.16}$$

Under mild regularity conditions (e.g., Deville, 1999), the variance estimator (1.2.16) is asymptotically unbiased for the approximate variance (1.2.14). Both variance estimators (1.2.9) and (1.2.16) depend on the second-order inclusion probabilities π_{ij}, which may be difficult to obtain for some unequal probability sampling designs. Moreover, the variance estimator (1.2.16) obtained through a first-order Taylor expansion requires separate derivations for different functions of estimated totals in order to obtain $\hat{z}$. In this context, resampling methods may prove useful. Commonly used resampling methods include the jackknife, the balanced repeated replication and the bootstrap.

1.3. Some resampling methods in survey sampling

In this section, we briefly discuss the jackknife and the balanced repeated replication methods. The bootstrap will be discussed in detail in Sections 1.4-1.7.

Balanced repeated replication (BRR) was first introduced in McCarthy (1969) for the specialized case of stratified simple random sampling with replacement, where two units are selected independently in each stratum. A set of R half-samples is formed by deleting one unit from the sample in each stratum in a balanced fashion: consider an $R \times L$ selection matrix $\boldsymbol{S}$ with $S_{rh} = \pm 1$, where L is the number of strata, indicating whether the first $(+)$ or the second $(-)$ sample unit in the h-th stratum is in the r-th half sample. In order to be balanced, $\boldsymbol{S}$ must satisfy two conditions: $\boldsymbol{S}\mathbf{1} = 0$ and $\boldsymbol{S}^\top \boldsymbol{S} = L\boldsymbol{I}$ where $\mathbf{1} = (1, \ldots, 1)^\top$ and $\boldsymbol{I}$ is the $L \times L$ identity matrix. A minimal set of balanced half samples may be constructed from an $R \times R$ Hadamard matrix (Wolter, 2007), where $L + 1 \leq R \leq L + 4$ by choosing L columns excluding the column of $+1$'s. Let $\hat{\theta}_r$ be the estimator of θ computed on the r-th half sample after doubling the survey weights of the resampled units. A BRR variance estimator of $\hat{\theta}$ is

$$\hat{V}_{BRR} = \frac{1}{R} \sum_{r=1}^{R} \left(\hat{\theta}_r - \hat{\theta}_{(\cdot)} \right)^2 , \tag{1.3.1}$$

where $\hat{\theta}_{(\cdot)} = R^{-1} \sum_{r=1}^{R} \hat{\theta}_r$. Several variations of $\hat{V}_{BRR}$ are also available. For instance, $\hat{\theta}_{(\cdot)}$ can be replaced by $\hat{\theta}$ in (1.3.1). The BRR variance estimator (1.3.1) reduces to the usual variance estimator in the case of a population total. The asymptotic consistency of the BRR variance estimators, as $L \to \infty$, was established by Krewski and Rao (1981) when θ is a function of totals and by Shao and Wu (1992) for quantiles.

The BRR method can be applied to the case of stratified multistage designs with two primary selected units per stratum by treating each cluster as a unit. The case of more than two primary sampling units was studied by Gurney and Jewett (1975). They extended the BRR method to the case of $n_h = p$ primary sampling units, for $h = 1, \ldots, L$, where p is a prime number, but the number of replications R is much larger than that in the case of two primary sampling units.

In practice, the case of equal n_h is not common. To construct balanced half samples for unequal n_h, Gupta and Nigam (1987) and Wu (1991) used mixed-level orthogonal arrays to select one primary sampling unit per stratum, which implies that the resulting variance estimator is inconsistent. A correct variance estimator can be obtained by adjusting the original weights (Wu, 1991), depending on whether the associated units are selected in the half-sample or not. Alternative methods for constructing BRR can be found in Sitter (1993). Note that constructing the balanced samples is not an easy task, especially when the number of units per stratum is large. To overcome this difficulty, Rao and Shao (1996, 1999) suggest to randomly divide the first stage sampling units into two groups of size $m_h = \lfloor n_h/2 \rfloor$ ($\lfloor \cdot \rfloor$ denotes the greatest integer smaller than) and $n_h - m_h$, respectively, and construct the balanced half samples as in the case of two primary selected units. However, the survey weights need to be modified when computing $\hat{\theta}_r$. If the first stage sampling is done with replacement, according to whether the units are selected in the half samples or not, the survey weights are rescaled by

$$1 + \varepsilon\sqrt{\frac{n_h - m_h}{m_h}} \quad \text{or} \quad 1 - \varepsilon\sqrt{\frac{m_h}{n_h - m_h}}$$

with any fixed $\varepsilon \in (0,1)$; see Rao and Shao (1999). The resulting BRR variance estimator is given by

$$\hat{V}_{BRR}(\varepsilon) = \frac{1}{\varepsilon^2 R}\sum_{r=1}^{R}\left(\hat{\theta}_r - \hat{\theta}\right)^2,$$

which reduces to the usual variance estimator in the linear case.

In classical statistics, the jackknife method was first proposed by Quenouille (1956) in order to reduce the bias of point estimators. Later, in an i.i.d. set-up, Tukey (1958) suggested that the jackknife method could also be used to produce variance estimates. The first application of jackknife variance estimation in the context of finite population sampling can be found in Durbin (1959). Jones (1974) extended the method to handle stratified sampling. The jackknife variance estimator can be applied to estimate the variance of a function of totals $\hat{\theta}$. In the case of stratified simple random sample with replacement, this estimator is computed based on the jackknife estimator of θ obtained by recalculating the estimator after deleting one unit from the original sample and rescaling the survey weights of the remaining units.

$$\hat{V}_J = \sum_{h=1}^{L}\frac{n_h - 1}{n_h}\sum_{i=1}^{n_h}\left(\hat{\theta}_{hi} - \hat{\theta}_{h(\cdot)}\right)^2, \tag{1.3.2}$$

where $\hat{\theta}_{hi}$ is the jackknife estimator computed after deleting the i-th unit from stratum h and rescaling the survey weights of the stratum h by the factor $n_h/(n_h - 1)$, and $\hat{\theta}_{h(\cdot)} = n_h^{-1}\sum_{i=1}^{n_h}\hat{\theta}_{hi}$. There are other variations of (1.3.2) in the literature, for example, $\hat{\theta}_{h(\cdot)}$ can be replaced by $\hat{\theta}$. Note that when the sampling is done without replacement within strata, the finite population correction factor $1-f_h$ must be inserted in (1.3.2) in order to account for the effect of sampling without replacement, i.e.

$$\hat{V}_J = \sum_{h=1}^{L} \frac{(1-f_h)(n_h-1)}{n_h} \sum_{i=1}^{n_h} \left(\hat{\theta}_{hi} - \hat{\theta}_{h(\cdot)}\right)^2 .$$

Properties of these resampling methods have been studied by Krewski and Rao (1981), Rao and Wu (1985), Wolter (2007), Kovar et al. (1988), Rao et al. (1992) and Shao and Tu (1995), among others.

The jackknife method can be easily extended to the case of stratified multistage sampling design by considering sample clusters as sample units and computing $\hat{\theta}_{hi}$ after omitting the data from the i-th sampled cluster in the h-th stratum. When the number of clusters within strata is large, Kott (1998, 2001) studied a delete-a-group jackknife method that was first suggested by Rust (1985). Under this method, the first-stage sampling units are first ordered in an appropriate manner; see Kott (1998, 2001) for more details. Then, a set of systematic samples are formed from the partitioned sample. This way, the number of needed replications is kept manageable, which is important from a practical point of view. However, the survey weights need to be modified to account for the grouping.

Campbell and Little (1980) proposed a generalized jackknife variance estimator for unequal probability sampling without replacement design. Berger and Skinner (2005) established its consistency for a single stage design under a set of regularity conditions. The generalized jackknife variance estimator is

$$\hat{V}_J = \sum_{i\in s}\sum_{j\in s} \frac{\Delta_{ij}}{\pi_{ij}} e_{(i)} e_{(j)}, \tag{1.3.3}$$

where

$$e_{(i)} = (1-\pi_i)\left(\hat{\theta} - \hat{\theta}_{(i)}\right),$$

and $\hat{\theta}_{(i)}$ is the jackknife estimator computed after removing the i-th unit. The estimator (1.3.3) has the same form as the linearized variance estimator (1.2.16) but the linearized variable $\hat{z}_i$ is replaced by the numerical residual $e_{(i)}$. The factor $(1-\pi_i)$ can be viewed as the finite population correction for unequal probability sampling designs. When the second order inclusion probabilities π_{ij} are not available, Berger (2007) suggested an approximation of (1.3.3) based on Hájek's approximation of the π_{ij}'s; see Hájek (1964). For two-stage sampling designs, a generalized jackknife variance estimator was developed by Escobar and Berger (2013). Their methods consists of deleting clusters and observations within clusters. As a result, the resulting variance estimator accounts for the variability in all stages and is consistent even if the sampling fraction is not negligible.

When θ is not a function of totals such as the sample quantiles, the delete one jackknife fails to provide a consistent variance estimator; see Miller (1974) for a review on the application of the jackknife variance estimator. To overcome this difficulty, Shao and Wu (1989) considered a more general jackknife method, called delete-d jackknife. The number of deleted observations d depends on the "smoothness" of the point estimator. In particular, for the sample quantiles, the delete-d jackknife variance estimator with d satisfying $n^{1/2}d^{-1} \to 0$ and $n-d \to \infty$ is consistent and asymptotically unbiased in the case of i.i.d. observations.

1.4. Bootstrap for independently and identically distributed data

The bootstrap method was first proposed by Efron (1979) in classical statistics, where data are i.i.d. from a distribution F. We start by presenting the bootstrap method in this context as it is important to understand how to generalize it to more complex problems.

Let $Y_1, \cdots, Y_n$ denote the i.i.d. data set from the unknown F and let θ be a given parameter which is estimated by $\hat{\theta}$ based on $Y_1, \cdots, Y_n$. The bootstrap estimates the variance of $\hat{\theta}$, $V\left(\hat{\theta}\right)$, by first estimating the unknown F by the sample distribution function

$$\hat{F}_n(z) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \leq z),$$

where z is a real number. Then, we obtain the bootstrap variance by $V^* = V^*(\hat{\theta}^*|Y_1, \cdots, Y_n)$, where $\hat{\theta}^*$ is the bootstrap analogue of $\hat{\theta}$ computed on $Y_1^*, \cdots, Y_n^*$, an i.i.d. sample from $\hat{F}_n$, called a bootstrap sample, and $V^*(\cdot|Y_1, \cdots, Y_n)$ denotes the conditional variance given $Y_1, \cdots, Y_n$. However, this bootstrap variance estimator is usually not a closed form function of $Y_1, \cdots, Y_n$. In practice, we use a Monte Carlo approximation of V^*. The bootstrap algorithm can be depicted as follows:

(1) Generate $Y_1^*, \cdots, Y_n^* \overset{i.i.d.}{\sim} \hat{F}_n$, which is equivalent to drawing a simple random sample $\{Y_1^*, \cdots, Y_n^*\}$ with replacement from $\{Y_1, \cdots, Y_n\}$. Let $\hat{\theta}^*$ be the bootstrap statistic computed on the resulting bootstrap sample.

(2) Repeat Step 1 a large number of times, B, to get $\hat{\theta}_1^*, \cdots, \hat{\theta}_B^*$.

(3) Estimate $V\left(\hat{\theta}\right)$ with

$$\hat{V}_B^* = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}_b^* - \hat{\theta}_{(\cdot)}^*\right)^2,$$

where $\hat{\theta}_{(\cdot)}^* = B^{-1} \sum_{b=1}^{B} \hat{\theta}_b^*$.

Conditional on the original sample, when the number of bootstrap sample B goes to infinity, the law of large numbers implies that $\hat{V}_B^*$ converges almost surely to V^*, which is a function of the original sample.

A straightforward extension of the bootstrap to survey problems is to apply the above i.i.d. bootstrap algorithm to draw s^*, a simple random sample with replacement (SRSWR) of size n, from the original sample s. For $\hat{\theta} = \hat{t}_{HT}$, the bootstrap variance estimator reduces to

$$V^* = N^2 \left(\frac{n-1}{n}\right) \frac{s^2}{n}. \tag{1.4.1}$$

Even in the case of simple random sampling without replacement, the bootstrap method leads to a biased estimator of the variance as (1.4.1) fails to account for the finite population correction, $1-f$; see expression (1.2.10). As a result, the bootstrap variance estimator V^* does not reduce to zero in the case of a census, $s = U$, which is somehow embarrassing; see Lahiri (2003). Of course, in this simple situation, a bias-adjusted variance is easily obtained as $(1-f)[n/(n-1)]V^*$ is consistent and unbiased for the true variance. However, for more complex survey designs, the variance estimator (1.4.1) is biased and adjusting for the bias may be a complex task unlike in the case of simple random sampling without replacement.

Successful application of the bootstrap in a finite population setting requires appropriate modifications. One approach consists of modifying the bootstrap procedure by taking into account the survey design. Instead of estimating the unknown distribution F and selecting i.i.d. samples from the estimated distribution $\hat{F}_n$, it estimates the unknown finite population U and takes bootstrap samples according to the sampling design. These methods will be presented in Section 1.5.1. Alternatively, modifications will be applied to the data so that bootstrap i.i.d. sampling from the modified data will reflect the variability found under the sampling design. These methods will be presented in Section 1.5.2. In addition, in Section 1.5.3, some bootstrap weights methods will be presented in which modifications are made on the survey weights rather than on the original data set. Note that most of the proposed methods are designed to capture the standard variance estimator of the population total estimator given by (1.2.9).

1.5. Design-based bootstrap methods for complete survey data

In this section, we study the bootstrap methods proposed so far for complete survey data. These methods can be classified into three main groups. In the first, a pseudo-population is first created by repeating the elements of the original sample, and bootstrap samples are then selected from the resulting pseudo-population mimicking the original sampling scheme (called pseudo-population bootstrap methods). The second one consists of selecting bootstrap samples from the original or a rescaled sample applying a with replacement sampling design that might be different from the original sampling design (called direct bootstrap methods). In the third group (called bootstrap weights methods), instead of generating a bootstrap sample by working on the original data set, as in the two first groups, a set of bootstrap survey weights is generated by making rescaling adjustments on the original survey weights. The resulting bootstrap weights with the original data set are used to compute bootstrap estimators.

It is important to note that most of these methods are constructed so that the resulting bootstrap expectation and variance in the case of the Horvitz-Thompson estimator of the total asymptotically coincide with the estimate $\hat{t}_{HT}$, and the usual variance estimator presented in (1.2.9), respectively.

1.5.1. Pseudo-population bootstrap methods

As seen in Section 1.4, in classical statistics the unknown is the distribution F. To perform the bootstrap procedure, F is first estimated by the empirical distribution function, and then the resampling method proceeds. Working with survey data, the unknown is the population U from which the sample was drawn. Therefore, under the pseudo-population bootstrap (PPB) approach, U is estimated by creating a pseudo-population via repeating the original sample using principles from the original sampling design. Then, the bootstrap sample is drawn from the resulting pseudo-population using the original sampling design. By obeying the original scheme to draw the bootstrap sample from the pseudo-population, the finite population correction factors, e.g., the $1 - f$ in the case of simple random sample without replacement (SRSWOR), are naturally captured by the bootstrap variance estimator. This important property has persuaded many researchers to widely study this approach.

The pseudo-population bootstrap methods for simple random sample without replacement (or stratified simple random sample) and that for unequal probability sampling designs are presented in the two following sections.

Pseudo-population bootstrap methods for simple random sampling without replacement

In this section, we discuss the proposed pseudo-population methods for the case of simple random sample without replacement: Booth et al. (1994) and Chao and Lo (1994) on the one hand, and Bickel and Freedman (1984), Chao and Lo (1985) and Sitter (1992a), on the other hand. To clarify the application of these bootstrap methods, we illustrate how a pseudo-population is constructed through a simple example. Assume that $N = 1000$ and a simple random sample s of size $n = 100$ is taken without replacement from U. A pseudo-population of size N

can be created by repeating the sample s, $N/n = 10$ times. This method was first proposed by Gross (1980). However, in reality, N/n is rarely an integer. In this case, a well-known method to build a pseudo-population of size N was proposed by Booth et al. (1994). In this method, they create a pseudo-population, U^*, by first repeating each unit of the original sample s, $k = \lfloor N/n \rfloor$ times. Then, U^* is completed by taking a simple random sample of size $N - nk$ without replacement from s. For example, assuming that $N = 1000$ and $n = 150$, to construct U^*, each unit in s is first repeated $k = \lfloor 1000/150 \rfloor = 6$ times. Then, U^* is completed by taking a simple random sample of size $N - nk = 100$ without replacement. Note that if N/n is an integer, the pseudo-population U^* created under the method of Booth et al. (1994) is exactly the same as that under the method of Gross (1980).

To construct the pseudo-population, all other pseudo-population methods work similarly to the Booth et al. (1994) method, but different designs are used to complete the pseudo-population. The following algorithm presents a general scheme of all existing methods in order to create the pseudo-population and to select the bootstrap sample. Elements in ***bold*** in the algorithm need to be specified for each method.

SRSWOR PPB Algorithm:

(1) Repeat each unit in the original sample s, $\boldsymbol{k}$ times to create, U^f, the fixed part of the pseudo-population.

(2) Draw $\boldsymbol{U^{c*}}$ from s to complete the pseudo-population, U^*. Therefore, $U^* = U^f \cup U^{c*}$.

(3) Take a simple random sample, s^*, of size $\boldsymbol{n'}$ without replacement from U^*.

(4) Compute the bootstrap statistic, $\hat{\theta}^*$, on the bootstrap sample s^*.

In Table 1.1, the number of repetitions k, the design to obtain U^{c*} and the bootstrap sample size n' are presented for all procedures.

Note that when N/n is not an integer, for the methods of Booth et al. (1994) and Chao and Lo (1994), the size of the pseudo-population is fixed at N, the original population size, but its (conditional) mean varies with each pseudo-population. On the other hand, for the methods of Bickel and Freedman (1984),

Table 1.1. Existing complete data PPB methods for the case of SRSWOR

Existing methods	k	U^{c*}	n'
Booth et al. (1994)	$\lfloor N/n \rfloor$	SRSWOR from s of size $N-nk$	n
Chao and Lo (1994)		SRSWR from s of size $N-nk$	
Bickel and Freedman (1984)		$\dagger \begin{cases} \emptyset, & \text{with } {q_{bf}}^{a}, \\ s, & \text{with } 1-q_{bf}. \end{cases}$	
Chao and Lo (1985)		Same as † with ${q_{cl}}^{b}$	
Sitter (1992a)	$\left\lfloor \frac{N}{n}\left(1-\frac{1-f}{n}\right)\right\rfloor$	Same as † with ${q_s}^{c}$	$n - I(U^{c*}=\emptyset)$

[a] $q_{bf} = \left(1-\frac{N-nk}{n}\right)\left(1-\frac{N-nk}{N-1}\right)$
[b] $q_{cl} = \frac{G(N)-G(n(k+1))}{G(nk)-G(n(k+1))}$ and $G(t) = \left(1-\frac{n}{t}\right)\frac{t(n-1)}{(t-1)n}$
[c] $q_s = \frac{\frac{1-f}{n(n-1)}-a_2}{a_1-a_2}$ with $a_1 = \frac{nk-n+1}{n(n-1)(nk-1)}$ and $a_2 = \frac{k}{n[n(k+1)-1]}$

Chao and Lo (1985) and Sitter (1992a), there is a randomization between two different pseudo-populations made up of either k or $k+1$ copies of the sample s so that in either case, the (conditional) mean of the pseudo-population is the mean of the sample.

In the SRSWOR PPB Algorithm, there are two random components in the bootstrap procedure: the sampling mechanism applied to complete the pseudo-population and the one to choose the bootstrap sample, indexed by $u*$ and $p*$, respectively. Considering both elements of randomness, the total bootstrap variance is

$$V^*\left(\hat{\theta}^*\right) = E_{u*}\left[V_{p*}\left(\hat{\theta}^*|U^*\right)\right] + V_{u*}\left[E_{p*}\left(\hat{\theta}^*|U^*\right)\right] = V_1^*\left(\hat{\theta}^*\right) + V_2^*\left(\hat{\theta}^*\right), \quad (1.5.1)$$

the first term representing the average, over the different pseudo-populations, of the sampling variability of the bootstrap estimator $\hat{\theta}^*$, whereas the second is the variability, over the different pseudo-populations, of the sampling mean of $\hat{\theta}^*$. As discussed above, in the case of the estimator of the mean, $\hat{\theta}^* = \bar{y}^* = n'^{-1}\sum_{i\in s^*} y_i^*$, for the methods of Bickel and Freedman (1984), Chao and Lo (1985) and Sitter (1992a), $E_{p*}\left(\hat{\theta}^*|U^*\right) = \bar{y}$ and so the second term in (1.5.1) is 0.

But this is not the case for the methods of Booth et al. (1994) and Chao and Lo (1994). When the goal is to estimate the sampling variance of the estimator, $V_p\left(\hat{\theta}\right)$, with these latter two bootstrap methods, two approaches are possible. The first one is to compute the total bootstrap variance of (1.5.1). The second one is to recognize that we are interested in estimating the sampling variance of the estimator and therefore the extra variability resulting from completing the pseudo-population so that it has the same size as the original population is viewed as a parasitic variance. Hence the (bootstrap) estimate of variance should be the first term in (1.5.1). We now look at these two possible bootstrap variance estimates in more detail.

So, the first bootstrap estimate of variance would be $V^*\left(\hat{\theta}^*\right)$, the total variance with respect to both random elements induced by creating U^{c*} and selecting s^*. This is what classical statisticians would naturally do. One might wonder about the extra randomness induced by the completion of the pseudo-population through U^{c*}, but there is an equivalent in classical statistics. Suppose that one estimates the unknown distribution F by $\hat{F}_n^{\kappa}$, a kernel distribution function estimate which gives a continuous estimate as opposed to the discrete empirical distribution function. For instance, if one uses a $N(0, \sigma_\kappa^2)$ kernel, resampling from $\hat{F}_n^{\kappa}$ is equivalent to adding independent $N(0, \sigma_\kappa^2)$ variables to each original observation, putting the resulting random variables in a hat and picking at random with replacement a sample of size n, generating new normal variables before picking each new bootstrap sample. Clearly, in this case a bootstrap estimate of variance would be based on the total variance with respect to both random elements. Returning to the survey sampling context, to make a Monte Carlo approximation to compute the total variance $V^*\left(\hat{\theta}^*\right)$, the following steps must be added to the SRSWOR PPB Algorithm.

5. Repeat Steps 2 to 4 a large number of times, B, to get $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$.

6. Estimate the variance of $\hat{\theta}$ by $V^*\left(\hat{\theta}^*\right)$ or by

$$\hat{V}_B^* = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_b^* - \hat{\theta}_{(\cdot)}^*\right)^2,$$

where $\hat{\theta}_{(\cdot)}^* = B^{-1}\sum_{b=1}^{B}\hat{\theta}_b^*$.

In the case of the population total, this bootstrap variance estimator for the method of Booth et al. (1994) is

$$\begin{aligned} V^*\left(\hat{t}_{HT}^*\right) =& E_{u*}\left[V_{p*}\left(\hat{t}_{HT}^*|U^*\right)\right] + V_{u*}\left[E_{p*}\left(\hat{t}_{HT}^*|U^*\right)\right] \\ =& \left[\frac{n-1}{n-f} - \frac{1-f\lfloor N/n\rfloor}{N-1}\left(1-\frac{N-n\lfloor N/n\rfloor}{n}\right)\right]N^2(1-f)\frac{s^2}{n} \\ &+ N\left(1-f\lfloor N/n\rfloor\right)\left(1-\frac{N-n\lfloor N/n\rfloor}{n}\right)s^2, \end{aligned} \tag{1.5.2}$$

where $\hat{t}_{HT}^* = (N/n)\sum_{i\in s^*} y_i^*$ is the bootstrap Horvitz-Thompson estimator of total computed on s^*. It is straightforward to see that the first term of the bootstrap variance estimator in (1.5.2) is asymptotically unbiased to estimate $V_p\left(\hat{t}_{HT}\right)$, i.e.

$$E_p\left\{E_{u*}\left[V_{p*}\left(\hat{t}_{HT}^*|U^*\right)\right]\right\} - V_p\left(\hat{t}_{HT}\right) = O\left(n^{-1}\right)V_p\left(\hat{t}_{HT}\right). \tag{1.5.3}$$

Moreover, the ratio of the expectation of each component to $V_p\left(\hat{t}_{HT}\right)$ is

$$\frac{E_p\left\{E_{u*}\left[V_{p*}\left(\hat{t}_{HT}^*|U^*\right)\right]\right\}}{V_p\left(\hat{t}_{HT}\right)} = O(1) \quad \text{and} \quad \frac{E_p\left\{V_{u*}\left[E_{p*}\left(\hat{t}_{HT}^*|U^*\right)\right]\right\}}{V_p\left(\hat{t}_{HT}\right)} = O(f).$$

As a result, the second term in (1.5.2) produces a bias and implies an overestimation of the variance. This bias can be ignored only when the sampling fraction f is negligible. Note that in the case of a negligible f, even the classical i.i.d. bootstrap method works well asymptotically, so there would be no need to consider more sophisticated resampling procedures.

It should be noted that Booth et al. (1994) were interested in constructing a confidence interval for a function of means and obtained asymptotic results for the distribution of the estimator, which is what is needed to study the confidence intervals. Even though they do provide an algorithm for the expected value of the bootstrap estimator, they are silent on estimating the variance of an estimator. In particular, we cannot infer from the paper whether they had in mind this first approach to estimate the variance or the second one which we now describe.

Survey samplers are much more interested in variance estimation than in confidence intervals, if only because of the emphasis on coefficient of variation as a measure of precision for estimators. Given that the interest is in estimating the *sampling* variability associated with simple random sampling, the extra variability associated with completing the pseudo-population is viewed as a parasitic variance. Hence the bootstrap estimate of variance is $E_{u*}\left[V_{p*}\left(\hat{t}^*_{HT}|U^*\right)\right]$. Incidentally, this is the point of view taken by Chauvet (2007). The following steps have to be added to the SRSWOR PPB Algorithm in order to get a Monte Carlo approximation of this bootstrap variance estimator.

5. Repeat Steps 3 and 4 a large number of times, B, to get $\hat{\theta}^*_1, \ldots, \hat{\theta}^*_B$. Let

$$\hat{V}^*_B = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}^*_b - \hat{\theta}^*_{(\cdot)}\right)^2,$$

where $\hat{\theta}^*_{(\cdot)} = B^{-1}\sum_{b=1}^{B}\hat{\theta}^*_b$.

6. Repeat Steps 2 to 5 a large number of times, D, to get $\hat{V}^*_{1B}, \ldots, \hat{V}^*_{DB}$.
7. Estimate the variance of $\hat{\theta}$ by $E_{u*}\left[V_{p*}\left(\hat{\theta}^*|U^*\right)\right]$ or by

$$\hat{V}^* = \frac{1}{D}\sum_{d=1}^{D}V^*_{dB}.$$

In the case of the population total, this bootstrap variance estimator for the method of Booth et al. (1994) becomes

$$E_{u*}\left[V_{p*}\left(\hat{t}^*_{HT}|U^*\right)\right] = \left[\frac{n-1}{n-f} - \frac{1-f\,\lfloor N/n\rfloor}{N-1}\left(1 - \frac{N-n\,\lfloor N/n\rfloor}{n}\right)\right]N^2(1-f)\frac{s^2}{n},$$

which is asymptotically unbiased to estimate $V_p\left(\hat{t}_{HT}\right)$ as it was shown in (1.5.3).

Like Booth et al. (1994), Chao and Lo (1994) attempt to create a pseudo-population of size N, the same as the original population size. However, Chao and Lo (1994) take a simple random sample *with replacement* to complete the pseudo-population. They construct their method through first principles, using ideas from the method of moments and maximum likelihood to show that in the case where N/n is an integer, the only natural thing to do is to repeat the original sample k times. When N/n is not an integer, they complete the pseudo-population with a simple random sample *with replacement* from the original sample, but while they argued why it should be completed by observations found in the sample, they do not argue why it should be by simple random sampling with replacement.

As with Booth et al. (1994), a bootstrap variance estimator could be obtained either using the total bootstrap variance or by taking only the first term of (1.5.1). In the case of the population total, we have

$$E_{u*}\left[V_{p*}\left(\hat{t}^*_{HT}|U^*\right)\right]=\left[\frac{n-1}{n-f}-\frac{1-f\lfloor N/n\rfloor}{N-1}\left(1-\frac{1}{n}\right)\right]N^2(1-f)\frac{s^2}{n}$$

and

$$V_{u*}\left[E_{p*}\left(\hat{t}^*_{HT}|U^*\right)\right]=N\left(1-f\lfloor N/n\rfloor\right)\left(1-\frac{1}{n}\right)s^2.$$

Since the first term is asymptotically unbiased for $V_p\left(\hat{t}_{HT}\right)$ and the second term cannot be ignored in the case of a non-negligible f, the first approach in which the bootstrap variance estimator is $E_{u*}\left[V_{p*}\left(\hat{t}^*_{HT}|U^*\right)\right]+V_{u*}\left[E_{p*}\left(\hat{t}^*_{HT}|U^*\right)\right]$ may lead to an overestimation of the variance $V_p\left(\hat{t}_{HT}\right)$. However, the second approach seems to be appropriate. It means that we should only consider the variability induced by selecting the bootstrap sample leading to the bootstrap variance estimator $E_{u*}\left[V_{p*}\left(\hat{t}^*_{HT}|U^*\right)\right]$.

We now return to the three other bootstrap procedures of Bickel and Freedman (1984), Chao and Lo (1985) and Sitter (1992a). As it was shown in Table 1.1, each bootstrap method uses a different randomization method to select the pseudo-population. In Bickel and Freedman (1984) and Chao and Lo (1985), the pseudo-population is constructed by randomly repeating the original sample $k=\lfloor N/n\rfloor$ or $\lfloor N/n\rfloor+1$ times. In Sitter (1992a) the number of repetitions k and the bootstrap sample size are different from those in the other methods. In this method, the randomization is done between two pairs of the number of repetitions k and the bootstrap sample size, i.e. between $(k,n-1)$ and $(k+1,n)$ where $k=\lfloor (N/n)\left[1-(1-f)/n\right]\rfloor$.

These three methods are designed to estimate the variance of a function of means. Writing the estimator $\hat{t}_{HT}$ of the population total as $\hat{t}_{HT}=N\bar{y}$, where N is the known population size, the bootstrap statistic is $\hat{t}^*_{HT}=N\bar{y}^*$ and for these three methods, the second term of the bootstrap variance in (1.5.1) is zero,

$$V_{u*}\left[E_{p*}\left(\hat{t}^*_{HT}|U^*\right)\right]=V_{u*}\left(N\bar{y}\right)=0.$$

Note that if the bootstrap statistic is defined using the usual Horvitz-Thompson estimator on a sample of size n' drawn from a pseudo-population of size N', i.e. $\hat{t}^*_{HT} = (N'/n)\sum_{i\in s^*} y^*_i = N'\bar{y}^*$, this result does not hold anymore. If this definition is used, $V_{u*}\left[E_{p*}\left(\hat{t}^*_{HT}|U^*\right)\right] = O\left(n^2\right)$ which is not negligible.

In Table 1.2, the ratio of the expectation of $V^*\left(\hat{t}^*_{HT}\right) = V^*_1\left(\hat{t}^*_{HT}\right)$ to $V_p\left(\hat{t}_{HT}\right)$ is presented for the last three methods.

TABLE 1.2. The ratio of $V^*\left(t^*_{HT}\right) = V^*_1\left(\hat{t}^*_{HT}\right)$ to $V_p\left(\hat{t}_{HT}\right)$ in the case of SRSWOR

Existing methods	$E_p\left\{E_{u*}\left[V_{p*}\left(\hat{t}^*_{HT}|U^*\right)\right]\right\}/V_p\left(\hat{t}_{HT}\right)$
Bickel and Freedman (1984)	$(n-1)/(n-f)$
Chao and Lo (1985)	$\left[q_{cl}\left(\frac{k-1}{nk-1}\right) + (1-q_{cl})\left(\frac{k}{n(k+1)-1}\right)\right]\frac{n-1}{1-f}$ [a]
Sitter (1992a)	1

[a] $k = \lfloor N/n \rfloor$, $q_{cl} = \frac{G(N)-G(n(k+1))}{G(nk)-G(n(k+1))}$ and $G(t) = \left(1-\frac{n}{t}\right)\frac{t(n-1)}{(t-1)n}$

There is quite a bit of confusion in the literature regarding the method of Bickel and Freedman (1984), especially the probability q_{bf} of using $\lfloor N/n \rfloor$ copies of the sample as the pseudo-population. Sitter (1992a) and Lahiri (2003) and others refer to McCarthy and Snowden (1985) who give an example where apparently $q_{bf} < 0$, which would make the procedure infeasible. But it is clear that the probability q_{bf} presented in Table 1.1 is always positive. The confusion comes from the fact that McCarthy and Snowden (1985) gave the example for the probability suggested in Bickel and Freedman (1983), which is an unpublished manuscript, rather than from Bickel and Freedman (1984). According to McCarthy and Snowden (1985), the suggested probability in Bickel and Freedman (1983) is

$$q'_{bf} = \frac{(1-f)/(n-1) - b_2}{b_1 - b_2},$$

where $b_1 = \frac{k-1}{nk-1}$ and $b_2 = \frac{k}{n(k+1)-1}$ with $k = \lfloor N/n \rfloor$. Using this probability to estimate $V_p\left(\bar{y}\right)$ leads to $E_{u*}\left[V_{p*}\left(\bar{y}^*|U^*\right)\right] = (1-f)s^2/n$, which is the usual variance estimator of the sample mean, and $V_{u*}\left[E_{p*}\left(\bar{y}^*|U^*\right)\right] = 0$. However, the probability

q'_{bf} can be negative in some cases as discussed in McCarthy and Snowden (1985), which is probably why the probability q_{bf} changed between the two versions. On the other hand, using probability q_{bf} leads to a biased estimator of variance as seen in Table 1.2.

To illustrate the accuracy of the five pseudo-population methods in estimating the variance of $\hat{t}_{HT}$, the ratio of the expectation of both terms of the bootstrap variance estimator, $E_p\left\{V_1^*\left(\hat{t}^*_{HT}\right)\right\}$ and $E_p\left\{V_2^*\left(\hat{t}^*_{HT}\right)\right\}$, to $V_p\left(\hat{t}_{HT}\right)$, which only depend on the population (N) and sample (n) sizes, are presented in Table 1.3. Four different scenarios made up of two population sizes $N_1 = 100$ and $N_2 = 10000$ with two sampling fractions $f_1 = 6\%$ and $f_2 = 60\%$ are considered.

Table 1.3. The ratio of the expectation of both components of the bootstrap variance estimator to $V_p\left(\hat{t}_{HT}\right)$ assuming $N_1 = 100$, $N_2 = 10000$, f_1=6% and f_2=60%.

PPB methods for SRSWOR	$E_p\left\{V_1^*\left(\hat{t}^*_{HT}\right)\right\}/V_p\left(\hat{t}_{HT}\right)$				$E_p\left\{V_2^*\left(\hat{t}^*_{HT}\right)\right\}/V_p\left(\hat{t}_{HT}\right)$			
	f_1=6%		f_2=60%		f_1=6%		f_2=60%	
	N_1	N_2	N_1	N_2	N_1	N_2	N_1	N_2
Booth et al. (1994)	0.842	0.998	0.992	1.0	0.001	0.001	0.2	0.2
Chao and Lo (1994)	0.841	0.998	0.989	1.0	0.002	0.003	0.59	0.6
Bickel and Freedman (1984)	0.842	0.998	0.993	1.0	0	0	0	0
Chao and Lo (1985)	0.842	0.998	0.993	1.0	0	0	0	0
Sitter (1992a)	1	1	1	1	0	0	0	0

In all methods, except Sitter (1992a) where $E_p\left\{V_1^*\left(\hat{t}^*_{HT}\right)\right\}/V_p\left(\hat{t}_{HT}\right)$ is exactly 1, this ratio is close to 1 confirming that the first term of bootstrap variance estimator is a good estimator of the variance $V_p\left(\hat{t}_{HT}\right)$. Only in the case of $N_1 = 100$ with $f_1 = 6\%$ are the ratios about 0.84. This is because the sample size in this scenario, $n = 6$, is very small and the results are much better when the sample size increases. In the case of $n = 6$, we can not improve the results even when the second term of bootstrap variance estimator is added to the first term.

In the case of Sitter (1992a), this ratio is exactly 1 because the probability q_s of Table 1.1 is constructed so that the bootstrap variance estimator is identical to the usual variance estimator in the case of the population mean (or total). The contribution of $V_2^*\left(\hat{t}_{HT}^*\right)$ to the total variance is significant in Booth et al. (1994) and Chao and Lo (1994) when the sampling fraction is large ($f_2 = 60\%$), as suggested by the theory above, while it is zero for the other methods as discussed earlier. We note that completing the pseudo-population using without replacement sampling as in Booth et al. (1994) leads to a much smaller bias than the with replacement sampling of Chao and Lo (1994). In both methods, the sum of the first and the second term of the bootstrap variance estimator implies an overestimation of the variance.

All methods for the case of simple random sample without replacement can be easily extended to stratified simple random sample without replacement by applying a resampling method independently within strata. In addition, the method of Sitter (1992a) was extended to more complicated sampling designs, such as two-stage cluster sampling and the Rao-Hartley-Cochran method for probability proportional to size sampling (Rao et al., 1962). These extensions give the usual variance estimates in the linear case. Later, in a similar manner, Saigo (2010) extended the Sitter (1992a) method for stratified three-stage sampling.

Chao and Lo (1994) also investigated the case of unequal probability sampling design. Again, they take the point of view of maximizing the likelihood of obtaining the original sample from the pseudo-population. Clearly, putting values in the pseudo-population which are not part of the original sample will lead to some samples with values different from the original sample. Therefore, the pseudo-population must only contain values from the original sample in order to maximize the likelihood that the bootstrap sample will be identical to the original sample. The case of two-stage sampling is also studied through an example.

Pseudo-population bootstrap methods for unequal probability sampling

We now study two procedures designed for unequal (single-stage) probability sampling designs (UEQPS). The methods of Chauvet (2007) and Holmberg (1998) try to follow the original sampling design as was the case with simple random sample without replacement. Letting π_i be the inclusion probability for the i-th unit in s, Chauvet (2007) for the case of Poisson sampling design and Holmberg (1998) for inclusion probability proportional to size sampling designs apply the following general algorithm to create the pseudo-population and to draw the bootstrap sample. The element in ***bold*** will be specified for each method.

UEQPS PPB Algorithm:

(1) Repeat the pair (y_i, π_i), $\left\lfloor \pi_i^{-1} \right\rfloor$ times for all i in s to create, U^f, the fixed part of the pseudo-population.

(2) To complete the pseudo-population, U^*, draw U^{c*} from $\{(y_i, \pi_i)\}_{i \in s}$ using Poisson sampling with inclusion probability $\pi_i^{-1} - \left\lfloor \pi_i^{-1} \right\rfloor$ for the i-th pair. Therefore, $U^* = U^f \ \cup \ U^{c*} = \{(\check{y}_i, \check{\pi}_i)\}_{i \in U^*}$.

(3) Take the bootstrap sample s^* from U^* using the same sampling design that led to s, but with inclusion probability $\boldsymbol{\pi}_i'$ for the i-th unit in U^*, as defined in the sequel.

We see that the way of constructing the pseudo-population is the same for both methods. However, to draw the bootstrap sample, the original sampling mechanism used to draw s from U is applied, but with inclusion probability π_i'. Note that π_i' may be different from the original inclusion probability. The sampling design and the inclusion probability π_i' in Step 3 are presented in the following for both methods.

