The Survey Statistician, 2025, Vol. 91, 14–25.

Ask the Experts

A review of bootstrap methods for survey sampling

Guillaume Chauvet[1]

[1]Univ Rennes, ENSAI, CNRS, IRMAR-UMR 6625, Rennes, France, chauvet@ensai.fr

Abstract

Estimators in surveys are accompanied by measures of imprecision, such as a confidence interval or a coefficient of variation. Two primary approaches for variance estimation are the analytical approach, which involves explicit estimators requiring detailed knowledge of survey strategies, and replication methods, which approximate variances through resampling techniques. While analytical methods demand expertise, bootstrap is user-friendly, relying only on recalculated weights to produce confidence intervals. This article focuses on reviewing bootstrap methods tailored for sample surveys.

Keywords: confidence interval, non-response, plug-in estimation, variance estimation.

1 Introduction

The estimators produced in surveys are accompanied by a measure of imprecision (coefficient of variation or confidence interval), which presupposes the ability to calculate variance estimators. This is often a difficult problem. The sampling design can be complex, and in particular include several sampling stages if the population of interest is widely dispersed over the study area. In addition, the final weights delivered with the survey include several statistical treatments (e.g., correction for unit non-response and calibration), and their impact on the variance must be taken into account. Finally, some variables may suffer from item non-response, which reduces the effective sample size and can be treated by random imputation methods, if the distribution of the imputed variable needs to be preserved. All the sampling and estimation steps of the survey strategy should be taken into account to provide variance estimators that reflect the total survey error.

Two approaches are used in surveys to produce variance estimators and confidence intervals: the analytical approach and replication methods. The analytical approach consists of using explicit variance estimators, incorporating all the sampling and estimation steps. Linearization can also be used for non-linear parameters (e.g. Särndal et al., 2003, Chapter 5). The main drawback is that it requires in-depth knowledge of the survey strategy (e.g., design variables, response and imputation models, calibration variables), which is often beyond the reach of survey data users. On the other hand, replication methods such as Jackknife (Quenouille, 1956), balanced repeated replication (McCarthy, 1969) or bootstrap (Efron, 1979) make it possible to replace explicit variance estimators with numerical approximations, based on recalculations on the sample. In the case of the bootstrap, the parameter of interest can be re-estimated using bootstrap weights, leading to a simulation-based variance estimator or confidence interval. The approach is particularly simple for users, since no information other than these bootstrap weights is required to produce measures of imprecision.

In this article, we review bootstrap methods proposed in sample surveys. Other literature reviews have been proposed by Deville (1987), Presnell and Booth (1994), Lahiri (2003), Shao (2003), Shao and Tu (1995, Chapter 6), Davison and Hinkley (1997, Section 3.7) Mashreghi et al. (2016) and Conti and Mecatti (2022). We present the bootstrap method on independent and identically distributed data in Section 2. We consider the case of finite population sampling in Section 3, and present the different bootstrap families for one-stage sampling designs. Multistage sampling is dealt with in Section 4. Taking into account practical aspects such as non-response and calibration is considered in Section 5. We conclude in Section 6.

2 Bootstrap on i. i. d. data

In this section, we are interested in a q-dimensional probability distribution, which we assimilate to its multivariate distribution function $F(\cdot)$. We wish to estimate a parameter $\theta(F) \equiv \theta$ of the distribution. A sample of n i.i.d. values $X_1, \dots, X_n$ is generated according to F. For all $t = (t_1, \dots, t_q) \in \mathbb{R}^q$, the empirical distribution function is defined by

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1\left(X_{i1} \leq t_1, \dots, X_{iq} \leq t_q\right),$$

with $1(\cdot)$ the indicator function. It is used to obtain the plug-in estimator $\theta(F_n) \equiv \hat{\theta}$. Under mild conditions (e.g. Shao and Tu, 1995), $\theta(F_n)$ is asymptotically unbiased and consistent for $\theta(F)$. For example, suppose that X follows a bivariate distribution F with expectation $\mu = (\mu_1, \mu_2)^\top$ and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2. \end{pmatrix}$$

Then $\mu \equiv \theta_1(F) = E_F(X)$ is estimated by

$$\theta_1(F_n) = E_{F_n}(X) = \frac{1}{n} \sum_{i=1}^{n} X_i \equiv \bar{X}_n,$$

and the correlation coefficient $\theta_2(F) = \text{Corr}_F(X_1, X_2) = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$ is estimated by

$$\theta_2(F_n) = \text{Corr}_{F_n}(X_1, X_2) = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_{1i} - \bar{X}_{1n})(X_{2i} - \bar{X}_{2n})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_{1i} - \bar{X}_{1n})^2 \times \frac{1}{n}\sum_{i=1}^{n}(X_{2i} - \bar{X}_{2n})^2}}.$$

