Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

With-replacement bootstrap variance estimation for household surveys Principles, examples and implementation

by Pascal Bessonneau, Gwennaëlle Brilhaut, Guillaume Chauvet
and Cédric Garcia

Release date: January 6, 2022

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-514-283-9350

Depository Services Program

• Inquiries line	1-800-635-7943
• Fax line	1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2021
Vol. 47, No. 2, pp. 313-347
Statistics Canada, Catalogue No. 12-001-X

With-replacement bootstrap variance estimation for household surveys Principles, examples and implementation

Pascal Bessonneau, Gwennaëlle Brilhaut, Guillaume Chauvet and Cédric Garcia[1]

Abstract

Variance estimation is a challenging problem in surveys because there are several nontrivial factors contributing to the total survey error, including sampling and unit non-response. Initially devised to capture the variance of non-trivial statistics based on independent and identically distributed data, the bootstrap method has since been adapted in various ways to address survey-specific elements/factors. In this paper we look into one of those variants, the with-replacement bootstrap. We consider household surveys, with or without sub-sampling of individuals. We make explicit the benchmark variance estimators that the with-replacement bootstrap aims at reproducing. We explain how the bootstrap can be used to account for the impact sampling, treatment of non-response and calibration have on total survey error. For clarity, the proposed methods are illustrated on a running example. They are evaluated through a simulation study, and applied to a French Panel for Urban Policy. Two SAS macros to perform the bootstrap methods are also developed.

Key Words: Bootstrap; Calibration; Variance estimation; Unit non-response.

1. Introduction

Variance estimation is a challenging problem in surveys. The final weights used at the estimation stage include several statistical treatments, including correction of unit non-response and calibration, and their impact on the variance is to be assessed. Bootstrap is a useful tool, leading to the creation of so-called bootstrap weights released with the survey data set. These weights can be used to compute repeatedly the bootstrap version of the parameter of interest, leading to a simulation-based variance estimator or confidence interval. The interest for practitioners is that no information other than the bootstrap weights is needed for variance estimation. In particular, a comprehensive description of the original sampling design and estimation process is not required, which would be the case under an analytic approach where the variance estimator needs to be worked out. And thus the same set of bootstrap weights is to be used to obtain a variance estimate regardless of whether the parameters of interest are totals, medians or regression coefficients. Even when a comprehensive description of the sampling design and estimation process is available, the analytic approach poses issues for important parameters for which linearization variance estimation is not straightforward; see for example Shao (1994) for L-statistics, and Shao and Rao (1993) for low income proportions.

There is an extensive literature on bootstrap in survey sampling, see for example Rao and Wu (1988), Rao, Wu and Yue (1992), Shao and Tu (1995, Chapter 6), Davison and Hinkley (1997, Section 3.7), Davison and Sardy (2007), Chauvet (2007) and Mashreghi, Haziza and Léger (2016) for detailed reviews.

[1] Pascal Bessonneau, Ined; Gwennaëlle Brilhaut, Ined; Guillaume Chauvet, Ensai (Irmar), Campus de Ker Lann, Bruz - France. E-mail: guillaume.chauvet@ensai.fr; Cédric Garcia, Université Gustave Eiffel.

One of these techniques is the so-called rescaled bootstrap proposed by Rao and Wu (1988), which may be summarized as follows. First, inside each first-stage sample S_h of size n_h selected in stratum *h*, a with-replacement simple random sample of size m_h is selected, leading to the initial bootstrap weights. Then, these weights may be rescaled so as to reproduce an unbiased variance estimator for the estimation of a total (linear case). As explained by Rao and Wu (1988), the rescaled bootstrap may be applied to a variety of sampling designs including two-stage sampling and with/without-replacement sampling at the first stage. However, it is not straightforward to account for some practical features of a survey such as the treatment of unit non-response. This is considered in Yeo, Mantel and Liu (1999) and Girard (2009). A related topic is treated in Kim, Navarro and Fuller (2006), who consider replication variance estimation for two-phase sampling.