Chauvet (2007) estimates the variance of the population total $V_p\left(\hat{t}_{HT}\right)$ for Poisson sampling design. To obtain the bootstrap variance estimator of Chauvet, Poisson sampling with the original inclusion probabilities $\pi_i' = \check{\pi}_i$ in Step 3 of the UEQPS PPB Algorithm is used and the following steps are added to complete the resampling procedure.

4. Compute the bootstrap statistic, $\hat{\theta}^*$, on the bootstrap sample s^*.

5. Repeat Steps 3 and 4 a large number of times, B, to get $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$. Let

   $$\hat{V}_B^* = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_b^* - \hat{\theta}_{(\cdot)}^*\right)^2,$$

   where $\hat{\theta}_{(\cdot)}^* = B^{-1}\sum_{b=1}^{B}\hat{\theta}_b^*$.
6. Repeat Steps 2 to 5 a large number of times, D, to get $\hat{V}_{1B}^*, \ldots, \hat{V}_{DB}^*$.
7. Estimate the variance of $\hat{\theta}$ by $E_{u*}\left[V_{p*}\left(\hat{\theta}^*|U^*\right)\right]$ or by

   $$\hat{V}^* = \frac{1}{D}\sum_{d=1}^{D} V_{dB}^*.$$

We see that Chauvet (2007) follows the same approach as with the second interpretation that we gave to Booth et al. (1994). Chauvet showed that under Poisson sampling, the bootstrap variance estimator, $E_{u*}\left[V_{p*}\left(\hat{\theta}^*|U^*\right)\right]$, reduces to the usual variance estimator of (1.2.11) in the case of the total estimator, as proven below.

$$\begin{aligned}
E_{u*}\left[V_{p*}\left(\hat{t}_{HT}^*|U^*\right)\right] &= E_{u*}\left[V_{p*}\left(\sum_{i\in s^*}\pi_i'^{-1}y_i^*|U^*\right)\right]\\
&= E_{u*}\left(\sum_{i\in U^*}\frac{1-\check{\pi}_i}{\check{\pi}_i}\check{y}_i^2\right)\\
&= \sum_{i\in s}\left\lfloor\pi_i^{-1}\right\rfloor\frac{1-\pi_i}{\pi_i}y_i^2 + E_{u*}\left(\sum_{i\in U^{c*}}\frac{1-\check{\pi}_i}{\check{\pi}_i}\check{y}_i^2\right)\\
&= \sum_{i\in s}\left\lfloor\pi_i^{-1}\right\rfloor\frac{1-\pi_i}{\pi_i}y_i^2 + \sum_{i\in s}\left(\pi_i^{-1} - \left\lfloor\pi_i^{-1}\right\rfloor\right)\frac{1-\pi_i}{\pi_i}y_i^2\\
&= \sum_{i\in s}\frac{1-\pi_i}{\pi_i^2}y_i^2.
\end{aligned}$$

Note that the resulting pseudo-population may not have the same size as the original population size, N. But, letting $\check{M}_i$ be the number of times unit i appears in U^*, we have $E_p E_{u*}\left(\sum_{i\in s}\check{M}_i\right) = N$.

If instead of using the random sample size Poisson design one uses the fixed size rejective sampling (or conditional Poisson sampling), Chauvet (2007) suggests using the same algorithm as before replacing Poisson sampling by rejective sampling to construct the pseudo-population and to generate the bootstrap sample. To show that the bootstrap estimate of variance works well in this case, he uses the Hájek approximation for the second order inclusion probability to derive an approximation to the variance of the Horvitz-Thompson estimator of the total and shows that $E_{u*}\left[V_{p*}\left(\hat{\theta}^*|U^*\right)\right]$ is asymptotically unbiased for $V_p\left(\hat{t}_{HT}\right)$. The Hájek approximation will be good for rejective sampling as it is a high-entropy design. We conjecture that the method of Chauvet (2007) will perform well for any sampling design belonging to the class of high entropy sampling designs, which includes the Rao-Sampford method (Rao, 1965; Sampford, 1967) and randomized proportional-to-size systematic sampling as special cases.

Note that applying the proposed method to rejective sampling where $\sum_{i\in U}\pi_i = n$, it is possible that the sum of the inclusion probabilities on U^* is not an integer, so the condition of exact fixed size may not be satisfied. When the original inclusion probabilities are proportional to size, the inclusion probabilities to select the bootstrap sample have to be recalculated on each resulting pseudo-population in the same way that the original inclusion probabilities were computed on U.

Chauvet (2007) also extended his pseudo-population procedure to the case of multistage sampling design.

Holmberg (1998) proposed his bootstrap method for inclusion probability proportional to size sampling designs, so the first order inclusion probability used in Step 3 of the UEQPS PPB Algorithm is $\pi_i' = n\check{\pi}_i/\sum_{j\in U^*}\check{\pi}_j$. Unlike Chauvet (2007), according to the theory done in Holmberg (1998), the total bootstrap variance estimator in (1.5.1) is captured under this method as in the second interpretation of Booth et al. (1994). Holmberg (1998) applied this procedure to Pareto sampling (Rosén, 1997), a special case of inclusion probability proportional to size sampling, which produces the smallest asymptotic variance for the population total estimator. He studied both terms in (1.5.1) for the case of the population total.

However, to compute the Monte Carlo variance estimator, he ignores the variability induced by creating the pseudo-population. In the case of Pareto sampling, the following steps must be added to the UEQPS PPB Algorithm to obtain his suggested Monte Carlo approximation of the bootstrap variance estimator.

4. Compute the bootstrap statistic, $\hat{\theta}^*$, on the bootstrap sample s^*.

5. Repeat Steps 3 and 4 a large number of times, B, to get $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$.
6. Estimate $V_p\left(\hat{\theta}\right)$ with

$$\hat{V}^* = \frac{n}{n-1}\hat{V}_B^* = \frac{n}{n-1}\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_b^* - \hat{\theta}_{(\cdot)}^*\right)^2,$$

where $\hat{\theta}_{(\cdot)}^* = B^{-1}\sum_{b=1}^{B}\hat{\theta}_b^*$.

As a result, since U_{c*} in Step 2 does not change, the pseudo-population is created once from which a large number of bootstrap samples are taken, so the second term in (1.5.1) is estimated by zero. It seems Holmberg believes that once the created pseudo-population is a good representative of the population, there is no need to create a new pseudo-population in each bootstrap iteration.

1.5.2. Direct bootstrap methods

The bootstrap methods in this category are based on the idea that the bootstrap samples can be directly drawn from the original data set as in Efron (1979) without requiring the creation of a pseudo-population and mimicking the original sampling design. However, some modifications have to be made in order to obtain correct bootstrap estimators which will reflect the appropriate sampling variability of the original sampling design. Some methods modify the observations while others concatenate independent smaller simple random samples without replacement. First, we focus on the procedures handling the case of simple random sampling without replacement.

The rescaling bootstrap (RSB) method proposed by Rao and Wu (1988) is one of the well-known bootstrap methods. In this procedure, a rescaling of the original data set is made before drawing the bootstrap sample leading to a valid estimator of the variance of $\hat{\theta} = g(\hat{t}_{1HT}, \ldots, \hat{t}_{JHT})$, a function of population totals such as a ratio, a correlation coefficient or the generalized regression estimator. Let n' be the bootstrap sample size and $y_i' = \bar{y} + C(y_i - \bar{y})$ be the rescaled y-value for unit i, where

$$C = \sqrt{\frac{n'(1-f)}{n-1}}. \tag{1.5.4}$$

The bootstrap sample, $s^* = \{y_i^*\}_{i=1}^{n'}$, of size n', is then taken *with replacement* from $s' = \{y_i'\}_{i=1}^{n}$ the set of rescaled data. Afterwards, the bootstrap statistic $\hat{\theta}^* =$

$g(\hat{t}^*_{1HT}, \ldots, \hat{t}^*_{JHT})$, where $\hat{t}^*_{jHT} = (N/n') \sum_{i \in s^*} y^*_{ji}$ for $j = 1, \ldots, J$, is computed. To illustrate how this bootstrap method performs for a function of totals, assume that the parameter of interest is the population variance which is a function of two totals:

$$\theta = N^{-1} \sum_{i \in U} y_i^2 - \left(N^{-1} \sum_{i \in U} y_i \right)^2 = N^{-1} t_1 - \left(N^{-1} t_2 \right)^2, \qquad (1.5.5)$$

with $(y_{1i}, y_{2i}) = (y_i^2, y_i)$. Therefore, the rescaled values of y_{1i} and y_{2i} are given by $(y'_{1i}, y'_{2i}) = (\bar{y}_1 + C\left(y_i^2 - \bar{y}_1\right), \bar{y}_2 + C(y_i - \bar{y}_2))$, where $\bar{y}_1 = n^{-1} \sum_{i \in s} y_i^2$ and $\bar{y}_2 = \bar{y}$. The bootstrap sample is now drawn from $\{(y'_{1i}, y'_{2i})\}_{i=1}^n$.

It is worth noting that s^* is drawn with replacement like in Efron (1979), but from a rescaled data set and with a size that may be different from n.

As shown below, the rescaling factor C is chosen so that the variance under resampling matches the usual variance estimator of the population total.

$$\begin{aligned} V_{p*}\left(\hat{t}^*_{HT}\right) &= V_{p*}\left(\frac{N}{n'} \sum_{i \in s^*} y_i^*\right) \\ &= \frac{N^2}{n'} \frac{1}{n} \sum_{i \in s} \left(y'_i - n^{-1} \sum_{j \in s} y'_j \right)^2 \\ &= \frac{N^2 C^2}{n' n} \sum_{i \in s} (y_i - \bar{y})^2 \\ &= N^2 (1-f) \frac{s^2}{n}. \end{aligned}$$

Rao and Wu (1988) showed that an improper choice of n' could lead to negative values of $\hat{\theta}^*$ even when $\hat{\theta} \geq 0$ and the parameter of interest is necessarily positive. For example, when the parameter of interest is the population variance given by (1.5.5), choosing $n' > (n-1)/(1-f)$ might lead to a negative value for $\hat{\theta}^*$. However, in this case, choosing $n' \leq (n-1)/(1-f)$, we have $\hat{\theta}^* \geq 0$.

When applying this method to estimate the variance of the GREG estimator given by (1.2.7), the auxiliary variables $\boldsymbol{x}$ also need to be rescaled the same way that the study variables are rescaled. The bootstrap samples are then selected from the rescaled version of the set of pairs $\{(y_i, \boldsymbol{x}_i)\}_{i \in s}$. The resulting bootstrap variance estimator is asymptotically unbiased for the linearization variance estimator given by (1.2.14). In addition, Kovar et al. (1988) applied the RSB method to the case of quantiles.

In the following, a general algorithm for the direct bootstrap methods is presented. In Table 1.4, the different notations used in this algorithm in ***bold*** are defined for each procedure. To put the various procedures in the same algorithm, we define three quantities. We let C be the rescaling factor of the observations. Also the method of Sitter (1992b), called the mirror-match bootstrap, involves the concatenation of k' simple random samples without replacement of size n''. For the methods involving a single i.i.d. sample of size n', we will use $n'' = 1$ and $k' = n'$. In other words, setting $n'' = 1$ in the algorithm described below is equivalent to selecting the bootstrap samples with replacement.

SRSWOR Direct Algorithm:

(1) Let $y_i' = \bar{y} + \boldsymbol{C}(y_i - \bar{y})$, for $i = 1, \cdots, n$. Define $s' = \{y_i'\}_{i=1}^{n}$.

(2) Take a simple random sample of size $\boldsymbol{n''}$ without replacement from s'.

(3) Repeat Step 2, $\boldsymbol{k'}$ times independently, concatenating all subsamples, to get $s^* = \{y_i^*\}_{i=1}^{n'}$, where $n' = k'n''$.

(4) Compute the bootstrap statistic, $\hat{\theta}^* = g(\hat{t}_{1HT}^*, \ldots, \hat{t}_{JHT}^*)$, where $\hat{t}_{jHT}^* = (N/n') \sum_{i \in s^*} y_{ji}^*$ for $j = 1, \ldots, J$.

(5) Repeat Steps 2 to 4 a large number of times, B, to get $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$.

(6) Estimate the variance of $\hat{\theta}$ by $V_{p*}\left(\hat{\theta}^*\right)$ or by $\hat{V}_B^* = (B-1)^{-1} \sum_{b=1}^{B} \left(\hat{\theta}_b^* - \hat{\theta}_{(\cdot)}^*\right)^2$, where $\hat{\theta}_{(\cdot)}^* = B^{-1} \sum_{b=1}^{B} \hat{\theta}_b^*$.

Table 1.4 shows that the i.i.d. bootstrap of Efron (1979) overestimates the variance because of failing to cover the without replacement correction factor. McCarthy and Snowden (1985) do the same as Efron (1979), but they recommended a new bootstrap sample size $n' = (n-1)/(1-f)$ to capture the finite population correction factor which yields the customary variance estimator of $\hat{t}$. If the recommended resample size $(n-1)/(1-f)$ is non-integer, the closest integer to this value is considered as n'.

Table 1.4. Existing complete data direct bootstrap methods for the case of SRSWOR

Existing methods	C	n''	k'	$\frac{E_p\left[V_{p*}\left(\hat{t}^*_{HT}\right)\right]}{V_p\left(\hat{t}_{HT}\right)}$
Efron (1979)	1	1	n	$\frac{n-1}{n(1-f)}$
McCarthy and Snowden (1985)	1	1	$\frac{n-1}{1-f}$ [a]	1
Rao and Wu (1988)	$\sqrt{\frac{k'(1-f)}{n-1}}$	1	Arbitrary [b]	1
Sitter (1992b)	1	$\leq \frac{n}{2-f}$	$\left\lfloor \frac{n(1-f'')}{n''(1-f)} \right\rfloor + I_q$ [c]	1

[a] It may be a non-integer. If so, $n' = \lfloor (n-1)/(1-f) + 0.5 \rfloor$.

[b] More conditions are required to have a positive $\hat{\theta}^*$ when $\hat{\theta}$ is necessarily positive.

[c] $I_q \sim Bernoulli(q)$ with $q = (\lfloor k \rfloor^{-1} - k^{-1})/(\lfloor k \rfloor^{-1} - \lceil k \rceil^{-1})$, $\lceil k \rceil = \lfloor k \rfloor + 1$, $k = n(1-f'')/[n''(1-f)]$ and $f'' = n''/n$

As mentioned above, the method of Sitter (1992b) consists of taking a resample without replacement, as in the original sampling scheme, but of size n'' smaller than the original sample size and then repeating this resampling independently $k = n(1-f'')/[n''(1-f)]$ times. The bootstrap sample is obtained by accumulating all these resamples. The number of repetitions k is chosen in such a way that the resulting bootstrap variance matches the usual variance estimate of the population total in (1.2.9), $V_{p*}\left(\hat{t}^*_{HT}\right) = \hat{V}\left(\hat{t}_{HT}\right)$. Since k is usually not an integer, a randomization between bracketing integers is available as shown in Table 1.4. Sitter (1992b) showed that this procedure remains valid for the case of a function of totals, but more study is required for more complex parameters such as a population quantile.

Sitter (1992b) also discussed an alternative choice of resample size with $n'' = fn$ such that the resampling fraction $f'' = n''/n$ is the same as the original sampling fraction f. However, this procedure is generally not feasible since both n'' and k are generally not integer values. In this case, two types of randomization between bracketing integers were suggested. In the first one, the bootstrap sample size $n'' = \lfloor fn \rfloor + I_{q'}$ is first fixed, where $I_{q'} \sim Bernoulli(q')$ with $q' = fn - \lfloor fn \rfloor$. Then, a randomization between the integer values of k is done, as presented in

Table 1.4, so that $E(f'') = f$ and $V_{p*}\left(\hat{t}^*_{HT}\right) = \hat{V}\left(\hat{t}_{HT}\right)$. Choosing n'' by this way may lead to $k < 1$. So, this randomization is not valid. In this case, another kind of randomization made between $(\lfloor fn \rfloor, \lfloor k \rfloor)$ and $(\lceil fn \rceil, \lceil k \rceil)$ is presented, where $\lceil \cdot \rceil$ denotes the smallest integer greater than.

All proposed methods can be easily extended to the case of stratified simple random sample without replacement by performing resampling independently within each stratum. Rao and Wu (1988) extended their method not only to stratified simple random sample with replacement, but also to two-stage cluster sampling without replacement and the Rao-Hartley-Cochran method. Different rescaling factors are used so that the resulting bootstrap variance estimators match the textbook variance estimator of the point estimator; see also Sitter (1992b) for an extension of the mirror-match method for these sampling designs. Saigo (2010) extended the Rao and Wu (1988) and Sitter (1992b) methods to stratified three-stage sampling. Drawing the bootstrap samples is of course performed in three stages independently across strata and the rescaling factors used at each stage for the rescaling bootstrap method as well as the number of replications needed at each stage in the mirror-match bootstrap are explicitly presented.

1.5.3. Bootstrap weights methods

As discussed in Section 1.2, an estimator of θ can be viewed as a function of the observations and the survey weights. Rao et al. (1992) developed the idea of creating bootstrap survey weights rather than drawing the bootstrap sample of observations to compute the bootstrap statistic. In the case of the sample mean, they noted that the bootstrap sample mean $\bar{y}^*$ of the RSB method of Rao and Wu (1988), the mean of the bootstrap observations y^*_i, is a weighted mean of the rescaled observations y'_i where the weights are the number of times that y'_i is in the bootstrap sample. But since y'_i is itself a weighted mean of the original observations y_i, $\bar{y}^*$ is therefore a weighted mean of the original observations y_i. To better understand this statement, let

$$I^*_{ji} = \begin{cases} 1, & \text{if } y^*_j = y'_i = \bar{y} + C(y_i - \bar{y}), \\ 0, & \text{otherwise,} \end{cases} \qquad j = 1, \ldots, n'; i = 1, \ldots, n.$$

As a result, $\sum_{j\in s^*} I^*_{ji}$ represents the number of times unit i in s is selected in the bootstrap sample under the RSB method. In the case of a population mean, the bootstrap estimator in Rao and Wu (1988) is $n'^{-1}\sum_{j\in s^*} y^*_i$. In the case of simple random sampling without replacement and applying the definition of I^*_{ji}, we have

$$\begin{aligned}\frac{1}{n'}\sum_{j\in s^*} y^*_j &= \frac{1}{n'}\sum_{j\in s^*}\sum_{i\in s} I^*_{ji} y'_i \\ &= \frac{1}{n'}\sum_{j\in s^*}\sum_{i\in s} I^*_{ji}[\bar{y}+C(y_i-\bar{y})] \\ &= \bar{y}+\frac{C}{n'}\sum_{i\in s} y_i \sum_{j\in s^*} I^*_{ji} - C\bar{y} \\ &= \frac{1}{N}\sum_{i\in s}\left[1+C\left(\frac{n\sum_{j\in s^*} I^*_{ji}}{n'}-1\right)\right] w_i y_i,\end{aligned}$$

where w_i is the weight of the observation, in this case N/n. Therefore, rather than selecting bootstrap observations, Rao et al. (1992) suggested to keep the original observations and create bootstrap weights. This method is attractive for users of public data files prepared by statistical agencies such as Statistics Canada. These agencies provide data sets consisting of columns with the original observations, a column with the original survey weights and B columns of bootstrap weights. As a result, the agencies do not need to provide certain details about the sampling design which may reveal enough information that could jeopardize confidentiality.

Bootstrap weights are of the general form

$$w^*_i = a^*_i w_i, \tag{1.5.6}$$

where a^*_i is computed based on the bootstrap sample. In Rao et al. (1992), the suggested bootstrap adjustments for the case of simple random sampling without replacement are

$$a^*_i = 1+\sqrt{\frac{n'(1-f)}{n-1}}\left(\frac{nm^*_i}{n'}-1\right),$$

where m^*_i is the number of times that the i-th element is appearing in the bootstrap sample of size n' selected with replacement from the original sample ($\sum_{i\in s} m^*_i = n'$). Therefore, according to the definition of the random variable $\sum_{j\in s^*} I^*_{ji}$ in the Rao and Wu (1988) method and that of m^*_i, it is clear that the number of times unit i in s is selected in the bootstrap sample has the same distribution in both Rao and Wu (1988) and Rao et al. (1992), i.e. $m_i^* \overset{D}{=} \sum_{j\in s^*} I_{ji}^*$ where $\overset{D}{=}$ indicates equality in distribution. Consequently, we have

$$a_i^* \overset{D}{=} 1 + C\left(\frac{n\sum_{j\in s^*} I_{ji}^*}{n'} - 1\right).$$

That is, both methods are equivalent for a function of means (or totals). Note that even if the i-th element is not selected, $m_i^* = 0$, the associated bootstrap survey weight is nonzero. This is because the rescaled observations y_i' are centered at $\bar{y}$ which involves all observations. If $w_i > 0$ for all $i \in s$ and n' is chosen to be less than or equal to $(n-1)/(1-f)$, then the bootstrap weights are all positive.

Rao et al. (1992) presented a similar method for the case of stratified multistage cluster sampling with replacement. An extension of this method for stratified three-stage sampling is considered in Saigo (2010).

Letting m_i^* be the number of times that the i-th element is appearing in a bootstrap sample selected according to a particular resampling design of size n', Table 1.5 displays the way of computing a_i^* in (1.5.6) for different bootstrap weights methods in the case of simple random sample without replacement: Rao et al. (1992), Chipperfield and Preston (2007), Beaumont and Patak (2012) and Antal and Tillé (2011a, 2014).

As shown in Table 1.5, the method of Chipperfield and Preston (2007) introduced a new set of bootstrap weights rescaled on the basis of the number of times that the original units are selected in a simple random sample of size $n' = \lfloor n/2 \rfloor$ drawn without replacement from s. So, unlike the method of Rao et al. (1992), bootstrap samples are drawn *without replacement*. As a result, $m_i^* = 0$ or 1. Chipperfield and Preston (2007) applied their method and the Rao et al. (1992) method to estimate the variance of GREG estimators. The bootstrap statistics are computed using the following GREG bootstrap weights:

$$w_i^* = a_i^* \pi_i^{-1}\left\{1 + (t_{\boldsymbol{x}} - \hat{t}_{\boldsymbol{x}}^*)^\top \hat{\boldsymbol{T}}^{*-1} c_i^{-1} \boldsymbol{x}_i\right\},$$

where $\hat{t}_{\boldsymbol{x}}^* = \sum_{i\in s} a_i^* \pi_i^{-1} \boldsymbol{x}_i$ and $\hat{\boldsymbol{T}}^* = \sum_{i\in s} a_i^* \pi_i^{-1} \boldsymbol{x}_i c_i^{-1} \boldsymbol{x}_i^\top$. Note that replacing a_i^* by 1 in the expression of w_i^* leads to the usual GREG weights given by (1.2.6). Both bootstrap variance estimators are asymptotically unbiased to estimate the linear approximation of the variance of total presented in (1.2.14). Based on

Table 1.5. Existing complete data bootstrap weights methods for SRSWOR

Existing methods	Resampling	n'	a_i^*
Rao et al. (1992)	SRSWR	Any [a]	$1+\sqrt{\frac{n'(1-f)}{n-1}}\left(\frac{nm_i^*}{n'}-1\right)$
Chipperfield and Preston (2007)	SRSWOR	$\lfloor n/2 \rfloor$	$1+\sqrt{\frac{\lfloor n/2 \rfloor(1-f)}{n-\lfloor n/2 \rfloor}}\left(\frac{nm_i^*}{\lfloor n/2 \rfloor}-1\right)$
Beaumont and Patak (2012)	–	–	Generate from a distribution with $E^*(\boldsymbol{a}^*)=\mathbf{1}$ and $V^*(\boldsymbol{a}^*-\mathbf{1})(\boldsymbol{a}^*-\mathbf{1})^\top=\boldsymbol{\Sigma}$ [b]
Antal and Tillé (2011a)	SRSWOR & one-one	n	m_i^*
Antal and Tillé (2014)	*Bernoulli* & one-one	n	m_i^*

[a] More conditions are required to have positive bootstrap weights.

[b] $\boldsymbol{a}^* = (a_1^*, \ldots, a_n^*)$ and $\boldsymbol{\Sigma} = (\Delta_{ij}\pi_i\pi_j/\pi_{ij})$ where $\Delta_{ij}\pi_i\pi_j/\pi_{ij} = -(1-f)/(n-1)$ if $i \neq j$ and $1-f$ if $i=j$.

empirical results, they showed that the Chipperfield and Preston (2007) method can be significantly more efficient than the bootstrap weights method of Rao et al. (1992) in terms of variance; see Preston and Chipperfield (2002). As the sample size n increases, Preston and Chipperfield (2002) showed empirically that the difference between both methods vanishes.

A closer look at the Rao et al. (1992) method for the case of SRSWOR reveals that the distribution of m_i^* is a $Multinomial(n', \frac{1}{n}, \ldots, \frac{1}{n})$, which implies that $E^*(a_i^*)=1$ and $E^*(a_i^*-1)(a_j^*-1)=\Delta_{ij}\pi_i\pi_j/\pi_{ij}$ with $\Delta_{ij}\pi_i\pi_j/\pi_{ij}=1-f$ if $i=j$, and $-(1-f)/(n-1)$ otherwise. Therefore, the bootstrap adjustments a_i^* are constructed so that the bootstrap expectation and the bootstrap variance estimator in the case of population total respectively capture the Horvitz-Thompson estimator of total $\hat{t}_{HT}$ and the usual variance estimator $\hat{V}\left(\hat{t}_{HT}\right)$ in (1.2.9). Beaumont and Patak (2012) indicate that if any appropriate distribution is used to generate a_i^* so that

$$E^*(a_i^*) = 1 \quad \text{and} \quad E^*(a_i^* - 1)(a_j^* - 1) = \frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}} = \frac{\Delta_{ij}\pi_i\pi_j}{\pi_{ij}}, \qquad (1.5.7)$$

the first two moments are captured. This type of bootstrap method belongs to the class of the generalized bootstrap method (e.g., Lo (1991), Mason and Newton (1992) and Barbe and Bertail (1995)) which was first presented by Bertail and Combris (1997) in survey sampling with unequal probability sampling. Bertail and Combris (1997) suggested generating the vector $\boldsymbol{a}^* = \mathbf{1} + \boldsymbol{\Sigma}^{1/2}\tilde{\boldsymbol{a}}^*$ where $\boldsymbol{a}^* = (a_1^*, \ldots, a_n^*)$, $\boldsymbol{\Sigma}$ is a $n \times n$ matrix containing $\Delta_{ij}\pi_i\pi_j/\pi_{ij}$ in its i-th row and j-th column and $\tilde{\boldsymbol{a}}^*$ is a n-vector of independent random variables with mean of 0 and variance of 1 for all its elements. A simple choice is to generate $\tilde{a}_i^*$ from the standard normal distribution. So, the vector $\boldsymbol{a}^*$ follows a multivariate normal distribution $\mathcal{N}(\mathbf{1}, \boldsymbol{\Sigma})$.

In the case of Poisson sampling, the pseudo-population bootstrap method of Chauvet (2007) (see Section 1.5.1) can be implemented using a bootstrap weights method; see Beaumont and Patak (2012). That is, the creation of a pseudo-population is not required. Rather, bootstrap weights are directly generated from some appropriate distributions so that (1.5.7) holds. They suggested generating $m_i^* \sim Binomial(\tilde{w}_i, \pi_i)$, where $\tilde{w}_i = \left\lfloor \pi_i^{-1} \right\rfloor + I_i^{bp}$ and $I_i^{bp} \sim Bernoulli\left(\pi_i^{-1} - \left\lfloor \pi_i^{-1} \right\rfloor\right)$, and they define $a_i^* = m_i^*$. The resulting bootstrap estimator of the population total from this method and that from Chauvet (2007) have the same distribution (see also Ranalli and Mecatti (2012) for π_i^{-1} integer, for all $i \in s$). Applying the method of Chauvet (2007), we have

$$\begin{aligned}
\hat{\theta}^* &= \sum_{i \in s^*} \pi_i^{'-1} y_i^* \\
&= \sum_{i \in U^*} I(i \in s^*) \check{\pi}_i^{-1} \check{y}_i \\
&= \sum_{i \in s} m_i^{'*} w_i y_i,
\end{aligned}$$

where $m_i^{'*}$ is the number of times that the i-th unit of s is selected in the bootstrap sample from the pseudo-population U^*. Since sample unit i is repeated $\left\lfloor \pi_i^{-1} \right\rfloor + I(i \in U_{c*})$ times in U^* and $I(i \in U_{c*}) \overset{D}{=} I_i^{bp}$, it is easy to see that $m_i^{'*} \overset{D}{=} m_i^*$ which confirms that both methods are equivalent in the case of the population total.

In the case of the GREG estimator of total, assuming $c_i = \boldsymbol{\lambda}^\top \boldsymbol{x}_i$ in (1.2.6) with $\boldsymbol{\lambda}$ a vector of known constants so that $c_i > 0$, the GREG survey weights become

$$w_i(s) = \pi_i^{-1} \boldsymbol{x}_i^\top c_i^{-1} \hat{\boldsymbol{T}}^{-1} t_{\boldsymbol{x}}. \tag{1.5.8}$$

In this case, to compute the corresponding bootstrap statistic, Beaumont and Patak (2012) suggest using their proposed bootstrap adjustments a_i^*, obtained on the basis of the original sampling design, and defining GREG bootstrap weights similar to (1.5.8) by

$$w_i^* = a_i^* \pi_i^{-1} \boldsymbol{x}_i^\top c_i^{-1} \hat{\boldsymbol{T}}^{*-1} t_{\boldsymbol{x}}, \tag{1.5.9}$$

where $\hat{\boldsymbol{T}}^* = \sum_{i \in s} a_i^* \pi_i^{-1} \boldsymbol{x}_i c_i^{-1} \boldsymbol{x}_i^\top$. The bootstrap estimator of total is then computed by $\hat{t}^* = \sum_{i \in s} w_i^* y_i$. They showed that the resulting bootstrap variance estimator is approximately equal to the usual variance estimator presented in (1.2.16). An alternative consists of replacing $t_{\boldsymbol{x}}$ in (1.5.9) by $\hat{t}_{\boldsymbol{x}HT}$.

In general, some bootstrap adjustments may be negative. To avoid negative bootstrap adjustments a_i^*, they suggested using the following bootstrap adjustments

$$\check{a}_i^* = \frac{a_i^* + \tau - 1}{\tau},$$

where $\tau \geq 1$ is a small number but large enough so that the scaled bootstrap adjustments are non-negative. Note that $E^*(\check{a}_i^*) = 1$ and $E^*(\check{a}_i^* - 1)(\check{a}_j^* - 1) = \tau^{-2} E^*(a_i^* - 1)(a_j^* - 1)$. Therefore, to have a valid bootstrap estimator for the variance, the resulting bootstrap variance estimator obtained after applying the new bootstrap adjustment $\check{a}_i^*$ must be multiplied by τ^2. So, this value must be provided to an ultimate user.

The methods of Antal and Tillé (2011a, 2014) are a different kind of bootstrap weights methods in which a new family of sampling designs, called one-one designs, are applied. Unlike the methods introduced so far, they were interested in building integer bootstrap adjustments a_i^* so that

$$E_{p*}(a_i^*) = 1 \quad \text{and} \quad V_{p*}(a_i^*) = 1 - \pi_i.$$

In fact, they only attempt to capture the diagonal of the matrix $\boldsymbol{\Sigma} = (\Delta_{ij} \pi_i \pi_j / \pi_{ij})$, $1 - \pi_i$, and not the entire matrix in the case of the population total. Therefore, the suggested bootstrap variance estimator is usually not equal to the usual estimator in (1.2.9) when $\theta = t$.

To better understand their method, we first briefly present one-one designs which are only used to construct one part of bootstrap samples. A sample $\tilde{s}$ drawn from s under a one-one design has the following properties.

$$E_{\tilde{p}}\left(\tilde{m}_i\right) = V_{\tilde{p}}\left(\tilde{m}_i\right) = 1 \quad \text{and} \quad \sum_{i \in s} \tilde{m}_i = n,$$

where $\tilde{m}_i$ is the number of times that unit i in s is selected in $\tilde{s}$ and the subscript $\tilde{p}$ denotes the one-one sampling design. Therefore, both the expectation and the variance of $\tilde{m}_i$ are 1. That is why these designs are called one-one. In addition, the resulting sample size is the same as the original sample size n. Antal and Tillé (2011a) first proposed a one-one design. They used a mixture of simple random sample with replacement and simple random sample with over replacement as proposed by Antal and Tillé (2011b). Another one-one sampling design, called repeated half-sample, is presented in Antal and Tillé (2014) which was previously used by Saigo et al. (2001) in the context of imputed survey data. Under repeated half-sampling, if the original sample size n is even, a simple random sample of size $n/2$ without replacement is first selected and then, it is repeated a second time to form the resample of size n. If n is odd, a resample of size n can be obtained in two different ways. The first one consists of choosing a simple random sample of size $(n-1)/2$ without replacement and repeating this sample twice, so we end up with $n-1$ units. An additional unit is obtained by selecting one at random from the $n-1$ units already resampled. In the second way, we choose a simple random sample of size $(n+1)/2$ without replacement and repeat each unit twice, leading to a sample of size $n+1$. One of these units is discarded at random. Finally, we select the resulting resample of method 1 with probability 1/4 and that of method 2 with probability 3/4. This design is used by Antal and Tillé (2014) to complete the bootstrap samples in the proposed procedures.

In the case of simple random sampling without replacement, Antal and Tillé (2011a) use a mixture of simple random sampling without replacement and one-one designs in the proposed resampling procedure. We will not present this complex method here. Instead we will present Antal and Tillé (2014) where they applied a Poisson sampling design and completed it with a repeated half-sampling design, which is a one-one design. Note that the bootstrap sample is not chosen using the original design, which is simple random sampling without replacement. The following algorithm shows all steps in Antal and Tillé (2014) needed to construct the bootstrap weights in the case of simple random sampling without replacement.

(1) Take a sample, s_1^*, from s under Poisson sampling with $\pi_i = n/N$ for all $i \in s$, which is equivalent to generating $I_1^*, \ldots, I_n^* \overset{i.i.d.}{\sim} Bernoulli(n/N)$ with I_i^* indicating if the i-th unit of s is selected in s_1^* or not. Define $n_1' = \sum_{i \in s} I_i^*$.

(2) To complete the bootstrap sample:

- If $n_2' = n - n_1' \geq 2$, select a repeated half-sample, s_2^*, from the non-selected units in s_1^* (i.e. from $s \setminus s_1^*$). Define $\tilde{m}_i^*$ the number of times that the i-th unit of s is selected in s_2^*. Let $s^* = s_1^* \cup s_2^*$. Therefore, $m_i^* = I_i^* + (1 - I_i^*)\tilde{m}_i^*$ for all $i \in s$.
- If $n_2' = n - n_1' = 1$, so only one unit was not selected in s_1^*, e.g. y_k. First, generate $m_k^* = 0, 1$ or 2 with probability $1/4, 1/2$ and $1/4$, respectively. Then, randomly select one unit from s_1^*, e.g. y_l. Define

$$m_i^* = \begin{cases} I_i^*, & \text{if } i \neq k, l, \\ m_k^*, & \text{if } i = k, \\ 2 - m_k^*, & \text{if } i = l. \end{cases}$$

Finally, define $a_i^* = m_i^*$ that is the number of times that unit i of s is appearing in the final bootstrap sample.

In both papers, the case of Poisson sampling is also studied. In both methods, s_1^* is taken from s under Poisson sampling with the original inclusion probability π_i. However, s_2^* is drawn differently in each paper. In Antal and Tillé (2011a), a Poisson distribution with parameter equal to 1 is generated for the non-selected units in s_1^*, $s \setminus s_1^*$, to form s_2^* while in Antal and Tillé (2014), independent *Bernoulli* trials with probability 1/2 is first generated for units in $s \setminus s_1^*$.

Then, the selected units are repeated twice to build s_2^*. In these cases, a_i^* is also the number of times that unit i of s shows up in $s_1^* \cup s_2^*$.

Similar methods are also proposed in both papers for unequal probability sampling without replacement.

In the case of two-phase sampling, Kim et al. (2006) applied bootstrap weights methods for estimating the variance of the double-expansion estimator and the reweighted expansion estimator of the population total.

For the case of multi-stage stratified designs where sampling fractions are large and simple random sample without replacement is used at each stage, a *Bernoulli*-type bootstrap method was proposed by Funaoka et al. (2006). Under this method, *Bernoulli* trials are applied in each stage of resampling procedure. Finally, the bootstrap adjustment for each ultimate unit is the number of times that this unit is selected in the final bootstrap sample.

1.6. Bootstrap methods for model parameters

Until now, we have focused on design-based bootstrap methods for finite population parameters. In practice, analysts are often interested in generalizing the conclusions to a universe larger than the finite population under study. For example, one may be interested in studying people's perception of discrimination in their experiences with health care services as a function of characteristics such as race, sex and age. Here, the analyst is not interested in the finite population U currently under study but rather in the process relating these variables. The interest lies in estimating model parameters, also called analytic parameters (e.g., regression coefficients) rather than finite population parameters. An important distinction between finite population parameters and model parameters is that the former may be estimated perfectly provided that a census is conducted and that non-sampling errors such as non-response, measurement errors and coverage errors are absent. In contrast, even with a perfect census, it is not possible to estimate a model parameter perfectly since one faces an infinite population.

In analytic studies, the selected sample can be viewed as the result of a two-stage process: (i) first, the finite population U of size N is generated according to a statistical model, called the superpopulation model. That is, the finite population of size N can be viewed as a realization of the superpopulation model. (ii) Then, from the population generated in (i), a sample s is selected according to a given sampling design $p(s)$. Estimators of model parameters are constructed using the sample observations. This begs the question: how to estimate the variance of estimators of model parameters? From the above, it is clear that the variance involves two sources of variability: the first due to the superpopulation model that has generated the finite population U and the second due to the selection of the sample s from U. Application of the bootstrap in this context has been considered in Beaumont and Charest (2012), Wang and Thompson (2012) and Kovacevic et al. (2006). In the sequel, we focus on the method of Beaumont and Charest (2012).

For simplicity, we consider the problem of estimating the regression coefficient $\boldsymbol{\beta}$ in a linear regression model

$$m : y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i,$$

where $\boldsymbol{x}_i$ is a l-vector of predictors and $\boldsymbol{\beta}$ is a l-vector of unknown parameters. We assume that $E_m(\varepsilon_i) = 0$, $E_m(\varepsilon_i \varepsilon_j) = 0$ if $i \neq j$ and $V_m(\varepsilon_i) = \sigma^2$. Had a census been conducted, an estimator of $\boldsymbol{\beta}$ would be given by

$$\boldsymbol{\beta}_U = \left(\sum_{i \in U} \boldsymbol{x}_i \boldsymbol{x}_i^\top \right)^{-1} \sum_{i \in U} \boldsymbol{x}_i y_i. \tag{1.6.1}$$

The estimator (1.6.1) is often called a census regression coefficient. Since the y-values are only observed for $i \in s$, it is not possible to compute (1.6.1). An estimator of $\boldsymbol{\beta}_U$ based on the sample units is given by

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i \in s} w_i \boldsymbol{x}_i \boldsymbol{x}_i^\top \right)^{-1} \sum_{i \in s} w_i \boldsymbol{x}_i y_i.$$

To derive the variance of $\hat{\boldsymbol{\beta}}$, we first express its total error as

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_U \right) + \left(\boldsymbol{\beta}_U - \boldsymbol{\beta} \right).$$

It follows that the total variance of $\hat{\boldsymbol{\beta}}$ is given by

$$V_{mp} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right) = E_m V_p \left(\hat{\boldsymbol{\beta}} \right) + V_m \left(\boldsymbol{\beta}_U \right),$$

which involves both the model variability and the sampling variability of $\boldsymbol{\beta}$. Under mild regularity conditions, the term $V_p\left(\hat{\boldsymbol{\beta}}\right)$ is of order $O(n^{-1})$, whereas the term $V_m\left(\boldsymbol{\beta}_U\right)$ is of order $O(N^{-1})$; e.g., see Binder (2011). Therefore, the contribution of the term $V_m\left(\boldsymbol{\beta}_U\right)$ to the total variance is negligible if the sampling fraction f is negligible. In this case, the term $V_m\left(\boldsymbol{\beta}_U\right)$ can be omitted and the total variance reduces to

$$V_{mp}\left(\hat{\boldsymbol{\beta}}\right) \approx E_m V_p\left(\hat{\boldsymbol{\beta}}\right). \tag{1.6.2}$$

In order to estimate $E_m V_p\left(\hat{\boldsymbol{\beta}}\right)$, it suffices to obtain a consistent estimator of $V_p\left(\hat{\boldsymbol{\beta}}\right)$, which represents the sampling variance of a function of totals. To that end, any bootstrap method presented in Section 5, which estimates the sampling variability, can be applied.