Bootstrapping (Efron, 1979) consists of using an empirical estimate of the distribution of the error $\theta(F_n) - \theta(F)$. The sampling and estimation mechanisms are repeated within the sample itself, as presented in Table 1. The principle can be summarized as follows: under mild conditions, the conditional distribution of $\theta(F_n^*) - \theta(F_n)$ converges to the distribution of $\theta(F_n) - \theta(F)$; see for example Bickel and Freedman (1981) for means and quantiles, and Shao and Tu (1995) for smooth functions of means. Note that the use of a resample size $n' = n - 1$ allows the usual unbiased variance estimator to be reproduced asymptotically for the estimation of the expectation μ.

Table 1: Sampling and estimation steps and their bootstrap counterpart

$F \xrightarrow{\text{Sampling}} F_n \xrightarrow{\text{Estimation}} \theta(F_n)$	$F_n \xrightarrow{\text{Resampling}} F_n^* \xrightarrow{\text{Reestimation}} \theta(F_n^*)$
Basic procedure: • Parameter of interest: $\theta(F)$ • Sampling: $(X_1, \ldots, X_{n'}) \sim_{iid} F$. Compute $F_n(t) = \frac{1}{n}\sum_{i=1}^n 1(X_{i1} \leq t_1, \ldots, X_{iq} \leq t_q)$. • Estimation by $\theta(F_n)$.	Bootstrap procedure: • Parameter of interest: $\theta(F_n)$ • Resampling: $(X_1^*, \ldots, X_{n'}^*) \sim_{iid} F_n$ with $m = n' - 1$. Compute $F_n^*(\cdot) = \frac{1}{n'}\sum_{i=1}^{n'} 1(X_{i1}^* \leq t_1, \ldots, X_{iq}^* \leq t_q)$. • Estimation by $\theta(F_n^*)$.

The bootstrap resampling procedure is repeated a large number of times (say B) in the sample $(X_1, \ldots, X_n)$. Each resampling can be summarised by the variable of bootstrap weights

$$G_i = \frac{n}{n-1} \times m_i,$$

with m_i the number of times the observation X_i is selected in the resample, also called the multiplicity. The bootstrap version of the parameter can then be recalculated using these weights. Returning to the previous example, we obtain

$$\theta_1(F_n^*) = \frac{\sum_{i=1}^n G_i X_i}{\sum_{i=1}^n G_i} \equiv \bar{X}_n^*,$$
$$\theta_2(F_n^*) = \frac{\sum_{i=1}^n G_i(X_{1i} - \bar{X}_{1n}^*)(X_{2i} - \bar{X}_{2n}^*)}{\sqrt{\sum_{i=1}^n G_i(X_{1i} - \bar{X}_{1n}^*)^2 \times \sum_{i=1}^n G_i(X_{1i} - \bar{X}_{2n}^*)^2}}.$$

We denote $\hat{\theta}^{1*}, \ldots, \hat{\theta}^{B*}$ the B bootstrap versions of the parameter of interest.

We can obtain a variance estimator for $\theta(F_n)$ by calculating the dispersion of these bootstrap values, which leads to

$$v_{boot}^B\{\theta(F_n)\} \quad = \quad \frac{1}{B-1}\sum_{b=1}^B \left(\hat{\theta}^{b*} - \frac{1}{B}\sum_{c=1}^B \hat{\theta}^{c*}\right)^2.$$

There are several ways of producing a confidence interval. If $\theta(F_n)$ asymptotically follows a normal distribution, the normality-based confidence interval of level $1-\alpha$ is given by

$$IC_{1-\alpha}^{nor}\{\theta(F)\} \quad = \quad \left[\theta(F_n) - u_{1-\frac{\alpha}{2}}\sqrt{v_{boot}^B\{\theta(F_n)\}}, \theta(F_n) + u_{1-\frac{\alpha}{2}}\sqrt{v_{boot}^B\{\theta(F_n)\}}\right],$$

with $u_{1-\frac{\alpha}{2}}$ the quantile of order $1 - \frac{\alpha}{2}$ of the standard normal distribution. The percentile confidence interval (Efron, 1981) is obtained by estimating the distribution of $\theta(F_n)$ by that of $\theta(F_n^*)$. By sorting the bootstrap values in ascending order $\hat{\theta}^{(1)*} \leq \cdots \leq \hat{\theta}^{(B)*}$, we obtain the interval

$$IC_{1-\alpha}^{per}\{\theta(F)\} \quad = \quad \left[\hat{\theta}^{(L)*}, \hat{\theta}^{(U)*}\right],$$

with $L = \alpha B$ and $U = (1-\alpha)B$. The reverse percentile confidence interval (e.g. Davison and Hinkley, 1997) is obtained by estimating the distribution of $\theta(F_n) - \theta(F)$ by that of $\theta(F_n^*) - \theta(F_n)$, which is more in line with the bootstrap principle. We obtain the interval

$$IC_{1-\alpha}^{rev}\{\theta(F)\} \quad = \quad \left[2\theta(F_n) - \hat{\theta}^{(U)*}, 2\theta(F_n) - \hat{\theta}^{(L)*}\right].$$