Applying the Rao-Wu bootstrap in the particular case when the resample sizes are $m_h = n_h - 1$ leads to the so-called bootstrap of Primary Sampling Units (PSUs) or with-replacement bootstrap (McCarthy and Snowden, 1985). The with-replacement bootstrap is fairly simple to implement; in particular, it requires to resample the primary sampling units only, and not the final units. Accounting for treatment of non-response and calibration is fairly natural, as explained in this paper. An important property of a bootstrap method is to match (at least, approximately) a known variance estimator in the linear case, which we call the benchmark variance estimator. For with-replacement bootstrap, it is possible to state precisely this benchmark variance estimator at any step of the method, which is helpful in understanding how the method works to assess the total survey error. The with-replacement bootstrap leads to conservative variance estimation, in the sense that the first-stage sampling variance is overestimated if the sampling designs used inside strata at first-stage are more efficient than multinomial sampling, which we assume to hold true in this paper. This is therefore a prudent approach in producing confidence intervals. The positive bias of the bootstrap variance estimator is expected to be negligible when the first-stage sampling rates inside strata are negligible, which is often the case in phone surveys. Also, if the survey is repeated over time, the contribution of the first-stage sampling variance is likely to fade while the variance due to attrition and unit non-response grows bigger.

Our paper, which examines the with-replacement bootstrap, is intended to be user-oriented. In particular, we do not propose particular modifications of the with-replacement bootstrap. Rather, we explain how this bootstrap method may be applied to account for sampling, treatment of non-response and calibration, and in so doing, what is the variance estimator that we aim at reproducing when estimating a total. We give some running examples to illustrate how bootstrap weights are computed in simple cases. Two SAS macros implementing the proposed bootstrap methods are presented, evaluated through a simulation study, and illustrated on a real survey dataset from the Panel for Urban Policy.

For simplicity of presentation, our terminology is that of household surveys, which is our original motivation for this paper. We consider two cases: first, when a sample of households only is selected; secondly, when a subsample of individuals is selected inside the selected households. Despite this specific terminology, our approach is general and may be applied to any other situation when a survey is performed by one-stage sampling (first case) or by two-stage sampling (second case).

We are in particular interested in household phone surveys, which have been extensively used at the French National Institute for Demographic Studies (INED) over the last decades. Originally, a sample of phone numbers was selected from a register of fixed-line numbers, and more recently the phone numbers used in the survey are randomly generated to account for households not covered in the registers (unlisted or cell numbers). In a second step, individuals are selected within the households, using classic selection methods (e.g., Kish individual). Phone surveys have proved to be efficient, specifically for sensitive subjects like sexuality, violence or addictions. Some examples of surveys performed by INED include the national survey on violence against women in France in 2000 (ENVEFF), the national survey on violence and gender exchange in 2015 and 2018 (VIRAGE and VIRAGE overseas, respectively), or the national survey on the context of sexuality in France in 2006. The same protocol is likely to be used in a near future for surveys on similar subjects, like the one on young adults' sexuality or the one on birth control, to begin between 2021 and 2023.

The paper is organized as follows. In Section 2, our main notations are defined, and we consider the estimation of a total by accounting for sampling, unit non-response and calibration. We treat in Section 2.1 the situation when a sample of households only is selected (one-stage case), and in Section 2.2 the case when individuals are sub-sampled within households (two-stage case). The basic bootstrap method is described in Section 3: the one-stage case is considered in Sections 3.1 and 3.2, and the two-stage case is considered in Sections 3.3 and 3.4. We explain in Section 3.5 how the basic bootstrap procedure may be applied to obtain an estimator of variance or a confidence interval. The proposed bootstrap methods are evaluated in Section 4 through a simulation study. We present in Section 5 an illustration on a sample of households and individuals from the French Panel for Urban Policy. We conclude in Section 6. The benchmark variance estimators for the sample of individuals are presented in Appendix A. The SAS program used to perform bootstrap variance estimation are presented in Appendices B and C. These SAS programs are available upon request to the corresponding author.

2. Notation and estimation

In this section, we define our main notations, and we describe the sampling and estimation process. We first consider in Section 2.1 the case when a sample of households only is selected, and we describe the estimation process which includes treatment of unit non-response and calibration. We indicate in each case what is the benchmark variance estimator considered, i.e. the variance estimator that we aim at reproducing for the estimation of a total with the bootstrap method proposed in Section 3. The case when individuals are sub-sampled inside households is covered in Section 2.2. The benchmark variance estimators for this second case are given in Appendix A.