We now turn to the case of non-negligible f. First, using a first-order Taylor expansion, we obtain

$$\begin{aligned} V_{mp}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) &\simeq \left\{E_{mp}\left(\hat{\boldsymbol{T}}\right)\right\}^{-1} V_{mp}\left(\sum_{i\in s} w_i \boldsymbol{x}_i e_i\right)\left\{E_{mp}\left(\hat{\boldsymbol{T}}\right)\right\}^{-1} \\ &= \left\{E_{mp}\left(\hat{\boldsymbol{T}}\right)\right\}^{-1} E_m V_p\left(\sum_{i\in s} w_i \boldsymbol{x}_i e_i\right)\left\{E_{mp}\left(\hat{\boldsymbol{T}}\right)\right\}^{-1} \\ &+ \left\{E_{mp}\left(\hat{\boldsymbol{T}}\right)\right\}^{-1} \sum_{i\in U} \boldsymbol{x}_i \boldsymbol{x}_i^\top E_m\left(e_i^2\right)\left\{E_{mp}\left(\hat{\boldsymbol{T}}\right)\right\}^{-1}, \end{aligned} \tag{1.6.3}$$

where $\hat{\boldsymbol{T}} = \sum_{i\in s} w_i \boldsymbol{x}_i \boldsymbol{x}_i^\top$ and $e_i = y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}$. In the case of non-negligible f, the last term on the right hand-side of (1.6.3) is no longer negligible and must be accounted for. A consistent linearization variance estimator of $V_{mp}(\hat{\boldsymbol{\beta}})$ is thus given by

$$\hat{V}\left(\hat{\boldsymbol{\beta}}\right) = \hat{\boldsymbol{T}}^{-1}\hat{V}\left(\sum_{i\in s} w_i \boldsymbol{x}_i e_i\right)\hat{\boldsymbol{T}}^{-1} + \hat{\boldsymbol{T}}^{-1}\left\{\sum_{i\in s} w_i \boldsymbol{x}_i \boldsymbol{x}_i^\top e_i^2\right\}\hat{\boldsymbol{T}}^{-1}. \tag{1.6.4}$$

This begs the question: how to apply the bootstrap method in order to capture both terms in (1.6.3)? It is clear that applying the bootstrap methods described in Section 1.5 may lead to an appreciable underestimation of the total variance as the model variability $V_m\left(\boldsymbol{\beta}_U\right)$ is ignored. To overcome this problem, Beaumont and Charest (2012) proposed a bootstrap weights method that accounts for both the sampling and the model variabilities when the sampling design is non informative. Note that a sampling design is non-informative if the distribution of the study variables in the sample is the same as the distribution of these variables in the population, after accounting for $\boldsymbol{x}$. Suppose that the sampling variance in (1.6.3) is to be estimated through a bootstrap weights method such as the method of Rao et al. (1992). Let $w_i^* = a_i^* w_i$ be the bootstrap weight defined as in Section 1.5.3 and which addresses the sampling variability. To account for the model variability Beaumont and Charest (2012) suggest making an additional adjustment on the w_i^*. The resulting bootstrap weights are of the form $w_i^{**} = \psi_i^* w_i^* = \psi_i^* a_i^* w_i$, with a_i^* being defined in Section 1.5.3 and ψ_i^* denoting a random bootstrap adjustment for unit i, whose role is to account for the model variability.

The bootstrap adjustments ψ_i^* are generated independently with expectation equal to 1 and variance equal to

$$V_{o^*}(\psi_i^*) = \sigma_{\psi i}^2 = \frac{w_i}{E_{p*}(w_i^{*2})}, \tag{1.6.5}$$

where the subscript o^* denotes the distribution of ψ_i^* in the bootstrap samples. To better understand the rationale behind the method of Beaumont and Charest (2012), we first express the bootstrap version of $\hat{\boldsymbol{\beta}}$ as

$$\hat{\boldsymbol{\beta}}^* = \left(\sum_{i\in s} w_i^{**} \boldsymbol{x}_i \boldsymbol{x}_i^\top\right)^{-1} \sum_{i\in s} w_i^{**} \boldsymbol{x}_i y_i.$$

Using a first-order Taylor expansion, we obtain

$$V_{p*o*}\left(\hat{\boldsymbol{\beta}}^*\right) \simeq \hat{\boldsymbol{T}}^{-1} V_{p*o*}\left(\sum_{i\in s} w_i^{**} \boldsymbol{x}_i e_i\right) \hat{\boldsymbol{T}}^{-1}, \tag{1.6.6}$$

where

$$V_{p*o*}\left(\sum_{i\in s} w_i^{**} \boldsymbol{x}_i e_i\right) = V_{p*}\left(\sum_{i\in s} w_i^{*} \boldsymbol{x}_i e_i\right) + E_{p*}\left(\sum_{i\in s} \sigma_{\psi i}^2 w_i^{*2} \boldsymbol{x}_i \boldsymbol{x}_i^\top e_i^2\right).$$

From (1.6.5), it becomes clear that the total bootstrap variance estimator (1.6.6) is asymptotically equivalent to the linearization variance estimator (1.6.4).

To generate ψ_i^*, Beaumont and Charest (2012) suggest using the distribution: $Prob(\psi_i^* = 1 - \sigma_{\psi i}) = 1/2$ and $Prob(\psi_i^* = 1 + \sigma_{\psi i}) = 1/2$. This ensures that ψ_i^* is always non-negative provided that $\sigma_{\psi i} \leq 1$. Note that, in order to compute $\sigma_{\psi i}$, $E_{p*}(w_i^{*2})$ in (1.6.5) can be easily approximated through a Monte Carlo approximation by taking the mean of the B generated w_i^{*2}.

It is worthwhile to mention that if all the weights w_i are large (implying a small f), $\sigma^2_{\psi i}$ is expected to be small, in which case the contribution of ψ_i^* is expected to be small and, as a result, may be ignored. This conclusion is consistent with the result in (1.6.2) that the model variability can be ignored if the sampling fraction is small.

1.7. Bootstrap for missing survey data

Virtually all the surveys must face the problem of missing observations due to various reasons. Survey statisticians distinguish unit non-response (when no information is collected on a sample unit) from item non-response (when the absence of information is limited to some variables only). Unit non-response occurs, for example, when the sampled unit is not at home or refuses to participate in the survey, while item non-response occurs when the sample unit refuses to respond to sensitive items, may not know the answer to some items, or because of edit failures. In this section, we focus on item non-response, which is typically treated by some forms of imputation. In the last two decades, the problem of variance estimation in the presence of imputed data has been widely studied in the literature; see, e.g., Haziza (2009) for a review. It is well known that treating the imputed values as if they were observed values leads to underestimation of the true variance, leading to invalid inferences. In this section, after presenting some useful concepts, some bootstrap methods for imputed survey data will be presented.

1.7.1. Some useful concepts

Let r_i be the response indicator associated with unit i such that $r_i = 1$ if unit i responds to item y and $r_i = 0$, otherwise. Let

$$y_i^I = r_i y_i + (1 - r_i)\tilde{y}_i,$$

where $\tilde{y}_i$ denotes the imputed value used to replace the missing y_i. Let θ be a finite population parameter, $\hat{\theta}$ be the complete data estimator of θ and $\hat{\theta}^I$ be the imputed estimator obtained after imputation. The imputed estimator $\hat{\theta}^I$ can be computed the same way as the complete data estimator $\hat{\theta}$ using y^I-values instead of the y-values. For example, in the case of a total t, an imputed estimator is

$$\hat{t}^I = \sum_{i \in s} w_i y_i^I.$$

In practice, various imputation methods are used. We distinguish between two classes of imputation methods: the deterministic methods, which are those that yield the same imputed values if the imputation process is repeated, and the random methods that may yield different imputed values if the imputation is repeated. A random method can be viewed as a deterministic method with an added random noise. Most imputation methods encountered in practice are motivated by the general model

$$m : y_i = f(\boldsymbol{x}_i; \boldsymbol{\beta}) + \varepsilon_i, \tag{1.7.1}$$

where $f(\cdot)$ is a given function, $\boldsymbol{x}$ is a vector of auxiliary variables recorded for all the sample units (respondents and non-respondents) and $\boldsymbol{\beta}$ is a vector of unknown parameters. The errors ε_i satisfy

$$E_m(\varepsilon_i) = 0,\ V_m(\varepsilon_i) = \sigma^2 c_i \text{ and } \ cov_m(\varepsilon_i, \varepsilon_j) = 0,\ \forall i \neq j,$$

where σ^2 is an unknown parameter and $c_i = v(\boldsymbol{x}_i)$ is a known function. For example, deterministic linear regression imputation is motivated by (1.7.1) with $f(\boldsymbol{x}_i; \boldsymbol{\beta}) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$ and $c_i = \boldsymbol{\lambda}^\top \boldsymbol{x}_i$ for a vector of specified constants $\boldsymbol{\lambda}$. In this case, the imputed value $\tilde{y}_i$ is given by

$$\tilde{y}_i = \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_r, \tag{1.7.2}$$

where

$$\hat{\boldsymbol{\beta}}_r = \left(\sum_{i \in s} w_i r_i \boldsymbol{x}_i {c_i}^{-1} \boldsymbol{x}_i^\top \right)^{-1} \sum_{i \in s} w_i r_i \boldsymbol{x}_i {c_i}^{-1} y_i$$

is the weighted least square estimator of $\boldsymbol{\beta}$ based on the responding units. Mean imputation, whereby the missing values are replaced by the mean of the respondents, $\bar{y}_r = \sum_{i \in s} w_i r_i y_i / \sum_{i \in s} w_i r_i$, is a special case of (1.7.2) with $\boldsymbol{x}_i = c_i = 1$ for all i.

A frequently used random method is random hot-deck imputation, which consists of imputing a missing value by the value of a respondent selected at random from the set of responding units. More specifically, the imputed values under random hot-deck imputation are

$$\tilde{y}_i = \bar{y}_r + \tilde{\varepsilon}_i, \tag{1.7.3}$$

where $\tilde{\varepsilon}_i$ takes a value in $\{e_1, \ldots, e_{n_r}\}$ such that $Prob(\tilde{\varepsilon}_i = e_j) = r_j w_j / \sum_{l \in s} r_l w_l$ with $e_j = y_j - \bar{y}_r$ and n_r denoting the number of respondents to item y.

In this section, we assume that the data are Missing At Random (MAR); (Rubin, 1976). The data are MAR if the probability of response to item y is independent of the error term in (1.7.1) after accounting for the vector of auxiliary variables $\boldsymbol{x}$.

There exist two theoretical frameworks for variance estimation: the customary two-phase framework and the reverse framework. In the two-phase framework, non-response is viewed as a second phase of selection. In the reverse framework, the order of sampling and response is reversed. First, the population is randomly divided into a population of respondents and a population of non-respondents according to the non-response mechanism. Then, a random sample is selected from the population (containing respondents and non-respondents) according to the sampling design. Unlike the two-phase framework, the reverse framework requires the additional assumption that the non-response mechanism does not depend on which sample is selected. The reverse framework is particularly useful in the context of bootstrap variance estimation in the presence of imputed data, as we argue in the next section.

1.7.2. Bootstrap methods for negligible sampling fraction

In this section, we focus on the case of negligible f. In this context, Shao and Sitter (1996) proposed a bootstrap method for handling imputed data. The rationale behind their method is to first select, using any complete data bootstrap method, a bootstrap sample of pairs of original or rescaled imputed data and their corresponding original response status. The bootstrap data with a missing status are then reimputed using the same imputation method that was used in the original sample. To illustrate the Shao-Sitter method, we consider the case of simple random sampling without replacement with the RSB method of Rao and Wu (1988) and mean imputation to compensate for the missing values. The algorithm proceeds as follows:

Shao-Sitter Algorithm:

(1) Let n' be the bootstrap sample size and $y_i' = \bar{y}^I + C(y_i^I - \bar{y}^I)$, for all i in s, where $\bar{y}^I = n^{-1}\sum_{i \in s} y_i^I$ and

$$C = \sqrt{\frac{n'(1-f)}{n-1}}.$$

(2) Draw a bootstrap sample of pairs $s^* = \{(y_i^*, r_i^*)\}_{i=1}^{n'}$ of size n' with replacement from $\{(y_i', r_i)\}_{i=1}^{n}$.

(3) Reimpute the missing values in the bootstrap sample s^* using the respondents in this sample, i.e. define y_i^{*I} as follows

$$y_i^{*I} = \begin{cases} y_i^*, & \text{if } r_i^* = 1, \\ \bar{y}_r^*, & \text{if } r_i^* = 0, \end{cases} \qquad \text{where } \bar{y}_r^* = \frac{\sum_{i \in s^*} r_i^* y_i^*}{\sum_{i \in s^*} r_i^*}, \text{ for } \; i \in s^*.$$

Let $\hat{\theta}^{*I}$ be the bootstrap statistic based on the observed and imputed bootstrap data.

(4) Repeat Steps 2 and 3 a large number of times, B, to get $\hat{\theta}_1^{*I}, \cdots, \hat{\theta}_B^{*I}$.

(5) Estimate $V\left(\hat{\theta}^I\right)$ with $V_{p*}\left(\hat{\theta}^{*I}\right)$ or its Monte Carlo approximation $\hat{V}_B^* = (B-1)^{-1}\sum_{b=1}^{B}\left(\hat{\theta}_b^{*I} - \hat{\theta}_{(\cdot)}^{*I}\right)^2$, where $\hat{\theta}_{(\cdot)}^{*I} = B^{-1}\sum_{b=1}^{B}\hat{\theta}_b^{*I}$.

Note that, for imputation methods using auxiliary information (e.g., regression imputation), the vector of auxiliary variables $\boldsymbol{x}_i$ also accompany the pairs (y_i, r_i) in the bootstrap sample and need to be rescaled as is done for y_i.

In the case of the population total, the bootstrap total estimator is $\hat{t}^{*I} = (N/n')\sum_{i=1}^{n'} y_i^{*I} = N\bar{y}_r^*$. Using a first order Taylor linearization, when the nonresponse mechanism is uniform, i.e. the response probability $p_i = Prob(r_i = 1) =$

p_0 for all $i \in s$, the bootstrap variance estimator $V_{p*}\left(\hat{t}^{*I}\right)$ is approximated by

$$\begin{aligned} V_{p*}\left(\hat{t}^{*I}\right) &\approx V_{p*}\left\{\frac{N}{\hat{p}_0 n'}\sum_{i=1}^{n}\left(m_i^* - \frac{n'}{n}\right)(y_i' - \bar{y}_r)\right\} \\ &= \frac{N^2}{\hat{p}_0^2}\frac{C^2}{n'n}\sum_{i \in s} r_i(y_i - \bar{y}_r)^2 \qquad (1.7.4) \\ &= N^2\left(\frac{1-f}{\hat{p}_0}\right)\frac{n_r - 1}{\hat{p}_0(n-1)}\frac{s_r^2}{n}, \end{aligned}$$

where m_i^* is the number of times that the i-th unit in s is selected in the bootstrap sample, $\hat{p}_0 = n_r/n$, the response rate, is the estimator of p_0 and $s_r^2 = (n_r - 1)^{-1}\sum_{i \in s} r_i(y_i - \bar{y}_r)^2$.

At this point, one may be wondering what quantity (1.7.4) is really estimating. To answer this question, one has to rely on the reverse framework for variance estimation mentioned above. The reverse framework, proposed by Fay (1991) and Shao and Steel (1999), can be used to express the variance of $\hat{\theta}^I$ as the sum of two terms in the case of deterministic imputation. Under this framework, the population is first randomly divided into a population of respondents and a population of non-respondents according to the non-response mechanism. Then, a sample (containing respondents and non-respondents) is selected from the population according to the sampling design $p(s)$. In this case, the total variance of $\hat{\theta}^I$ under deterministic imputation is given by

$$V_{RV}^{NR}\left(\hat{\theta}^I\right) = EV_p\left(\hat{\theta}^I|\boldsymbol{y}, \boldsymbol{r}\right) + VE_p\left(\hat{\theta}^I|\boldsymbol{y}, \boldsymbol{r}\right), \qquad (1.7.5)$$

where $\boldsymbol{r} = (r_1, \ldots, r_N)^\top$ is the vector of response indicators and $\boldsymbol{y} = (y_1, \ldots, y_N)^\top$. Under mild regularity conditions, the contribution of the second component to the total variance in (1.7.5), $VE_p\left(\hat{\theta}^I|\boldsymbol{y}, \boldsymbol{r}\right)/V^{NR}\left(\hat{\theta}^I\right)$, is of order $O(f)$, which is negligible when the sampling fraction, f, is negligible. Therefore, when f is negligible, this component can be omitted from the calculations and it remains to estimate the first component $EV_p\left(\hat{\theta}^I|\boldsymbol{y}, \boldsymbol{r}\right)$. To that end, it suffices to estimate $V_p\left(\hat{\theta}^I|\boldsymbol{y}, \boldsymbol{r}\right)$ in an (approximately) unbiased fashion. Suppose that we are interested in estimating a population total t. Noting that the imputed estimator $\hat{t}^I$ can be expressed as a function of totals, estimating $V_p\left(\hat{t}^I|\boldsymbol{y}, \boldsymbol{r}\right)$ reduces to the classical problem of estimating the sampling variance of a function of totals. To that end any complete data variance estimation methods can be used, including Taylor expansion procedures and resampling methods. The bootstrap variance estimator (1.7.4) is an estimator of $V_p\left(\hat{t}^I|\boldsymbol{y},\boldsymbol{r}\right)$ as the Shao-Sitter method simulates the effect of sampling conditionally on the vector of response indicators $\boldsymbol{r}$ and since the bootstrap method reflects the sampling variability. This can be explained by the fact that non-response is not generated in each bootstrap sample before the imputation process is performed; see Mashreghi, Léger, and Haziza (2014). As a result, the bootstrap variance estimator (1.7.4) can be used if the sampling fraction f is negligible. Also, it is worth noting that (1.7.4) is approximately unbiased for $V_p\left(\hat{t}^I|\boldsymbol{y},\boldsymbol{r}\right)$ regardless of the validity of the underlying imputation model. The problem of bootstrap variance estimation in the case of quantiles is discussed in Shao and Chen (1998). The method of Shao-Sitter may lead to a biased estimator in the case of very small stratum sizes. To overcome the problem, Saigo et al. (2001) proposed a modification of the method of Shao and Sitter (1996). Instead of using any complete data bootstrap method like Shao and Sitter (1996), they proposed a new sampling design, called the repeated half-sample bootstrap, which is actually identical to that of Antal and Tillé (2014); see Section 1.5.3.

1.7.3. Bootstrap methods for non-negligible sampling fraction

When the sampling fraction is appreciable, the Shao-Sitter method may lead to a significant underestimation of the variance as the term $VE_p\left(\hat{\theta}^I|\boldsymbol{y},\boldsymbol{r}\right)$ in (1.7.5) is not accounted for. To overcome this problem Mashreghi, Léger, and Haziza (2014) proposed a method called the independent bootstrap in the special case of stratified simple random sample without replacement with uniform non-response in each stratum. Their method consists of selecting bootstrap samples according to a direct bootstrap method (see Section 5.2) and then regenerating non-response within each bootstrap sample, mimicking the initial non-response mechanism, i.e., independent *Bernoulli* trials with the observed response rate. Afterwards, the non-respondents in the bootstrap sample are reimputed using the same imputation method that was used on the original data. Since direct bootstrap methods involve some constants, e.g., C and k' in Table 1.4, Mashreghi, Léger, and Haziza (2014) showed how to modify these constants to obtain an approximately unbiased estimator of the total variance. The modified constants explicitly depend on the response rate as well as the imputation method. For example, in the case of mean imputation with uniform non-response mechanism, the rescaling factor in the method of Rao and Wu (1988) presented in (1.5.4) has to be replaced by

$$C^I = \sqrt{\frac{n'[1-(n_r/N)]}{n_r - 1}}.$$

Comparing C^I with C in (1.5.4), we see that to compute C^I, n in C is replaced by n_r, i.e. the number of respondents is used instead of the sample size as the information contained in the sample only comes from the observed values. In this case, the following algorithm leads to the creation of samples of bootstrap imputed data:

(1) Let n' be the bootstrap sample size and $y_i' = \bar{y}^I + C^I(y_i^I - \bar{y}^I)$, for all i in s.

(2) Draw a bootstrap sample $\{y_i^*\}_{i=1}^{n'}$ of size n' with replacement from $\{y_i'\}_{i=1}^{n}$.

(3) Generate the bootstrap sample of response indicators, $\{r_i^*\}_{i=1}^{n'} \overset{i.i.d.}{\sim} Bernoulli(\hat{p}_0)$. Let $s^* = \{(y_i^*, r_i^*)\}_{i=1}^{n'}$.

(4) Identify the missing and observed bootstrap data using the regenerated r_i^* and reimpute the bootstrap missing values using the bootstrap respondents. Let $\hat{\theta}^{*I}$ be the bootstrap statistic based on the bootstrap imputed data.

Unlike the Shao-Sitter algorithm presented in the previous section, the previous algorithm includes an additional step in order to generate non-response within each bootstrap sample.

In order to handle more complex sampling designs and/or more general non-response mechanism, Mashreghi, Haziza, and Léger (2014a) developed pseudo-population bootstrap methods that lead to approximately unbiased variance estimator in the case of non-negligible sampling fractions. The key idea is to recognize that the set of respondents to a specific item can be viewed as a random sample obtained by a Poisson sampling design using the (unknown) response probabilities as the inclusion probabilities. Therefore, a pseudo-population can be created in two distinct steps: in the first, one applies the pseudo-population bootstrap method appropriate for Poisson sampling (see Section 1.5.1), which leads to a "pseudo-sample". Then, from the pseudo-sample, the pseudo-population is created using a complete data pseudo-population bootstrap method depending on the original sampling mechanism; e.g., the method of Booth et al. (1994) for the case of simple random sampling without replacement. Bootstrap samples are then selected from the pseudo-population by applying the original sampling design and non-respondents are regenerated in each bootstrap sample using the *Bernoulli* distribution with the original estimated response probabilities. Imputation within each bootstrap sample is performed according to the same imputation method that was used in the original sample. Finally the bootstrap statistic is computed on the reimputed data. Mashreghi, Haziza, and Léger (2014a) showed that their method leads to an approximately unbiased estimator of the total variance.

Chapter 2

BOOTSTRAP METHODS FOR IMPUTED DATA FROM REGRESSION, RATIO AND HOT DECK IMPUTATION

Résumé

La non-réponse partielle en échantillonnage est habituellement traitée par imputation. Une méthode bootstrap traitant les valeurs imputées comme si elles avaient été observées conduit généralement à des estimations de la variance qui sont trop petites. Shao et Sitter (1996) ont introduit une méthode bootstrap menant à des estimateurs convergents de la variance lorsque la fraction de sondage est faible. Dans le contexte d'un plan stratifié aléatoire simple, nous introduisons le bootstrap indépendant qui est valide même si la fraction de sondage est grande. Elle consiste à modifier une méthode de bootstrap applicable aux enquêtes, à générer indépendamment le statut de la réponse de chaque unité, et à imputer les non-répondants dans l'échantillon bootstrap. Une attention particulière est portée à l'approche des poids bootstrap de Rao, Wu et Yue (1992).

Abstract

Item non-response in sample surveys is usually addressed by imputation. A bootstrap method that treats the imputed values as if they were observed generally leads to variance estimates that are too small. Shao and Sitter (1996) introduced a bootstrap method in this context, which leads to consistent variance estimators when the sampling fraction is small. In the context of stratified simple random sampling, we introduce the independent bootstrap which is valid even when the sampling fraction is large. It consists of modifying a bootstrap method for sample surveys, of independently generating the response status of each unit, and of imputing the non-respondents in the bootstrap sample. We pay special attention to the bootstrap survey weights approach of Rao et al. (1992).

Key words and phrases: bootstrap, non-response, imputation and bootstrap weights.

2.1. Introduction

Statistical agencies, such as the Census Bureau, the National Center for Health Statistics, and Statistics Canada among others, provide access to detailed micro-level data through their research data centres allowing researchers in social and health sciences to advance research in their fields. The files provided to the researchers are usually rectangular, each row corresponding to an ultimate unit in the survey and columns corresponding to the different variables under study, plus other columns for survey weights. To estimate the variance of estimators, columns of bootstrap survey weights are often added following the method of Rao et al. (1992). Non-response is an important practical problem in statistical surveys. Unit non-response is usually dealt with through reweighting of the respondents, whereas item non-response is generally addressed by imputation. It should be noted that bootstrap weights only account for the *sampling* variability in the observations (including unit non-response adjustments), but not the added variability due to item non-response and imputation, leading to underestimation of the variance. And the underestimation of this method which we call the naive bootstrap can be substantial as will be illustrated in our real-life example in Section 2.7. Shao and Sitter (1996) introduced a bootstrap method to deal with imputed data. It consists of using any (complete) data bootstrap method to select a bootstrap sample of imputed data while keeping their corresponding original response status, and then to reimpute the bootstrap data with a missing status using the same imputation method that was used on the original data. The estimator is computed on the imputed bootstrap data, leading to a bootstrap estimate of variance.

This bootstrap method requires the presence of a missing value indicator variable for each item under study, variables which are usually not present in the files in research data centres making the Shao-Sitter unapplicable in practice; as of this writing, no file in the Canadian network of research data centres contains missing value flags. While they claim that their method works well without any restriction on the sampling design or on the imputation method, a detailed analysis of their method through the reverse framework of Fay (1991) and Shao and Steel (1999) shows that their variance estimate is consistent only when the sampling fraction f is negligible. The example in Section 2.7 will show that this condition does not always hold in practice, *even if missing data status was available.*

In this paper we introduce the independent bootstrap method to overcome the two drawbacks of the Shao-Sitter method. It leads to an asymptotically consistent bootstrap estimator of the variance of an estimator defined as a function of means under stratified simple random sample without replacement even for a large overall sampling fraction. The theory applies to the case of uniform non-response in each stratum and the method only requires information about the response rate of the item under study in each stratum rather than the detailed information about response status for each sample unit. The procedure is applied independently in each stratum and consists of first regenerating bootstrap response indicators mimicking the initial non-response mechanism, i.e., independent Bernoulli trials with the observed response rate. Then, independently, bootstrap observations are regenerated using one of the bootstrap methods. We call it the independent bootstrap method because the sample of observations and the response status are generated independently whereas in the Shao-Sitter method it is as if these two components were treated as a pair and were generated jointly. While the Shao-Sitter method simply uses one of the bootstrap methods designed for complete data without any modification, for the independent bootstrap, we need to modify them. Since the sampling mechanism used in most bootstrap methods differs from simple (or stratified) random sampling, they all involve some kind of a constant which guarantees that when they are applied to the mean (or total) estimator, they consistently estimate the variance of the estimator, if only to account for the sampling fraction. For instance, in the case of the rescaling bootstrap method of Rao and Wu (1988), the constant is the rescaling factor. But in the presence of item non-response not only the sampling fraction but also the non-response mechanism and the method of imputation both influence the variance of the estimators. So unlike in the Shao-Sitter method, the constant of the bootstrap method used is modified in the independent bootstrap. Afterwards, we reimpute the non-respondents in the bootstrap sample using the same imputation technique that was used on the original data. Note that the modified constants explicitly depend on the response rate as well as the imputation method.

This article is organized as follows. After introducing some notation in Section 2.2, we study the properties of the Shao-Sitter bootstrap method through the reverse framework in Section 2.3. Then we introduce the new independent bootstrap procedure for imputed data and present the modified constants for the different combination of bootstrap and imputation methods in Section 2.4. The case of bootstrap weights receives special attention as it is the method of choice of the data sets in many research data centres. A modification of the original Shao-Sitter method is possible provided that the response status of each observation is available and is introduced in Section 2.5. To compare the different methods, Section 2.6 presents a simulation study which supports the theory. To illustrate some of the practical difficulties in the estimation of the variance of imputed estimators, Section 2.7 presents some results of a case study from the Research and Development in Canadian Industry survey conducted at Statistics Canada. The Appendix A concludes with some theoretical justifications for the results.

2.2. Preliminaries

Throughout this article, we consider a stratified simple random sampling design where the population U consists of L non-overlapping strata with N_h units in the h-th stratum, $h = 1, \cdots, L$. In stratum h, a sample s_h of size n_h is selected from U_h according to simple random sampling without replacement. The selection is independent across strata. We denote the full sample by $s = \cup_{h=1}^{L} s_h$. Associated with the i-th unit in stratum h is a characteristic y_{hi} and a t-vector[1] of auxiliary variables $\mathbf{x}_{hi}$. The sampling fraction in the h-th stratum is defined by $f_h = n_h/N_h$, and the overall sampling fraction is $f = n/N$, where $n = \sum_{h=1}^{L} n_h$ and $N = \sum_{h=1}^{L} N_h$.

To illustrate the concepts, we consider the case of a population mean, $\theta = \bar{Y} = \sum_{h=1}^{L} \sum_{i \in U_h} y_{hi}/N$. Having a complete data set (i.e., full response), a design-unbiased estimator of $\bar{Y}$ is the Horvitz-Thompson estimator

$$\hat{\theta} = \bar{y} = \sum_{h=1}^{L} W_h \bar{y}_h,$$

where $W_h = N_h/N$ and $\bar{y}_h = \sum_{i \in s_h} w_{hi} y_{hi}/N_h$ is the sample mean in the h-th stratum with $w_{hi} = N_h/n_h$ denoting the survey weight associated with the (hi)-th unit.

We now turn to the case of missing y-values. We assume that the vector $\mathbf{x}$ is observed for all sample units (respondents and non-respondents). Let r_{hi} be the response indicator associated with the (hi)-th unit. Let $y_{hi}^{I} = y_{hi}$ if $r_{hi} = 1$, and $y_{hi}^{I} = \tilde{y}_{hi}$ if $r_{hi} = 0$, where $\tilde{y}_{hi}$ denotes the imputed value used to replace missing y_{hi}. An imputed estimator of $\bar{Y}$ based on observed and imputed data is

$$\hat{\theta}^{I} = \bar{y}^{I} = \sum_{h=1}^{L} W_h \, \bar{y}_h^{I},$$

where $\bar{y}_h^{I} = \sum_{i \in s_h} w_{hi} \, y_{hi}^{I}/N_h$. In this paper, we consider the case of deterministic ratio and linear regression imputations as well as random hot-deck imputation.

Deterministic ratio imputation within stratum consists of imputing the missing value y_{hi} by

$$\tilde{y}_{hi} = \hat{R}_h x_{hi}, \tag{2.2.1}$$

[1] In this chapter, t represents the size of the auxiliary variables rather than a total.

where x_{hi} is an auxiliary variable (in this case we assume that the vector of auxiliary variables $\mathbf{x}_{hi}$ is one-dimensional) and

$$\hat{R}_h = \left(\sum_{i \in s_h} w_{hi}\ r_{hi} y_{hi} \right) \Big/ \left(\sum_{i \in s_h} w_{hi}\ r_{hi} x_{hi} \right).$$

For the linear regression imputation method within stratum, we assume that a vector of auxiliary variables of the form $\tilde{\mathbf{x}}'_{hi} = (1, \mathbf{x}'_{hi})$ is available and a missing value y_{hi} is imputed using a regression model as follows:

$$\tilde{y}_{hi} = \tilde{\mathbf{x}}'_{hi}\ \tilde{\boldsymbol{\beta}}_{hr}, \tag{2.2.2}$$

where

$$\tilde{\boldsymbol{\beta}}_{hr} = \left(\sum_{i \in s_h} w_{hi}\ r_{hi} \tilde{\mathbf{x}}_{hi} \tilde{\mathbf{x}}'_{hi} \right)^{-1} \left(\sum_{i \in s_h} w_{hi}\ r_{hi} \tilde{\mathbf{x}}_{hi} y_{hi} \right).$$

Mean imputation (MI), which consists of imputing by using the mean of the respondents, is a special case of (2.2.1) and (2.2.2) obtained by setting $x_{hi} = 1$ and $\tilde{\mathbf{x}}_{hi} = 1$ for all (hi), respectively.

The most common random imputation method used in practice is random hot-deck imputation (RHDI). It consists of selecting a respondent (donor) at random from the set of respondents with probability proportional to the sampling weight, and then using the donor's item value to “fill in" for the missing value of a non-respondent (recipient). In this paper, we consider the case of RHDI within stratum for which RHDI is performed independently within each stratum. That is, if $r_{hi} = 0$, then

$$\tilde{y}_{hi} = y_{hj} \quad \text{with} \quad Prob(\tilde{y}_{hi} = y_{hj}) = \frac{w_{hj}\ r_{hj}}{\sum_{l \in s_h} w_{hl}\ r_{hl}}.$$

Note that in the case of stratified simple random sampling, as is the case here, this is equivalent to replacing the non-respondents by a simple random sample with replacement from the respondents.

2.3. The Shao-Sitter Method for Missing Data

To the best of our knowledge, it seems that the only existing bootstrap method for imputed data was proposed by Shao and Sitter (1996). It assumes that the data set carries the original response status for each individual variable and each unit in the sample. To evaluate the variance of point estimators, we use the reverse framework, studied by Fay (1991) and Shao and Steel (1999), where the population is first randomly divided into a population of respondents and a population of non-respondents according to the non-response mechanism and a sample is selected from the population (containing respondents and non-respondents) according to the sampling design. The total variance of $\hat{\theta}^I$, based on deterministic and random imputation methods, can be respectively written as

$$V(\hat{\theta}^I) = E_q V_p(\hat{\theta}^I|\mathbf{r}) + V_q E_p(\hat{\theta}^I|\mathbf{r}) = V_1 + V_2 \tag{2.3.1}$$

and

$$V(\hat{\theta}^I) = E_q V_p E_I(\hat{\theta}^I|s,\mathbf{r}) + V_q E_p E_I(\hat{\theta}^I|s,\mathbf{r}) + E_q E_p V_I(\hat{\theta}^I|s,\mathbf{r}) = \widetilde{V}_1 + \widetilde{V}_2 + \widetilde{V}_3, \tag{2.3.2}$$

where $\mathbf{r}$ is the vector of response indicators, and the subscripts p, q, and I refer to the randomness induced by the sampling, non-response, and random imputation mechanisms, respectively. Throughout, we assume that the non-response mechanism is uniform, where the response probability is constant for all units in each stratum, a special case of uniform non-response within imputation classes. Under mild regularity conditions, the components V_1, $\widetilde{V}_1$ and $\widetilde{V}_3$ in (2.3.1) and (2.3.2) are of order $O(1/n)$, whereas the components V_2 and $\widetilde{V}_2$ are of order $O(1/N)$. As a result, the contribution of the second component to the total variance in both (2.3.1) and (2.3.2), $V_2/V(\hat{\theta}^I)$ and $\widetilde{V}_2/V(\hat{\theta}^I)$, is negligible when the overall sampling fraction, f, is negligible. Note that the individual sampling fractions f_h are not required to be negligible.

The Shao-Sitter method consists of taking a "paired bootstrap" sample in the h-th stratum from the pairs $\{(y^I_{hi}, r_{hi})\}_{i=1}^{n_h}$ using any complete bootstrap method applicable to simple random sampling. Non-respondents in the bootstrap sample are reimputed using the same method that was used in the original sample. The process is repeated independently in each stratum. To better understand the rationale behind the Shao-Sitter method, consider a deterministic imputation method. Suppose that we want to estimate the first term of the variance decomposition of (2.3.1), i.e., $E_q V_p(\hat{\theta}^I\ |\mathbf{r})$. If we use an estimator $\widehat{V}_p(\hat{\theta}^I\ |\mathbf{r})$ that accounts for the (sampling) variability of the imputed estimator $\hat{\theta}^I$ conditional on the observed response indicators $\mathbf{r}$, we will have a valid estimator for the first term of (2.3.1). This is exactly what the Shao-Sitter method does: the response indicators are fixed (respondents and non-respondents in the original sample remain respectively respondents and non-respondents in the bootstrap sample) and in this paired bootstrap, the only variability reflected in the bootstrap mechanism is the *sampling* variability. It therefore ignores the second term in the variance decomposition (2.3.1). Denoting the Shao-Sitter variance estimator of $\hat{\theta}^I = \bar{y}^I$ by $V^*_{Sh.S.}$, we show in the Appendix A that

$$E_p E_q(\widehat{V}^*_{Sh.S.}) \approx V_1 = V(\hat{\theta}^I) - V_2.$$

If f is negligible, then $V^*_{Sh.S.}$ provides a valid estimator of $V(\hat{\theta}^I)$ as the contribution of V_2 to the total variance is negligible.

We now consider the case of RHDI. In this case, the Shao-Sitter method involves two sources of randomness, one reflecting the sampling variability, and one reflecting the donor variability. Complete bootstrap methods usually include some kind of rescaling to account for the without replacement sampling of the original sample. But this rescaling also affects the bootstrap distribution of donors. Consequently, if the sampling fraction is important the *joint* bootstrap distribution of sampling and donor imputation conditional on the response status will lead to a poor approximation of the *joint* distribution of sampling and donor imputation in the original sample. As a result, the Shao-Sitter procedure will not consistently estimate the sum of the first and third terms in (2.3.2) while completely ignoring the second one. For simplicity, let $L = 1$. In the Appendix A, we show that

$$E_p E_q E_I(V^*_{Sh.S.}) \approx \widetilde{V}_1 + \lambda_f \widetilde{V}_3 = V(\hat{\theta}^I) - [\widetilde{V}_2 + (1 - \lambda_f)\widetilde{V}_3], \qquad (2.3.3)$$

where $\lambda_f = (1 - f)$ for the bootstrap rescaling method (BRS) of Rao and Wu (1988) and the bootstrap weights approach (BW) of Rao et al. (1992) with $C = [n'/(n-1)][1-f]$ where n' is the bootstrap sample size. Thus, the variance in

(2.3.2) can be written as

$$V(\hat{\theta}^I) = [\widetilde{V}_1 + \lambda_f \widetilde{V}_3] + [\widetilde{V}_2 + (1 - \lambda_f)\widetilde{V}_3]. \qquad (2.3.4)$$

Comparing (2.3.3) with (2.3.4), it becomes clear that the Shao-Sitter method consistently estimates $V(\hat{\theta}^I)$ only when f is small.

2.4. The Independent Bootstrap Method

In general, there are two difficulties with the Shao-Sitter method: the need to have the response status for each unit in s and the underestimation of $V(\hat{\theta}^I)$ for large f. The independent bootstrap will overcome these difficulties. Let p_{0h} be the true response probability in the h-th stratum which we estimate by $\hat{p}_{0h} = n_{hr}/n_h$, where n_{hr} is the number of respondents in the h-th stratum. Recall that we assume uniform non-response within stratum and that the response rate $\hat{p}_{0h}$ in each stratum is available.