Many other methods have been proposed for calculating bootstrap confidence intervals, like the t-bootstrap (Efron, 1982), the bootstrap bias-corrected percentile (e.g., Schenker, 1985) or the bootstrap accelerated bias-corrected percentile (e.g., Efron, 1987).

3 Bootstrap for sample surveys

We now study the situation where the observed data come from a survey. We are interested in a finite population $U = \{1, \ldots, N\}$, in which a random sample S is selected according to a sampling design $p(\cdot)$. We denote $E_p(\cdot)$ and $V_p(\cdot)$ the expectation and variance under the sampling design. We assume that the first-order inclusion probabilities $\pi_k \equiv Pr(k \in S)$ are strictly positive. We are interested in a q-vector of variables of interest y_k. The total $t_y = \sum_{k \in U} y_k$ can be estimated without bias by the Horvitz-Thompson (HT) estimator

$$\hat{t}_{y\pi} = \sum_{k \in S} d_k y_k \tag{1}$$

with $d_k = \pi_k^{-1}$ the sampling weights. Let us denote

$$v(\hat{t}_{y\pi}) = \sum_{k,l \in S} a_{kl}(d_k y_k)(d_l y_l) \tag{2}$$

a variance estimator for $\hat{t}_{y\pi}$. For example, if we denote $\pi_{kl} \equiv Pr(k, l \in S)$ the second-order inclusion probabilities, the choice $a_{kl} = 1 - \frac{\pi_k \pi_l}{\pi_{kl}}$ leads to the HT variance estimator $v_{HT}(\hat{t}_{y\pi})$, which is unbiased for $V_p(\hat{t}_{y\pi})$ if $\pi_{kl} > 0$ for all $k, l \in U$. Suppose now that we wish to estimate the parameter $\theta = f(t_y)$, with $f(\cdot)$ a known function. We can use the plug-in/substitution estimator

$$\hat{\theta} = f(\hat{t}_{y\pi}),$$

which can be shown to be consistent for θ if $f(\cdot)$ verifies some regularity assumptions, see for example Särndal et al. (2003, Section 5.5) and Tillé (2020, Section 15).

In the search for a bootstrap method adapted to survey data, the main difficulty lies in reproducing the sampling step in the bootstrap. On the other hand, reproducing the estimation steps does not generally pose any particular difficulty. In this section, we consider bootstrap methods for one-stage sampling designs.

3.1 Pseudo-population bootstrap

We first deal with the case where S is selected according to simple random sampling without replacement (SRSWOR) of size n. Gross (1980) proposed to use the sample S to recreate a pseudo-population in which resampling is applied. Assuming that N/n is integer, each unit $k \in S$ is duplicated N/n times to create the pseudo-population U^*. A resample S^* is then selected from U^* according to an SRSWOR of size n, allowing us to calculate

$$\hat{t}^*_{y\pi} = \frac{N}{n} \sum_{k \in S^*} y_k \quad \text{and} \quad \hat{\theta}^* = f(\hat{t}^*_{y\pi}). \tag{3}$$

The sampling and estimation steps are then repeated B times to obtain the bootstrap values $\hat{\theta}^{1*}, \ldots, \hat{\theta}^{B*}$. This method is known as the pseudo-population bootstrap or bootstrap without replacement (BWO). We note E^* and V^* the expectation and variance under the resampling mechanism, conditional on the sample S.

The BWO is intuitively close to the basic bootstrap principle on i.i.d. data. For a SRSWOR, it leads to

$$E^*(\hat{t}^*_{y\pi}) = \hat{t}_{y\pi} \quad \text{et} \quad V^*(\hat{t}^*_{y\pi}) = v_{HT}(\hat{t}_{y\pi}) \left\{1 + O(n^{-1})\right\}, \tag{4}$$

which means that the first two moments of the HT estimator are estimated approximately without bias. However, it is assumed that N/n is an integer, which is rarely the case in practice. Many adaptations have been proposed to cover the general case, see for example Bickel and Freedman (1984), McCarthy and Snowden (1985), Booth et al. (1994) or Presnell and Booth (1994).