2.1 Case of a sample of households only

We consider estimation for a population U_{hh} of households. We let y_k denote the value taken by some variable of interest for the household $k.$ We are interested in the estimation of the total

$$Y_{hh} = \sum_{k \in U_{hh}} y_k. \tag{2.1}$$

2.1.1 Sampling design

We suppose that a sample S_{hh} is selected in U_{hh} by means of a stratified one-stage sampling design. The population U_{hh} is partitioned into H strata $U_{hh}^1, \ldots, U_{hh}^H,$ the samples $S_{hh}^1, \ldots, S_{hh}^H$ are selected inside independently, and the sample S_{hh} is the union of these samples. We let π_k denote the inclusion probability of a given household $k.$ The design weight is

$$d_k = \frac{1}{\pi_k}. \tag{2.2}$$

In case of full response, the estimator of Y_{hh} is

$$\hat{Y}_{hh} = \sum_{k \in S_{hh}} d_k y_k. \tag{2.3}$$

We consider as a benchmark variance estimator

$$v_{\text{mult}}(\hat{Y}_{hh}) = \sum_{h=1}^{H} \left[\frac{n_h}{n_h - 1} \sum_{k \in S_{hh}^h} \left(d_k y_k - \frac{1}{n_h} \sum_{k' \in S_{hh}^h} d_{k'} y_{k'} \right)^2 \right], \tag{2.4}$$

with n_h the size of the sample $S_{hh}^h.$ This variance estimator is unbiased if the samples are selected inside strata by multinomial sampling (Tillé, 2011, Section 5.4), a.k.a. sampling with replacement. It is conservative if the sampling designs used inside strata are more efficient than multinomial sampling (Särndal, Swensson and Wretman, 1992, Section 4.6), which we assume to hold true in the rest of the paper. The positive bias of this variance estimator is expected to be negligible when the sampling rates inside strata are themselves negligible, which is often the case in phone surveys. This is illustrated by the results of our simulation study, see Section 4.

2.1.2 Treatment of non-response

In practice, the sample S_{hh} is prone to unit non-response, which leads to the observation of a sub-sample of respondents $S_{r,hh}$ only. We let r_k denote the response indicator of a household $k,$ and p_k denote the response probability of the household $k.$ We suppose that the households respond independently of one another. Also, we suppose that unit non-response is handled through the method of Response Homogeneity Groups (RHGs), which is popular in practice (e.g. Brick, 2013; Juillard and Chauvet, 2018). Under this framework, it is assumed that the sample S_{hh} may be partitioned into C RHGs denoted as $S_{1,hh}, \ldots, S_{C,hh}$ such that the response probability p_k is constant inside a RHG.

For $c = 1, \ldots, C,$ we let p_c denote the common response probability inside the RHG $S_{c,hh}.$ It is estimated by

$$\hat{p}_c = \frac{\sum_{k \in S_{c,hh}} \theta_k r_k}{\sum_{k \in S_{c,hh}} \theta_k}, \tag{2.5}$$

with θ_k some weight attached to the household k. The choice $\theta_k = 1$ leads to estimating p_c by the unweighted response rate inside the RHG. The choice $\theta_k = d_k$ leads to estimating p_c by the response rate inside the RHG, weighted by the sampling weights (e.g. Kott, 2012).

Accounting for the estimated response probabilities leads to the weights corrected for non-response

$$d_{rk} = \frac{d_k}{\hat{p}_{c(k)}}, \tag{2.6}$$

with $c(k)$ the RHG of the household k. The estimator of Y_{hh} adjusted for non-response is

$$\hat{Y}_{r,hh} = \sum_{k \in S_{r,hh}} d_{rk} y_k. \tag{2.7}$$

Building on the multinomial variance estimator in (2.4) and on linearization for estimators reweighted for unit-non-response (Kim and Kim, 2007, Section 2), our benchmark variance estimator is

$$v_{\text{mult}}(\hat{Y}_{r,hh}) = \sum_{h=1}^{H} \left[\frac{n_h}{n_h - 1} \sum_{k \in S_{hh}^h} \left(d_k u_{1k} - \frac{1}{n_h} \sum_{k' \in S_{hh}^h} d_{k'} u_{1k'} \right)^2 \right], \tag{2.8}$$

with

$$u_{1k} = \theta_k \pi_k \bar{y}_{rc(k)} + \frac{r_k}{\hat{p}_{c(k)}} \left\{ y_k - \theta_k \pi_k \bar{y}_{rc(k)} \right\},$$

and

$$\bar{y}_{rc} = \frac{\sum_{k \in S_{c,hh}} d_k r_k y_k}{\sum_{k \in S_{c,hh}} \theta_k r_k}.$$

This is a conservative estimator for the asymptotic variance of $\hat{Y}_{r,hh}$. A key assumption for this to hold is that the response indicators r_k are mutually independent.