2.4.1. Independent Bootstrap Method

In Section 2.3, we argued that the Shao-Sitter method can be seen as a paired bootstrap method. The independent bootstrap consists of choosing a sample and, independently, of generating the response status for each unit in this sample. The first problem of the Shao-Sitter method is solved by generating the response status in the bootstrap samples independently from its selection using the estimated original non-response model, i.e., independent Bernoulli random variable with probability $\hat{p}_{0h}$. The second problem is taken care by modifying the constant of the complete data bootstrap method being used to get a consistent variance estimator in the case where the statistic is a smooth function of means. Once the bootstrap sample and response status are independently generated, the nonrespondents in the bootstrap data are reimputed using the original imputation method. Finally, the bootstrap statistic, $\hat{\theta}^{*I}$, is computed on the sample of reimputed data and the bootstrap estimator of $V(\hat{\theta}^I)$ is $V^*(\hat{\theta}^{*I}) = E^*[\hat{\theta}^{*I} - E^*(\hat{\theta}^{*I})]^2$.

To illustrate the proposed method, we consider the case $\theta = \bar{Y}$ under mean imputation and apply the independent bootstrap method with the bootstrap rescaling method (BRS) of Rao and Wu (1988):

(1) Let $z_{hi} = \bar{y}_h^I + \sqrt{C_h^I}\ (y_{hi}^I - \bar{y}_h^I)$, for $i = 1, \cdots, n_h$, where $C_h^I = \frac{n'_h}{n_{hr}-1}\ [1 - \hat{p}_{0h} f_h]$ and n'_h is the bootstrap sample size in the h-th stratum.

(2) Generate the bootstrap sample of response indicators,

$$\{r_{hi}^*\}_{i=1}^{n'_h} \overset{i.i.d.}{\sim} Bernoulli(\hat{p}_{0h}).$$

(3) Independently draw a simple random sample of size n'_h, $s_h^* = \{z_{hi}^*\}_{i=1}^{n'_h}$, with replacement from $\{z_{hi}\}_{i=1}^{n_h}$. Afterward, using mean imputation, reimpute the missing values in s_h^*, and let

$$z_{hi}^{*I} = \begin{cases} z_{hi}^*, & \text{if } r_{hi}^* = 1, \\ \bar{z}_{hr}^*, & \text{if } r_{hi}^* = 0, \end{cases}$$

where $\bar{z}_{hr}^* = \sum_{i=1}^{n'_h} r_{hi}^* z_{hi}^* / \sum_{i=1}^{n'_h} r_{hi}^*$ for $i = 1, \cdots, n'_h$.

(4) Repeat Steps 2 and 3 independently across the strata to get $\{z_{11}^{*I}, \cdots, z_{Ln'_L}^{*I}\}$.

(5) The bootstrap estimator of the mean is defined as $\hat{\theta}^{*I} = \sum_{h=1}^{L} W_h\ \bar{z}_h^{*I}$, where $\bar{z}_h^{*I} = \sum_{i=1}^{n'_h} z_{hi}^{*I} / n'_h$ and we use them to estimate $V(\hat{\theta}^I)$.

Comparing the complete data rescaling factor $C_h = [n'_h/(n_h - 1)][1 - f_h]$ (Rao and Wu, 1988) with the modified factor $C_h^I = [n'_h/(n_{hr} - 1)][1 - \hat{p}_{0h} f_h]$, we note that since $\hat{p}_{0h} f_h = n_{hr}/N_h$ then C_h^I is similar to C_h but using the size of the respondents in the sample rather than the sample size.

We show in the Appendix A that the bootstrap variance of $\hat{\theta}^{*I}$ under the proposed procedure is approximately equal to the usual consistent estimator of $V(\hat{\theta}^I)$ obtained from a first-order Taylor expansion and so

$$E_p E_q [V^*(\hat{\theta}^{*I})] \approx V(\hat{\theta}^I) \quad \text{or} \quad E_p E_q E_I [V^*(\hat{\theta}^{*I})] \approx V(\hat{\theta}^I),$$

in the case of deterministic or random imputation, respectively.

2.4.2. Modified Constants for the BRS and BMM Methods

Applying the independent bootstrap method with a complete data bootstrap procedure requires some modifications of its constants. In this section, we present the modified constants for the BRS and the mirror match bootstrap method (BMM) of Sitter (1992b) when using the independent bootstrap method to guarantee a valid variance estimator in the case of the mean. These constants depend on the original imputation method, the response rate, and the auxiliary variables, when appropriate.

For simplicity of notation, the subscript h will be suppressed in the sequel. Suppose ratio imputation (RI) based on x is used. We define

$$\rho_I = \frac{s_{xy^I}}{s_x\ s_{y^I}}, \quad CV(x) = \frac{s_x}{\bar{x}}, \quad CV(y^I) = \frac{s_{y^I}}{\bar{y}^I} \quad \text{and} \quad R_I = \frac{CV(x)}{CV(y^I)},$$

where s_x^2 and $\bar{x}$ are the sample variance and mean of the auxiliary variable, respectively, computed on the original sample s, and $s_{xy^I} = \sum_{i\in s}(x_i - \bar{x})(y_i^I - \bar{y}^I)/(n-1)$. For linear regression imputation (LRI), let $CC_I^2 = s'_{\mathbf{x}y^I}\ s_{\mathbf{xx}}^{-1}\ s_{\mathbf{x}y^I}/s_{y^I}^2$ with

$$s_{\mathbf{xx}} = \frac{1}{n-1}\sum_{i\in s}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' \quad \text{and} \quad s_{\mathbf{x}y^I} = \frac{1}{n-1}\sum_{i\in s}(\mathbf{x}_i - \bar{\mathbf{x}})(y_i^I - \bar{y}^I)$$

denote the square correlation coefficient between the y_i^I's and $\mathbf{x}_i$'s on the imputed sample where $\bar{\mathbf{x}} = \sum_{i\in s}\mathbf{x}_i/n$.

The following table shows the original and modified constants to be used in the BRS or BMM methods for complete and imputed data sets, respectively.

TABLE 2.1. Rescaling factors for complete data (C. D.) and for imputed data using the independent bootstrap

	C^I in BRS (C for complete case)	k^I in BMM (k for complete case)
C. D.	$\frac{n'}{n-1}\,[1-f]$	$\frac{n}{n''}\left[\frac{1-f'}{1-f}\right]$
MI	$\frac{n'}{n_r-1}\,[1-\hat{p}_0 f]$	$\frac{n_r}{n''}\left[\frac{1-\hat{p}_0 f'}{1-\hat{p}_0 f}\right]$
RI	$\frac{n'}{n_r-1}\left[1-\frac{\hat{p}_0 f+\hat{p}_0(1-\hat{p}_0)(1-f)R_I^2}{1+(1-\hat{p}_0)R_I(R_I-2\rho_I)}\right]$	$\frac{n_r}{n''}\left[\frac{1-\hat{p}_0 f'+(1-\hat{p}_0)R_I\{R_I-2\rho_I\}}{1-\hat{p}_0 f+(1-\hat{p}_0)R_I[\{1-(1-f)\hat{p}_0\}R_I-2\rho_I]}\right]$
LRI	$\frac{n'}{n_r-1}\left[1-\hat{p}_0 f-\frac{\hat{p}_0(1-\hat{p}_0)CC_I^2}{1+(1-\hat{p}_0)CC_I^2}\right]$	$\frac{n_r}{n''}\left[\frac{1-\hat{p}_0 f'+(1-\hat{p}_0)CC_I^2}{1-\hat{p}_0 f+(1-\hat{p}_0)\{1-\hat{p}_0(1+f)\}CC_I^2}\right]$
RHDI	$\frac{n'\hat{p}_0}{n_r-1}\left[\frac{(1-\hat{p}_0 f)+\hat{p}_0(1-\hat{p}_0)}{\left\{1-\frac{1}{n}(1-\hat{p}_0)\right\}\{1+\hat{p}_0(1-\hat{p}_0)\}}\right]$	$\frac{n}{n''}\left[\frac{\left\{1-f'+\hat{p}_0^{-1}(1-\hat{p}_0^2)\right\}\left\{1-\frac{1}{n}(1-\hat{p}_0)\right\}}{1-f+\hat{p}_0^{-1}\left(1-\hat{p}_0^2\right)}\right]$

All the modified constants depend on the response rate, $\hat{p}_0 = n_r/n$. It is straightforward to see that these constants converge to that of complete data when $\hat{p}_0$ tends toward 1.

2.4.3. Bootstrap Weights Approach

We now present the bootstrap weights approach to the independent bootstrap for imputed data. As with complete data, the approach involves the same constant as with the BRS method. So the constants C_h^I for the BRS presented in Table 2.1 can be used. Let us see how to implement the method in practice. We let $B_h^* = [b_{hij}^*]$ be the $n_h \times n_h'$ selection matrix with $b_{hij}^* = 1$ if y_{hi} is selected in the j-th trial, and $b_{hij}^* = 0$ otherwise, for $i = 1, \cdots, n_h$ and $j = 1, \cdots, n_h'$. We also let m_{hi}^* be the sum of the i-th row of B_h^* indicating the number of times the hi-th unit is selected in the bootstrap sample. After drawing the matrix B_h^*, response indicators are regenerated according to $\{r_{hj}^*\}_{j=1}^{n_h'} \overset{i.i.d.}{\sim} Bernoulli(\hat{p}_{0h})$. Define the $n_h \times n_h'$ respondents matrix $B_{hr}^* = [b_{hrij}^*]$ by $b_{hrij}^* = r_{hj}^* b_{hij}^*$. Let R_{hi}^* be the sum of the i-th row of B_{hr}^*, representing the number of times the (hi)-th unit is selected in the bootstrap sample of *respondents*, and let $R_h^* = \sum_{i=1}^{n_h} R_{hi}^*$ be the size of the bootstrap sample of respondents. Let

$$a_{hi}^* = 1 + \sqrt{C_h^I}\left(\frac{n_h m_{hi}^*}{n_h'} - 1\right) \quad \text{and} \quad c_{hi}^* = 1 + \sqrt{C_h^I}\left(\frac{n_h R_{hi}^*}{R_h^*} - 1\right).$$

To define the bootstrap weights, we let $\mathbf{u}_h = \bar{\mathbf{x}}_h + \sqrt{C_h^I}\,(\mathbf{x}_h - \bar{\mathbf{x}}_h)$,

$$\mathbf{k}^* = (k_1^*, \cdots, k_{n_h}^*)' = \left(\frac{1}{N_h}\sum_{i \in s_h}[a_{hi}^* - c_{hi}^*] w_{hi}\, \mathbf{x}_{hi}'\right)\, \left(\mathbf{u}_h' B_{hr}'^* B_{hr}^* \mathbf{u}_h\right)^{-1} \mathbf{u}_h' B_{hr}'^* B_{hr}^*,$$

and define $\mathbf{1}_h$ to be a n_h vector of 1. The bootstrap estimator of $\hat{\theta}^I$ is $\hat{\theta}^{*I} = \sum_{h=1}^{L} W_h \left[\sum_{i \in s_h} w_{hi}^* y_{hi}^I / N_h\right]$, where w_{hi}^* is presented in the following table.

To determine w_{hi}^* under random hot-deck imputation, we need to consider an additional random step in the bootstrap procedure to identify the donors among the respondents in the bootstrap sample. To apply RHDI and to draw the bootstrap sample of donors, a simple random sample of size $n_h' - R_h^*$ is taken with replacement from the set of the index positions of the bootstrap sample of respondents in the original sample, in which the index position i, corresponding

Table 2.2. Bootstrap weights for the independent bootstrap

	w^*_{hi} in BW
MI	$c^*_{hi}w_{hi}$
RI	$\left[\left(\sum_{i\in s_h} a^*_{hi}w_{hi}\, x_{hi}\right) \Big/ \left(\sum_{i\in s_h} c^*_{hi}w_{hi}\, x_{hi}\right)\right]\ c^*_{hi}w_{hi}$
LRI	$\left[c^*_{hi} + \mathbf{1}'_h\mathbf{k}^* + \sqrt{C^I_h}\,(nk^*_i - \mathbf{1}'_h\mathbf{k}^*)\right] w_{hi}$

to the (hi)-th unit in the original sample s_h is repeated R^*_{hi} times, the number of times this observation appears in the bootstrap sample of respondents. Define by D^*_{hi} the number of times the index position i shows up in the selected bootstrap sample of *donors*, for $i = 1, \cdots, n_h$, i.e., the number of times the (hi)-th unit in s_h is selected in the bootstrap sample of donors. The bootstrap weights are given by

$$w^*_{hi} = \left[1 + \sqrt{C^I_h}\left(\frac{n_h(R^*_{hi} + D^*_{hi})}{n'_h} - 1\right)\right] w_{hi}.$$

Note that $R^*_{hi} + D^*_{hi}$ represents the number of times the (hi)-th unit appears in the bootstrap sample of observed and reimputed data.

Not only are these bootstrap weights methods easy to use in practice, but they also estimate the variance of $\hat{\theta}^I$ consistently even when the sampling fraction is not negligible.

As mentioned earlier, the data sets in research data centers often include bootstrap weights which are appropriate only for complete data. We now show how these complete data bootstrap weights can be transformed in order to compute independent bootstrap weights appropriate for missing data in the case of the mean, ratio and random hot-deck imputation methods. Three elements are always needed to compute the independent weights: C^I_h, m^*_{hi} and R^*_{hi}. For RDHI, D^*_{hi} is also needed. Table 2.1 contains C^I_h which must be selected according to the imputation method used. Since the complete data bootstrap weights are given by $w^*_{hi} = \left[1 + \sqrt{C_h}\left(\frac{n_h m^*_{hi}}{n'_h} - 1\right)\right] w_{hi}$ they can easily be transformed to obtain m^*_{hi}, i.e.,

$$m^*_{hi} = \frac{n'_h}{n_h}\left[1 + \frac{1}{\sqrt{C_h}}\left(1 - \frac{w^*_{hi}}{w_{hi}}\right)\right].$$

It is straightforward to check that

$$R^*_{hi} \sim Bin(m^*_{hi}; \hat{p}_{0h}) \quad \text{and} \quad D^*_{hi} \sim Multinomial\left(n'_h - R^*_h; 1/R^*_h, \cdots, 1/R^*_h\right).$$

Therefore, if one generates these two quantities, all that is needed to compute the independent bootstrap weights will be available.

In the Appendix A, we show that applying the independent bootstrap with BRS or BW leads to identical estimators when the parameter of interest is a function of means.

2.5. A Modified Shao-Sitter Method for Non-negligible Sampling Fraction

Assuming that the response status is available in the data file, an interesting question is whether the Shao-Sitter method, i.e., a paired bootstrap, can be modified to work even in the case of a large sampling fraction f.

The original Shao-Sitter method uses the complete data bootstrap methods whose constants only account for sampling variability and not the variability due to non-response and imputation. As will be shown in Equation (2.A.3) of the appendix A, the second term V_2 in the decomposition of the variance of $\hat{\theta}^I$, an estimator based on deterministic imputation, is a multiple of the first one, so that the variance is a multiple of the first term V_1. Since the Shao-Sitter method provides an estimate of the first term only, by modifying the constants of the complete data bootstrap methods, it will be possible to estimate the total variance, even for large f. For simplicity, we consider the case $L = 1$.

Consider for instance mean imputation. Using a first-order Taylor expansion, the variance of $\hat{\theta}^I = \bar{y}^I$ under MI can be approximated by

$$\begin{aligned} V(\hat{\theta}^I) &= V_1 + V_2 \\ &\approx \frac{1-f}{p_0 f} \frac{1}{N^2} \sum_{i \in U} (y_i - \bar{Y})^2 + \frac{1-p_0}{p_0} \frac{1}{N^2} \sum_{i \in U} (y_i - \bar{Y})^2 \qquad (2.5.1) \\ &= \alpha^{MI} V_1, \end{aligned}$$

where $\alpha^{MI} = (1 - p_0 f)/(1 - f)$. Note that $\alpha^{MI} \approx 1$ if f is negligible. Now, an estimate of V_1, such as the Shao-Sitter estimator, can be used to estimate the total variance by multiplying it by an estimator of α^{MI}. For instance, let

$$\hat{\alpha}^{MI} = \frac{1 - \hat{p}_0 f}{1 - f}$$

and let $\widehat{V}^*_{Sh.S.}$ be the Shao-Sitter estimator of $V(\hat{\theta}^I)$. Then $\hat{\alpha}^{MI}\widehat{V}^*_{Sh.S.}$ is a consistent estimate of the variance, even for large f. Alternatively, let C'^{MI} be a modified constant for the BRS (or BW) defined by $C'^{MI} = \hat{\alpha}^{MI} C$ where C is the complete data constant. Then applying the paired bootstrap (i.e., the Shao-Sitter method) using this constant and the same reimputation method on the bootstrap nonrespondents will lead to a consistent variance estimator. Similar adjustments can be made for other deterministic imputation methods.

As discussed in Section 2.3 in the case of RHDI, the variance of the imputed estimator $V(\hat{\theta}^I)$ is the sum of three terms $\widetilde{V}_1$, $\widetilde{V}_2$, and $\widetilde{V}_3$, and the Shao-Sitter method (with the usual complete data bootstrap constants) estimates $\widetilde{V}_1 + \lambda_f \widetilde{V}_3$, where $\widetilde{V}_1$ and $\widetilde{V}_2$ are defined as V_1 and V_2 in (2.5.1) and

$$\widetilde{V}_3 = \frac{1 - p_0}{nN} \sum_{i \in U} (y_i - \bar{Y})^2.$$

It is straightforward to see that

$$V(\hat{\theta}^I) \approx \left[\frac{\{1 + p_0 (1 - p_0)\} - p_0 f}{\{1 + p_0 (1 - p_0)\} (1 - f)} \right] [\widetilde{V}_1 + \lambda_f \widetilde{V}_3] = \alpha^{RHDI} \, [\widetilde{V}_1 + \lambda_f \widetilde{V}_3]. \tag{2.5.2}$$

Therefore, multiplying the Shao-Sitter variance estimator by $\hat{\alpha}^{RHDI}$, where $\hat{\alpha}^{RHDI}$ is as in (2.5.2) with p_0 replaced by $\hat{p}_0$, leads to a valid estimator for $V(\hat{\theta}^I)$. Again replacing the constant C in the BRS (BW) method by $\hat{\alpha}^{RHDI} C$ and applying the Shao-Sitter method results in a consistent variance estimator.

2.6. Simulation Study

To compare the performance of the proposed methods with the existing methods, we performed a simulation study. A description of this simulation experiment is presented in Section 2.6.1. A discussion of the results follows in Section 2.6.2.

2.6.1. Description of the Simulation Study

Given that the behaviour of an estimator in survey sampling critically depends on the sample size and the sampling fraction, we have designed our simulation experiment in a factorial way with two levels for the sample size ($n_1 = 100$ and $n_2 = 400$) and two levels for the sampling fraction ($f_1 = 5\%$ and $f_2 = 50\%$). Combining the two levels of the two factors in a factorial way leads to four population sizes N, i.e., 200, 800, 2 000, and 8 000. Rather than generating four separate populations, we generated a single population of size 8 000 consisting of a single stratum (i.e., $L = 1$) and we considered the first 200, 800 or 2 000 units from that population when such populations were needed depending on the combination of sample size and sampling fraction. An auxiliary variable x was first generated from a gamma distribution with scale and shape parameters equal to 3 and 7 (with mean of 21), respectively. Given the x-values, the characteristic of interest y was generated according to the model $y_i = 0.1\ x_i + \varepsilon_i$, $i = 1, \cdots, 8000$, where ε_i follows a standard normal distribution. The correlation between x and y is 0.77. For each simulation, the goal is to estimate the variance and compute 95% confidence intervals for the population mean estimator using ratio imputation and for the estimator of the population median using RHDI.

The bootstrap weights point of view was considered in all of the following bootstrap methods: the independent bootstrap, the modified Shao-Sitter estimator presented in Section 2.5, the original Shao-Sitter method and finally the naive method, where the imputed data are treated as true observations. In addition, in the case of the population mean estimator with ratio imputation, we computed the variance estimators using the linearization method. Note that since the population median is not a function of totals, this method is not applicable. For the two parameters, we also computed 95% bootstrap percentile confidence intervals (see Efron and Tibshirani, 1993) as well as normal-based confidence intervals using the bootstrap estimate of variance for all bootstrap methods (with the exception of the linearization method).

Along with the factors sample size and sampling fraction, we have crossed them with the factor response rate of the uniform non-response mechanism, with two levels: $f_{r1} = 60\%$ and $f_{r2} = 80\%$, leading to a total of eight scenarios. In each scenario, we drew $S = 2000$ random samples from the corresponding population. To apply any bootstrap method, we drew $B = 1000$ bootstrap samples from each sample with bootstrap sample size $n' = n - 3$, as suggested by Rao and Wu (1988).

We computed the following quantities in each scenario. Suppose that $\hat{\theta}^I_j$ is the estimator of the parameter of interest (mean or median) calculated on the j-th sample, $j = 1, \cdots, S$. The Monte Carlo variance estimator of $\hat{\theta}^I$ is

$$V_{MC}(\hat{\theta}^I) = \frac{1}{S-1} \sum_{j=1}^{S} \left(\hat{\theta}^I_j - \hat{\theta}^I_{(\cdot)} \right)^2, \quad \text{where } \hat{\theta}^I_{(\cdot)} = \frac{1}{S} \sum_{j=1}^{S} \hat{\theta}^I_j,$$

which is used as a consistent estimator of $V(\hat{\theta}^I)$. Fixing a method to estimate $V(\hat{\theta}^I)$, the Monte Carlo average and the Monte Carlo variance of the variance estimator of $\hat{\theta}^I$ denoted by $\hat{V}$ is respectively

$$E_{MC}(\hat{V}) = \frac{1}{S} \sum_{j=1}^{S} \hat{V}_j \quad \text{and} \quad V_{MC}(\hat{V}) = \frac{1}{S-1} \sum_{j=1}^{S} \left(\hat{V}_j - E_{MC}(\hat{V}) \right)^2,$$

where $\hat{V}_j$ is the variance estimator on the j-th sample. As a measure of bias of a variance estimator $\hat{V}$, we use the Monte Carlo percent relative bias (RB) defined by

$$RB_{MC}(\hat{V}) = 100 \times \frac{E_{MC}(\hat{V}) - V_{MC}(\hat{\theta}^I)}{V_{MC}(\hat{\theta}^I)}.$$

Another measure used in the next section is the Monte Carlo percent relative efficiency (RE) which is

$$RE_{MC}(\hat{V}) = 100 \times \frac{MSE_{MC}(V^*_{Sh.S.})}{MSE_{MC}(\hat{V})},$$

where $MSE_{MC}(\hat{V}) = V_{MC}(\hat{V}) + \left[E_{MC}(\hat{V}) - V_{MC}(\hat{\theta}^I) \right]^2$. A value greater than 100 means that $\hat{V}$ is more precise than the Shao-Sitter method. We also report the coverage probability of the 95% confidence intervals which were computed. Note that at the 5% level, the coverage probability is statistically different from the nominal level if it falls outside the interval $[94.04, 95.96]$. In the next section, we use these measures to compare the performance of different methods to estimate the variance.

2.6.2. Simulation Results

The Monte Carlo RB and RE of the variance estimators and the coverage probability of the two 95% confidence intervals are shown in Tables 2.3-2.6. In the case of the Shao-Sitter method, since its RE is by definition 100, we present its $MSE_{MC}(V^*_{Sh.S.})$ instead. Tables 2.3 and 2.4 contain the results for the population mean estimator with ratio imputation for the non-respondents while Tables 2.5 and 2.6 are for the population median estimator with RHDI imputation since it preserves the distribution of observations unlike a deterministic imputation method.

Table 2.3. RE (in parenthesis) and RB of the variance estimators for the population mean and ratio imputation using 2000 samples of size $n_1 = 100$ and $n_2 = 400$. The italic numbers of line * are $MSE_{MC}(V^*_{Sh.S.}) \times 10^7$.

	f_1=5%				f_2=50%			
	$f_{r1} = 60\%$		$f_{r2} = 80\%$		$f_{r1} = 60\%$		$f_{r2} = 80\%$	
	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2
Independent	-0.32	-0.84	-3.18	0.67	-8.67	-2.26	-3.10	-1.82
	(103.51)	(108.47)	(102.79)	(99.56)	(288.73)	(570.94)	(154.80)	(270.80)
Mod. Sh.-S.	-0.96	-0.82	-3.57	0.48	-9.22	-2.25	-3.43	-1.86
	(98.60)	(100.71)	(100.63)	(99.02)	(265.62)	(574.90)	(148.84)	(263.10)
Linearization	-1.46	-1.16	-3.94	0.36	-9.76	-2.48	-3.58	-1.88
	(94.96)	(119.96)	(99.38)	(119.39)	(238.90)	(689.45)	(137.34)	(352.87)
Shao-Sitter	-2.07	-1.93	-4.03	-0.0034	-25.48	-19.69	-11.64	-10.28
*	*(307.67)*	*(6.19)*	*(192.58)*	*(3.68)*	*(373.15)*	*(10.08)*	*(75.30)*	*(2.11)*
Naive	-36.44	-35.37	-20.34	-16.46	-51.64	-47.95	-26.91	-25.59
	(21.35)	(8.06)	(43.79)	(25.72)	(28.66)	(18.56)	(35.34)	(21.92)

We begin with the case of the mean estimator and a negligible sampling fraction of $f_1 = 5\%$. The naive method, *which is what is used whenever someone*

Table 2.4. Coverage probability of the 95% bilateral percentile bootstrap and normal confidence intervals based on a standard error computed from the corresponding bootstrap method (the latter coverage probability in parenthesis) for the population mean with ratio imputation using 2000 samples of size $n_1 = 100$ and $n_2 = 400$.

	f_1=5%				f_2=50%			
	$f_{r1} = 60\%$		$f_{r2} = 80\%$		$f_{r1} = 60\%$		$f_{r2} = 80\%$	
	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2
Independent	94.50	94.00	94.25	94.50	94.10	94.75	94.25	94.60
	(94.60)	(94.05)	(94.35)	(94.70)	(94.25)	(94.65)	(94.40)	(94.85)
Mod. Sh.-S.	94.55	94.20	94.20	94.70	93.75	95.05	94.00	94.70
	(94.80)	(93.90)	(94.05)	(94.50)	(93.85)	(95.35)	(93.85)	(95.10)
Shao-Sitter	94.35	93.95	94.15	94.70	90.70	91.85	92.85	93.75
	(94.65)	(93.85)	(94.05)	(94.50)	(90.60)	(91.95)	(92.90)	(93.70)
Naive	87.95	87.20	91.45	92.75	81.05	84.85	90.10	91.35
	(88.40)	(87.65)	(91.40)	(92.70)	(81.45)	(85.05)	(90.15)	(91.25)

uses the bootstrap weights included in a dataset of a research data centre along with imputed data, treating them as if they were true observations, has very large negative biases leading to very poor efficiency and is the worst method, as was expected. All other methods have small relative biases and all relative efficiencies are around 100 meaning that they all have the same level of efficiency. For the non-negligible sampling fraction $f_2 = 50\%$, we can see the good performance of the independent, modified Shao-Sitter, and linearization methods in terms of bias and efficiency. The high relative bias of the Shao-Sitter method and its inefficiency compared with the first three methods confirms its poor performance for a large sampling fraction. The two confidence intervals for the independent bootstrap are not significantly different from 95% except for the small sampling fraction, smaller response rate and larger sample size where the coverage probability is slightly outside the interval $[94.05, 95.96]$; most of the confidence intervals for the other bootstrap methods are significantly different in this case. The confidence intervals for the modified Shao-Sitter method are mostly not different from 95%, but more of them are different than with the independent method. For the Shao-Sitter some intervals are significantly different for a 5% sampling fraction, but all intervals with a 50% sampling fraction are different, especially those with a 60% response rate where the coverage probability is in the 90-92% range. The bad behavior of the naive method in terms of bias and efficiency translates itself in very bad coverage (between 81% and 93%).

TABLE 2.5. RE (in parenthesis) and RB of the variance estimators for the population median and RHDI using 2000 samples of size $n_1 = 100$ and $n_2 = 400$. The italic numbers of line * are $MSE_{MC}(V^*_{Sh.S.}) \times 10^5$.

	f_1=5%				f_2=50%			
	$f_{r1} = 60\%$		$f_{r2} = 80\%$		$f_{r1} = 60\%$		$f_{r2} = 80\%$	
	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2
Independent	22.08	13.38	20.13	10.82	11.87	8.77	2.72	10.89
	(69.36)	(71.02)	(83.60)	(80.56)	(57.65)	(89.45)	(73.96)	(69.97)
Mod. Sh.-S.	18.41	11.22	18.45	9.31	8.30	6.85	1.44	9.20
	(94.26)	(89.65)	(97.29)	(93.09)	(81.32)	(119.97)	(91.98)	(87.03)
Shao–Sitter	14.90	6.50	16.96	6.19	-25.66	-28.90	-19.01	-16.52
*	*(157.49)*	*(4.16)*	*(62.49)*	*(1.84)*	*(47.73)*	*(2.81)*	*(16.41)*	*(0.653)*
Naive	-39.004	-45.31	-15.97	-24.33	-60.23	-63.57	-41.38	-41.03
	(108.02)	(53.66)	(137.88)	(89.41)	(46.17)	(35.60)	(59.74)	(47.74)

In the case of the median and a negligible sampling fraction of $f_1 = 5\%$, all methods are biased with the naive method being negatively biased (between -16% and -45%), as expected, while the other methods are positively biased. The independent method has a larger bias than the modified Shao-Sitter method, followed by the Shao-Sitter method with biases among the three methods between 6% and 22%. As is often the case with variance estimators, when their mean is smaller, their variance is also smaller. Consequently, the Shao-Sitter method has the best relative efficiency except for small sample size where the naive method,

TABLE 2.6. Coverage probability of the 95% bilateral percentile bootstrap and normal confidence intervals based on a standard error computed from the corresponding bootstrap method (the latter coverage probability in parenthesis) for the population median with RHDI using 2000 samples of size $n_1 = 100$ and $n_2 = 400$.

	f_1=5%				f_2=50%			
	$f_{r1} = 60\%$		$f_{r2} = 80\%$		$f_{r1} = 60\%$		$f_{r2} = 80\%$	
	n_1	n_2	n_1	n_2	n_1	n_2	n_1	n_2
Independent	95.80	95.80	95.15	95.45	94.90	94.65	95.00	94.85
	(94.60)	(94.60)	(94.60)	(94.45)	(94.10)	(93.30)	(91.80)	(94.65)
Mod. Sh.-S.	97.30	97.00	96.55	96.25	97.80	97.60	96.80	97.15
	(95.80)	(95.00)	(94.85)	(94.35)	(95.35)	(93.90)	(92.90)	(94.90)
Shao–Sitter	97.10	96.70	96.40	95.90	92.95	93.65	94.15	94.45
	(95.05)	(94.40)	(94.45)	(94.30)	(88.10)	(89.00)	(88.15)	(92.65)
Naive	83.15	81.80	89.40	88.90	73.45	73.10	84.60	85.45
	(83.30)	(83.55)	(89.25)	(89.90)	(69.90)	(76.85)	(80.20)	(86.60)

which leads to small variance estimates (because of its large negative bias), has the largest relative efficiency; more on that method when we discuss the confidence intervals. For the larger sampling fraction of 50%, the relative bias of all methods decreases leading to very large negative bias for the naive method (the best being −41%), and relatively large negative bias for the Shao-Sitter method (between −17% and −29%). The other two methods have smaller bias, but it is positive (the worst being 12%), resulting in worse relative efficiency than the Shao-Sitter method, as previously discussed. It should be noted that even with 100% response, the bootstrap estimate of the variance of the median can be quite biased, see e.g., Sitter (1992b). In fact for our four scenarios, in results that we do not include in our tables, the bias for complete response with samples of size 100 oscillates between −7% and 17% depending on the sampling fraction, whereas it still oscillates between 3% and 12% for samples of size 400.

Estimation of the variance is important, but often the ultimate goal is a confidence interval and the estimation of the variance is sometimes just a step towards the construction of a normal-based confidence interval. If the bootstrap is used, then it is possible to compute percentile intervals rather than compute a variance estimator to use with normal quantiles. While the independent bootstrap method did not perform as well as the Shao-Sitter method in terms of the relative efficiency of its variance estimator, the coverage probability of its bootstrap percentile intervals are never significantly different from the claimed level of 95%. In two cases with large sampling fraction, its normal-based intervals are significantly different with coverage probabilities of 91.8% and 93.3%. In the same two cases, the normal-based intervals using the modified Shao-Sitter variance estimate are also different from the claimed level, but all other cases are good. On the other hand, all bootstrap percentile intervals using this method have larger coverage between 96.25% and 97.80%. In cases where the sampling fraction is small, the original Shao-Sitter method has good coverage for its normal-based intervals and some overcoverage for three of its four bootstrap percentile intervals. But for large sampling fraction (50%), all of its normal-based intervals undercover (between 88.10% and 92.65%) while the bootstrap percentile intervals are better with two not significantly different from 95%. Finally, as expected, all intervals from the naive method drastically undercover with coverage probabilities between 73.10% and 89.90%.

2.7. Application: Research and Development in Canadian Industry Survey for 2008

In this section, we present results obtained using data from the Research and Development in Canadian Industry (RDCI) survey conducted at Statistics Canada. The RCDI is used to analyze the relationship between the size of the firm and the proportion of expenditures spent on research and development (R&D). A stratified simple random sample without replacement design was selected from the Canadian Business Register. All must-take enterprises formed one stratum. The remaining strata were defined at the NAICS5 (North American Industry Classification System 5-digit) level. Then the smallest enterprises making no more than 5% of the total SIZE (the sum of extramural payments or contracting-out – EXTOT – and the total intramural spending – TIE – variables) were put in take-none strata within each NAICS5, to reduce the response burden on the smallest enterprises. The remaining enterprises in each stratum were divided into three substrata: a take-all stratum, which would consist of the largest enterprises that were clearly larger than the remainder, and two take-some strata with the medium-size enterprises put into a substratum with a higher sampling fraction than the one containing the smallest take-some enterprises.

To apply the bootstrap methods, we dropped the take-none and the fully observed take-all strata since both types of strata do not contribute to the variance of point estimators. In addition, to avoid small numbers of respondents within strata, strata were collapsed with other strata belonging to the same NAICS group to have at least 3 respondents per stratum. At the end, the number of strata was equal to 122 and the population and sample sizes were 13,289 and 1,562, respectively. As the variable of interest, we chose *Expenditures in Canada planned for 2009 for R&D.*

In this study, we were interested in estimating several population parameters: the mean, the first quartile (Q_1), the median and the third quartile (Q_3). To replace the missing values when estimating the mean, we used mean and random hot-deck imputation. For the median and the quartiles, we only considered hot-deck imputation. We estimated the variance of the resulting imputed estimators using several bootstrap procedures: the independent bootstrap of Section 2.4, the Shao-Sitter method of Section 2.3 and the naive procedure of Section 2.6. In the case of the population mean with mean imputation, we also computed the linearization variance estimator given by $\widehat{V}(\bar{y}^I) = \hat{V}_1 + \hat{V}_2$, where the two terms are computed by appropriately modifying formulas (2.A.4) and (2.A.5).

The first column of Table 2.7 shows the different variance estimates for the mean estimator under mean imputation. Note that the linearization variance estimates $\hat{V}_1$ and $\hat{V}_2$ were respectively equal to 1,581.47 and 6,055.63 for a total variance estimate of 7,637.10. It is interesting to note that, in our example, the second term was considerably larger than the first term even though the overall

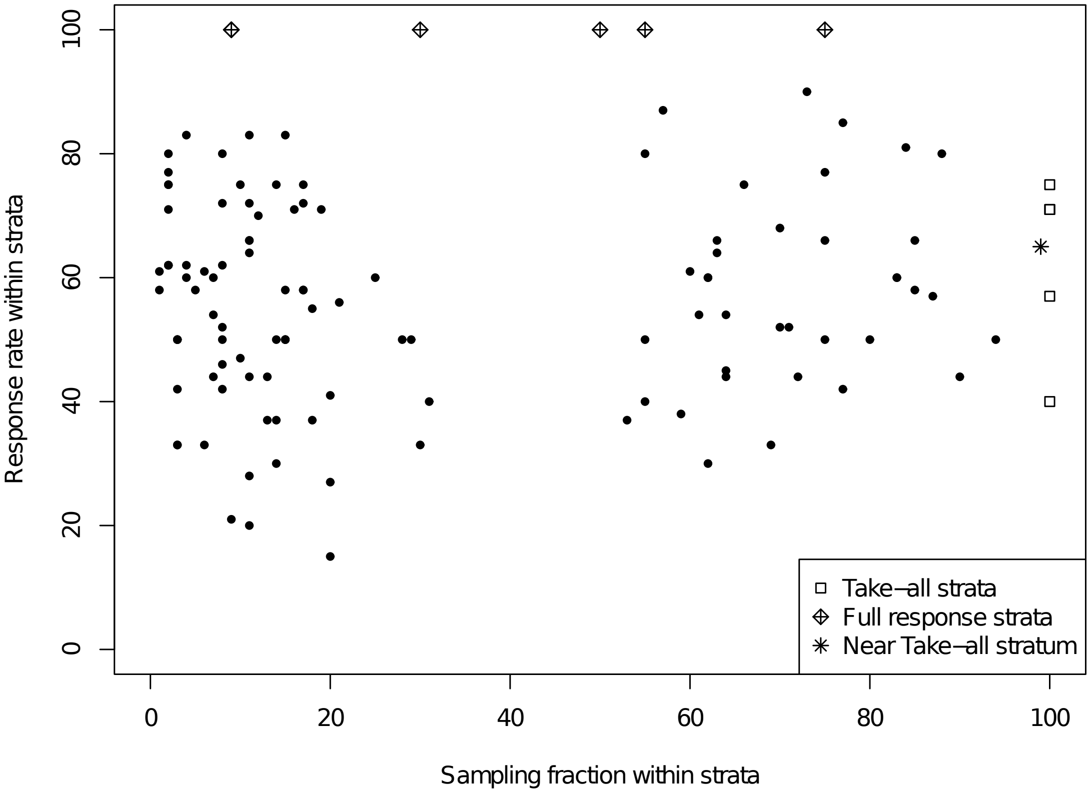
Table 2.7. Bootstrap Variance Estimates of the Mean under Mean and Random Hot Deck Imputation

	Mean Imputation		RHDI	
Bootstrap Method	All Strata	Without the Near Take-all	All Strata	Without the Near Take-all
Independent	7,264.15	2,057.59	10,065.99	2,409.84
Shao-Sitter	1,552.17	1,477.64	1,697.68	1,587.53
Naive	421.26	381.13	765.81	716.86
Linearization	7,637.10	1,802.48	11,899.66	2,302.09

sampling fraction at 12% is not particularly large. This can be explained by the presence of take-all or near take-all strata, with relatively large non-response which do not contribute to the first term $\hat{V}_1$ (or contribute little in near take-all strata), but can contribute largely to $\hat{V}_2$, especially if the stratum is large. Moreover, in this type of economic study, some units will be particularly large. Figure 2.1 shows the sampling fraction and the response rate in each of the 122 strata. The overall response rate is 58%, but we see that many strata have low response rates, including some with a large sampling fraction. We distinguish between three sets of strata. The strata identified by a square are take-all strata of size between 4 and 10 with non-response. Such strata do not contribute to the first term of variance as there is no sampling, but they do contribute to the second term because of the non-response. For these strata, there is no contribution to the variance estimate of the Shao-Sitter or naive methods. In the case of the Shao-Sitter estimator based on the bootstrap weights estimator, since the sampling fraction f_h is 1, then the rescaling constant C_h of the BRS method is 0 so that the bootstrap weights are always $w^*_{hi} = w_{hi}$ for all bootstrap samples leading to no variability. The strata identified by a triangle are the only six full-response strata. These strata only contribute to the first term of variance since there is no non-response variation in this case and so all variance estimators are similar.