The BWO can be extended to the case where S is selected according to a single-stage unequal probability sampling design with inclusion probabilities π_k. Assuming that the quantities $1/\pi_k$ take integer values, each unit k is duplicated $1/\pi_k$ times to create the pseudo-population U^* (Deville, 1987; Holmberg, 1998). Then, the resample S^* is selected from U^* according to the original sampling design. For sampling designs with large entropy, such as Poisson sampling or conditional Poisson sampling (Hájek, 1964), this method once again allows the first two moments of the HT estimator to be approximately unbiasedly estimated (Chauvet, 2007, Chapter 3). On the other hand, the method can be inconsistent for sampling designs with smaller entropy, such as systematic sampling. The method must also be adapted to the general case of non-integer $1/\pi_k$ quantities, see for example Chauvet (2007), Mashreghi et al. (2014), Barbiero et al. (2015) and Chen et al. (2019).

3.2 Weighted bootstrap

Another possibility is to use a weighted bootstrap (WB) method (Bertail and Combris, 1997). We look for random weights G_k for each unit $k \in S$, and the estimators are recalculated using these weights as

$$\hat{t}^*_{y\pi} = \sum_{k \in S} G_k d_k y_k \quad \text{and} \quad \hat{\theta}^* = f(\hat{t}^*_{y\pi}). \tag{5}$$

The sampling and estimation steps are then repeated B times to obtain the bootstrap values $\hat{\theta}^{1*}, \ldots, \hat{\theta}^{B*}$. These weights are generated in such a way as to respect the equalities

$$E^*(G_k) = 1, \tag{6}$$

$$\text{Cov}^*(G_k, G_l) = a_{kl}, \tag{7}$$

at least approximately. This means that the weights are generated so that the bootstrap version of the HT estimator is equal in expectation to the HT estimator, and has a variance equal to the variance estimator in equation (2). Most of the weighted bootstrap methods developed in surveys aim to verify these two equalities.

These methods can be divided into three families. The first consists of using the basic bootstrap method presented in Section 2, and then modifying the multiplicities obtained to match equations (6) and (7). This is known as the rescaled bootstrap (RSB) method Rao and Wu (1988); Rao et al. (1992). This method is commonly used in surveys, and we will consider it in more details in Section 4 when multistage sampling is treated. Chipperfield and Preston (2007) have also proposed a rescaled bootstrap method for SRSWOR, but where the basic resampling procedure is done without replacement. Fuller et al. (2017) have studied Poisson sampling. With-replacement bootstrap with appropriate selection probabilities is applied, followed by a rescaling of the empirical variance of the bootstrap replicates. They suggest an extension for stratified conditional Poisson sampling, which is evaluated through simulations.

The second family of methods consists of finding a resampling method in S leading to G_k draws for unit k, so that the bootstrap weights respect the equations (6) and (7). In addition to the pseudo-population method of Gross (1980), there is the mirror-match bootstrap (MMB) of Sitter (1992b) or the direct methods of Antal and Tillé (2011). The third family of methods consists of generating bootstrap G_k weights directly, according to a probability distribution with the appropriate moments. It has been studied in particular by Bertail and Combris (1997), and by Beaumont and Patak (2012) who have focused more particularly on Poisson sampling.

4 Multistage sampling

Now we consider the case where S is selected according to a multistage sampling design, which is very common in sample surveys. Let U_I be the population of primary sampling units (PSUs). We consider the case of a single stratum at the first-stage, but if the PSUs are stratified beforehand, the methods presented in this section are simply applied stratum by stratum. A sample S_I of PSUs is selected in U_I with inclusion probabilities $\pi_{Ii} > 0$, leading to sampling weights $d_{Ii} = \pi_{Ii}^{-1}$ for the PSUs. In each $u_i \in S_I$, a sub-sample S_i of secondary sampling units (SSUs) is selected (possibly, itself according to a multi-stage sampling design). We denote $\pi_{k|i}$ the conditional inclusion probability of k in u_i, and $d_{k|i} = \pi_{k|i}^{-1}$ the associated weight. The sample S is equal to the union of the sub-samples S_i for $u_i \in S_I$, and the units of S are given the weights

$$d_k = d_{Ii} d_{k|i} \quad \text{for any} \quad k \in u_i. \tag{8}$$

The RSB (Rao and Wu, 1988; Rao et al., 1992) consists of selecting a resample S_I^* of size n_I^* from S_I, with replacement and with equal probabilities. We denote m_{Ii} the multiplicity of the primary unit u_i in S_I^*. Then, we take