2.1.3 Calibration

Lastly, the weights adjusted for non-response are calibrated on auxiliary totals known on the population. For simplicity, we describe only the Generalized REGression estimator (GREG, Särndal et al., 1992, Chapter 6). Let x_k denote the vector of calibration variables at the household level, and X_{hh} the total on the population U_{hh}. For the sample $S_{r,hh}$, this leads to the linear calibrated weights

$$w_k = d_{rk} \left(1 + x_k^\top \lambda_{hh} \right),$$

with

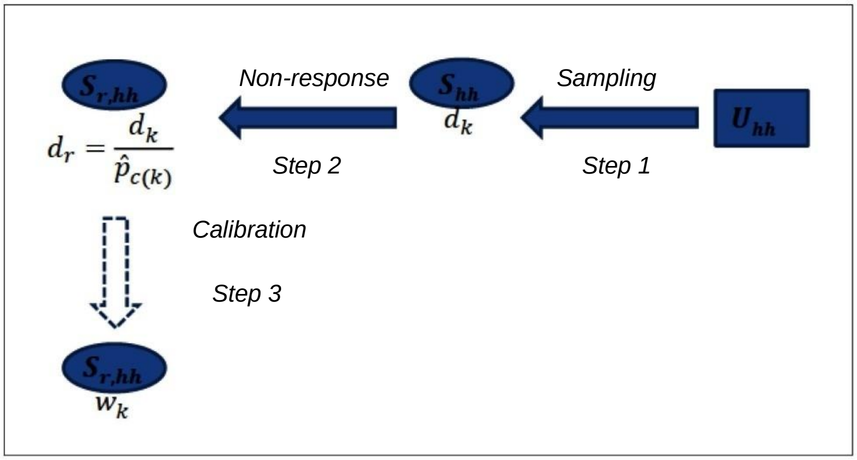
$$\lambda_{hh} = \left(\sum_{k \in S_{r,hh}} d_{rk} x_k x_k^{\mathrm{T}} \right)^{-1} \left(X_{hh} - \hat{X}_{r,hh} \right), \tag{2.9}$$

and where $\hat{X}_{r,hh}$ is the estimator of X_{hh}, obtained by plugging x_k into (2.7). The calibrated estimator is

$$\hat{Y}_{\text{cal},hh} = \sum_{k \in S_{r,hh}} w_k y_k. \tag{2.10}$$

The sampling and estimation steps are summarized in Figure 2.1.

Figure 2.1 Sampling and estimation steps for a household sample.



Using linearization for estimators reweighted for unit-non-response and calibrated (Kim and Kim, 2007, Section 5), our benchmark variance estimator is

$$v_{\text{mult}}(\hat{Y}_{\text{cal},hh}) = \sum_{h=1}^{H} \left[\frac{n_h}{n_h - 1} \sum_{k \in S_{hh}^h} \left(d_k u_{2k} - \frac{1}{n_h} \sum_{k' \in S_{hh}^h} d_{k'} u_{2k'} \right)^2 \right], \tag{2.11}$$

with

$$u_{2k} = \theta_k \pi_k \bar{e}_{rc(k)} + \frac{r_k}{\hat{p}_{c(k)}} \left\{ e_k - \theta_k \pi_k \bar{e}_{rc(k)} \right\},$$

and

$$\bar{e}_{rc} = \frac{\sum_{k \in S_{c,hh}} d_k r_k e_k}{\sum_{k \in S_{c,hh}} \theta_k r_k},$$

where we let

$$e_k = y_k - \hat{B}_{r,hh}^{\top} x_k \quad \text{with} \quad \hat{B}_{r,hh} = \left(\sum_{k \in S_{r,hh}} d_{rk} x_k x_k^{\top} \right)^{-1} \sum_{k \in S_{r,hh}} d_{rk} x_k y_k \tag{2.12}$$

denote the estimated regression residuals of the variable of interest on the calibration variables. This is a conservative estimator for the asymptotic variance of $\hat{Y}_{\text{cal},hh}$.

2.1.4 Computation of household weights on an example

To fix ideas, we describe a small example. We consider a population U_{hh} of $N_{hh} = 100$ households. We suppose without loss of generality that a single stratum is used, and that a sample of $n_{hh} = 10$ households is selected.