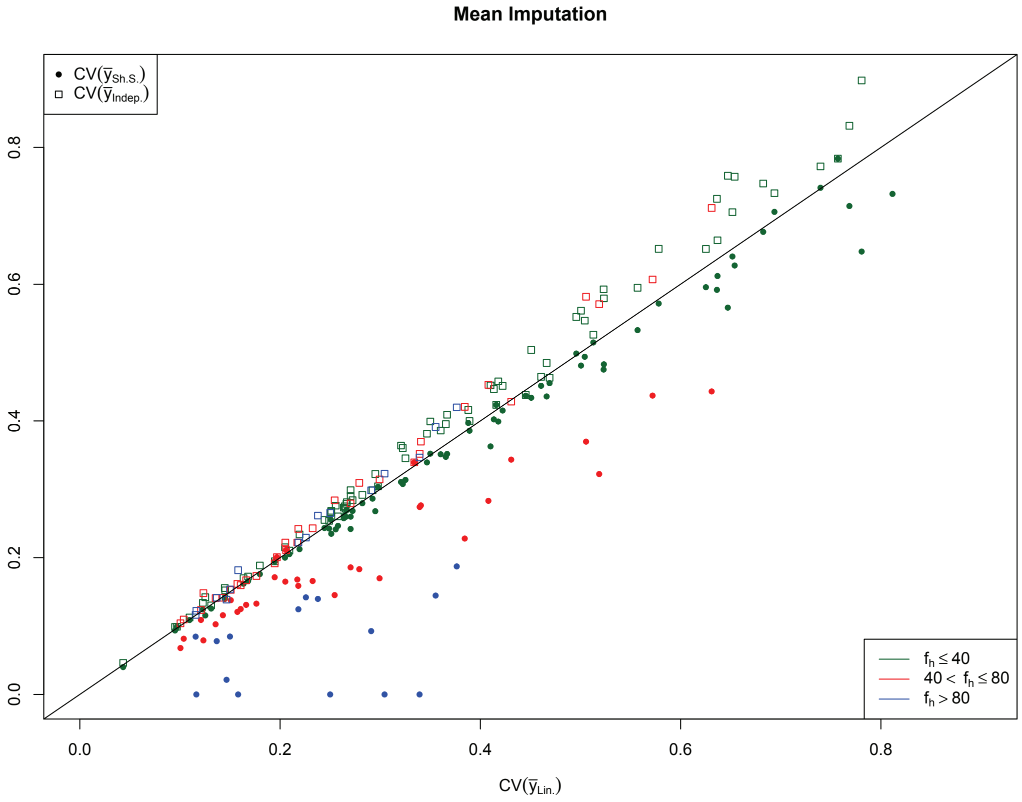
The stratum identified by a star is a near take-all stratum in that although all 237 units were contacted, only 235 were reached and 153 responded to that

FIGURE 2.1. Sampling fraction and response rates of the 122 strata



question. Since f_h is near 1, C_h is near 0 and there is very little variability in the bootstrap weights so that the Shao-Sitter variance estimator for that stratum is small, as is the naive estimator. Moreover, the stratum was planned to be take-all because of the importance of its units and so it is not surprising that some of its values are large. Hence, because of the large within-stratum variance, its relatively large size, and a relatively small response rate, the contribution of this stratum to the overall variance is very important. The second column of Table 2.7 shows the variance of the mean estimator under mean imputation once we remove the near take-all stratum. We see that the independent bootstrap and linearized estimators are relatively close whereas the Shao-Sitter is somewhat smaller, but the difference is much less important than it is when all strata are included. In this case, the Shao-Sitter estimator is about 20% of the independent bootstrap or linearized variance estimators. Finally, it is worth reminding the reader that the naive variance estimator is what would be obtained if we used the bootstrap weights method with the weights computed to reflect the sampling uncertainty, precisely the weights that would be provided to users. The underestimation of the variance is tremendous as it reflects only 6% of the independent bootstrap estimate with all strata or 19% without the near take-all stratum.

FIGURE 2.2. CV of the three estimators in terms of the sampling fraction



To better understand the effect of the sampling fraction on the estimation of the variance within each stratum, Figure 2.2 shows the coefficient of variation (CV) of within stratum point estimates $\bar{Y}_h$ obtained using the linearized variance estimator on the x-axis and the corresponding CV of the Shao-Sitter (circles) or the independent (squares) bootstrap methods on the y-axis for all strata. The first thing to notice is that most squares are close to the line, but usually slightly higher, resulting in a variance estimate for the independent bootstrap which is slightly larger than the linearized one once the near take-all is removed (the independent bootstrap variance estimate in that stratum is less than the linearized variance estimate reversing the trend when all strata are considered). We used color to identify the corresponding sampling fraction in the stratum: green corresponds to small sampling fractions ($f_h \leq .4$), red moderate f_h (between .4 and .8) and blue large f_h (greater than .8). Notice that while the sampling fraction does not really change the CV of the independent bootstrap, it has a large effect on the Shao-Sitter method with points being further away from the line as the sampling fraction increases.

Columns 3 and 4 of Table 2.7 present the corresponding results for estimating the mean under random hot-deck imputation. The results are qualitatively similar to the mean imputation case. In this case, the linearization variance estimator also accounts for the imputation variance due to the random selection of donors:

$$\widehat{V}(\bar{y}^I) = \tilde{V}_1 + \tilde{V}_2 + \tilde{V}_3.$$

In our example, the linearization variance estimates $\tilde{V}_1$, $\tilde{V}_2$, and $\tilde{V}_3$ were respectively equal to 1,581.5, 6,055.6 and 4,262.6 for a total variance estimate of 11,899.7. The second and third terms are much more important than the first one, again mostly because of the near take-all stratum. If we remove it, the corresponding terms are 1,493.8, 308.7 and 499,6 for a total of 2,302.1, agreeing with the theory developed earlier whereby the third term is not negligible, unlike the second one.

We now turn to the estimation of quartiles under random hot-deck imputation where Table 2.8 presents the different bootstrap variance estimates. While the naive variance estimate is clearly smaller than the other two, it is worth noting that the Shao-Sitter estimates are slightly higher than those obtained under the independent bootstrap procedure in the case of the median and the third quartile. In this case we don't have a linearization variance estimate to compare to. More research is needed to better understand variance estimation under imputation for quartile estimation. However, we saw that the independent method was closer to the target (linearization method) in the case of the mean and based on the simulation results, we would trust the results of the independent method.

Table 2.8. Bootstrap variance estimators for the quartiles under random hot-deck imputation

Bootstrap Method	Q_1	Median	Q_3
Independent	116.4	435.8	3575.5
Shao-Sitter	101.8	553.5	3657.5
Naive	63.7	224.0	2164.8

2.8. Conclusion

Item non-response is an important practical problem in survey sampling. Through the reverse framework, we have shown that the Shao-Sitter bootstrap method only estimates the first term in the variance of an imputed estimator leading to a biased estimator of variance whenever the sampling fraction is large. Under the assumption of uniform non-response within stratum, we introduce the independent bootstrap which consists of choosing a bootstrap sample according a survey sampling bootstrap method and independently generating response indicators from Bernoulli random variables with the estimated response rate as the success probability. The survey sampling bootstrap methods generally involve some type of constants and these constants need to be modified to obtain a consistent estimator of the variance of an imputed estimator, regardless of the sampling fraction. The modifications depend on the bootstrap method, the response rate, and the imputation method. We have illustrated the method for the BRS and bootstrap weights approaches combined with mean, ratio, regression, and hot-deck imputation. The simulation and the application show the strengths of the method and also illustrate how poorly the naive approach – which consists of using the ordinary bootstrap weights provided by statistical agencies, as if we had complete data – can perform.

In this paper we are making the strong assumption (from a practical standpoint) of uniform non-response within stratum. In the case where the probability of response p_i depends on the unit rather than uniformly being equal to p_0, it is no longer possible to find a constant C for the bootstrap method that will lead to a consistent estimator of the variance of the imputed estimator. In Section

2.5, we showed that the second variance term of the reverse framework approach, V_2, is a multiple of the first one, V_1, in the case of uniform non-response so that the total variance is a multiple of V_1. Given that the bootstrap methods used in survey sampling, such as the BRS and bootstrap weights, essentially attempt to estimate the first term V_1, we have been able to consistently estimate the variance by choosing the appropriate constant C in the method. If non-response is no longer uniform, this is not possible anymore. For instance, looking back to Equation (2.5.1) corresponding to mean imputation, it is relatively easy to see that the first term in the variance is a multiple of $\sum p_i(y_i - \bar{Y}_p)^2$ whereas the second term is a multiple of $\sum p_i(1-p_i)(y_i - \bar{Y}_p)^2$ where $\bar{Y}_p = \sum_{i \in U} p_i y_i / \sum_{i \in U} p_i$. Note that the Shao-Sitter approach will continue to succeed in estimating the first term V_1 and will do well provided that the sampling fraction is small. Some limited simulations not reported here show that if the violation to uniform non-response is not too large, the independent bootstrap will do reasonably well, but if the hypothesis clearly does not hold, it will not do well. Research on the case of non-uniform non-response, based on pseudo-population bootstrap methods instead of the BRS and bootstrap weights methods, is ongoing and will be reported elsewhere.

2.9. Appendix A

We begin by showing the claims we made for the Shao-Sitter and the independent bootstrap variance estimators to estimate the variance of the mean estimator based on an imputed data set. We assume that $N \approx N-1$, $n \approx n-1$ and $n_r \approx n_r - 1$. In the following, let $L = 1$.

To illustrate that the independent bootstrap method consistently estimates the variance of an imputed estimator of the population mean via a deterministic method, we develop the theory in detail for linear regression imputation method using the BRS method. For ratio imputation, a similar argument can be used. We will show that the independent bootstrap variance estimator is approximately equal to the estimator obtained using a first-order linearization method in the case of the population mean.

In the case of linear regression imputation, we applied the first-order Demnati and Rao (2004) linearization method and the same arguments used in Kim and Rao (2009) to find a linearization variance estimation of $\bar{y}^I$, but instead of using the population-model approach where a model for the distribution of y_i is used without specifying the distribution of r_i, we used another approach in which y_i is treated as fixed and the uniform non-response mechanism for r_i is assumed. Therefore, assuming $(\bar{\mathbf{X}} - \bar{\mathbf{X}}_r)' s_{\mathbf{XX}}^{-1}$ goes to zero, where $\bar{\mathbf{X}} = \sum_{i\in U} \mathbf{x}_i/N$, and $\bar{\mathbf{X}}_r = \left(\sum_{i\in U} r_i \mathbf{x}_i\right)/\left(\sum_{i\in U} r_i\right)$, we obtain the following approximations:

$$V_1 = E_q V_p\left(\bar{y}^I\,|\mathbf{r}\right) \approx (1-f)\left[1-(1-p_0)\,CC^2\right]\frac{s_U^2}{np_0}, \tag{2.A.1}$$

and

$$V_2 = V_q E_p\left(\bar{y}^I\,|\mathbf{r}\right) \approx \frac{N-1}{N}\,f(1-p_0)\left[1-CC^2\right]\frac{s_U^2}{np_0}, \tag{2.A.2}$$

where $CC^2 = s_{\mathbf{X}Y}' s_{\mathbf{XX}}^{-1} s_{\mathbf{X}Y}/s_U^2$, $s_U^2 = \sum_{i\in U}(y_i - \bar{Y})^2/(N-1)$, $s_{\mathbf{XX}} = \sum_{i\in U}(\mathbf{x}_i - \bar{\mathbf{X}})(\mathbf{x}_i - \bar{\mathbf{X}})'/(N-1)$ and $s_{\mathbf{X}Y} = \sum_{i\in U}(\mathbf{x}_i - \bar{\mathbf{X}})(y_i - \bar{Y})/(N-1)$. As a result, according to (2.A.1) and (2.A.2), the total variance of the imputed estimator under the reverse framework is given by

$$\begin{aligned} V(\bar{y}^I) =&\ E_q V_p\left(\bar{y}^I\,|\mathbf{r}\right) + V_q E_p\left(\bar{y}^I\,|\mathbf{r}\right) \\ \approx&\ \left[1 - p_0 f - (1-p_0)\ CC^2\right]\frac{s_U^2}{np_0} =\ \alpha^{LRI} V_1, \end{aligned} \tag{2.A.3}$$

where $\alpha^{LRI} = (1 - p_0 f - (1-p_0)\ CC^2)/((1-f)\left[1-(1-p_0)\,CC^2\right])$. To estimate $V(\bar{y}^I)$, it suffices to estimate the two components in (2.A.1) and (2.A.2). Asymptotically consistent estimators of V_1 and of V_2 are respectively given by

$$\widehat{E_q V_p}\left(\bar{y}^I\,|\mathbf{r}\right) \approx (1-f)\left[1-(1-\hat{p}_0)\,CC_r^2\right]\frac{s_{ry}^2}{n_r}, \tag{2.A.4}$$

and

$$\widehat{V_q E_p}\left(\bar{y}^I\,|\mathbf{r}\right) \approx f(1-\hat{p}_0)[1-CC_r^2]\frac{s_{ry}^2}{n_r}, \tag{2.A.5}$$

where $CC_r^2 = s_{r\mathbf{x}y}'\ s_{r\mathbf{xx}}^{-1}\ s_{r\mathbf{x}y}/s_{ry}^2$, $s_{ry}^2 = \sum_{i\in s} r_i(y_i - \bar{y}_r)^2/(n_r-1)$, $s_{r\mathbf{xx}} = \sum_{i\in s} r_i(\mathbf{x}_i - \bar{\mathbf{x}}_r)(\mathbf{x}_i - \bar{\mathbf{x}}_r)'/(n_r-1)$ and $s_{r\mathbf{x}y} = \sum_{i\in s} r_i(\mathbf{x}_i - \bar{\mathbf{x}}_r)(y_i - \bar{y}_r)/(n_r-1)$, $\bar{y}_r = \sum_{i\in s} r_i y_i/n_r$ and $\bar{\mathbf{x}}_r = \sum_{i\in s} r_i \mathbf{x}_i/n_r$. As a result, the asymptotically unbiased estimator of $V(\bar{y}^I)$ is

$$\widehat{V}(\bar{y}^I) = \left[1 - \hat{p}_0 f - (1-\hat{p}_0)\ CC_r^2\right]\ \frac{s_{ry}^2}{n_r}. \tag{2.A.6}$$

This estimator is based on the sample of respondents, so the response status is required for all units. However, assuming

$$s_{r\mathbf{xx}}^{-1} s_{\mathbf{xx}} \approx 1, \tag{2.A.7}$$

it is straightforward to see that

$$CC_r^2 \approx \frac{\hat{p}_0 CC_I^2}{1-(1-\hat{p}_0)CC_I^2} \quad \text{and} \quad s_{ry}^2 \approx \frac{1}{\hat{p}_0}\left[1-(1-\hat{p}_0)CC_I^2\right] s_{y^I}^2.$$

where $CC_I^2 = s'_{\mathbf{x}y^I}\ s_{\mathbf{xx}}^{-1}\ s_{\mathbf{x}y^I}/s_{y^I}^2$. Consequently, the estimator in (2.A.6) can be rewritten as follows:

$$\widehat{V}(\bar{y}^I) = \left[(1-\hat{p}_0 f) - \frac{\hat{p}_0\,(1-\hat{p}_0)}{1-(1-\hat{p}_0)CC_I^2}\right]\left[1-(1-\hat{p}_0)CC_I^2\right]\frac{s_{y^I}^2}{\hat{p}_0 n_r}, \tag{2.A.8}$$

which is computable on the imputed data set without requiring the response status.

Now we move on to the Shao-Sitter bootstrap variance estimator. We show that it only estimates the first component of the variance, $E_q V_p\left(\bar{y}^I\,|\mathbf{r}\right)$. Suppose that $s^* = \{(z_i^*, \tilde{\mathbf{u}}_i^*, r_i^*)\}_{i=1}^{n'}$ is the bootstrap sample drawn with replacement from the sample of rescaled data and response status, $\{(z_i, \tilde{\mathbf{u}}_i, r_i)\}_{i=1}^{n}$, where $\tilde{\mathbf{u}}_i' = (1, \mathbf{u}_i')$ and $\mathbf{u}_i' = \bar{\mathbf{x}}' + \sqrt{C}(\mathbf{x}_i - \bar{\mathbf{x}})'$. The bootstrap sample of non-respondents is reimputed using linear regression imputation. Therefore, the bootstrap data after reimputation are

$$z_i^{*I} = \begin{cases} z_i^*, & \text{if } r_i^* = 1, \\ \tilde{\mathbf{u}}_i^{*\prime}\tilde{\boldsymbol{\beta}}_r^*, & \text{if } r_i^* = 0, \end{cases} \qquad \text{for } i = 1, \cdots, n',$$

where

$$\tilde{\boldsymbol{\beta}}_r^* = \left(\sum_{i=1}^{n'} r_i^* \tilde{\mathbf{u}}_i^* \tilde{\mathbf{u}}_i^{*\prime}\right)^{-1}\left(\sum_{i=1}^{n'} r_i^* \tilde{\mathbf{u}}_i^* z_i^*\right).$$

The bootstrap statistic based on the imputed data is $\hat{\theta}^{*I} = \sum_{i=1}^{n'} z_i^{*I}/n'$. Assuming $(\bar{\mathbf{u}} - \bar{\mathbf{u}}_r)' s_{r\mathbf{uu}}^{-1}$ tends to 0, where $\bar{\mathbf{u}} = \sum_{i=1}^{n} \mathbf{u}_i/n$, $\bar{\mathbf{u}}_r = \sum_{i=1}^{n} r_i \mathbf{u}_i/n_r$ and $s_{r\mathbf{uu}} = \sum_{i=1}^{n} r_i\,(\mathbf{u}_i - \bar{\mathbf{u}}_r)\,(\mathbf{u}_i - \bar{\mathbf{u}}_r)'/(n_r - 1)$, that m_i^* is the number of times the i-th unit in the sample of rescaled data is selected in the bootstrap sample and using a first-order linearization, we have

$$\hat{\theta}^{*I} - \bar{y}^I \approx \sum_{i=1}^{n}\left(m_i^* - \frac{n'}{n}\right)\left\{\frac{r_i}{n' n_r/n}(z_i - \bar{z}_r) + \left(\frac{\mathbf{u}_i}{n'} - \frac{r_i}{n' n_r/n}(\mathbf{u}_i - \bar{\mathbf{u}}_r)\right)'\hat{\boldsymbol{\beta}}_r\right\},$$

where $\bar{z}_r = \sum_{i=1}^{n} r_i z_i / n_r$ and $\hat{\boldsymbol{\beta}}_r = s_{r\mathbf{xx}}^{-1} s_{r\mathbf{x}y}$. To compute the bootstrap variance estimator, we have only one source of randomness: the sampling mechanism indexed by $p*$. Using the linearized variable, we have

$$\begin{aligned} V^*_{Sh.S.}\left(\hat{\theta}^{*I}\right) &\approx V_{p*}\left[\frac{1}{n'}\sum_{i=1}^{n'}\left\{\frac{n}{n_r}\, r_i^*(z_i^* - \bar{z}_r) + \mathbf{u}_i^{*\prime}\,\hat{\boldsymbol{\beta}}_r - \frac{n}{n_r}\, r_i^*(\mathbf{u}_i^* - \bar{\mathbf{u}}_r)'\,\hat{\boldsymbol{\beta}}_r\right\}\right] \\ &= \frac{C\,(n_r-1)}{n' n_r \hat{p}_0}\left[s_{ry}^2 + \hat{p}_0^2\,\frac{n-1}{n_r-1}\hat{\boldsymbol{\beta}}_r' s_{\mathbf{xx}}\,\hat{\boldsymbol{\beta}}_r - s_{r\mathbf{x}y}'\hat{\boldsymbol{\beta}}_r\right]. \end{aligned}$$

Replacing the rescaling factor by $C = [n'/(n-1)][1-f]$ and assuming (2.A.7), we obtain

$$V^*_{Sh.S.}\left(\hat{\theta}^{*I}\right) \approx (1-f)\left[1 - (1-\hat{p}_0)\; CC_r^2\right]\frac{s_{ry}^2}{n_r}, \qquad (2.A.9)$$

which is equal to the asymptotically unbiased estimator of the first component of the variance, $\widehat{E_q V_p}\left(\bar{y}^I\,|\mathbf{r}\right)$, presented in (2.A.4). Therefore, as we claimed, in the case of non-negligible sampling fraction, the Shao-Sitter method underestimates the total variance of $\hat{\theta}^I$.

We now move to the independent bootstrap variance estimator and show that it is a consistent estimator for $V(\bar{y}^I)$. Suppose that the bootstrap sample, $s^* = \{(z_i^*, \tilde{\mathbf{u}}_i^*)\}_{i=1}^{n'}$ is drawn with replacement from $\{(z_i, \tilde{\mathbf{u}}_i)\}_{i=1}^{n}$, the rescaled sample with C^I presented in Table 2.1. Then, using the response rate, we independently regenerate the response indicators, $\{r_i^*\}_{i=1}^{n'}$. The bootstrap statistic based on the reimputed bootstrap data set using linear regression imputation is $\hat{\theta}^{*I} = \sum_{i=1}^{n'} z_i^{*I}/n'$ which has the same form as the Shao-Sitter bootstrap statistic, but the regenerated response status is used. In this bootstrap procedure, there exist two sources of randomness: the sampling and the non-response mechanisms indexed by $p*$ and $q*$, respectively. We study the independent bootstrap variance estimator, $V^*(\hat{\theta}^{*I})$, under the two-phase framework which implies

$$V^*(\hat{\theta}^{*I}) = E_{p*}V_{q*}\left(\hat{\theta}^{*I}\,|s^*\right) + V_{p*}E_{q*}\left(\hat{\theta}^{*I}\,|s^*\right). \qquad (2.A.10)$$

To compute $V_{q*}\left(\hat{\theta}^{*I}\ |s^*\right)$ and $E_{q*}\left(\hat{\theta}^{*I}\ |s^*\right)$, we apply a first-order Demnati-Rao linearization as follows:

$$\begin{aligned} V_{q*}\left(\hat{\theta}^{*I}\ |s^*\right) &\approx\ V_{q*}\left[\frac{1}{n'\hat{p}_0}\ \sum_{i=1}^{n'} r_i^*\ \left(z_i^* - \bar{z}^* - (\mathbf{u}_i^* - \bar{\mathbf{u}}^*)'\ \hat{\boldsymbol{\beta}}^*\right)\right] \\ &= \frac{1-\hat{p}_0}{n'^2\hat{p}_0}\ \sum_{i=1}^{n'}\left(z_i^* - \bar{z}^* + (\bar{\mathbf{u}}^* - \mathbf{u}_i^*)'\ \hat{\boldsymbol{\beta}}^*\right)^2, \end{aligned}$$

and

$$E_{q*}\left(\hat{\theta}^{*I}\ |s^*\right) \approx \bar{z}^*, \tag{2.A.11}$$

where $\bar{z}^* = \sum_{i=1}^{n'} z_i^*/n'$, $\bar{\mathbf{u}}^* = \sum_{i=1}^{n'} \mathbf{u}_i^*/n'$, and

$$\hat{\boldsymbol{\beta}}^* = \left[\sum_{i=1}^{n'}\left(\mathbf{u}_i^* - \bar{\mathbf{u}}^*\right)\left(\mathbf{u}_i^* - \bar{\mathbf{u}}^*\right)'\right]^{-1}\left[\sum_{i=1}^{n'}\left(\mathbf{u}_i^* - \bar{\mathbf{u}}^*\right)\left(z_i^* - \bar{z}^*\right)\right].$$

Using a first-order Taylor linearization, we have

$$\begin{aligned} E_{p*}V_{q*}(\hat{\theta}^{*I}\ |s^*) &\approx\ E_{p*}\left[\frac{1-\hat{p}_0}{n'^2\hat{p}_0}\ \sum_{i=1}^{n'}\left(z_i^* - \bar{z}^* - (\mathbf{u}_i^* - \bar{\mathbf{u}}^*)'\,\hat{\boldsymbol{\beta}}^*\right)^2\right] \\ &\approx\ \frac{1-\hat{p}_0}{n'n\hat{p}_0}\ \sum_{i=1}^{n}\left(z_i - \bar{z} - (\mathbf{u}_i - \bar{\mathbf{u}})'\, s_{\mathbf{xx}}^{-1}s_{\mathbf{x}y^I}\right)^2 \qquad (2.A.12) \\ &= \ \frac{C^I(n-1)}{n'}\frac{1-\hat{p}_0}{\hat{p}_0}\left[1 - CC_I^2\right]\frac{s_{y^I}^2}{n}. \end{aligned}$$

To compute the second term, the result in (2.A.11) implies that

$$V_{p*}E_{q*}[\hat{\theta}^{*I}\ |s^*] \approx \frac{1}{n'n}\ \sum_{i=1}^{n}(z_i - \bar{z})^2 =\ \frac{C^I(n-1)}{n'}\frac{s_{y^I}^2}{n}. \tag{2.A.13}$$

As a result, the bootstrap variance estimator presented in (2.A.10) is obtained by combining the results in (2.A.12) and (2.A.13):

$$V^*(\hat{\theta}^{*I}) = \frac{C^I(n-1)}{n'\hat{p}_0}\left[1 - (1-\hat{p}_0)\,CC_I^2\right]\frac{s_{y^I}^2}{n}.$$

Using the proposed rescaling factor for linear regression imputation, this estimator equals the consistent estimator of the variance of $\hat{\theta}^I$ based on the imputed data set presented in (2.A.8).

We now look at the claims regarding the bootstrap estimators under random hot-deck imputation and begin with a linearization of the variance. Under RHDI, a missing value is imputed by selecting completely at random from the set of respondents with probability $1/n_r$ for each unit. Therefore, an estimator of the population mean estimator based on imputed data is

$$\bar{y}^I = \frac{1}{N} \sum_{i \in s} w_i \, [r_i y_i + (1 - r_i) \, \tilde{y}_i].$$

Under the reverse framework, the total variance of $\bar{y}^I$ is given by

$$V(\bar{y}^I) = E_q V_p E_I \left(\bar{y}^I \mid s, \mathbf{r} \right) + V_q E_p E_I \left(\bar{y}^I \mid s, \mathbf{r} \right) + E_q E_p V_I \left(\bar{y}^I \mid s, \mathbf{r} \right) = \widetilde{V}_1 + \widetilde{V}_2 + \widetilde{V}_3. \tag{2.A.14}$$

The fact that $E_I(\bar{y}^I) = \bar{y}_r$, which is the estimator of $\bar{Y}$ under mean imputation, implies that $E_q V_p E_I \left(\bar{y}^I \mid s, \mathbf{r} \right) + V_q E_p E_I \left(\bar{y}^I \mid s, \mathbf{r} \right)$ can be approximated by (2.A.3) and estimated by (2.A.6) assuming $\tilde{\mathbf{x}} = 1$. Consequently, we only need to estimate the third component,

$$\begin{aligned} E_q E_p V_I \left(\bar{y}^I \mid s, \mathbf{r} \right) &= E_q E_p \left(\frac{1}{N^2} \sum_{i \in s} w_i^2 (1 - r_i) \, V_I(\tilde{y}_i) \, \middle| \, \mathbf{r} \right) \\ &= \frac{1 - p_0}{n} \, E_q E_p \left(\frac{n_r - 1}{n_r} \, s_{ry}^2 \, \middle| \, \mathbf{r} \right) \approx \frac{1 - p_0}{nN} \sum_{i \in U} \left(y_i - \bar{Y} \right)^2. \end{aligned}$$

It is easy to check that an asymptotically unbiased estimator for this term can be given by

$$\widehat{E_q E_p V_I} \left(\bar{y}^I \mid s, \mathbf{r} \right) = \hat{p}_0 \, (1 - \hat{p}_0) \, \frac{s_{ry}^2}{n_r}.$$

As a result, the total variance of $\bar{y}^I$ in (2.A.14) is approximated by

$$V(\hat{\theta}^I) = \frac{1 - p_0 f + p_0 \, (1 - p_0)}{n N p_0} \sum_{i \in U} (y_i - \bar{Y})^2.$$

Assuming $\lambda_f = 1 - f$, we can easily see that

$$V(\hat{\theta}^I) = \left[\frac{\{1 + p_0 \, (1 - p_0)\} - p_0 f}{\{1 + p_0 \, (1 - p_0)\} \, (1 - f)} \right] \, [\widetilde{V}_1 + \lambda_f \widetilde{V}_3] = \alpha^{RHDI} \, [\widetilde{V}_1 + \lambda_f \widetilde{V}_3].$$

The asymptotically unbiased estimator of $V(\hat{\theta}^I)$ can be given by

$$\widehat{V}(\hat{\theta}^I) = [1 - \hat{p}_0 f + \hat{p}_0 \, (1 - \hat{p}_0)] \, \frac{s_{ry}^2}{n_r}.$$

We note that this estimator is obtained using the sample of respondents, so the response status is required for all units.

We now study the Shao-Sitter bootstrap variance estimator under RHDI. In this case, the bootstrap statistic after reimputation is defined by

$$\hat{\theta}^{*I} = \frac{1}{n'} \sum_{i=1}^{n'} \left[r_i^* z_i^* + (1 - r_i^*)\ \tilde{z}_i^* \right].$$

where $\tilde{z}_i^*$ is selected completely at random from the bootstrap sample of respondents with probability $1/R^*$, where $R^* = \sum_{i=1}^{n'} r_i^*$. The variance of $\hat{\theta}^{*I}$ is given by

$$V_{Sh.S.}^*(\hat{\theta}^{*I}) = E_{p*} V_{I*}\left(\hat{\theta}^{*I}\ |s^*\right) + V_{p*} E_{I*}\left(\hat{\theta}^{*I}\ |s^*\right),$$

where index $I*$ indicates the random imputation mechanism in the bootstrap procedure. Using a first-order Taylor linearization, the first component of $V_{Sh.S.}^*(\hat{\theta}^{*I})$ is approximated as follows:

$$\begin{aligned} E_{p*} V_{I*}\left(\hat{\theta}^{*I}\ |s^*\right) &= E_{p*}\left[\frac{1}{n'^2}\sum_{i=1}^{n'}(1 - r_i^*)\ V_{I*}(\tilde{z}_i^*)\right] \\ &= E_{p*}\left[\frac{1}{n'^2}\sum_{i=1}^{n'}(1 - r_i^*)\ \frac{1}{\sum_{i=1}^{n'} r_i^*}\ \sum_{i=1}^{n'} r_i^*(z_i^* - \bar{z}_r^*)^2\right] \\ &\approx \frac{1}{n'^2}\ \frac{(n'/n)\ \sum_{i=1}^{n}(1 - r_i)}{(n'/n)\ \sum_{i=1}^{n} r_i}\ \frac{n'}{n}\ \sum_{i=1}^{n} r_i(z_i - \bar{z}_r)^2 \\ &= \frac{C\ (n-1)}{n'}\ \hat{p}_0\ \ (1 - \hat{p}_0)\ \frac{s_{ry}^2}{n_r}. \end{aligned}$$

Using the fact that $E_{I*}\left(\hat{\theta}^{*I}\ |s^*\right) = \bar{z}_r^* = \sum_{i=1}^{n'} r_i^* z_i^* / R^*$ which is the bootstrap statistic using mean imputation, the approximation in (2.A.9) in the case of $\tilde{\mathbf{x}} = 1$ can be used to compute the second component. As a result, we have

$$V_{p*} E_{I*}\left(\hat{\theta}^{*I}\ |s^*\right) = \frac{C\ (n-1)}{n'}\ \frac{s_{ry}^2}{n_r}.$$

Finally, applying the complete data rescaling factor, $C = [n'/(n-1)][1-f]$, the Shao-Sitter variance estimator is

$$\begin{aligned} V_{Sh.S.}^*(\hat{\theta}^{*I}) &\approx \frac{C\ (n-1)}{n'}\ \hat{p}_0\ \ (1 - \hat{p}_0)\ \frac{s_{ry}^2}{n_r} + \frac{C\ (n-1)}{n'}\ \frac{s_{ry}^2}{n_r} \\ &= (1-f)\ \hat{p}_0\ \ (1 - \hat{p}_0)\ \frac{s_{ry}^2}{n_r} + (1-f)\ \frac{s_{ry}^2}{n_r}, \end{aligned}$$

which is an asymptotically unbiased estimator for $\widetilde{V}_1 + \lambda_f \widetilde{V}_3$, or for $V(\hat{\theta}^I)$ when f is negligible.

Consider the independent bootstrap under RHDI. The bootstrap statistic after reimputing the non-respondents in the bootstrap sample is defined similarly to the case of the Shao-Sitter method, but with the response status regenerated according to the binomial distribution. This bootstrap procedure implies that

$$V^*(\hat{\theta}^{*I}) = V_{p*}E_{q*}E_{I*}\left(\hat{\theta}^{*I}\,|\mathbf{r}^*, s^*\right) + E_{p^*}V_{q^*}E_{I^*}\left(\hat{\theta}^{*I}\,|\mathbf{r}^*, s^*\right) + E_{p^*}E_{q^*}V_{I^*}\left(\hat{\theta}^{*I}\,|\mathbf{r}^*, s^*\right).$$

Because $E_{I*}\left(\hat{\theta}^{*I}\,|s^*\right) = \bar{z}_r^*$, the first two components can be simply obtained using (2.A.12) and (2.A.13) in the case of $\tilde{\mathbf{x}} = 1$.

$$V_{p*}E_{q*}E_{I*}\left(\hat{\theta}^{*I}\,|\mathbf{r}^*, s^*\right) + E_{p^*}V_{q^*}E_{I^*}\left(\hat{\theta}^{*I}\,|\mathbf{r}^*, s^*\right) = \frac{C^I}{n'n\hat{p}_0}\sum_{i=1}^{n}\left(y_i^I - \bar{y}^I\right)^2. \tag{2.A.15}$$

Using a first-order Demnati-Rao linearization and a first-order Taylor linearization, the third term is approximated by

$$\begin{aligned} E_{p*}E_{q*}V_{I*}\left(\hat{\theta}^{*I}\,|\mathbf{r}^*, s^*\right) &= E_{p*}E_{q*}\left[\frac{\sum_{i=1}^{n'}(1-r_i^*)}{n'^2\,\sum_{i=1}^{n'} r_i^*}\sum_{i=1}^{n'} r_i^*\,\left(z_i^* - \frac{\sum_{i=1}^{n'} r_i^* z_i^*}{\sum_{i=1}^{n'} r_i^*}\right)^2\right] \\ &\approx E_{p*}\left[\frac{1-\hat{p}_0}{n'^2}\sum_{i=1}^{n'}\left(z_i^* - \frac{1}{n'}\sum_{i=1}^{n'} z_i^*\right)^2\right] \\ &\approx \frac{C^I\,(1-\hat{p}_0)}{n'n}\sum_{i\in s}\left(y_i^I - \bar{y}^I\right)^2. \end{aligned} \tag{2.A.16}$$

As a result, (2.A.15) and (2.A.16) imply that

$$\begin{aligned} V^*(\hat{\theta}^{*I}) \approx&\ \frac{C^I}{n'n_r}\left[1 + \hat{p}_0\,(1-\hat{p}_0)\right]\sum_{i\in s}\left(y_i^I - \bar{y}^I\right)^2 \\ =&\ \frac{\hat{p}_0}{n_r(n_r-1)}\,\frac{(1-\hat{p}_0 f) + \hat{p}_0\,(1-\hat{p}_0)}{1-\frac{1}{n}\,(1-\hat{p}_0)}\sum_{i\in s}\left(y_i^I - \bar{y}^I\right)^2. \end{aligned} \tag{2.A.17}$$

It remains to show that this estimator is asymptotically consistent for $V(\hat{\theta}^I)$. Under random hot-deck imputation, we have that

$$\begin{aligned} E_I\left[\sum_{i\in s}(y_i^I - \bar{y}^I)^2\right] &= E_I\left(\sum_{i\in s} y_i^{I2}\right) - nE_I\left(\bar{y}^{I2}\right) \\ &= \frac{n}{n_r}\sum_{i\in s} r_i y_i^2 - n\left[\frac{1}{nn_r}\,\left(1-\frac{n_r}{n}\right)\sum_{i\in s} r_i\,(y_i - \bar{y}_r)^2 + \bar{y}_r^2\right] \\ &\approx \frac{n_r - 1}{\hat{p}_0}\,\left[1 - \frac{1}{n}\,(1-\hat{p}_0)\right] s_{ry}^2, \end{aligned}$$

which can be directly used to show that

$$E_I\left[V^*(\hat{\theta}^{*I})\right] \approx \left[(1-\hat{p}_0 f) + \hat{p}_0\ (1-\hat{p}_0)\right]\ \frac{s^2_{ry}}{n_r}.$$

Consequently, the bootstrap variance estimator is asymptotically consistent. Note that the independent bootstrap variance estimator in (2.A.17) is computed on the sample of imputed data without requiring the response indicators.

We conclude by showing the equivalence of the independent bootstrap statistics of the BRS and BW methods for a function of means, as is the case for BRS and BW statistics for complete survey data. We do this in the case of RHDI imputation; simpler arguments can be used for a deterministic imputation method.

Under the BRS approach, we define

$$I^*_{ij} = \begin{cases} 1, & \text{if } z_i = z^*_j, \\ 0, & \text{otherwise.} \end{cases} \tag{2.A.18}$$

According to this definition $\sum_{j=1}^{n'} I^*_{ij}$ and $\sum_{j=1}^{n'} I^*_{ij} r^*_j$ are the numbers of times that unit i in s is selected in the bootstrap sample s^* and in the bootstrap sample of respondents, respectively. From this sampling procedure and the one presented in Section 2.4.3, it is straightforward to see that

$$m^*_i \overset{D}{=} \sum_{j=1}^{n'} I^*_{ij} \quad \text{and} \quad R^*_i \overset{D}{=} \sum_{j=1}^{n'} I^*_{ij} r^*_j,$$

where $\overset{D}{=}$ indicates equality in distribution.

Using random hot-deck imputation method in the bootstrap procedure, we obtain

$$\hat{\theta}^{*I} = \frac{1}{n'} \sum_{j=1}^{n'} \left[r^*_j z^*_j + (1-r^*_j)\ \tilde{z}^*_j\right] = \frac{1}{n'} \sum_{j=1}^{n'} \left[r^*_j z^*_j + (1-r^*_j) \sum_{l=1}^{n'} \tilde{I}^*_{jl} r^*_l z^*_l\right],$$

where

$$\tilde{I}^*_{jl} = \begin{cases} 1, & \text{if } \tilde{z}^*_j = z^*_l \text{ with } r^*_l = 1, \\ 0, & \text{otherwise,} \end{cases}$$

which indicates which unit in the bootstrap sample of respondents is selected for imputing the missing data z^*_j. This definition and that in (2.A.18) imply that

$$\begin{aligned}
\hat{\theta}^{*I} &= \frac{1}{n'}\sum_{l=1}^{n'} r_l^* \left[1+\sum_{j=1}^{n'}(1-r_j^*)\tilde{I}_{jl}^*\right]\ z_l^* \\
&= \frac{1}{n'}\sum_{l=1}^{n'} r_l^* \left[1+\sum_{j=1}^{n'}(1-r_j^*)\tilde{I}_{jl}^*\right]\ \sum_{i\in s} I_{il}^* z_i \\
&= \frac{1}{n'}\sum_{i\in s}\left\{\sum_{l=1}^{n'} I_{il}^* r_l^* + \sum_{l=1}^{n'} I_{il}^* r_l^* \left[\sum_{j=1}^{n'}(1-r_j^*)\tilde{I}_{jl}^*\right]\right\}\ z_i \\
&\overset{D}{=} \frac{1}{n'}\sum_{i\in s}\left\{R_i^*+D_i^*\right\}\ z_i \\
&= \frac{1}{N}\sum_{i\in s}\left[1+\sqrt{C^I}\left(\frac{n(R_i^*+D_i^*)}{n'}-1\right)\right] w_i\ y_i^I.
\end{aligned}$$

As a result, the bootstrap statistics via the two different approaches have the same distribution.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of grants from the Natural Sciences and Engineering Research Council. Moreover, they would like to sincerely thank the Associate Editor and the reviewers for the useful comments and suggestions made on the first version of the paper. We believe that these comments have contributed to improving the overall quality of the paper.