$$G_k = \left[1 + \sqrt{\frac{n_I^*}{n_I - 1}} \left(\frac{n_I}{n_I^*} m_{Ii(k)} - 1 \right) \right], \tag{9}$$

with $u_{i(k)}$ the PSU containing the SSU k. This choice allows equation (6) to be respected, and the covariance term a_{kl} obtained in equation (7) corresponds to the unbiased variance estimator for a with-replacement sample S_I. It is important to note that S_I^* is selected with equal probabilities, even though S_I may be selected with unequal probabilities. This choice is crucial for obtaining the desired variance estimator. The most common choice for the resample size is $n_I^* = n_I - 1$, which leads to

$$G_k = \frac{n_I}{n_I - 1} m_{Ii(k)}. \tag{10}$$

In this case, the classic bootstrap method (see Section 2) is applied to the sample S_I, treated as if it had been drawn with replacement. This is known as the bootstrap of PSUs or with-replacement bootstrap (BWR) (McCarthy and Snowden, 1985). In the case of HT estimation, the variance estimator that we aim at matching under BWR is very close to the ultimate cluster variance estimator (Kalton, 1979) which is standard in statistical softwares. The main difference is that the BWR does not include a finite population correction at the first-stage. As a result, it generally leads to an overestimation of the variance, if the sampling design used at the first-stage is more efficient than with-replacement sampling. The positive bias of the bootstrap variance estimator is limited if the inclusion probabilities π_{Ii} used in the first stage are small, but can be appreciable otherwise. Nigam and Rao (1996) have also proposed a balanced modification of the RSB for stratified multistage sampling. Kolenikov (2007) and Aidara (2013) have suggested to replace the Monte Carlo method by quasi Monte Carlo, using the Halton sequence or the shuffled Halton sequence, with a faster rate of convergence in the latter case.

Other bootstrap methods have been proposed for multistage sampling. The generalisation of the BWO to multi-stage sampling is quite complex, see Sitter (1992a), Funaoka et al. (2006) and Chauvet (2007, Chapter 5). Chaudhuri and Saha (2004) have proposed an extension of the MMB for two-stage sampling, when the Rao-Hartley-Cochran method is employed at both stages. Funaoka et al. (2006) have proposed the Bernoulli bootstrap for stratified multistage sampling with large sampling fractions at each stage. The validity of the method for smooth and non-smooth estimators was evaluated through a simulation study. Preston (2009) has considered an extension of the RSB to stratified multistage sampling where units are selected by SRSWOR at each stage. The bootstrap weights include a finite population correction for the variance to be unbiasedly estimated. Beaumont and Émond (2022) have proposed a weighted bootstrap method for multistage sampling, suitable when the first-stage sampling inclusion probabilities are not small. Saigo (2010) and Chen et al. (2022) have compared by simulations the performances of bootstrap methods for multistage sampling. Wang et al. (2022) have proposed bootstrap methods for a variety of one-stage or multistage sampling designs, which enable to obtain second-order correct methods; i.e., which catch the two first terms in the Edgeworth expansion of the distribution function of the studentized HT estimator. Rubin and Chauvet (2024) have considered bootstrap for cross-classified sampling (CCS) designs (e.g. Ohlsson, 1996). They have demonstrated the suitability of weighted bootstrap techniques for CCS, given the availability of a weighted bootstrap technique in each dimension.

5 Practical aspects

In practice, the sample S is rarely fully observed, due to non-response. The first cause is unit non-response, i.e. failure to obtain a response to a survey as a whole, which leads to a sub-sample of respondents S_r. It is usually accounted for by reweighting, where the response probabilities are estimated, leading to inverse probability weights. Multiplying them by the sampling weights d_k leads to the weights adjusted for unit non-response. The goal is to reduce, as much as possible, the non-response bias (e.g. Haziza, 2009). On the other hand, these adjusted weights are finally calibrated on known totals, which is helpful to reduce the variance. Post-stratification and generalized regression estimation (GREG) are commonly used for calibration (Deville and Särndal, 1992).

Rust and Rao (1996) have underlined the need to account for non-response and calibration in the bootstrap, see also Canty and Davison (1999). Bessonneau et al. (2021) explain how these aspects are taken into account in the RSB, which may be summarized by two broad principles. Firstly, the sampling and non-response steps after the first stage are not bootstrapped. That is, the approach is conditional on the response and selection indicators following the first stage. Secondly, all the estimation steps (including correction of unit non-response and calibration) are reproduced during the bootstrap. Bessonneau et al. (2021) also proposed SAS programs to implement the RSB in such cases. The method proposed by Beaumont and Émond (2022) is also suitable for two-phase sampling with Poisson sampling at the second phase, which is commonly used to model non-response and is sometimes called the quasi-randomization approach. Related works include Schreuder et al. (1987) and Sitter (1997), who proposed a bootstrap procedure for two-phase sampling with simple random sampling at each phase. Saegusa (2015) has considered the BWO for variance estimation of the second-phase variance of a weighted likelihood estimator, under two-phase sampling. Handling calibration in the bootstrap process is relatively straightforward, as explained in Bessonneau et al. (2021). Wu and Rao (2010) have considered bootstrap procedures for constructing pseudo empirical likelihood ratio confidence intervals. Stefan and Hidiroglou (2023) have proposed a bootstrap procedure for the GREG estimator, where a working model is used to build the pseudo-population.

Another practical aspect in surveys is the item non-response, arising from the fact that a sampled unit may answer to some, but not all questions in a survey. Item non-response is usually handled by imputation motivated by a model, namely a supposed relationship between the imputed variable and the covariates observed for both respondents and non-respondents. Deterministic imputation methods include mean imputation within classes, deterministic regression imputation, and nearest neighbour imputation. Random imputation methods include random regression imputation, and hot-deck within classes, see for example Haziza (2009). Shao and Sitter (1996) have shown that imputation can be accounted for in the bootstrap, by imputing the bootstrap data sets in exactly the same way as the original data. They provided examples with the BWO, the MMB and the RSB; see also Saigo et al. (2001). Shao and Chen (1998) have considered stratified multistage sampling with item non-response, treated by hot-deck imputation. Assuming with-replacement of PSUs inside strata at the first stage, they prove the consistency of bootstrap estimators for sample quantiles, and of associated bootstrap variance estimators. Chen et al. (2019) have considered the BWO for imputation, and proved that it leads to a consistent variance estimator for Poisson sampling and large entropy sampling designs for smooth functions of totals under linear regression imputation, and for quantiles in the case of hot-deck imputation; see also Chen et al. (2021).

6 Conclusion

In this literature review, we have considered the application of bootstrap for survey sampling. We have not considered bootstrap methods for small area estimation, which is the subject of an extensive literature; see for example Lahiri (2003) and Rao and Molina (2015). Recently, the bootstrap has also been used in survey sampling to obtain robust estimators. Beaumont et al. (2023) consider using bootstrap weights to obtain an estimator of the conditional bias, when it is desired to measure the influence of sampled units; see also Chen et al. (2024). Moya et al. (2023) have considered RSB confidence intervals in case of outliers.

Bootstrap can also be used when survey data is used for modeling purposes. Beaumont and Charest (2012) consider the estimation of model parameters, and propose a modified weighted bootstrap that accounts for both model and sampling variability; see also Beaumont (2014), who considers tests of hypotheses. Kim et al. (2024) also consider hypotheses testing, and propose bootstrap methods to approximate the limiting distributions of quasi-likelihood ratio statistics, and quasi-score statistics. Hypothesis testing for categorical data is also considered.

The study of the properties of bootstrap confidence intervals in surveys has been relatively understudied in the literature, apart from the normality-based confidence interval, which only requires a consistent variance estimator. The bootstrap has also received little theoretical attention for quantile estimation. Shao and Chen (1998) have shown the convergence of quantile estimators under multistage sampling designs with random imputation, but under the assumption of with-replacement sampling of PSUs. Conti and Marella (2015) have considered an extension of the RSB for variance estimation on quantiles, and shown that it leads to a consistent variance estimator. However, the correction is applied to the quantiles instead of the weights, which makes it more difficult to use in practice. The bootstrap has also been little studied for continuous sampling designs, despite their interest for environmental surveys and forest inventories in particular. Further work on these aspects would be of both practical and theoretical interest.

References

Aidara, C. A. T. (2013). Bootstrap variance estimation for complex survey data: a quasi monte carlo approach. *Sankhya B*, 75(1):29–41.

Antal, E. and Tillé, Y. (2011). A direct bootstrap method for complex sampling designs from a finite population. *Journal of the American Statistical Association*, 106(494):534–543.

Barbiero, A., Manzi, G., and Mecatti, F. (2015). Bootstrapping probability-proportional-to-size samples via calibrated empirical population. *Journal of Statistical Computation and Simulation*, 85(3):608–620.

Beaumont, J.-F. (2014). The analysis of survey data using the bootstrap. *Contributions to Sampling Statistics*, pages 53–63.

Beaumont, J.-F., Bocci, C., and St-Louis, M. (2023). Bootstrap estimation of the conditional bias for measuring influence in complex surveys. *Journal of Survey Statistics and Methodology*, 11(2):393–411.