Chapter 3

PSEUDO-POPULATION BOOTSTRAP METHODS FOR IMPUTED SURVEY DATA

Abstract

Item non-response in surveys is usually dealt with through single imputation. It is well known that treating the imputed values as if they were observed values may lead to serious underestimation of the variance of point estimators. Two approaches are used for studying the properties of point and variance estimators: the non-response model approach that requires assumptions about the non-response mechanism and the imputation model approach that requires assumption about the distribution of the variable being imputed. In this paper, we propose three pseudo-population bootstrap schemes: the first two lead to an approximately unbiased variance estimator with respect to the non-response model approach and the imputation model approach, respectively. The third scheme leads to a doubly robust bootstrap variance estimator. That is, the latter is approximately unbiased for the true variance if one model or the other is correctly specified. The proposed bootstrap procedures can be used even for large sampling fraction. Results from a simulation study suggest that the resulting variance estimators perform well in terms of relative bias.

Key words and phrases: Bootstrap; Imputation, Imputation model approach, Non-response model approach, Pseudo-population approach, Variance estimation.

3.1. Introduction

Item non-response in surveys is usually dealt with through single imputation. That is, a missing value is replaced by a single artificial value, which is constructed on the basis of auxiliary information recorded for both respondents and non-respondents. Variance estimation in the presence of imputed data has been widely treated in the literature; e.g., Särndal (1992), Rao and Shao (1992), Rao (1996), Shao and Sitter (1996), Shao and Steel (1999), Haziza (2009) and Kim and Rao (2009), among others. It is well known that treating the imputed values as if they were observed values may lead to serious underestimation of the variance of the point estimators, leading to confidence intervals that are too narrow.

In the absence of non-response, bootstrap procedures can be classified into two main groups. In the first, bootstrap samples are selected from the original sample; e.g., Rao and Wu (1988), Sitter (1992b) and Rao et al. (1992), among others. Rao and Wu (1988) applied a scale adjustment directly to the survey data values so as to recover the usual variance formulae. Rao et al. (1992) presented a modification of the method of Rao and Wu (1988), where the scale adjustment is applied to the survey weights rather than to the data values. The second group of procedures consists of 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 Gross (1980), Bickel and Freedman (1984), Booth et al. (1994), Chauvet (2007) and Wang and Thompson (2012), among others. Most bootstrap procedures can 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 Antal and Tillé (2011a) and Beaumont and Patak (2012).

Shao and Sitter (1996) introduced a bootstrap method to deal with imputed data. It consists of using any (complete) data bootstrap method to select a bootstrap sample of imputed data while keeping their corresponding original response status, and then to re-impute the bootstrap data with a missing status using the same imputation method that was used on the original data. The Shao-Sitter bootstrap variance estimator is consistent for the true variance *provided* *that the sampling fraction is negligible*; see also Davison and Sardy (2007). More recently, Mashreghi et al. (2014) considered a bootstrap procedure that works for non-negligible sampling fractions in the case of stratified simple random sample without replacement and uniform non-response within strata.

In order to assess the properties of point and variance estimators and to derive variance estimators, two inferential approaches are used: (i) the non-response model (NRM) approach that requires explicit assumptions about the unknown non-response mechanism and (ii) the imputation model (IM) approach that requires the specification of a model describing the distribution of the variable under study being imputed. In this paper, we consider the class of linear regression imputation, which includes mean and ratio imputation as special cases. We focus on doubly robust regression imputation, which makes explicit use of both the non-response model and the imputation model. The resulting imputed estimator is doubly robust. That is, it remains asymptotically unbiased and consistent for the true parameter if either model (non-response or imputation) is true. This type of procedures offers some protection against misspecification of one model or the other; e.g., Haziza and Rao (2006) and Kim and Haziza (2014).

Assuming that the data are MAR (Rubin, 1976), we develop pseudo-population bootstrap procedures that can be used even for large sampling fractions unlike the Shao-Sitter procedure. We present three bootstrap schemes: the first (called NRM Scheme) leads to an asymptotically unbiased bootstrap variance estimator with respect to the NRM approach. The IM Scheme leads to an asymptotically unbiased bootstrap variance estimator with respect to the IM approach. Finally, the third Scheme (called DR Scheme), which is a combination of the first two, leads to a doubly robust bootstrap variance estimator. That is, the latter is asymptotically unbiased for the true variance if either the non-response or imputation model is correctly specified. In this paper, we focus on simple random sampling without replacement and Poisson sampling. The extension to stratified simple random sampling is relatively straightforward as sampling is performed independently within each stratum. We show that our methods lead to valid estimators for simple random sampling without replacement and Poisson sampling. In the literature, complete data pseudo-population methods have been proposed (Chauvet, 2007) for high entropy sampling designs that includes the Rao-Sampford procedure (Rao, 1965; Sampford, 1967) and the Chao procedure (Chao, 1982) as special cases (Berger, 1998); see also Wang and Thompson (2012) for unequal probability sampling designs. Our bootstrap procedures can be naturally extended to handle these sampling design although we do not provide a formal proof in this paper.

In the case of complete data, bootstrap pseudo-population approaches rely on the sampling design and on the (observed) values in the sample to create the pseudo-population from which bootstrap samples will be selected. In presence of item non-response, a further random mechanism is present which breaks the sample into respondents and non-respondents, the values of the latter group being unobserved of course. The challenge in designing bootstrap pseudo-population approaches in presence of item non-response is to use models, specified up to unknown constants, to compensate for the unobserved values in the sample. In the NRM Scheme, we use a postulated model for the non-response random mechanism based on auxiliary variables while in the IM Scheme, we use a model linking the variable under study to auxiliary variables.

We begin with the NRM Scheme. The key idea here is to recognize that the set of respondents to a specific item can be viewed as a random sample obtained by a Poisson sampling design using the (unknown) response probabilities as the inclusion probabilities. In the NRM Scheme, we therefore begin by considering the sample as a “population” from which a Poisson sample was taken leading to the sample of respondents and then we apply a pseudo-population approach appropriate for Poisson sampling to create a “pseudo-population” which we will call a pseudo-sample. The pseudo-sample will have properties similar to what should be expected from a sample without item non-response. Then we simply use a complete data approach to create the pseudo-population for the vector made up of the variable under study and the auxiliary variables from which bootstrap samples will be taken and the response status generated. More specifically, we create the pseudo-population in two distinct steps: we first apply the complete data bootstrap method of Chauvet (2007) for Poisson sampling to obtain a pseudo-sample for the vector of the variable under study and the auxiliary variables. Then, from the pseudo-sample, we create a pseudo-population of vectors according to a complete data pseudo-population bootstrap procedure. For example, if the original sample was selected according to simple random sampling without replacement, the method of Booth et al. (1994) may be used whereas we use Chauvet (2007) if it was Poisson sampling. Bootstrap samples are then selected from the pseudo-population by applying the original sampling design and generating non-response in each bootstrap sample using Poisson sampling with the estimated inclusion probabilities as the inclusion probabilities. Imputation within each bootstrap sample is performed according to the same imputation method that was used in the original sample.

For the IM Scheme, the model linking the variable under study to the auxiliary variables will be estimated from the respondents in the sample and the empirical distribution function of its standardized residuals can be used to generate errors to be added to predicted values to represent the distribution of the population of the variable under study. But since we do not have the values of the auxiliary variables for the elements of the population outside the sample, we cannot create a pseudo-population directly from the model. Hence, as in the NRM Scheme, we begin by creating a pseudo-population of vectors made up of the auxiliary variables *and* the response status from the values in the sample. And using the estimated model, we create a pseudo-population of the variable under study and its response status, by adding bootstrap errors from the standardized residuals to the predicted values. Bootstrap samples of pairs of the variable under study and its response status are taken from the pseudo-population according to the original sampling design. The missing data in bootstrap samples are imputed using the original imputation method. Finally, the DR Scheme combines both NRM Scheme and IM Scheme to lead to a doubly robust bootstrap variance estimator.

This article is organized as follows. After introducing some notation in Section 3.2, we briefly describe the complete data pseudo-population bootstrap procedures of Booth et al. (1994) and that of Chauvet (2007) in Section 3.3. In Section 3.4, we present a bootstrap procedure that leads to an asymptotically unbiased estimator of the true variance with respect to the NRM approach while a valid bootstrap procedure under the IM approach is presented in Section 3.5. A doubly robust bootstrap procedure is discussed in Section 3.6. Finally, the results of a simulation study, assessing the performance of several bootstrap variance estimators in terms of bias, are presented in Section 3.7.

3.2. Preliminaries

Let U be a population of size N. We are interested in estimating the population total, $t = \sum_{i \in U} y_i$, of a study variable y. A sample s, of (expected) size n, is selected from U according to a given sampling design $p(s)$. We assume that the sampling design is non-informative. A complete data estimator of t is the expansion estimator

$$\hat{t} = \sum_{i \in s} w_i y_i, \tag{3.2.1}$$

where $w_i = \pi_i^{-1}$ denotes the survey weight associated with the ith unit and π_i denotes its inclusion probability. The estimator $\hat{t}$ is design-unbiased (or p-unbiased) for t; i.e., $E_p(\hat{t}) = t$, where the subscript p denotes the sampling design.

We now turn to the case of missing y-values. Let r_i be the response indicator associated with unit i such that $r_i = 1$ if unit i is a respondent to item y and $r_i = 0$, otherwise. Let $y_i^I = y_i$ if $r_i = 1$, and $y_i^I = \tilde{y}_i$ if $r_i = 0$, where $\tilde{y}_i$ denotes the imputed value used to replace missing y_i. An imputed estimator of t based on observed and imputed data is

$$\hat{t}^I = \sum_{i \in s} w_i y_i^I. \tag{3.2.2}$$

To replace the missing y-values, we consider linear regression imputation based on a vector of auxiliary variables, $\boldsymbol{x}$, recorded for all the sample units (respondents and non-respondents). Linear regression imputation is motivated by the following imputation model:

$$m : y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \tag{3.2.3}$$

where $\boldsymbol{\beta}$ is a vector of unknown parameters and the errors ε_i satisfy

$$E_m(\varepsilon_i) = 0,\ V_m(\varepsilon_i) = \sigma^2 c_i \text{ and } \ cov_m(\varepsilon_i, \varepsilon_j) = 0,\ \forall i \neq j,$$

where σ^2 is an unknown parameter. We assume that $c_i = \boldsymbol{\lambda}^\top \boldsymbol{x}_i$, where $\boldsymbol{\lambda}$ is a vector of known constants.

Let $p_i = Prob(r_i = 1)$ be the response probability for unit i. We assume that units respond independently of one another. Further, we assume that the response probability to item y can be parametrically modeled:

$$p_i = Prob(r_i = 1) = m(\boldsymbol{x}_i; \boldsymbol{\gamma}) \tag{3.2.4}$$

for some known function $m(\boldsymbol{x}_i; \cdot)$, where $\boldsymbol{\gamma}$ is a vector of unknown parameters. Model (3.2.4) is called a non-response model. Throughout this article, we assume that the data are missing at random (MAR) (Rubin, 1976). That is, we assume that the probability that y is missing does not depend on y as long as we account for $\boldsymbol{x}$. Formally, we have

$$Prob(r_i = 1 | \boldsymbol{x}, y) = Prob(r_i = 1 | \boldsymbol{x}).$$

The estimated response probability for unit i is $\hat{p}_i = m(\boldsymbol{x}_i; \hat{\boldsymbol{\gamma}})$, where $\hat{\boldsymbol{\gamma}}$ is an estimator of $\boldsymbol{\gamma}$ (e.g., the maximum likelihood estimator).

In the case of linear regression imputation, missing y_i is replaced by the imputed value

$$\tilde{y}_i = \boldsymbol{x}_i^\top \ \hat{\boldsymbol{\beta}}_r, \tag{3.2.5}$$

where

$$\hat{\boldsymbol{\beta}}_r = \left\{ \sum_{i \in s} w_i r_i \left(\frac{1 - \hat{p}_i}{\hat{p}_i} \right) \frac{\boldsymbol{x}_i \boldsymbol{x}_i^\top}{c_i} \right\}^{-1} \sum_{i \in s} w_i r_i \left(\frac{1 - \hat{p}_i}{\hat{p}_i} \right) \frac{\boldsymbol{x}_i y_i}{c_i} \tag{3.2.6}$$

is the weighted least square estimator of $\boldsymbol{\beta}$ using $w_i \hat{p}_i^{-1}(1 - \hat{p}_i) c_i^{-1}$ as the weight for unit i. If the non-response model contains only the intercept, then $\hat{p}_i = \hat{p}_0$, the overall response rate, and (3.2.5) reduces to the customary deterministic regression imputation. The imputed estimator (3.2.2) that uses the imputed values (3.2.5) is doubly robust in the sense that it remains consistent for t if either the imputation model (3.2.3) or the non-response model (3.2.4) is correctly specified; e.g., Haziza and Rao (2006) and Kim and Haziza (2014).

In order to derive variance estimators for (3.2.2), we consider two approaches: the NRM approach and the IM approach. In the NRM approach, inference is made with respect to the joint distribution induced by the sampling design and the assumed non-response model given by (3.2.4). In the IM approach, inference is made with respect to the joint distribution induced by the imputation model (3.2.3), the sampling design, and the non-response mechanism. In the latter approach, explicit assumptions about the non-response mechanism are not required except for the MAR assumption.

To express the variance of (3.2.2), we use the standard decomposition of the total error, $\hat{t}^I - t$:

$$\hat{t}^I - t = \left(\hat{t} - t\right) + \left(\hat{t}^I - \hat{t}\right), \tag{3.2.7}$$

where $\hat{t}$ in the above expression is the complete data estimator (3.2.1). The first term on the right hand side of (3.2.7) is the sampling error, whereas the second term represents the non-response error.

Using decomposition (3.2.7), the variance of (3.2.2) with respect to the NRM approach can be expressed as

$$\begin{aligned} V^{NRM} \simeq E_{pq}\left(\hat{t}^I - t\right)^2 &= V_p\left(\hat{t}\right) + E_p E_q\left\{\left(\hat{t}^I - \hat{t}\right)^2 \mid s\right\} \\ &= V_1^{NRM} + V_2^{NRM}, \end{aligned} \tag{3.2.8}$$

where the subscript q denotes the non-response mechanism. The term V_1^{NRM} in (3.2.8) is the sampling variance of the complete data estimator $\hat{t}$, whereas the term V_2^{NRM} represents the non-response variance.

To express the variance of (3.2.2) with respect to the IM approach, we use, once again, decomposition (3.2.7):

$$\begin{aligned} V^{IM} &\simeq E_{mpq}\left(\hat{t}^I - t\right)^2 \\ &= E_m V_p\left(\hat{t}\right) + E_{pq} E_m\left\{\left(\hat{t}^I - \hat{t}\right)^2 \mid s, s_R\right\} + 2E_{pq}E_m\left\{\left(\hat{t} - t\right)\left(\hat{t}^I - \hat{t}\right) \mid s, s_R\right\} \\ &= V_1^{IM} + V_2^{IM} + V_3^{IM}; \end{aligned} \tag{3.2.9}$$

e.g., Särndal (1992). From (3.2.9), the variance of (3.2.2) is the sum of three terms: the anticipated sampling variance, V_1^{IM}, of the complete data estimator $\hat{t}$, the non-response variance, V_2^{IM}, and a mixed component, V_3^{IM}.

3.3. Complete data pseudo-population bootstrap methods

In the case of simple random sampling without replacement, Gross (1980) proposed the without replacement bootstrap method in which, assuming that $k = {\pi_i}^{-1} = N/n$ is an integer, a pseudo-population is first created by replicating each element in the original sample, s, k times. A bootstrap sample is then drawn from the created pseudo-population according to the sampling design utilized for selecting the original sample. However, in practice, k is rarely an integer. In the case of simple random sampling without replacement, Booth et al. (1994) proposed duplicating each unit in the original sample $k = \lfloor {\pi_i}^{-1} \rfloor = \lfloor N/n \rfloor$ times ($\lfloor \cdot \rfloor$ denotes the greatest integer smaller than), and completing the pseudo-population by taking a simple random sample without replacement of size $N - nk$ from s. Chauvet (2007) extended the method of Booth et al. (1994) to the class of high entropy sampling designs, which includes Poisson sampling as a special case. In the case of Poisson sampling, the pseudo-population is created by first replicating the ith unit $k_i = \lfloor {\pi_i}^{-1} \rfloor$ times, for all i in s. Then, the pseudo-population is completed by taking a Poisson sample from s with inclusion probability ${\pi_i}^{-1} - \lfloor {\pi_i}^{-1} \rfloor$ for unit i. Bootstrap samples are then selected from the pseudo-population according to the original sampling design with the original inclusion probabilities.

A general pseudo-population bootstrap algorithm can be described as follows:

(1) A pseudo-population U_p^* is constructed by duplicating unit i in the original sample, s, $k_i = \lfloor {\pi_i}^{-1} \rfloor$ times and adding a further sample from s according to the original sampling design with inclusion probability ${\pi_i}^{-1} - \lfloor {\pi_i}^{-1} \rfloor$ for unit i.

(2) A bootstrap sample, s^*, is selected from U_p^* mimicking the original sampling design using the original inclusion probabilities. Define the bootstrap statistic $\hat{t}^* = \sum_{i \in s^*} w_i^* y_i^*$, where $w_i^* = \pi_i^{*-1}$ and π_i^* denote the survey

weight and the original inclusion probability associated with the ith unit in s^*, respectively.

(3) Repeat Step 2 a large number of times, B (say), to get $\hat{t}_1^*, \ldots, \hat{t}_B^*$. Compute

$$\widehat{V}_B^* = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{t}_b^* - \hat{t}_{(\cdot)}^* \right)^2 ,$$

where $\hat{t}_{(\cdot)}^* = \sum_{b=1}^{B} \hat{t}_b^* / B$.

(4) Repeat Steps 1–3 a large number of times, C (say), to get $\widehat{V}_{1B}^*, \ldots, \widehat{V}_{CB}^*$.

(5) A bootstrap variance estimator of $V_p(\hat{t})$ is $E_{u*} \left\{ V_{p*} \left(\hat{t}^* \mid U_p^* \right) \mid s \right\}$, where the subscripts u^* and $p*$ indicates the sampling mechanisms in Step 1 and Step 2, respectively. In practice, we use its Monte Carlo approximation

$$\frac{1}{C} \sum_{c=1}^{C} \widehat{V}_{cB}^*,$$

Booth et al. (1994) and Chauvet (2007) showed that the above scheme leads to an asymptotically unbiased bootstrap variance estimator for simple random sampling without replacement and Poisson sampling, respectively.

3.4. Bootstrap method with respect to NRM approach

Under the NRM approach, there are two random mechanisms: the selection of the sample according to the sampling design and the non-response mechanism, which we assume known up to some constants that we can estimate from the respondents. To construct the pseudo-population, we use the fact that the set of respondents to item y can be viewed as a sample that would have been selected by a Poisson sampling design with (unknown) inclusion probabilities p_i. In practice, the p_i's being unknown, the estimated response probabilities $\hat{p}_i$ are used in the bootstrap procedures. The pseudo-population is created in two distinct steps: first, a pseudo-sample s_p^* of size n_p^* is created from the set of respondents s_R by applying the method of Chauvet (2007) for Poisson sampling. Then, the pseudo-population U_p^* is constructed from s_p^* according to the method of Booth et al. (1994) if the original sample was selected under simple random sampling without replacement, or according to the method of Chauvet (2007) in the case of Poisson sampling. Bootstrap samples s^* are selected from the pseudo-population

U_p^* according to the original sampling design. Note that U_p^* will be made of vectors $(y_i, \boldsymbol{x}_i, \pi_i, \hat{p}_i)$.

Afterwards, non-response is generated in s^* according to Poisson sampling with the estimated original response probabilities $\hat{p}_i$ as the inclusion probabilities. The resulting bootstrap set of respondents is denoted by s_R^*. Missing values (i.e., units belonging in $s^* \backslash s_R^*$) are filled in using the same imputation method that was utilized in the original sample. Finally, the bootstrap imputed estimator $\hat{t}^{*I}$ is computed from the imputed bootstrap data set. A bootstrap variance estimator of V^{NRM} is

$$V^{NRM*} = E_{s*u*} \left(E_{p*q*} \left[\left\{ \hat{t}^{*I} - E_{p*q*} \left(\hat{t}^{*I} \mid s_R, U_p^* \right) \right\}^2 \mid s_R, U_p^* \right] \mid s_R \right), \qquad (3.4.1)$$

where the subscripts $s*$, $u*$, $p*$ indicate, respectively, the sampling mechanisms for generating s_p^* and U_p^* and for selecting s^*, whereas the subscript $q*$ indicates the mechanism used to generate s_R^*.

The NRM Scheme can be described as follows:

NRM Scheme:

(1) For each $i \in s_R$, make $k_{2i} = \left\lfloor \hat{p}_i^{-1} \right\rfloor$ copies of $(y_i, \boldsymbol{x}_i, \pi_i, \hat{p}_i)$ to construct a partial pseudo-sample s_1^*. Use Poisson sampling with inclusion probabilities $\hat{p}_i^{-1} - \left\lfloor \hat{p}_i^{-1} \right\rfloor$ for $i \in s_R$ to draw a further pseudo-sample s_2^* of vectors $(y_i, \boldsymbol{x}_i, \pi_i, \hat{p}_i)$ from s_R. Combine s_1^* and s_2^* to create the pseudo-sample s_p^* of size $n_p^* = \sum_{i \in s_R} \left[k_{2i} + I_i(s_2^*) \right]$, where $I(\cdot)$ denotes the usual indicator function.

(2) If the original sample was selected according to simple random sampling, first, make $k_1^* = \left\lfloor N/n_p^* \right\rfloor$ copies of the vectors in s_p^* to build a partial pseudo-population U_1^*. Then, draw a simple random sample of vectors, U_2^*, of size $N - n_p^* k_1^*$ from s_p^*. In the case of Poisson sampling, the pseudo-population is created by duplicating unit i in the pseudo-sample, s_p^*, $k_{1i}^* = \left\lfloor \pi_i^{*-1} \right\rfloor$ times to build U_1^* and taking a sample, U_2^*, from s_p^* according to Poisson sampling with the inclusion probability $\pi_i^{*-1} - \left\lfloor \pi_i^{*-1} \right\rfloor$ for unit i. In both cases, the pseudo-population, U_p^*, is obtained by combining U_1^* and U_2^*.

(3) The bootstrap sample $s^* = \{(y_i^*, \boldsymbol{x}_i^*, \pi_i^*, p_i^*)\}_{i=1}^n$ is drawn from U_p^* using the original sampling design.

(4) Generate the bootstrap sample of response indicators, $\{r_i^*\}_{i=1}^n$, using the original estimated response probabilities, i.e., $r_i^* \sim Bernoulli(p_i^*)$ for all $i \in s^*$. Afterwards, impute the bootstrap missing values in $s^* \setminus s_R^*$ so that the vector of imputed values is

$$y_i^{*I} = \begin{cases} y_i^*, & \text{if } r_i^* = 1, \\ \boldsymbol{x}_i^{*\top} \hat{\boldsymbol{\beta}}_r^*, & \text{if } r_i^* = 0, \end{cases}$$

where

$$\hat{\boldsymbol{\beta}}_r^* = \left\{ \sum_{i \in s^*} w_i^* r_i^* \left(\frac{1 - \hat{p}_i^*}{\hat{p}_i^*} \right) \frac{\boldsymbol{x}_i^* \boldsymbol{x}_i^{*\top}}{c_i^*} \right\}^{-1} \sum_{i \in s^*} w_i^* r_i^* \left(\frac{1 - \hat{p}_i^*}{\hat{p}_i^*} \right) \frac{\boldsymbol{x}_i^* y_i^*}{c_i^*}$$

with $\hat{p}_i^*$ denoting the estimated response probability for unit i in s^* (recomputed from the bootstrap values), $w_i^* = \pi_i^{*-1}$ and $c_i^* = \boldsymbol{\lambda}^\top \boldsymbol{x}_i^*$. The bootstrap estimator of t is defined as

$$\hat{t}^{*I} = \sum_{i \in s^*} w_i^* y_i^{*I}.$$

(5) Repeat Steps 3 and 4 a large number of times, B, to get $\hat{t}_1^{*I}, \ldots, \hat{t}_B^{*I}$. Compute

$$\widehat{V}_B^{NRM*} = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{t}_b^{*I} - \hat{t}_{(\cdot)}^{*I} \right)^2,$$

where $\hat{t}_{(\cdot)}^{*I} = \sum_{b=1}^B \hat{t}_b^{*I} / B$.

(6) Repeat Steps 1–5 a large number of times, C, to get $\widehat{V}_{1B}^{NRM*}, \ldots, \widehat{V}_{CB}^{NRM*}$.

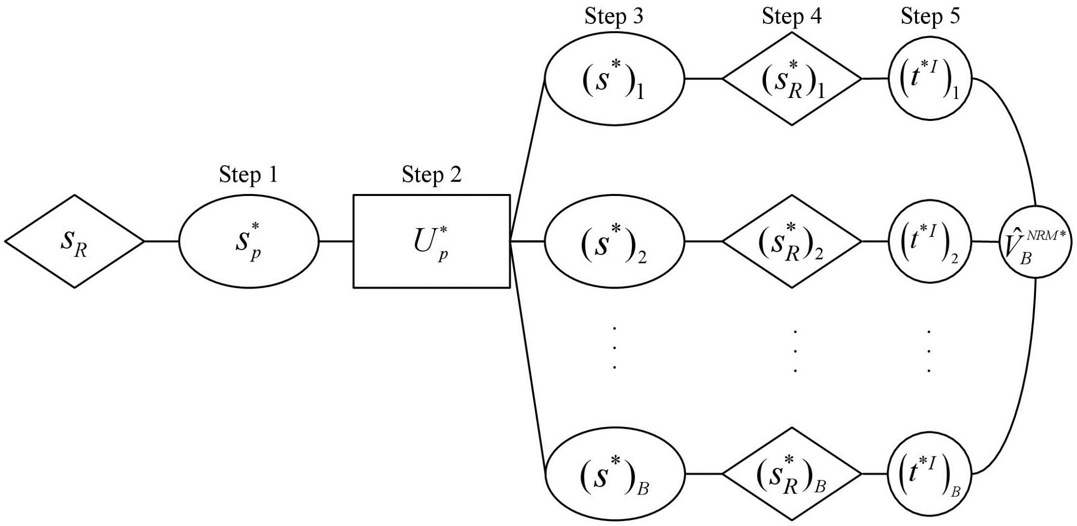
(7) A bootstrap variance estimator of V^{NRM} is (3.4.1). In practice, we use its Monte Carlo approximation

$$\widehat{V}^{NRM*} = \frac{1}{C} \sum_{c=1}^{C} \widehat{V}_{cB}^{NRM*}.$$

Figure 3.1 describes how $\widehat{V}_B^{NRM*}$ is obtained via Steps 1 to 5 of the above algorithm. The complete algorithm requires repeating these steps C times. We will discuss the choice of B and C in section 3.7. We will argue that $C = 1$ is often sufficient.

In the Appendix B, we show that V^{NRM*} in (3.4.1) is asymptotically unbiased for the true variance, V^{NRM} under the NRM scheme when the original sample is

Figure 3.1. One cycle of the pseudo-population bootstrap pattern under the NRM approach.



selected according to simple random sampling without replacement and Poisson sampling. That is,

$$E_{pq}\left(V^{NRM*}\right) \simeq V^{NRM}. \tag{3.4.2}$$

3.5. Bootstrap method with respect to IM approach

Under the IM approach, there are three random mechanisms: the selection of the sample according to the sampling design, a non-response mechanism which is totally unknown to us, and a known model (up to unknown constants that can be estimated from the respondents) generating the variable under study. Given that the non-response mechanism is unknown, the only hope to get an unbiased estimator of the variance of the estimator of t is to keep the response indicators from the sample fixed and to use them, along with an estimate of the model generating y to construct the pseudo-population. Define the standardized centered residual for unit i, $\tilde{e}_i$, as

$$\tilde{e}_i = \frac{e_i}{\sqrt{c_i}} - \frac{1}{n_R} \sum_{j \in s_R} \frac{e_j}{\sqrt{c_j}},$$

where $e_i = y_i - \tilde{y}_i$ with $\tilde{y}_i$ given by (3.2.5) and $n_R = \sum_{i \in s} r_i$ denotes the number of respondents.

A pseudo-population $\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}$ of size $\tilde{N}$ of auxiliary variables, of inclusion probabilities, and of response indicators is first created from $s_{\boldsymbol{x};\boldsymbol{r}} = \{(\boldsymbol{x}_i, \pi_i, r_i)\}_{i=1}^n$. In the case of simple random sampling without replacement, we have $\tilde{N} = N$. For Poisson sampling, we have $E_p E_{\tilde{u}*}(\tilde{N}) = N$, where the subscript $\tilde{u}*$ denotes the sampling mechanism used for creating $\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}$. Unlike in the case of the NRM approach, this pseudo-population is built in a single step by applying a complete data pseudo-population method on the sample $s_{\boldsymbol{x};\boldsymbol{r}}$. Then, an i.i.d. sample of size $\tilde{N}$, $U^*_\varepsilon = \{\varepsilon^*_i\}_{i=1}^{\tilde{N}}$, is selected from the set of standardized residuals $\{\tilde{e}_i\}_{i \in s_R}$. Finally, the bootstrap pseudo-population $\tilde{U}^*_p$ of vectors $\{(y^*_i, \boldsymbol{x}^*_i, \pi^*_i, r^*_i)\}_{i=1}^{\tilde{N}}$ is obtained using the predicted values in equation (3.2.5), the auxiliary variables $\boldsymbol{x}^*$ in $\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}$ and the selected bootstrap errors U^*_ε; that is, the y-values in $\tilde{U}^*_p$ are generated according to $y^*_i = \boldsymbol{x}^{*\top}_i \hat{\boldsymbol{\beta}}_r + \sqrt{c^*_i}\varepsilon^*_i$, where $c^*_i = \boldsymbol{\lambda}^\top \boldsymbol{x}^*_i$, $i = 1, \ldots, \tilde{N}$. The bootstrap sample $\tilde{s}^*$ is drawn from $\tilde{U}^*_p$ according to the original sampling design. The bootstrap set of respondents, $\tilde{s}^*_R$, is immediately identified through the response indicators r^*_i obtained from the original response indicators when constructing $\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}$. A bootstrap variance estimator of V^{IM} is

$$V^{IM*} = E_{\tilde{u}*}\left[E_{m*p*}\left\{\left(\hat{t}^{*I} - t^*\right)^2 \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}}, \tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}\right\} \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}}\right], \tag{3.5.1}$$

where the subscripts $m*$ and $p*$ denote the bootstrap imputation model and the sampling mechanism used for selecting $\tilde{s}^*$, respectively.

The IM Scheme can be described as follows:

IM Scheme:

(1) First, make $k_i = \lfloor {\pi_i}^{-1} \rfloor$ copies of unit i in $s_{\boldsymbol{x};\boldsymbol{r}} = \{(\boldsymbol{x}_i, \pi_i, r_i)\}_{i=1}^n$, for all i, to build a partial pseudo-population $\tilde{U}^*_1$. Then, draw a random sample, $\tilde{U}^*_2$, from $s_{\boldsymbol{x};\boldsymbol{r}}$ according to the original sampling design with inclusion probability ${\pi_i}^{-1} - \lfloor {\pi_i}^{-1} \rfloor$ for unit i. The pseudo-population, $\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}$ of size $\tilde{N}$, is obtained by combining $\tilde{U}^*_1$ and $\tilde{U}^*_2$.

(2) Draw an i.i.d. sample of size $\tilde{N}$, $U^*_\varepsilon = \{\varepsilon^*_i\}_{i=1}^{\tilde{N}}$, from the sample of centered standardized residuals $\{\tilde{e}_i\}_{i \in s_R}$. Combining $\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}$ and U^*_ε, define the

bootstrap pseudo-population $\tilde{U}_p^* = \{(y_i^*, \boldsymbol{x}_i^*, \pi_i^*, r_i^*)\}_{i=1}^{\tilde{N}}$, where

$$y_i^* = \boldsymbol{x}_i^{*\top}\hat{\boldsymbol{\beta}}_r + \sqrt{c_i^*}\varepsilon_i^*$$

and $c_i^* = \boldsymbol{\lambda}^\top \boldsymbol{x}_i^*$ for $i = 1, \ldots, \tilde{N}$.

(3) The bootstrap sample $\tilde{s}^* = \{(y_i^*, \boldsymbol{x}_i^*, \pi_i^*, r_i^*)\}_{i=1}^{n}$ is drawn from $\tilde{U}_p^*$ using the original sampling design. The set of respondents $\tilde{s}_R^*$ is defined as those units in the bootstrap sample for which the response indicators r_i^* is 1. Then impute the bootstrap missing values in $\tilde{s}^* \setminus \tilde{s}_R^*$ so that the vector of imputed values is

$$y_i^{*I} = \begin{cases} y_i^*, & \text{if } r_i^* = 1, \\ \boldsymbol{x}_i^{*\top}\hat{\boldsymbol{\beta}}_r^*, & \text{if } r_i^* = 0, \end{cases}$$

where

$$\hat{\boldsymbol{\beta}}_r^* = \left\{ \sum_{i \in \tilde{s}^*} w_i^* r_i^* \left(\frac{1 - \hat{p}_i^*}{\hat{p}_i^*} \right) \frac{\boldsymbol{x}_i^* \boldsymbol{x}_i^{*\top}}{c_i^*} \right\}^{-1} \sum_{i \in \tilde{s}^*} w_i^* r_i^* \left(\frac{1 - \hat{p}_i^*}{\hat{p}_i^*} \right) \frac{\boldsymbol{x}_i^* y_i^*}{c_i^*}$$

with $\hat{p}_i^*$ denoting the estimated response probability for $i \in \tilde{s}^*$ and $w_i^* = \pi_i^{*-1}$. The bootstrap statistics are defined as

$$\hat{t}^{*I} = \sum_{i \in \tilde{s}^*} w_i^* y_i^{*I} \quad \text{and} \quad t^* = \sum_{i \in \tilde{U}_p^*} y_i^*.$$

(4) Repeat Steps 2 and 3 a large number of times, B, to get $\left(\hat{t}_1^{*I} - t_1^*\right)^2, \ldots, \left(\hat{t}_B^{*I} - t_B^*\right)^2$. Compute

$$\widehat{V}_B^{IM*} = \frac{1}{B} \sum_{b=1}^{B} \left(\hat{t}_b^{*I} - t_b^*\right)^2.$$

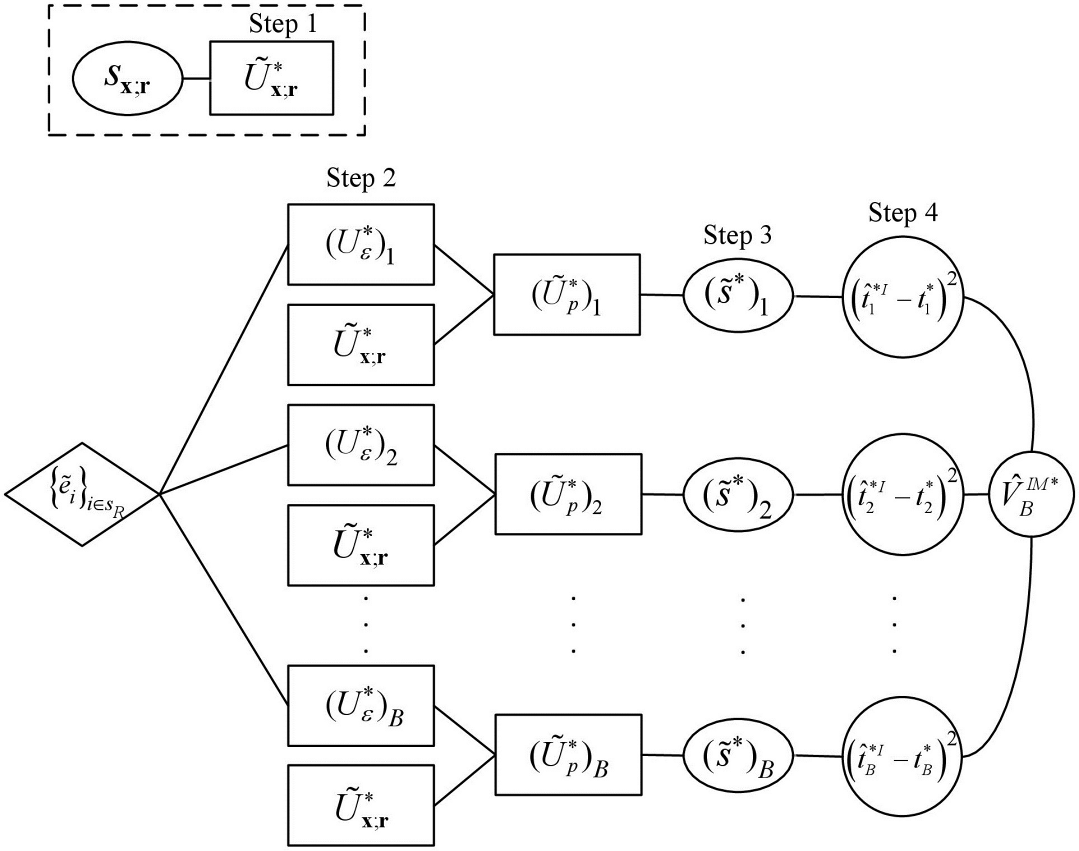
(5) Repeat Steps 1–4 a large number of times, C, to get $\widehat{V}_{1B}^{IM*}, \ldots, \widehat{V}_{CB}^{IM*}$.

(6) A bootstrap variance estimator of V^{IM} is (3.5.1). In practice, we use its Monte Carlo approximation

$$\widehat{V}^{IM*} = \frac{1}{C} \sum_{c=1}^{C} \widehat{V}_{cB}^{IM*}.$$

Figure 3.2 illustrates how the IM Scheme works when $C = 1$. The complete algorithm requires repeating the steps presented in this figure, C times.

Figure 3.2. One cycle of the pseudo-population bootstrap pattern under the IM approach.