Beaumont, J.-F. and Charest, A.-S. (2012). Bootstrap variance estimation with survey data when estimating model parameters. *Computational Statistics and Data Analysis*, 56(12):4450–4461.

Beaumont, J.-F. and Émond, N. (2022). A bootstrap variance estimation method for multistage sampling and two-phase sampling when poisson sampling is used at the second phase. *Stats*, 5(2):339–357.

Beaumont, J.-F. and Patak, Z. (2012). On the generalized bootstrap for sample surveys with special attention to Poisson sampling. *International Statistical Review.*, 80(1):127–148.

Bertail, P. and Combris, P. (1997). Bootstrap généralisé d'un sondage. *Annales d'Économie et de Statistique*, (46):49–83.

Bessonneau, P., Brilhaut, G., Chauvet, G., and Garcia, C. (2021). With-replacement bootstrap variance estimation for household surveys principles, examples and implementation. *Survey Methodology*, 47(2):313–348.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9(6):1196–1217.

Bickel, P. J. and Freedman, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *The Annals of Statistics*, 12(2):470–482.

Booth, J. G., Butler, R. W., and Hall, P. (1994). Bootstrap methods for finite populations. *Journal of the American Statistical Association*, 89(428):1282–1289.

Canty, A. J. and Davison, A. C. (1999). Resampling-based variance estimation for labour force surveys. *Journal of the Royal Statistical Society Series D*, 48(3):379–391.

Chaudhuri, A. and Saha, A. (2004). Extending sitter's mirror-match bootstrap to cover rao-hartley-cochran sampling in two-stages with simulated illustrations. *Sankhya Series B*, pages 791–802.

Chauvet, G. (2007). *Méthodes de bootstrap en population finie*. PhD thesis, ENSAI.

Chen, S., Haziza, D., Léger, C., and Mashreghi, Z. (2019). Pseudo-population bootstrap methods for imputed survey data. *Biometrika*, 106(2):369–384.

Chen, S., Haziza, D., and Mashreghi, Z. (2021). Multiply robust bootstrap variance estimation in the presence of singly imputed survey data. *Journal of Survey Statistics and Methodology*, 9(4):810–832.

Chen, S., Haziza, D., and Mashreghi, Z. (2022). A comparison of existing bootstrap algorithms for multi-stage sampling designs. *Stats*, 5(2):521–537.

Chen, S., Haziza, D., and Michal, V. (2024). Efficient multiply robust imputation in the presence of influential units in surveys. *Canadian Journal of Statistics*, 52(3):829–854.

Chipperfield, J. and Preston, J. (2007). Efficient bootstrap for business surveys. *Survey Methodology*, 33(2):167–172.

Conti, P. L. and Marella, D. (2015). Inference for quantiles of a finite population: asymptotic versus resampling results. *Scandinavian Journal of Statistics*, 42(2):545–561.

Conti, P. L. and Mecatti, F. (2022). Resampling under complex sampling designs: Roots, development and the way forward. *Stats*, 5(1):258–269.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*. Number 1. Cambridge university press.

Deville, J. (1987). Réplications d'échantillons, demi-échantillons, jackknife, bootstrap dans les sondages. *Economica, Ed. Droesbeke, Tassi, Fichet*.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American statistical Association*, 87(418):376–382.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Efron, B. (1981). Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics*, 9(2):139–172.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*, volume 38 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397):171–200.

Fuller, W. A., Legg, J. C., and Li, Y. (2017). Bootstrap variance estimation for rejective sampling. *Journal of the American Statistical Association*, 112(520):1562–1570.

Funaoka, F., Saigo, H., Sitter, R. R., and Toida, T. (2006). Bernoulli bootstrap for stratified multistage sampling. *Survey Methodology*, 32(2):151.

Gross, S. (1980). Median estimation in sample surveys. In *Proceedings of the section on survey research methods*, volume 1814184. American Statistical Association Alexandria, VA.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics*, 35:1491–1523.

Haziza, D. (2009). Imputation and inference in the presence of missing data. In *Handbook of statistics*, volume 29, pages 215–246. Elsevier.

Holmberg, A. (1998). A bootstrap approach to probability proportional-to-size sampling. In *Proceedings of the Survey Research Methods Section of the ASA*, pages 378–383.

Kalton, G. (1979). Ultimate cluster sampling. *Journal of the Royal Statistical Society Series D*, 142(2):210–222.

Kim, J. K., Rao, J., and Wang, Z. (2024). Hypotheses testing from complex survey data using bootstrap weights: A unified approach. *Journal of the American Statistical Association*, 119(546):1229–1239.