Step 1
$S_{\mathbf{x};\mathbf{r}}$
$\tilde{U}^*_{\mathbf{x};\mathbf{r}}$
Step 2
Step 3
Step 4
$\{\tilde{e}_i\}_{i\in s_R}$
$(U^*_\varepsilon)_1$
$\tilde{U}^*_{\mathbf{x};\mathbf{r}}$
$(\tilde{U}^*_p)_1$
$(\tilde{s}^*)_1$
$\left(\hat{t}^{*I}_1 - t^*_1\right)^2$
$(U^*_\varepsilon)_2$
$\tilde{U}^*_{\mathbf{x};\mathbf{r}}$
$(\tilde{U}^*_p)_2$
$(\tilde{s}^*)_2$
$\left(\hat{t}^{*I}_2 - t^*_2\right)^2$
$\hat{V}^{IM*}_B$
$(U^*_\varepsilon)_B$
$\tilde{U}^*_{\mathbf{x};\mathbf{r}}$
$(\tilde{U}^*_p)_B$
$(\tilde{s}^*)_B$
$\left(\hat{t}^{*I}_B - t^*_B\right)^2$

We show in the Appendix B that V^{IM*} in (3.5.1) is asymptotically unbiased for V^{IM} under the IM Scheme when the original sample is selected according to simple random sampling without replacement and Poisson sampling. That is,

$$E_{mpq}\left(V^{IM*}\right) \simeq V^{IM}.$$

3.6. Doubly robust bootstrap method

In this section, we present a doubly robust bootstrap variance estimator which remains asymptotically unbiased for the true variance if either the non-response model or the imputation model is correctly specified. From (3.4.2), the variance estimator V^{NRM*} is asymptotically unbiased for V^{NRM} if the non-response model is correctly specified. We now express the bias of V^{NRM*} with respect to the IM approach:

$$B^{IM} = E_{mpq}\left(V^{NRM*}\right) - E_{mpq}\left(\hat{t}^{I} - t\right)^{2}. \tag{3.6.1}$$

By combining the IM and NRM Schemes, we first estimate $E_{mpq}\left(V^{NRM*}\right)$ in (3.6.1) through a double bootstrap procedure. The outer bootstrap will be done under the IM Scheme so as to provide a bootstrap sample with responses which are consistent with the generating model and response indicators which remain fixed so that they satisfy the (unknown) non-response mechanism, as was done in Section 3.5. This will lead to a bootstrap sample $\tilde{s}^{*}$, which includes the set of bootstrap respondents $\tilde{s}_{R}^{*}$. Using this bootstrap sample generated from the estimated IM model, an inner bootstrap loop based on the NRM Scheme of Section 3.4 is applied on the outer bootstrap sample. From the inner loop, we get an estimator V^{NRM*} whereas from the outer loop, we get an estimator of its mean under the imputation model, that is

$$E^{IM*}\left(V^{NRM*}\right) = E_{\tilde{u}*}\left\{E_{m*p*}\left(V^{NRM*} \mid s_{R}, s_{\boldsymbol{x};\boldsymbol{r}}, \tilde{U}_{\boldsymbol{x};\boldsymbol{r}}^{*}\right) \mid s_{R}, s_{\boldsymbol{x};\boldsymbol{r}}\right\}. \tag{3.6.2}$$

Note that in the outer loop, the set of respondents s_R comes from the original sample and satisfies the true generating non-response mechanism, so that the resulting bootstrap estimator is valid even when the non-response model postulated in computing V^{NRM*} is misspecified. In the Appendix B, we show that

$$E_{mpq}\left\{E^{IM*}\left(V^{NRM*}\right)\right\} \simeq E_{mpq}\left(V^{NRM*}\right).$$

The algorithm for obtaining a Monte Carlo approximation of $E^{IM*}\left(V^{NRM*}\right)$ can be described as follows.

(1) Do Steps 1–3 in the IM Scheme to obtain the bootstrap sample of respondents $\tilde{s}_{R}^{*}$. But note that the estimated response probabilities $\hat{p}_i$ must be added to the pseudo-population so that the pseudo-population is $\tilde{U}_{p}^{*} = \{(y_i^{*}, \boldsymbol{x}_i^{*}, \pi_i^{*}, r_i^{*}, p_i^{*})\}_{i=1}^{\tilde{N}}$. The bootstrap sample of respondents $\tilde{s}_{R}^{*}$ is determined by the response indicators r_i^{*} only; the estimated response probabilities will be used in the next step only.

(2) Apply the NRM Scheme to the bootstrap sample of respondents $\tilde{s}_{R}^{*}$ by assuming $C = C_2$ and $B = B_2$ to get $\widehat{V}^{NRM*}$.

(3) Repeat Steps 1 and 2 a large number of times, B_1, to get $\left(\widehat{V}^{NRM*}\right)_1, \ldots, \left(\widehat{V}^{NRM*}\right)_{B_1}$. Let

$$\widehat{E}^{IM*}_{B_1}\left(\widehat{V}^{NRM*}\right) = \frac{1}{B_1}\sum_{b_1=1}^{B_1}\left(\widehat{V}^{NRM*}\right)_{b_1}.$$

(4) Repeat Steps 1–3 a large number of times, C_1, to get

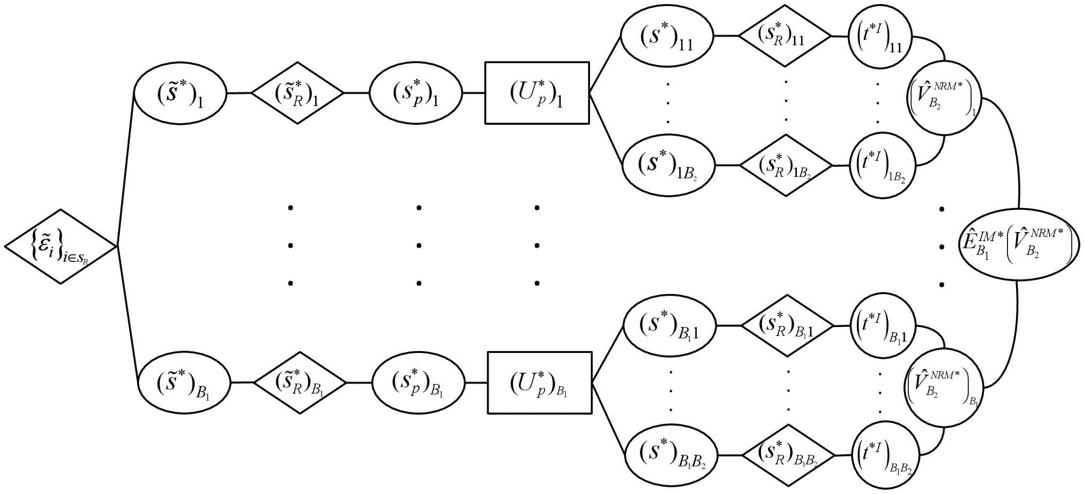
$$\widehat{E}^{IM*}_{1B_1}\left(\widehat{V}^{NRM*}\right), \ldots, \widehat{E}^{IM*}_{C_1B_1}\left(\widehat{V}^{NRM*}\right).$$

(5) An estimator of $E_{mpq}\left(V^{NRM*}\right)$ is (3.6.2). In practice, we use its Monte Carlo approximation

$$\widehat{E}^{IM*}\left(\widehat{V}^{NRM*}\right) = \frac{1}{C_1}\sum_{c_1=1}^{C_1}\widehat{E}^{IM*}_{c_1B_1}\left(\widehat{V}^{NRM*}\right).$$

Figure 3.3 displays this algorithm for $C_1 = 1$ and $C_2 = 1$.

FIGURE 3.3. The bootstrap procedure pattern to estimate $E_{mpq}\left(V^{NRM*}\right)$ assuming $C_1 = C_2 = 1$.



For the second term on the right hand side of (3.6.1), we simply use V^{IM*} given by (3.5.1). A bias-adjusted bootstrap variance estimator is defined as

$$\begin{aligned} V^{DR*} &= V^{NRM*} - B^{IM*} \\ &= V^{NRM*} - E_{\tilde{u}*}\left\{E_{m*p*}\left(V^{NRM*} \mid s_R, s_{\mathbf{x};\mathbf{r}}, \tilde{U}^*_{\mathbf{x};\mathbf{r}}\right) \mid s_R, s_{\mathbf{x};\mathbf{r}}\right\} + V^{IM*}. \end{aligned} \tag{3.6.3}$$

Clearly, the variance estimator V^{DR*} is asymptotically unbiased for the true variance provided that the imputation model (3.2.3) holds. We also have $E_{pq}\left(V^{NRM*}\right)$

$\simeq E_{pq}\left(\hat{t}^I - t\right)^2$ which implies that the adjusted bias B^{IM} is approximately equal to zero if the non-response model is correctly specified regardless of the validity of the imputation model. As a result, our bias-adjusted variance estimator V^{DR*} is doubly robust.

3.7. Simulation Study

To assess the performance of the proposed methods, we performed a limited simulation study. We generated a population of size $N = 2000$ with three variables: a study variable y and two auxiliary variables x_1 and x_2. The x_1-values were generated from a gamma distribution with shape and scale parameters set to 2 and 1, respectively, whereas the x_2-values were generated from a gamma distribution with shape and scale parameters set to 2 and 0.5, respectively. Given x_1 and x_2, the y-values were generated according to the following linear regression model

$$y_i = 0.1 + 2x_{1i} - 3x_{2i} + \varepsilon_i, \quad i = 1, \ldots, 2000, \tag{3.7.1}$$

where ε_i follows a normal distribution with mean 0 and standard deviation 2.7, which led to a coefficient of determination of this regression model approximately equal to 0.66. Note that since the distribution of x_1 and that of x_2 are asymmetric, the distribution of the resulting study variable y is also asymmetric.

From the population, we drew $K = 2000$ samples, s, of size n, according to simple random sampling without replacement. The sample size n was set to 120 and 800 which corresponds to a sampling fraction, $f = n/N$, equal to 6% and 40%, respectively.

In each sample, non-response for the variable under study y for unit i was generated from a Bernoulli distribution with parameter p_i, where p_i follows a logistic regression model:

$$\text{logit}(p_i) = 0.2 + 0.4x_{1i} - 0.3x_{2i}. \tag{3.7.2}$$

The parameters were chosen so that the overall response rate was approximately equal to 65%. In practice, it is not rare to observe non-response rate between 30% and 50%.

To replace the missing values, we used linear regression imputation given by (3.2.5) based on different working models. We considered three distinct scenarios:

Scenario 1: Both the imputation and the non-response model are correctly specified.

Scenario 2: Only the non-response model is correctly specified.

Scenario 3: Only the imputation model is correctly specified.

The different scenarios as well as the corresponding working models are shown in Table 3.1.

TABLE 3.1. Working models used for imputation

Scenario	Non-response working model	Outcome regression working model
1	logit $(p_i) = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i}$	$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$
2	logit $(p_i) = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i}$	$y_i = \beta_0 + \beta_1 x_{1i}$
3	logit $(p_i) = \gamma_0 + \gamma_1 x_{1i}$	$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$

In each sample consisting of observed and imputed data, we computed the imputed estimator $\hat{t}^I$ given by (3.2.2). Also, in each sample, we computed the following bootstrap variance estimators in order to estimate the variance of $\hat{t}^I$:

(i) $\widehat{V}_B^{NRM*}$ based on $C = 1$ pseudo-population and $B = 10000$ bootstrap samples;

(ii) $\widehat{V}_B^{IM*}$ based on $C = 1$ pseudo-population and $B = 10000$ bootstrap samples;

(iii) $\widehat{V}_B^{DR*}$ based on $C_1 = 1$ and $C_2 = 1$ pseudo-population, $B_1 = 100$ bootstrap samples and $B_2 = 1000$ bootstrap iterations;

As a measure of bias of a variance estimator $\widehat{V}$, we computed the Monte Carlo percent relative bias (RB)

$$RB_{MC}(\widehat{V}) = 100 \times \left(\frac{E_{MC}(\widehat{V}) - V_{MC}(\hat{t}^I)}{V_{MC}(\hat{t}^I)} \right),$$

where

$$E_{MC}(\widehat{V}) = \frac{1}{K} \sum_{k=1}^{K} \widehat{V}_k \quad \text{and} \quad V_{MC}(\hat{t}^I) = \frac{1}{K'-1} \sum_{k=1}^{K'} \left\{ \hat{t}^I_k - E_{MC}(\hat{t}^I) \right\}^2$$

with $\hat{t}_k^I$ and $\widehat{V}_k$ denoting the estimators $\hat{t}^I$ and $\widehat{V}$ in the kth sample, respectively, and $E_{MC}(\hat{t}^I) = (K')^{-1} \sum_{k=1}^{K'} \hat{t}_k^I$. The Monte Carlo variance of $\hat{t}^I$, $V_{MC}(\hat{t}^I)$, was obtained through $K' = 25000$ independent runs. In Table 3.3, the latter corresponds to either V^{NRM} or V^{IM}. The Monte Carlo variance V^{NRM} was obtained by fixing the population and simulating the effect of sampling and non-response, whereas V^{IM} was obtained by generating a new population at each iteration and then simulating the effect of sampling and non-response.

For the first two variance estimators, we used $C = 1$, while we used $C_1 = C_2 = 1$ for the third one corresponding to creating a single pseudo-population. This choice is justified by the fact that different choices of the pair (C, B) seem to make little difference. We ran preliminary simulations to assess the performance of $\widehat{V}_B^{NRM*}$ and $\widehat{V}_B^{IM*}$ with four choices of (C, B) such that $C \times B = 10000$. We selected $K = 2000$ simple random samples without replacement of size $n = 800$, which corresponds to a sampling fraction of 40%. Based on the $K' = 25000$ independent runs, the Monte Carlo variance V_{MC} under the NRM approach was 79637.71 and that under the IM approach was 81725.91. Table 3.2 shows the Monte Carlo average of $\widehat{V}_B^{NRM*}$ and $\widehat{V}_B^{IM*}$ as well as their stability for different choices of (C, B). Both $\widehat{V}_B^{NRM*}$ and $\widehat{V}_B^{IM*}$ exhibited very similar behaviour in terms of both average and stability regardless of the choice of (C, B). It is worth noting that for the IM Scheme, a sample of size 800 out of a population of size 2000 will induce the maximum variability from pseudo-population to pseudo-population since the sample of size $n = 800$ will be repeated twice and a simple random sample without replacement of size $(n/2) = 400$ from it is needed to complete the pseudo-population. Hence, since a sample size of 800 is the worst case for a population of size 2000 in terms of variability from one pseudo-population to the other one, then if varying the number of pseudo-populations for a fixed amount of computing effort (i.e., the product $C \times B$) does not really change the results much in this case, it suggests that going beyond one pseudo-population (i.e., $C = 1$) is probably not needed.

In order to reduce the processing time, we used $B_2 = 1000$ instead of 10000 in the case of the doubly robust variance estimator, $\widehat{V}_B^{DR*}$.

Table 3.2. Monte Carlo expectation and variance of $\widehat{V}^{NRM*}$ and $\widehat{V}^{IM*}$ based on C pseudo-populations and B bootstrap samples based on 2000 iterations in the case of $f = 40\%$

	$E_{MC}(\widehat{V}^{NRM*})$	$V_{MC}(\widehat{V}^{NRM*})$	$E_{MC}(\widehat{V}^{IM*})$	$V_{MC}(\widehat{V}^{IM*})$
$(C, B) = (1, 10000)$	80117.24	23367509	80244.62	17500868
$(C, B) = (10, 1000)$	80133.35	20551672	80290.77	17368245
$(C, B) = (20, 500)$	79870.77	20661606	80107.54	17143604
$(C, B) = (100, 100)$	79985.16	21549946	80205.43	17605525

Table 3.3 shows the relative bias of three bootstrap variance estimators. The three bootstrap estimators exhibited small bias in Scenario 1. In this case, V^{NRM} and V^{IM} were approximately equal; see expression (3.B.3). In Scenario 2, the estimator $\widehat{V}_B^{NRM*}$ showed good performance with an absolute relative bias less than 3%, which was expected as the non-response model was correctly specified. On the other hand, the estimator $\widehat{V}_B^{IM*}$ was biased, especially for $f = 40\%$. This results is not surprising as $\widehat{V}_B^{IM*}$ is approximately unbiased for V^{IM} provided that the imputation model holds, which was not the case in Scenario 2. The doubly robust estimator $\widehat{V}_B^{DR*}$ showed small bias in Scenario 2. Finally, in Scenario 3, for which the imputation model was correctly specified, we note that $\widehat{V}_B^{IM*}$ performed well with an absolute relative bias less than 3.5%. As expected, the estimator $\widehat{V}_B^{NRM*}$ was biased as the non-response model was misspecified. Once again, the doubly robust estimator $\widehat{V}_B^{DR*}$ showed small relative bias.

3.8. Discussion

In this paper, we considered the class of deterministic regression imputation procedures. While this type of procedure leads to asymptotically unbiased estimators of simple parameters such as population totals or means, it may lead to considerably biased estimators of more complex parameters such as quantiles, as deterministic regression imputation tends to distort the distribution of the variable being imputed. To overcome this problem, we may use a doubly robust

Table 3.3. Monte Carlo percent relative bias of several bootstrap variance estimators based on 2000 iterations

		$f = 6\%$		$f = 40\%$	
		V^{NRM}	V^{IM}	V^{NRM}	V^{IM}
Scenario 1	$\widehat{V}_B^{NRM*}$	-0.73	-0.39	0.70	-1.86
	$\widehat{V}_B^{IM*}$	-0.31	0.02	0.81	-1.76
	$\widehat{V}_B^{DR*}$	-0.30	0.03	0.81	-1.76
Scenario 2	$\widehat{V}_B^{NRM*}$	-0.47	-2.81	0.93	-0.14
	$\widehat{V}_B^{IM*}$	3.60	1.16	12.56	11.35
	$\widehat{V}_B^{DR*}$	0.29	-2.06	0.95	-0.12
Scenario 3	$\widehat{V}_B^{NRM*}$	-4.93	-7.80	-2.31	-3.25
	$\widehat{V}_B^{IM*}$	-0.25	-3.27	1.54	0.57
	$\widehat{V}_B^{DR*}$	0.15	-2.87	1.72	0.74

version of random regression imputation, which can be viewed as the deterministic imputation (3.2.5) plus a random noise $\tilde{\varepsilon}_i$ is added; see Haziza and Rao (2006). That is, missing y_i is replaced by the imputed value $\tilde{y}_i$

$$\tilde{y}_i = \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_r + \sqrt{c_i}\tilde{\varepsilon}_i,$$

where $\hat{\boldsymbol{\beta}}_r$ is given by (3.2.6). The residuals $\tilde{\varepsilon}_i$ are selected independently and with replacement from the set, $\{\tilde{u}_j\}_{j \in s_R}$, of standardized centered residuals observed from the responding units, with probabilities

$$pr\left(\tilde{\varepsilon}_i = \tilde{u}_j\right) = \frac{w_j \hat{p}_j^{-1}(1-\hat{p}_j)}{\sum_{l \in s} w_l r_l p_l^{-1}(1-\hat{p}_l)},$$

where $\tilde{u}_j = u_j - \bar{u}_r$ with $u_j = c_j^{-1/2}\left(y_j - \boldsymbol{x}_j^\top \hat{\boldsymbol{\beta}}_r\right)$ and $\bar{u}_r = \sum_{i \in s} w_i r_i \hat{p}_i^{-1}(1-\hat{p}_i)u_i / \sum_{l \in s} w_l r_l \hat{p}_l^{-1}(1-\hat{p}_l)$. The properties of our bootstrap procedures in the context of quantile estimation is currently under investigation.

Finally, throughout the paper, we assumed the linear regression model (3.2.6). Our results can be extended to a more general model of the form

$$y_i = m(\boldsymbol{x}_i; \boldsymbol{\beta}) + \varepsilon_i$$

for a known function $m(\cdot)$. This topic requires further research.

3.9. Appendix B

In order to establish our results, we assume in the sequel that the response probabilities are known, $(N-1)^{-1}N \simeq 1$ and $(n-1)^{-1}n \simeq 1$. In addition, we assume that no survey weight is disproportionately large and no response probability is disproportionately small; i.e.,

$$\max_i\{w_i\} = O\left(\frac{N}{n}\right) \quad \text{and} \quad \max_i\left\{\frac{1}{p_i}\right\} = O\left(\frac{n}{n_R}\right).$$

Linearization variance estimators

In this section, we give expressions of the approximate variance of $\hat{t}^I$ under both simple random sampling without replacement and Poisson sampling.

We start by the NRM approach for which the total variance of $\hat{t}^I$ is given by (3.2.8). For simple random sampling without replacement, the first term on the right hand side of (3.2.8) is

$$V_{1srs}^{NRM} = N^2\left(\frac{1-f}{n}\right)S_U^2,$$

where $S_U^2 = (N-1)^{-1}\sum_{i\in U}(y_i - \bar{Y})^2$ with $\bar{Y} = N^{-1}\sum_{i\in U} y_i$. For Poisson sampling, we have

$$V_{1Pois}^{NRM} = \sum_{i\in U}\frac{1-\pi_i}{\pi_i}y_i^2.$$

The second term on the right hand side of (3.2.8) is the non-response variance and does not depend on the sampling design. Using a first-order Taylor expansion, it can be approximated by

$$V_2^{NRM} \simeq \sum_{i\in U} w_i\left(\frac{1-p_i}{p_i}\right)\left(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}_p\right)^2, \tag{3.B.1}$$

where

$$\boldsymbol{\beta}_p = \left\{\sum_{i\in U}(1-p_i)\frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top}{c_i}\right\}^{-1}\sum_{i\in U}(1-p_i)\frac{\boldsymbol{x}_i y_i}{c_i}.$$

Therefore, the total variance of $\hat{t}^I$ with respect to the NRM approach can be approximated by $V_{1srs}^{NRM}+V_2^{NRM}$ for simple random sampling without replacement and by $V_{1Pois}^{NRM}+V_2^{NRM}$ for Poisson sampling.

We now turn to the IM approach for which the variance of $\hat{t}^I$ is given by (3.2.9). For simple random sampling without replacement, the first component of (3.2.9) is

$$V_{1srs}^{IM} = N^2\left(\frac{1-f}{n}\right)\left(\frac{1}{N}\sum_{i\in U}c_i\sigma^2 + \boldsymbol{\beta}^\top S_{\boldsymbol{xx}}\boldsymbol{\beta}\right), \tag{3.B.2}$$

where $S_{\boldsymbol{xx}} = (N-1)^{-1}\sum_{i\in U}\left(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_U\right)\left(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_U\right)^\top$ and $\bar{\boldsymbol{x}}_U = N^{-1}\sum_{i\in U}\boldsymbol{x}_i$. For Poisson sampling, we have

$$V_{1Pois}^{IM} = \sigma^2\sum_{i\in U}\frac{1-\pi_i}{\pi_i}c_i + \boldsymbol{\beta}^\top\left(\sum_{i\in U}\frac{1-\pi_i}{\pi_i}\boldsymbol{x}_i\boldsymbol{x}_i^\top\right)\boldsymbol{\beta}.$$

The second and third terms on the right hand side of (3.2.9) do not depend on the sampling design. They are given by

$$\begin{aligned} V_2^{IM} &= \sigma^2\sum_{i\in U}w_i(1-p_i)c_i \\ &+\ \sigma^2 E_{pq}\left[\left\{\sum_{j\in s}w_j(1-r_j)\boldsymbol{x}_j^\top\right\}\hat{\boldsymbol{T}}_r^{-1}\hat{\boldsymbol{K}}_r\hat{\boldsymbol{T}}_r^{-1}\left\{\sum_{j\in s}w_j(1-r_j)\boldsymbol{x}_j\right\}\right] \end{aligned}$$

and

$$V_3^{IM} = 2\sigma^2 E_{pq}\left[\left\{\sum_{j\in s}w_j(1-r_j)\boldsymbol{x}_j^\top\right\}\hat{\boldsymbol{T}}_r^{-1}\hat{\boldsymbol{L}}_r\right] - 2\sigma^2\sum_{i\in U}(w_i-1)(1-p_i)c_i,$$

where

$$\hat{\boldsymbol{T}}_r = \sum_{i\in s}w_i r_i\left(\frac{1-p_i}{p_i}\right)\frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top}{c_i},\ \hat{\boldsymbol{K}}_r = \sum_{i\in s}w_i^2 r_i\left(\frac{1-p_i}{p_i}\right)^2\frac{\boldsymbol{x}_i\boldsymbol{x}_i^\top}{c_i},$$

and

$$\hat{\boldsymbol{L}}_r = \sum_{i\in s}w_i(w_i-1)r_i\left(\frac{1-p_i}{p_i}\right)\boldsymbol{x}_i.$$

Therefore, the total variance of $\hat{t}^I$ with respect to the IM approach can be approximated by $V_{1srs}^{IM}+V_2^{IM}+V_3^{IM}$ for simple random sampling without replacement and by $V_{1Pois}^{IM}+V_2^{IM}+V_3^{IM}$ for Poisson sampling.

If both the non-response and imputation models are correctly specified, it can be shown that

$$\frac{E_m\left(V^{NRM}\right) - V^{IM}}{V^{IM}} = O\left(N^{-1}\right). \tag{3.B.3}$$

In this case, the variance (3.2.9) with respect to the IM reduces to

$$V^{IM} \simeq E_m V_p\left(\hat{t}\right) + \sigma^2 \sum_{i \in U} w_i \left(\frac{1-p_i}{p_i}\right) c_i$$

which implies that

$$E_m\left(V^{NRM}\right) - V^{IM} \simeq \sum_{i \in U} w_i \left(\frac{1-p_i}{p_i}\right) \left\{ E_m\left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_p\right)^2 - \sigma^2 c_i \right\}.$$

Expression (3.B.3) follows from the fact that $E_m\left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_p\right)^2 = \sigma^2 c_i + O(N^{-1})$ and $V^{IM} = O\left(N^2 {n_R}^{-1}\right).$

Bootstrap variance estimator with respect to NRM approach

In this section, we show that the bootstrap variance estimator V^{NRM*} given by (3.4.1) is approximately unbiased for V^{NRM} with respect to the NRM approach. It can be expressed as

$$\begin{aligned} V^{NRM*} &= E_{s*u*}\left\{V_{p*}E_{q*}\left(\hat{t}^{I*} \mid s_R, U_p^*, s^*\right) \mid s_R\right\} + E_{s*u*}\left\{E_{p*}V_{q*}\left(\hat{t}^{I*} \mid s_R, U_p^*, s^*\right) \mid s_R\right\} \\ &= V_1^{NRM*} + V_2^{NRM*}. \end{aligned} \tag{3.B.4}$$

Using a first-order Taylor expansion, the first and the second terms on the right hand side of (3.B.4) are respectively

$$\begin{aligned} V_{1srs}^{NRM*} &= E_{s*u*}\left\{N^2\left(\frac{1-f}{n}\right) S^2_{U_p^*} \mid s_R\right\} + O\left(\frac{N^2}{n_R^2}\right) \\ &\simeq N^2\left(\frac{1-f}{n}\right)\left\{\frac{\sum_{i\in s} r_i y_i^2/p_i}{\sum_{i\in s} r_i/p_i} - \left(\frac{\sum_{i\in s} r_i y_i/p_i}{\sum_{i\in s} r_i/p_i}\right)^2\right\} \end{aligned}$$

and

$$\begin{aligned} V_{2srs}^{NRM*} &= E_{s*u*}\left\{\sum_{i \in U_p^*} w_i^* \left(\frac{1-p_i^*}{p_i^*}\right)\left(y_i^* - \boldsymbol{x}_i^{*\top}\boldsymbol{T}_p^{*-1}\boldsymbol{t}_p^*\right)^2 \mid s_R\right\} + O\left(\frac{N^2}{n_R^2}\right) \\ &\simeq \frac{N}{\sum_{i\in s} r_i/p_i} \sum_{i \in s} w_i \frac{r_i}{p_i}\left(\frac{1-p_i}{p_i}\right)\left(y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_r\right)^2, \end{aligned}$$

for simple random sampling without replacement, where

$$S_{U_p^*}^2 = (N-1)^{-1} \sum_{i \in U_p^*} \left(y_i^* - \bar{Y}_p^* \right)^2$$

with $\bar{Y}_p^* = N^{-1} \sum_{i \in U_p^*} y_i^*$ and

$$\boldsymbol{T}_p^* = \sum_{i \in U_p^*} (1-p_i^*) \frac{\boldsymbol{x}_i^* \boldsymbol{x}_i^{*\top}}{c_i^*} \quad \text{and} \quad \boldsymbol{t}_p^* = \sum_{i \in U_p^*} (1-p_i^*) \frac{\boldsymbol{x}_i^* y_i^*}{c_i^*}.$$

For Poisson sampling, a first-order Taylor expansion leads to

$$\begin{aligned} V_{1Pois}^{NRM*} &= E_{s*u*} \left(\sum_{i \in U_p^*} \frac{1-\pi_i}{\pi_i} y_i^{*2} \mid s_R \right) + O\left(\frac{N^2}{n_R^2} \right) \\ &\simeq \sum_{i \in s} w_i \frac{r_i}{p_i} \frac{1-\pi_i}{\pi_i} y_i^2, \end{aligned}$$

and

$$\begin{aligned} V_{2Pois}^{NRM*} &= E_{s*u*} \left\{ \sum_{i \in U_p^*} w_i^* \left(\frac{1-p_i^*}{p_i^*} \right) \left(y_i^* - \boldsymbol{x}_i^{*\top} \boldsymbol{T}_p^{*-1} \boldsymbol{t}_p^* \right)^2 \mid s_R \right\} + O\left(\frac{N^2}{n_R^2} \right) \\ &\simeq \sum_{i \in s} w_i^2 \frac{r_i}{p_i} \left(\frac{1-p_i}{p_i} \right) \left(y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}_r \right)^2. \end{aligned}$$

In addition, it is easily seen that

$$E_{pq}\left(V_{1srs}^{NRM*} \right) + E_{pq}\left(V_{2srs}^{NRM*} \right) \simeq N^2 \left(\frac{1-f}{n} \right) S_U^2 + \sum_{i \in U} w_i \left(\frac{1-p_i}{p_i} \right) \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_p \right)^2$$

and

$$E_{pq}\left(V_{1Pois}^{NRM*} \right) + E_{pq}\left(V_{2Pois}^{NRM*} \right) \simeq \sum_{i \in U} \frac{1-\pi_i}{\pi_i} y_i^2 + \sum_{i \in U} w_i \left(\frac{1-p_i}{p_i} \right) \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_p \right)^2.$$

Therefore, for both simple random sampling without replacement and Poisson sampling, we have

$$E_{pq}\left(V^{NRM*} \right) \simeq V^{NRM}.$$

Bootstrap variance estimator with respect to IM approach

In order to express the bootstrap variance estimator V^{IM*}, we use the following decomposition of the total bootstrap error, $\hat{t}^{I*} - t^*$:

$$\hat{t}^{I*} - t^* = \left(\hat{t}^* - t^* \right) + \left(\hat{t}^{I*} - \hat{t}^* \right),$$

where $\hat{t}^* = \sum_{i \in \tilde{s}^*} w_i^* y_i^*$ is the bootstrap estimator of $\hat{t}$. It follows that

$$\begin{aligned} V^{IM*} &= E_{\tilde{u}*} \left\{ E_{m*} V_{p*} \left(\hat{t}^* \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}}, \tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}, U^*_{\varepsilon} \right) \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}} \right\} \\ &+ E_{\tilde{u}*} \left[E_{p*} E_{m*} \left\{ \left(\hat{t}^{*I} - \hat{t}^* \right)^2 \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}}, \tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}, \tilde{s}^* \right\} \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}} \right] \\ &+ 2 E_{\tilde{u}*} \left[E_{p*} E_{m*} \left\{ \left(\hat{t}^* - t^* \right) \left(\hat{t}^{*I} - \hat{t}^* \right) \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}}, \tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}, \tilde{s}^* \right\} \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}} \right] \\ &= V_1^{IM*} + V_2^{IM*} + V_3^{IM*}. \end{aligned} \tag{3.B.5}$$

In the second step of the IM Scheme, an i.i.d. sample of size $\tilde{N}$, $\{\varepsilon_i^*\}_{i=1}^{\tilde{N}}$, is taken from the standardized centered residuals $\{\tilde{e}_i\}_{i \in s_R}$. We have $E_{m*}(\varepsilon_i^*) = n_R^{-1} \sum_{i \in s_R} \tilde{e}_i = 0$ and

$$\begin{aligned} V_{m*}(\varepsilon_i^*) &= \frac{1}{n_R} \sum_{i \in s_R} \tilde{e}_i^2 \\ &= \frac{1}{n_R} \sum_{i \in s_R} \left\{ \left(\frac{y_i}{\sqrt{c_i}} - \frac{1}{n_R} \sum_{j \in s_R} \frac{y_j}{\sqrt{c_j}} \right) - \left(\frac{\boldsymbol{x}_i}{\sqrt{c_i}} - \frac{1}{n_R} \sum_{j \in s_R} \frac{\boldsymbol{x}_j}{\sqrt{c_j}} \right)^{\top} \hat{\boldsymbol{\beta}}_r \right\}^2 \\ &= \tilde{\sigma}^2. \end{aligned}$$

It follows that $E_{m*}(y_i^*) = \boldsymbol{x}_i^{*\top} \hat{\boldsymbol{\beta}}_r$ and $V_{m*}(y_i^*) = c_i^* V_{m*}(\varepsilon_i^*) = c_i^* \tilde{\sigma}^2$ for $i = 1, \ldots, \tilde{N}$. If the original sample, s, is selected according to simple random sampling without replacement, we have

$$\begin{aligned} V_{1srs}^{IM*} &= E_{\tilde{u}*} \left[N^2 \left(\frac{1-f}{n} \right) \frac{1}{N-1} E_{m*} \left\{ \sum_{i \in \tilde{U}_p^*} y_i^{*2} - \frac{1}{N} \left(\sum_{i \in \tilde{U}_p^*} y_i^* \right)^2 \right\} \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}} \right] \\ &= E_{\tilde{u}*} \left\{ N^2 \left(\frac{1-f}{n} \right) \left[\frac{1}{N} \sum_{i \in \tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} c_i^* \tilde{\sigma}^2 \right. \right. \\ &\left. \left. + \hat{\boldsymbol{\beta}}_r^{\top} \left\{ \frac{1}{N-1} \sum_{i \in \tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} \left(\boldsymbol{x}_i^* - \bar{\boldsymbol{x}}_p^* \right) \left(\boldsymbol{x}_i^* - \bar{\boldsymbol{x}}_p^* \right)^{\top} \right\} \hat{\boldsymbol{\beta}}_r \right] \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}} \right\} \\ &\simeq N^2 \left(\frac{1-f}{n} \right) \left(\frac{\tilde{\sigma}^2}{N} \sum_{i \in s} w_i c_i + \hat{\boldsymbol{\beta}}_r^{\top} s_{\boldsymbol{x}\boldsymbol{x}} \hat{\boldsymbol{\beta}}_r \right), \end{aligned}$$

noting that $\tilde{s}^*$ is also selected according to simple random sampling without replacement and $\tilde{N} = N$, where $\bar{\boldsymbol{x}}_p^* = N^{-1} \sum_{i \in \tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} \boldsymbol{x}_i^*$ and

$$s_{\boldsymbol{x}\boldsymbol{x}} = (n-1)^{-1} \sum_{i \in s} \left(\boldsymbol{x}_i - \bar{\boldsymbol{x}} \right) \left(\boldsymbol{x}_i - \bar{\boldsymbol{x}} \right)^{\top}$$

with $\bar{\boldsymbol{x}} = n^{-1} \sum_{i \in s} \boldsymbol{x}_i$. For Poisson sampling, we have

$$\begin{aligned}
V_{1Pois}^{IM*} &= E_{\tilde{u}*} \left\{ E_{m*} \left(\sum_{i \in \tilde{U}_p^*} \frac{1 - \pi_i^*}{\pi_i^*} y_i^{*2} \right) \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}} \right\} \\
&= E_{\tilde{u}*} \left\{ \sum_{i \in \tilde{U}_{\boldsymbol{x};\boldsymbol{r}}^*} \frac{1 - \pi_i^*}{\pi_i^*} c_i^* \tilde{\sigma}^2 + \hat{\boldsymbol{\beta}}_r^\top \left(\sum_{i \in \tilde{U}_{\boldsymbol{x};\boldsymbol{r}}^*} \frac{1 - \pi_i^*}{\pi_i^*} \boldsymbol{x}_i^* \boldsymbol{x}_i^{*\top} \right) \hat{\boldsymbol{\beta}}_r \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}} \right\} \\
&= \tilde{\sigma}^2 \sum_{i \in s} w_i \frac{1 - \pi_i}{\pi_i} c_i + \hat{\boldsymbol{\beta}}_r^\top \left(\sum_{i \in s} w_i \frac{1 - \pi_i}{\pi_i} \boldsymbol{x}_i \boldsymbol{x}_i^\top \right) \hat{\boldsymbol{\beta}}_r,
\end{aligned}$$

The second and third terms on the right hand side of (3.B.5) are identical for both sampling designs. Using a first-order Taylor expansion, they can be approximated by

$$\begin{aligned}
& V_2^{IM*} = E_{\tilde{u}*} \Bigg\{ \tilde{\sigma}^2 \sum_{i \in \tilde{U}_{\boldsymbol{x};\boldsymbol{r}}^*} w_i^* (1 - r_i^*) c_i^* \\
& + E_{p*} \Bigg[\tilde{\sigma}^2 \Bigg\{ \sum_{j \in \tilde{s}^*} w_j^* (1 - r_j^*) \boldsymbol{x}_j^{*\top} \Bigg\} \hat{\boldsymbol{T}}_r^{*-1} \hat{\boldsymbol{K}}_r^* \hat{\boldsymbol{T}}_r^{*-1} \Bigg\{ \sum_{j \in \tilde{s}^*} w_j^* (1 - r_j^*) \boldsymbol{x}_j^* \Bigg\} \Bigg] \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}} \Bigg\} \\
& \simeq E_{\tilde{u}*} \Bigg[\tilde{\sigma}^2 \sum_{i \in \tilde{U}_{\boldsymbol{x};\boldsymbol{r}}^*} w_i^* (1 - r_i^*) c_i^* \\
& + \tilde{\sigma}^2 \Bigg\{ \sum_{j \in \tilde{U}_{\boldsymbol{x};\boldsymbol{r}}^*} (1 - r_j^*) \boldsymbol{x}_j^{*\top} \Bigg\} \tilde{\boldsymbol{T}}_r^{*-1} \tilde{\boldsymbol{K}}_r^* \tilde{\boldsymbol{T}}_r^{*-1} \Bigg\{ \sum_{j \in \tilde{U}_{\boldsymbol{x};\boldsymbol{r}}^*} (1 - r_j^*) \boldsymbol{x}_j^* \Bigg\} \Bigg] \\
& \simeq \tilde{\sigma}^2 \sum_{i \in s} w_i^2 (1 - r_i) c_i + \tilde{\sigma}^2 \Bigg\{ \sum_{j \in s} w_j (1 - r_j) \boldsymbol{x}_j^\top \Bigg\} \hat{\boldsymbol{T}}_r^{-1} \hat{\boldsymbol{K}}_r \hat{\boldsymbol{T}}_r^{-1} \Bigg\{ \sum_{j \in s} w_j (1 - r_j) \boldsymbol{x}_j \Bigg\},
\end{aligned}$$

and

$$\begin{aligned}
& V_3^{IM*} = 2 E_{\tilde{u}*} \Bigg(\tilde{\sigma}^2 E_{p*} \Bigg[\Bigg\{ \sum_{j \in \tilde{s}^*} w_j^* (1 - r_j^*) \boldsymbol{x}_j^{*\top} \Bigg\} \hat{\boldsymbol{T}}_r^{*-1} \hat{\boldsymbol{L}}_r^* \\
& - \sum_{i \in \tilde{s}^*} w_i^* (w_i^* - 1)(1 - r_i^*) c_i^* \Bigg] \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}} \Bigg) \\
& \simeq 2 \tilde{\sigma}^2 E_{\tilde{u}*} \Bigg[\Bigg[\Bigg\{ \sum_{j \in \tilde{U}_{\boldsymbol{x};\boldsymbol{r}}^*} (1 - r_j^*) \boldsymbol{x}_j^{*\top} \Bigg\} \tilde{\boldsymbol{T}}_r^{*-1} \tilde{\boldsymbol{L}}_r^* - \sum_{i \in \tilde{U}_{\boldsymbol{x};\boldsymbol{r}}^*} (w_i^* - 1)(1 - r_i^*) c_i^* \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}} \Bigg] \\
& \simeq 2 \tilde{\sigma}^2 \Bigg\{ \sum_{j \in s} w_j (1 - r_j) \boldsymbol{x}_j^\top \Bigg\} \hat{\boldsymbol{T}}_r^{-1} \hat{\boldsymbol{L}}_r - 2 \tilde{\sigma}^2 \sum_{i \in s} w_i (w_i - 1)(1 - r_i) c_i,
\end{aligned}$$

where

$$\hat{\boldsymbol{T}}_r^* = \sum_{i \in \tilde{s}^*} w_i^* r_i^* \left(\frac{1 - p_i^*}{p_i^*} \right) \frac{\boldsymbol{x}_i^* \boldsymbol{x}_i^{*\top}}{c_i^*}, \quad \tilde{\boldsymbol{T}}_r^* = \sum_{i \in \tilde{U}_{\boldsymbol{x};\boldsymbol{r}}^*} r_i^* \left(\frac{1 - p_i^*}{p_i^*} \right) \frac{\boldsymbol{x}_i^* \boldsymbol{x}_i^{*\top}}{c_i^*},$$

$$\hat{\boldsymbol{K}}_r^* = \sum_{i \in \tilde{s}^*} w_i^{*2} r_i^* \left(\frac{1 - p_i^*}{p_i^*} \right)^2 \frac{\boldsymbol{x}_i^* \boldsymbol{x}_i^{*\top}}{c_i^*}, \quad \tilde{\boldsymbol{K}}_r^* = \sum_{i \in \tilde{U}_{\boldsymbol{x};\boldsymbol{r}}^*} w_i^* r_i^* \left(\frac{1 - p_i^*}{p_i^*} \right)^2 \frac{\boldsymbol{x}_i^* \boldsymbol{x}_i^{*\top}}{c_i^*},$$

$$\hat{\boldsymbol{L}}_r^* = \sum_{i \in \tilde{s}^*} w_i^* (w_i^* - 1) r_i^* \left(\frac{1 - p_i^*}{p_i^*} \right) \boldsymbol{x}_i^* \quad \text{and} \quad \tilde{\boldsymbol{L}}_r^* = \sum_{i \in \tilde{U}_{\boldsymbol{x};\boldsymbol{r}}^*} (w_i^* - 1) r_i^* \left(\frac{1 - p_i^*}{p_i^*} \right) \boldsymbol{x}_i^*.$$