Kolenikov, S. (2007). Applications of quasi-monte carlo methods in inference for complex survey data. *Proceedings of the Survey Reseacrh Methods Section of ASA*.

Lahiri, P. (2003). On the impact of bootstrap in survey sampling and small-area estimation. *Statistical Science*, 18(2):199–210.

Mashreghi, Z., Haziza, D., and Léger, C. (2016). A survey of bootstrap methods in finite population sampling. *Statistics Surveys*, 10:1–52.

Mashreghi, Z., Léger, C., and Haziza, D. (2014). Bootstrap methods for imputed data from regression, ratio and hot-deck imputation. *Canadian Journal of Statistics*, 42(1):142–167.

McCarthy, P. and Snowden, C. (1985). The bootstrap and finite population sampling. *Vital and Health statistics. Series 2, Data Evaluation and Methods Research*, (95):1–23.

McCarthy, P. J. (1969). Pseudo-replication: Half samples. *Bulletin of the International Statistical Institute*, pages 239–264.

Moya, P. J., Munoz, J. F., Verdejo, E. Á., and Blanco-Encomienda, F. J. (2023). Rescaled bootstrap confidence intervals for the population variance in the presence of outliers or spikes in the distribution of a variable of interest. *Communications in Statistics - Simulation and Computation*, 52(2):466–481.

Nigam, A. and Rao, J. (1996). On balanced bootstrap for stratified multistage samples. *Statistica Sinica*, pages 199–214.

Ohlsson, E. (1996). Cross-classified sampling. *Journal of Official Statistics*, 12(3):241.

Presnell, B. and Booth, J. G. (1994). Resampling methods for sample surveys. *Technical Report 470. Department of Statistics*.

Preston, J. (2009). Rescaled bootstrap for stratified multistage sampling. *Survey Methodology*, 35(2):227–234.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43(3/4):353–360.

Rao, J., Wu, C., and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18(2):209–217.

Rao, J. N. and Molina, I. (2015). *Small area estimation*. John Wiley & Sons.

Rao, J. N. K. and Wu, C.-F. J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83(401):231–241.

Rubin, J. and Chauvet, G. (2024). Asymptotic properties of cross-classified sampling designs. preprint.

Rust, K. F. and Rao, J. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5(3):283–310.

Saegusa, T. (2015). Variance estimation under two-phase sampling. *Scandinavian Journal of Statistics*, 42(4):1078–1091.

Saigo, H. (2010). Comparing four bootstrap methods for stratified three-stage sampling. *Journal of Official Statistics*, 26(1):193–207.

Saigo, H., Shao, J., and Sitter, R. R. (2001). A repeated half-sample bootstrap and balanced repeated replications for randomly imputed data. *Survey Methodology*, 27(2):189–196.

Särndal, C. E., Swensson, B., and Wretman, J. (2003). *Model assisted survey sampling*. Springer Verlag.

Schenker, N. (1985). Qualms about bootstrap confidence intervals. *Journal of the American Statistical Association*, 80(390):360–361.

Schreuder, H., Li, H., and Scott, C. (1987). Jackknife and bootstrap estimation for sampling with partial replacement. *Forest Science*, 33(3):676–689.

Shao, J. (2003). Impact of the bootstrap on sample surveys. *Statistical Science*, 18(2):191–198.

Shao, J. and Chen, Y. (1998). Bootstrapping sample quantiles based on complex survey data under hot deck imputation. *Statistica Sinica*, pages 1071–1085.

Shao, J. and Sitter, R. R. (1996). Bootstrap for imputed survey data. *Journal of the American Statistical Association*, 91(435):1278–1288.

Shao, J. and Tu, D. S. (1995). *The jackknife and bootstrap*. Springer Series in Statistics. Springer-Verlag, New York.

Sitter, R. (1997). Variance estimation for the regression estimator in two-phase sampling. *Journal of the American Statistical Association*, 92(438):780–787.

Sitter, R. R. (1992a). Comparing three bootstrap methods for survey data. *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.

Stefan, M. and Hidiroglou, M. A. (2023). A bootstrap variance procedure for the generalised regression estimator. *International Statistical Review*, 91(2):294–317.

Tillé, Y. (2020). *Sampling and estimation from finite populations*. John Wiley & Sons.

Wang, Z., Peng, L., and Kim, J. K. (2022). Bootstrap inference for the finite population mean under complex sampling designs. *Journal of the Royal Statistical Society Series B*, 84(4):1150–1174.

Wu, C. and Rao, J. (2010). Bootstrap procedures for the pseudo empirical likelihood method in sample surveys. *Statistics and Probability Letters*, 80(19-20):1472–1478.