To show that V^{IM*} is approximately unbiased for V^{IM}, we need to check that $E_{mpq}(V^{IM*}) \simeq V^{IM}$. Noting that

$$\begin{aligned} E_m \left(\hat{\boldsymbol{\beta}}_r^\top s_{\boldsymbol{xx}} \hat{\boldsymbol{\beta}}_r \right) &= \boldsymbol{\beta}^\top s_{\boldsymbol{xx}} \boldsymbol{\beta} + \sigma^2 \sum_{i \in s} r_i w_i^2 \left(\frac{1 - p_i}{p_i} \right)^2 \frac{\boldsymbol{x}_i^\top \hat{\boldsymbol{T}}_r^{-1} s_{\boldsymbol{xx}} \hat{\boldsymbol{T}}_r^{-1} \boldsymbol{x}_i}{c_i} \\ &= \boldsymbol{\beta}^\top s_{\boldsymbol{xx}} \boldsymbol{\beta} + O\left(\frac{\sigma^2}{n_R} \right), \end{aligned} \tag{3.B.6}$$

and

$$\begin{aligned} E_m \left(\tilde{\sigma}^2 \right) &= \frac{1}{n_R} \sum_{i \in s_R} \left(\frac{e_i}{\sqrt{c_i}} - \frac{1}{n_R} \sum_{j \in s_R} \frac{e_j}{\sqrt{c_j}} \right)^2 \\ &= \sigma^2 - \frac{\sigma^2}{n_R} \left\{ 1 - \sum_{i \in s} w_i^2 \left(\frac{1 - p_i}{p_i} \right)^2 \frac{\boldsymbol{x}_i^\top \hat{\boldsymbol{T}}_r^{-1} ss_{r\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}} \hat{\boldsymbol{T}}_r^{-1} \boldsymbol{x}_i}{c_i} \right. \\ &\left. + 2 \sum_{i \in s_R} w_i \left(\frac{1 - p_i}{p_i} \right) \left(\frac{\boldsymbol{x}_i}{\sqrt{c_i}} - \frac{1}{n_R} \sum_{j \in s_R} \frac{\boldsymbol{x}_j}{\sqrt{c_j}} \right)^\top \hat{\boldsymbol{T}}_r^{-1} \frac{\boldsymbol{x}_i}{\sqrt{c_i}} \right\} \\ &= \sigma^2 + O\left(\frac{\sigma^2}{n_R} \right), \end{aligned} \tag{3.B.7}$$

the first component of the bootstrap variance estimator, V_{1srs}^{IM*}, is unbiased for (3.B.2) in the case of simple random sampling without replacement; i.e.

$$\begin{aligned} E_{mpq}(V_{1srs}^{IM*}) &\simeq N^2 \left(\frac{1 - f}{n} \right) E_p \left(\frac{\sigma^2}{N} \sum_{i \in s} w_i c_i + \boldsymbol{\beta}^\top s_{\boldsymbol{xx}} \boldsymbol{\beta} \right) \\ &= N^2 \left(\frac{1 - f}{n} \right) \left(\frac{\sigma^2}{N} \sum_{i \in U} c_i + \boldsymbol{\beta}^\top S_{\boldsymbol{xx}} \boldsymbol{\beta} \right), \end{aligned}$$

where

$$ss_{r\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}} = \sum_{i \in s_R} r_i \left(\frac{\boldsymbol{x}_i}{\sqrt{c_i}} - \frac{1}{n_R} \sum_{j \in s_R} \frac{\boldsymbol{x}_j}{\sqrt{c_j}} \right) \left(\frac{\boldsymbol{x}_i}{\sqrt{c_i}} - \frac{1}{n_R} \sum_{j \in s_R} \frac{\boldsymbol{x}_j}{\sqrt{c_j}} \right)^\top .$$

For Poisson sampling, the unbiasedness of V_{1Pois}^{IM*} follows from results similar to (3.B.6) and (3.B.7).

$$
\begin{aligned}
E_{mpq}(V_{1Pois}^{IM*}) &\simeq E_p \left\{ \sigma^2 \sum_{i \in s} w_i \frac{1-\pi_i}{\pi_i} c_i + \boldsymbol{\beta}^\top \left(\sum_{i \in s} w_i \frac{1-\pi_i}{\pi_i} \boldsymbol{x}_i \boldsymbol{x}_i^\top \right) \boldsymbol{\beta} \right\} \\
&= \sigma^2 \sum_{i \in U} \frac{1-\pi_i}{\pi_i} c_i + \boldsymbol{\beta}^\top \left(\sum_{i \in U} \frac{1-\pi_i}{\pi_i} \boldsymbol{x}_i \boldsymbol{x}_i^\top \right) \boldsymbol{\beta}.
\end{aligned}
$$

Using a first-order Taylor expansion and (3.B.7), we obtain

$$
\begin{aligned}
E_{mpq}(V_2^{IM*}) &\simeq E_{pq} \Bigg[\sigma^2 \sum_{i \in s} w_i^2 (1-r_i) c_i \\
&+ \sigma^2 \left\{ \sum_{j \in s} w_j (1-r_j) \boldsymbol{x}_j^\top \right\} \hat{\boldsymbol{T}}_r^{-1} \hat{\boldsymbol{K}}_r \hat{\boldsymbol{T}}_r^{-1} \left\{ \sum_{j \in s} w_j (1-r_j) \boldsymbol{x}_j \right\} \Bigg] \\
&= \sigma^2 \sum_{i \in U} w_i (1-p_i) c_i + \sigma^2 E_{pq} \left[\left\{ \sum_{j \in s} w_j (1-r_j) \boldsymbol{x}_j^\top \right\} \hat{\boldsymbol{T}}_r^{-1} \hat{\boldsymbol{K}}_r \hat{\boldsymbol{T}}_r^{-1} \left\{ \sum_{j \in s} w_j (1-r_j) \boldsymbol{x}_j \right\} \right],
\end{aligned}
$$

and

$$
E_{mpq}(V_3^{IM*}) \simeq 2\sigma^2 E_{pq} \left[\left\{ \sum_{j \in s} w_j (1-r_j) \boldsymbol{x}_j^\top \right\} \hat{\boldsymbol{T}}_r^{-1} \hat{\boldsymbol{L}}_r \right] - 2\sigma^2 \sum_{i \in U} (w_i - 1)(1-p_i) c_i.
$$

It follows that V^{IM*} is approximately unbiased for V^{IM}.

Doubly robust bootstrap variance estimator

In this section, we show that the suggested bootstrap estimator $E^{IM*}\left(V^{NRM*}\right)$ is asymptotically unbiased for $E_{mpq}\left(V^{NRM*}\right)$. Using the results obtained for the

NRM Scheme, we first derive an expression of $E_{mpq}\left(V^{NRM*}\right)$:

$$
\begin{aligned}
& E_{mpq}\left(V^{NRM*}\right) \simeq E_{pq}\Bigg\{N^2\left(\frac{1-f}{n}\right)\Bigg[\frac{\sigma^2\sum_{i\in s} r_i c_i/p_i}{\sum_{i\in s} r_i/p_i} \\
& + \boldsymbol{\beta}^\top \left\{\frac{\sum_{i\in s} r_i \boldsymbol{x}_i \boldsymbol{x}_i^\top/p_i}{\sum_{i\in s} r_i/p_i} - \frac{\left(\sum_{i\in s} r_i \boldsymbol{x}_i/p_i\right)\left(\sum_{i\in s} r_i \boldsymbol{x}_i^\top/p_i\right)}{\left(\sum_{i\in s} r_i/p_i\right)^2}\right\} \boldsymbol{\beta} - \frac{\sigma^2 \sum_{i\in s} r_i c_i/p_i^2}{\left(\sum_{i\in s} r_i/p_i\right)^2}\Bigg] \\
& + \frac{N\sigma^2}{\sum_{i\in s} r_i/p_i}\Bigg[\sum_{i\in s} w_i \frac{r_i}{p_i}\left(\frac{1-p_i}{p_i}\right) c_i \\
& + \sum_{j\in s} w_j^2 r_j^2 \left(\frac{1-p_j}{p_j}\right)^2 \boldsymbol{x}_j^\top \hat{\boldsymbol{T}}_r^{-1}\left\{\sum_{i\in s} w_i \frac{r_i}{p_i}\left(\frac{1-p_i}{p_i}\right)\boldsymbol{x}_i \boldsymbol{x}_i^\top\right\}\hat{\boldsymbol{T}}_r^{-1}\frac{\boldsymbol{x}_j}{c_j} \\
& - 2\sum_{i\in s} w_i^2 \frac{r_i}{p_i}\left(\frac{1-p_i}{p_i}\right)^2 \boldsymbol{x}_i^\top \hat{\boldsymbol{T}}_r^{-1} \boldsymbol{x}_i\Bigg]\Bigg\},
\end{aligned}
$$

for simple random sampling without replacement and

$$
\begin{aligned}
& E_{mpq}\left(V^{NRM*}\right) \simeq \sigma^2 \sum_{i\in U} \frac{1-\pi_i}{\pi_i} c_i + \boldsymbol{\beta}^\top \left(\sum_{i\in U} \frac{1-\pi_i}{\pi_i} \boldsymbol{x}_i \boldsymbol{x}_i^\top\right)\boldsymbol{\beta} \\
& + \sigma^2 \sum_{i\in U} w_i \left(\frac{1-p_i}{p_i}\right) c_i \\
& + \sigma^2 E_{pq}\Bigg[\sum_{j\in s} w_j^2 r_j^2 \left(\frac{1-p_j}{p_j}\right)^2 \boldsymbol{x}_j^\top \hat{\boldsymbol{T}}_r^{-1}\left\{\sum_{i\in s} w_i^2 \frac{r_i}{p_i}\left(\frac{1-p_i}{p_i}\right)\boldsymbol{x}_i \boldsymbol{x}_i^\top\right\}\hat{\boldsymbol{T}}_r^{-1}\frac{\boldsymbol{x}_j}{c_j} \\
& -2\sum_{i\in s} w_i^3 \frac{r_i}{p_i}\left(\frac{1-p_i}{p_i}\right)^2 \boldsymbol{x}_i^\top \hat{\boldsymbol{T}}_r^{-1} \boldsymbol{x}_i\Bigg],
\end{aligned}
$$

for Poisson sampling.

Now, for simple random sampling without replacement, applying a first-order Taylor expansion, we obtain

$$
\begin{aligned}
&E_{m*p*}\left(V^{NRM*} \mid s_R, s_{\boldsymbol{x};\boldsymbol{r}}, \tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}\right) \simeq N^2\left(\frac{1-f}{n}\right)\Bigg[\frac{\tilde{\sigma}^2\sum_{i\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} r_i^* c_i^*/p_i^*}{\sum_{i\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} r_i^*/p_i^*}\\
&+\hat{\boldsymbol{\beta}}_r^\top\left\{\frac{\sum_{i\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} r_i^*\boldsymbol{x}_i^*\boldsymbol{x}_i^{*\top}/p_i^*}{\sum_{i\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} r_i^*/p_i^*}-\frac{\sum_{i\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} r_i^*\boldsymbol{x}_i^*/p_i^*\sum_{j\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} r_j^*\boldsymbol{x}_j^{*\top}/p_j^*}{\left(\sum_{i\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} r_i^*/p_i^*\right)^2}\right\}\hat{\boldsymbol{\beta}}_r\\
&-\frac{N\tilde{\sigma}^2\sum_{i\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} r_i^* c_i^*/p_i^{*2}}{n\left(\sum_{i\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} r_i^*/p_i^*\right)^2}\Bigg]+\frac{N^2\tilde{\sigma}^2}{n\sum_{i\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} r_i^*/p_i^*}\Bigg[\sum_{i\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}}\frac{r_i^*}{p_i^*}\left(\frac{1-p_i^*}{p_i^*}\right)c_i^*\\
&+\sum_{j\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} w_j^* r_j^{*2}\left(\frac{1-p_j^*}{p_j^*}\right)^2\boldsymbol{x}_j^{*\top}\tilde{\boldsymbol{T}}_r^{*-1}\left\{\sum_{i\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}}\frac{r_i^*}{p_i^*}\left(\frac{1-p_i^*}{p_i^*}\right)\boldsymbol{x}_i^*\boldsymbol{x}_i^{*\top}\right\}\tilde{\boldsymbol{T}}_r^{*-1}\frac{\boldsymbol{x}_j^*}{c_j^*}\\
&-2\sum_{i\in\tilde{U}^*_{\boldsymbol{x};\boldsymbol{r}}} w_i^*\frac{r_i^*}{p_i^*}\left(\frac{1-p_i^*}{p_i^*}\right)^2\boldsymbol{x}_i^{*\top}\tilde{\boldsymbol{T}}_r^{*-1}\boldsymbol{x}_i^*\Bigg].
\end{aligned}
$$

Finally, applying again a first-order Taylor expansion, the proposed bootstrap estimator is

$$
\begin{aligned}
&E_{srs}^{IM*}\left(V^{NRM*}\right) \simeq N^2\left(\frac{1-f}{n}\right)\Bigg[\frac{\tilde{\sigma}^2\sum_{i\in s} w_i r_i c_i/p_i}{\sum_{i\in s} w_i r_i/p_i}\\
&+\hat{\boldsymbol{\beta}}_r^\top\left\{\frac{\sum_{i\in s} w_i r_i\boldsymbol{x}_i\boldsymbol{x}_i^\top/p_i}{\sum_{i\in s} w_i r_i/p_i}-\frac{\left(\sum_{i\in s} w_i r_i\boldsymbol{x}_i/p_i\right)\left(\sum_{i\in s} w_i r_i\boldsymbol{x}_i^\top/p_i\right)}{\left(\sum_{i\in s} w_i r_i/p_i\right)^2}\right\}\hat{\boldsymbol{\beta}}_r\\
&-\frac{N\tilde{\sigma}^2\sum_{i\in s} w_i r_i c_i/p_i^2}{n\left(\sum_{i\in s} w_i r_i/p_i\right)^2}\Bigg]+\frac{N^2\tilde{\sigma}^2}{n\sum_{i\in s} w_i r_i/p_i}\Bigg[\sum_{i\in s} w_i\frac{r_i}{p_i}\left(\frac{1-p_i}{p_i}\right)c_i\\
&+\sum_{j\in s} w_j^2 r_j^2\left(\frac{1-p_j}{p_j}\right)^2\boldsymbol{x}_j^\top\hat{\boldsymbol{T}}_r^{-1}\left\{\sum_{i\in s} w_i\frac{r_i}{p_i}\left(\frac{1-p_i}{p_i}\right)\boldsymbol{x}_i\boldsymbol{x}_i^\top\right\}\hat{\boldsymbol{T}}_r^{-1}\frac{\boldsymbol{x}_j}{c_j}\\
&-2\sum_{i\in s} w_i^2\frac{r_i}{p_i}\left(\frac{1-p_i}{p_i}\right)^2\boldsymbol{x}_i^\top\hat{\boldsymbol{T}}_r^{-1}\boldsymbol{x}_i\Bigg].
\end{aligned}
$$

For Poisson sampling, we use similar arguments and obtain

$$\begin{aligned}
E^{IM^*}_{Pois}\left(V^{NRM*}\right) &\simeq \sum_{i\in s} w_i \frac{r_i}{p_i}\frac{1-\pi_i}{\pi_i} c_i \tilde{\sigma}^2 + \hat{\boldsymbol{\beta}}_r^\top \left(\sum_{i\in s} w_i \frac{r_i}{p_i}\frac{1-\pi_i}{\pi_i}\boldsymbol{x}_i\boldsymbol{x}_i^\top\right)\hat{\boldsymbol{\beta}}_r \\
&+ \tilde{\sigma}^2 \sum_{i\in s} w_i^2 \frac{r_i}{p_i}\left(\frac{1-p_i}{p_i}\right) c_i \\
&+ \tilde{\sigma}^2 \sum_{j\in s} w_j^2 r_j \left(\frac{1-p_j}{p_j}\right)^2 \boldsymbol{x}_j^\top \hat{\boldsymbol{T}}_r^{-1}\left\{\sum_{i\in s} w_i^2 \frac{r_i}{p_i}\left(\frac{1-p_i}{p_i}\right)\boldsymbol{x}_i\boldsymbol{x}_i^\top\right\}\hat{\boldsymbol{T}}_r^{-1}\frac{\boldsymbol{x}_j}{c_j} \\
&- 2\tilde{\sigma}^2 \sum_{i\in s} w_i^3 \frac{r_i}{p_i}\left(\frac{1-p_i}{p_i}\right)^2 \boldsymbol{x}_i^\top \hat{\boldsymbol{T}}_r^{-1}\boldsymbol{x}_i.
\end{aligned}$$

To prove the unbiasedness of the bootstrap estimator for $E_{mpq}\left\{E_{su}(V^{NRM*})\right\}$, we have to show that

$$E_{mpq}\left\{E^{IM^*}\left(V^{NRM*}\right)\right\} \simeq E_{mpq}\left(V^{NRM*}\right),$$

which can be done for both sampling designs using (3.B.7) and arguments similar to those that were used to obtain (3.B.6).

APPENDIX C

We assume that the parameter of interest θ can be written as a smooth function of totals, $\theta = g(t_1, \ldots, t_J)$ with $t_j = \sum_{i\in U} y_{ji}$ for $j = 1, \ldots, J$. Let $\hat{\theta} = g(\hat{t}_{1HT}, \ldots, \hat{t}_{JHT})$ be an estimator of θ, where $\hat{t}_{jHT}$ is the Horvitz-Thompson estimator of t_j, for $j = 1, \ldots, J$. The method of Demnati and Rao (2004) can be applied to linearize the non-linear statistic $\hat{\theta} = g(\hat{t}_{1HT}, \ldots, \hat{t}_{JHT})$. The basic idea behind the Demnati-Rao approach is to express $\hat{\theta} = g(\hat{t}_{1HT}, \ldots, \hat{t}_{JHT})$ as a function of the design weights $w_i(s) = w_i I_i(s)$, where $I_i(s)$ is the sample selection indicator for the i-th unit in U, instead of the customary approach that consists of regarding $\hat{\theta}$ as a function of the estimated totals, $\hat{t}_{1HT}, \ldots, \hat{t}_{JHT}$. Under this method, we have

$$\hat{\theta} - \theta \approx \sum_{i\in U} \left.\frac{\partial g(\hat{t}_{1HT}, \ldots, \hat{t}_{JHT})}{\partial w_i(s)}\right|_{\boldsymbol{w}(s)=\mathbf{1}} (w_i(s) - 1),$$

where $\boldsymbol{w}(s) = (w_1(s), \ldots, w_N(s))$.

In Chapter 2, to study the variance estimators under different imputation methods, the reverse framework was applied. To approximate the second term of the variance in (2.3.1), $V_q E_p(\hat{\theta}^I|\mathbf{r})$, the following theorem, which is an extension of the Demnati-Rao method, was applied. This theorem introduces an approximation for the estimator which is a function of totals on the population of respondents. This leads to an asymptotically unbiased approximation for the parameter under study under the non–response mechanism.

Theorem C.1 We suppose that $\theta_r = g(t_{1r}, \cdots, t_{Jr}) = g(\boldsymbol{t}_r)$, where $t_{jr} = \sum_{i\in U} r_i y_{ij}$, is the estimator of θ based on the population of respondents. Let

$\theta_p = g(t_{1p}, \cdots, t_{Jp}) = g(\boldsymbol{t}_p)$ where $t_{jp} = \sum_{i \in U} p_i y_{ij}$ and p_i is the probability of response for i-th unit. A first-order approximation of θ_r is

$$\theta_r - \theta_p \approx \sum_{i \in U} \left. \frac{\partial g(\boldsymbol{t}_r)}{\partial r_i} \right|_{\boldsymbol{r}=\boldsymbol{p}} (r_i - p_i). \tag{C.1}$$

Proof. Let $\boldsymbol{r} = (r_1, \cdots, r_N)$ and $\boldsymbol{p} = (p_1, \cdots, p_N)$. Using the chain rule, we obtain

$$\frac{\partial g(\boldsymbol{t}_r)}{\partial r_i} = \left(\frac{\partial g(\boldsymbol{t}_r)}{\partial \boldsymbol{t}_r} \right)^{\top} \cdot \frac{\partial \boldsymbol{t}_r}{\partial r_i},$$

where

$$\frac{\partial g(\boldsymbol{t}_r)}{\partial \boldsymbol{t}_r} = \left(\frac{\partial g(\boldsymbol{t}_r)}{\partial t_{1r}}, \cdots, \frac{\partial g(\boldsymbol{t}_r)}{\partial t_{Jr}} \right)^{\top} \quad \text{and} \quad \frac{\partial \boldsymbol{t}_r}{\partial r_i} = \left(\frac{\partial \boldsymbol{t}_{1r}}{\partial r_i}, \cdots, \frac{\partial \boldsymbol{t}_{Jr}}{\partial r_i} \right)^{\top} = (y_{1i}, \cdots, y_{Ji})'.$$

Using the above statements and the fact that $\boldsymbol{r} = \boldsymbol{p}$ if and only if $\boldsymbol{t}_r = \boldsymbol{t}_p$, we have

$$\begin{aligned}
\sum_{i \in U} \left. \frac{\partial g(\boldsymbol{t}_r)}{\partial r_i} \right|_{\boldsymbol{r}=\boldsymbol{p}} (r_i - p_i) &= \sum_{i \in U} \left(\frac{\partial g(\boldsymbol{t}_r)}{\partial \boldsymbol{t}_r} \right)^{\top} \cdot \left. \frac{\partial \boldsymbol{t}_r}{\partial r_i} \right|_{\boldsymbol{r}=\boldsymbol{p}} (r_i - p_i) \\
&= \sum_{i \in U} \left. \left(\frac{\partial g(\boldsymbol{t}_r)}{\partial \boldsymbol{t}_r} \right)^{\top} \right|_{\boldsymbol{r}=\boldsymbol{p}} (y_{1i}, \cdots, y_{Ji})' \, (r_i - p_i) \\
&= \left. \left(\frac{\partial g(\boldsymbol{t}_r)}{\partial \boldsymbol{t}_r} \right)^{\top} \right|_{\boldsymbol{r}=\boldsymbol{p}} (\boldsymbol{t}_r - \boldsymbol{t}_p) \\
&= \sum_{j=1}^{J} \left. \frac{\partial g(\boldsymbol{t}_r)}{\partial t_{jr}} \right|_{\boldsymbol{t}_r=\boldsymbol{t}_p} (t_{jr} - t_{jp}).
\end{aligned}$$

The proof is completed by using the above equality and a first–order Taylor expansion

$$g(t_{1r}, \cdots, t_{Jr}) = g(t_{1p}, \cdots, t_{Jp}) + \sum_{j=1}^{J} \left. \frac{\partial g(\boldsymbol{t}_r)}{\partial t_{jr}} \right|_{\boldsymbol{t}_r=\boldsymbol{t}_p} (t_{jr} - t_{jp}) + R,$$

where R is the remainder term.

□

CONCLUSION

Estimating the variance of a parameter of interest while dealing with imputed survey data is an important subject in survey methodology. The resampling bootstrap procedures presented in this thesis address the problems with the existing bootstrap method in this context proposed by Shao and Sitter (1996). These problems have been extensively discussed within the thesis, here we briefly mention again the specific achievements of this thesis.

In Chapter 1, we have studied all existing bootstrap methods for complete as well as imputed survey data. We classified the bootstrap methods for complete survey data into three groups: the pseudo-population bootstrap, the direct bootstrap and the bootstrap weights methods. We unified and compared the methods in each category to better see the strengths and weaknesses of these methods. This contribution is very helpful for researchers who would like to use bootstrap methods for survey data as well as develop new ones.

In the context of imputed data, the existing bootstrap method of Shao and Sitter (1996) requires the response status of each item under study and leads to a valid variance estimation only when the sampling fraction is negligible.

In Chapter 2, we proposed bootstrap methods for imputed data from regression, ratio and hot deck imputation. We assumed that the data came from stratified simple random sampling without replacement with uniform non-response in each stratum. To perform these methods, only the response rate within each stratum is needed. The resulting bootstrap variance estimators are asymptotically unbiased under the non-response model approach even for a large sampling fraction.

To work with more complex sampling designs and non-response mechanisms, we introduced bootstrap procedures for imputed data under the pseudo-population bootstrap approach in Chapter 3. These methods are designed to estimate the variance under the non-response model and under the imputation model approaches leading to a valid estimator even in the case of a non-negligible sampling fraction.

In this thesis, we developed different ideas, but there are many more avenues that remain unexplored. Studying the complete data bootstrap methods in Chapter 1 brings out the fact that there is not a considerable difference between the pseudo-population bootstrap methods. Comparing these methods leads us to develop a pseudo-population method in which an appropriate random mechanism is applied to create a pseudo-population with the same size as the original finite population. Such a random mechanism should have the property that the mean of the selected sample to complete the pseudo-population is asymptotically unbiased for the sample mean. This property leads to a negligible variability induced by creating the pseudo-population in the bootstrap statistics. In addition, an extension to a bootstrap weights method is very helpful in practice.

The independent bootstrap methods proposed in Chapter 2 are based on the assumption of uniform non-response mechanism. Under these methods, the constants of the direct bootstrap methods are modified depending on the response rate, the estimator of the response probability in the case of uniform non-response, and the imputation method. Studying the asymptotic behavior of the independent bootstrap methods for the case of the population quantiles assuming the uniform non-response mechanism has not been done yet and is very worthwhile doing. In this case, since deterministic imputation does not preserve the distribution of the variable being imputed, a random imputation method, such as random hot deck imputation, is used.

To the best of our knowledge, in the case of unequal response probabilities, it is not obvious how the constants of the direct bootstrap methods have to be modified even in the case of the population total. A pseudo-population bootstrap approach seems to be more appropriate in these cases.

The proposed pseudo-population bootstrap methods in Chapter 3 are built assuming the doubly robust deterministic regression imputation for the case of the population total (or mean). In reality, developing such methods suggests the possibility of applying a pseudo-population bootstrap method for the case of a population quantile which is not possible under the Kim and Haziza (2014) method. In this case, we believe that doubly robust random regression imputation should be used to compensate item non-response. A series of simulations is ongoing to check the behavior of the pseudo-population methods while applying the doubly robust random regression imputation method for the case of the population median.

Moreover, the nature of the pseudo-population bootstrap methods suggests that these methods work well in the case of more complex sampling designs. Developing the theory behind this claim is not an easy task and can be a great subject for further research.

Bibliography

Antal, E. and Y. Tillé (2011a). A direct bootstrap method for complex sampling designs from a finite population. *Journal of the American Statistical Association 106*(494), 534–543.

Antal, E. and Y. Tillé (2011b). Simple random sampling with over-replacement. *Journal of Statistical Planning and Inference 141*(1), 597–601.

Antal, E. and Y. Tillé (2014). A new resampling method for sampling designs without replacement: the doubled half bootstrap. *To appear*.

Barbe, P. and P. Bertail (1995). *The weighted bootstrap, Lecture notes in statistics*, Volume 98. Springer-Verlag, New York.

Beaumont, J.-F. and A.-S. Charest (2012). Bootstrap variance estimation with survey data when estimating model parameters. *Computational Statistics and Data Analysis 56*(12), 4450–4461.

Beaumont, J.-F. and Z. Patak (2012). On the generalized bootstrap for sample surveys with special attention to Poisson sampling. *International Statistical Review 80*(1), 127–148.

Berger, Y. G. (1998). Rate of convergence for asymptotic variance of the Horvitz-Thompson estimator. *Journal of Statistical Planning and Inference 74*(1), 149–168.

Berger, Y. G. (2007). A jackknife variance estimator for unistage stratified samples with unequal probabilities. *Biometrika 94*(4), 953–964.

Berger, Y. G. and C. J. Skinner (2005). A jackknife variance estimator for unequal probability sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*(1), 79–89.

Bertail, P. and P. Combris (1997). Bootstrap généralisé d'un sondage. *Annales d'économie et de statistique 46*, 49–83.

Bickel, P. J. and D. A. Freedman (1983). Asymptotic normality and the bootstrap in stratified sampling. *Unpublished manuscript. Department of Statistics, University of California, Berkeley.*

Bickel, P. J. and D. A. Freedman (1984). Asymptotic normality and the bootstrap in stratified sampling. *The Annals of Statistics 12*(2), 470–482.

Binder, D. A. (2011). Estimating model parameters from a complex survey under a model-design randomization framework. *Pakistan Journal of Statistics 27*(4), 371–390.

Booth, J. G., R. W. Butler, and P. Hall (1994). Bootstrap methods for finite populations. *Journal of the American Statistical Association 89*(428), 1282–1289.

Campbell, C. and A. D. Little (1980). A different view of finite population estimation. In *Proceedings of the Section on Survey Research Methods*, pp. 319–324.

Cassel, C. M., C. E. Särndal, and J. H. Wretman (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika 63*(3), 615–620.

Chao, M. T. (1982). A general purpose unequal probability sampling plan. *Biometrika 69*(3), 653–656.

Chao, M. T. and S.-H. Lo (1985). A bootstrap method for finite population. *Sankhyā: The Indian Journal of Statistics, Series A 47*, 399–405.

Chao, M. T. and S.-H. Lo (1994). Maximum likelihood summary and the bootstrap method in structured finite populations. *Statistica Sinica 4*(2), 389–406.

Chauvet, G. (2007). *Méthodes de bootstrap en population finie.* Ph. D. thesis, Université de Rennes 2.

Chipperfield, J. and J. Preston (2007). Efficient bootstrap for business surveys. *Survey Methodology 33*(2), 167–172.

Davison, A. C. and S. Sardy (2007). Resampling variance estimation in surveys with missing data. *Journal of Official Statistics 23*(3), 371–386.

Demnati, A. and J. N. K. Rao (2004). Linearization variance estimators for survey data. *Survey Methodology 30*(1), 17–26.

Deville, J. C. (1999). Variance estimation for complex statistics and estimators: Linearization and residual techniques. *Survey Methodology 25*(2), 193–203.

Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika 46*(3-4), 477–480.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics 7*(1), 1–26.

Efron, B. and R. J. Tibshirani (1993). *An introduction to the bootstrap*. Chapman & Hall, New York.

Escobar, E. L. and Y. G. Berger (2013). A jackknife variance estimator for self-weighted two-stage samples. *Statistica Sinica 23*(2), 595–613.

Fay, R. E. (1991). A design-based perspective on missing data variance. In *Proceedings of the 1991 Annual Research Conference, US Bureau of the census*, pp. 429–440.

Funaoka, F., H. Saigo, R. R. Sitter, and T. Toida (2006). Bernoulli bootstrap for stratified multistage sampling. *Survey Methodology 32*(2), 151–156.

Gross, S. (1980). Median estimation in sample surveys. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 181–184.

Gupta, V. K. and A. K. Nigam (1987). Mixed orthogonal arrays for variance estimation with unequal numbers of primary selections per stratum. *Biometrika 74*(4), 735–742.

Gurney, M. and R. S. Jewett (1975). Constructing orthogonal replications for variance estimation. *Journal of the American Statistical Association 70*(352), 819–821.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics 35*(4), 1491–1523.

Haziza, D. (2009). *Imputation and inference in the presence of missing data*. Handbook of Statistics 29A, 1th Edition Sample Surveys: Design, Methods and Applications, Elsevier, pp. 215-246.

Haziza, D. and J. N. K. Rao (2006). A non-response model approach to inference under imputation for missing survey data. *Survey Methodology 32*(1), 53–64.

Holmberg, A. (1998). A bootstrap approach to probability proportional to size sampling. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 378–383.

Horvitz, D. G. and D. J. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association 47*(260), 663–685.

Isaki, C. T. and W. A. Fuller (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association 77*(377), 89–96.

Jones, H. L. (1974). Jackknife estimation of functions of stratum means. *Biometrika 61*(2), 343–348.

Kim, J. K. and D. Haziza (2014). Doubly robust inference with missing data in survey sampling. *Statistica Sinica 24*, 375–394.

Kim, J. K., A. Navarro, and W. A. Fuller (2006). Replication variance estimation for two-phase stratified sampling. *Journal of the American Statistical Association 101*(473), 312–320.

Kim, J. K. and J. N. K. Rao (2009). A unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika 96*(4), 917–932.

Kott, P. S. (1998). Using the delete-a-group jackknife variance estimator in practice. In *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 763–768.

Kott, P. S. (2001). The delete-a-group jackknife. *Journal of Official Statistics 17*(4), 521–526.

Kovacevic, M. S., R. Huang, and Y. You (2006). Bootstrapping for variance estimation in multi-level models fitted to survey data. In *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 3260–3269.

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

Krewski, D. and J. N. K. Rao (1981). Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods. *The Annals of Statistics 9*(5), 1010–1019.

Lahiri, P. (2003). On the impact of bootstrap in survey sampling and small-area estimation. *Statistical Science 18*(2), 199–210.

Lo, A. Y. (1991). Bayesian bootstrap clones and a biometry function. *Sankhyā: The Indian Journal of Statistics, Series A 53*(3), 320–333.

Mashreghi, Z., D. Haziza, and C. Léger (2014a). Pseudo-population bootstrap methods for imputed survey data. *In preparation.*

Mashreghi, Z., D. Haziza, and C. Léger (2014b). A survey of bootstrap methods in finite population sampling. *In preparation.*

Mashreghi, Z., C. Léger, and D. Haziza (2014). Bootstrap methods for imputed data from regression, ratio and hot-deck imputation. *The Canadian Journal of Statistics 42*(1), 142–167.

Mason, D. M. and M. A. Newton (1992). A rank statistics approach to the consistency of a general bootstrap. *The Annals of Statistics 20*(3), 1611–1624.

McCarthy, P. J. (1969). Pseudo-replication: Half samples. *Review of the International Statistical Institute 37*(3), 239–264.

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

Miller, R. G. (1974). The jackknife-a review. *Biometrika 61*(1), 1–15.

Preston, J. and J. Chipperfield (2002). Using a generalised estimation methodology for ABS business surveys. *Methodology Advisory Committee, ABS, Belconnen, Australia (available at www.abs.gov.au).*

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika 43*(3-4), 353–360.

Ranalli, M. G. and F. Mecatti (2012). Comparing recent approaches for bootstrapping sample survey data: A first step toward a unified approach. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 4088–4099.

Rao, J. N. K. (1965). On two simple schemes of unequal probability sampling without replacement. *Journal of the Indian Statistical Association 3*, 173–180.

Rao, J. N. K. (1996). On variance estimation with imputed survey data. *Journal of the American Statistical Association 91*(434), 499–506.

Rao, J. N. K., H. O. Hartley, and W. G. Cochran (1962). On a simple procedure of unequal probability sampling without replacement. *Journal of the Royal Statistical Society. Series B (Methodological) 24*(2), 482–491.

Rao, J. N. K. and J. Shao (1992). Jackknife variance estimation with survey data under hot deck imputation. *Biometrika 79*(4), 811–822.

Rao, J. N. K. and J. Shao (1996). On balanced half-sample variance estimation in stratified random sampling. *Journal of the American Statistical Association 91*(433), 343–348.

Rao, J. N. K. and J. Shao (1999). Modified balanced repeated replication for complex survey data. *Biometrika 86*(2), 403–415.

Rao, J. N. K. and C. F. J. Wu (1985). Inference from stratified samples: second-order analysis of three methods for nonlinear statistics. *Journal of the American Statistical Association 80*(391), 620–630.

Rao, J. N. K. and C. F. J. Wu (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association 83*(401), 231–241.

Rao, J. N. K., C. F. J. Wu, and K. Yue (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology 18*(2), 209–217.

Rosén, B. (1997). Asymptotic theory for order sampling. *Journal of Statistical Planning and Inference 62*(2), 135–158.

Rubin, D. B. (1976). Inference and missing data. *Biometrika 63*(3), 581–592.

Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics 1*(4), 381–397.

Saigo, H. (2010). Comparing four bootstrap methods for stratified three-stage sampling. *Journal of Official Statistics 26*(1), 193–207.

Saigo, H., J. Shao, and R. R. Sitter (2001). A repeated half-sample bootstrap and balanced repeated replications for randomly imputed data. *Survey Methodology 27*(2), 189–196.

Sampford, M. R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika 54*(3-4), 499–513.

Särndal, C.-E. (1992). Methods for estimating the precision of survey estimates when imputation has been used. *Survey Methodology 18*(2), 241–252.

Shao, J. and Y. Chen (1998). Bootstrapping sample quantiles based on complex survey data under hot deck imputation. *Statistica Sinica 8*(4), 1071–1085.

Shao, J. and R. R. Sitter (1996). Bootstrap for imputed survey data. *Journal of the American Statistical Association 91*(435), 1278–1288.

Shao, J. and P. Steel (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association 94*(445), 254–265.

Shao, J. and D. Tu (1995). *The Jackknife and Bootstrap.* Springer Series in Statistics, New York.

Shao, J. and C. F. J. Wu (1989). A general theory for jackknife variance estimation. *The Annals of Statistics 17*(3), 1176–1197.

Shao, J. and C. F. J. Wu (1992). Asymptotic properties of the balanced repeated replication method for sample quantiles. *The Annals of Statistics 20*(3), 1571–1593.

Sitter, R. R. (1992a). Comparing three bootstrap methods for survey data. *The Canadian Journal of Statistics 20*(2), 135–154.

Sitter, R. R. (1992b). A resampling procedure for complex survey data. *Journal of the American Statistical Association 87*(419), 755–765.

Sitter, R. R. (1993). Balanced repeated replications based on orthogonal multi-arrays. *Biometrika 80*(1), 211–221.

Tukey, J. W. (1958). Bias and confidence in not quite large samples. Abstract. *The Annals of Mathematical Statistics 29*, 614.

Wang, Z. and M. E. Thompson (2012). A resampling approach to estimate variance components of multilevel models. *The Canadian Journal of Statistics 40*(1), 150–171.

Wolter, K. M. (2007). *Introduction to Variance Estimation.* Springer Series in Statistics, New York.

Wu, C. F. J. (1991). Balanced repeated replications based on mixed orthogonal arrays. *Biometrika 78*(1), 181–188.