The sample is $S = \{A, B, \ldots, J\}$. The inclusion probabilities of the selected units are (say)

$$\pi_A = \pi_B = \pi_C = \pi_D = \frac{1}{4} \quad \text{and} \quad \pi_E = \pi_F = \pi_G = \pi_H = \pi_I = \pi_J = \frac{1}{16}, \tag{2.13}$$

resulting in the design weights

$$d_A = d_B = d_C = d_D = 4 \quad \text{and} \quad d_E = d_F = d_G = d_H = d_I = d_J = 16. \tag{2.14}$$

Among the 10 selected households, 7 only are surveyed due to non-response. It is accounted for by using the method of RHGs, with two groups: the units $A,$ $B,$ F and J in the first one, and the units $C,$ $D,$ $E,$ $G,$ $H,$ and I in the second one. The units $B,$ C and G are non-respondents. Inside each RHG, we compute estimated response probabilities, weighted by the design weights $(\theta_k = d_k)$. This leads to

$$\begin{aligned} \hat{p}_1 &= \frac{\sum_{k \in S_{1,hh}} d_k r_k}{\sum_{k \in S_{1,hh}} d_k} = \frac{d_A + d_F + d_J}{d_A + d_B + d_F + d_J} = \frac{9}{10}, \\ \hat{p}_2 &= \frac{d_D + d_E + d_H + d_I}{d_C + d_D + d_E + d_G + d_H + d_I} = \frac{13}{18}. \end{aligned} \tag{2.15}$$

The weights accounting for non-response are obtained for the respondents by dividing the sampling weights by the estimated response probabilities. This leads to the weights

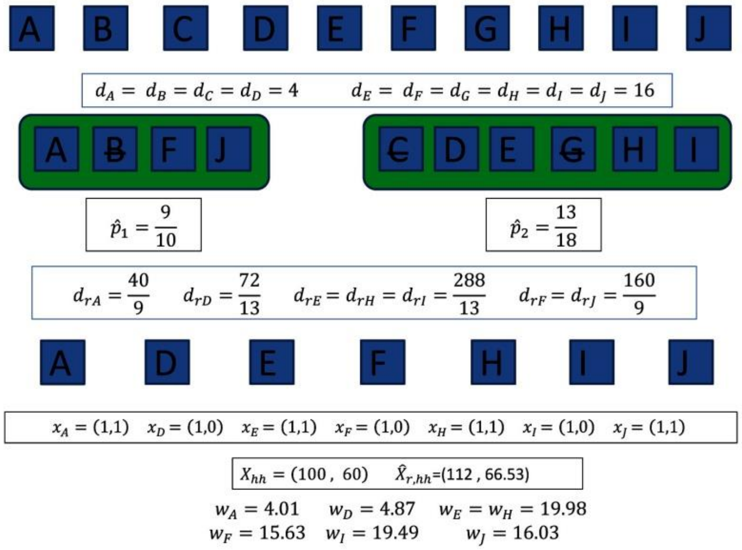
$$d_{rA} = \frac{40}{9} \quad d_{rD} = \frac{72}{13} \quad d_{rE} = d_{rH} = d_{rI} = \frac{288}{13} \quad d_{rF} = d_{rJ} = \frac{160}{9}. \tag{2.16}$$

Finally, the weights are calibrated to match exactly the population size $N_{hh} = 100$ and an auxiliary total $X_{1,hh} = 60.$ Note that, using the sample of respondents, we obtain $\hat{N}_{r,hh} = 112$ and $\hat{X}_{1r,hh} = 66.53.$ The calibrated weights are

$$\begin{aligned} &w_A = 4.01, \quad w_D = 4.87, \quad w_E = w_H = 19.98, \\ &w_F = 15.63, \quad w_I = 19.49, \quad w_J = 16.03. \end{aligned} \tag{2.17}$$

The sampling and estimation steps are summarized in Figure 2.2.

Figure 2.2 Estimation steps for the weighting of households.


A B C D E F G H I J
$d_A = d_B = d_C = d_D = 4$ $d_E = d_F = d_G = d_H = d_I = d_J = 16$
A ~~B~~ F J
~~C~~ D E ~~G~~ H I
$\hat{p}_1 = \frac{9}{10}$
$\hat{p}_2 = \frac{13}{18}$
$d_{rA} = \frac{40}{9}$ $d_{rD} = \frac{72}{13}$ $d_{rE} = d_{rH} = d_{rI} = \frac{288}{13}$ $d_{rF} = d_{rJ} = \frac{160}{9}$
A D E F H I J
$x_A = (1,1)$ $x_D = (1,0)$ $x_E = (1,1)$ $x_F = (1,0)$ $x_H = (1,1)$ $x_I = (1,0)$ $x_J = (1,1)$
$X_{hh} = (100, 60)$ $\hat{X}_{r,hh} = (112, 66.53)$
$w_A = 4.01$ $w_D = 4.87$ $w_E = w_H = 19.98$
$w_F = 15.63$ $w_I = 19.49$ $w_J = 16.03$

2.2 Case of a sample of households and individuals

We are interested in the population U_{ind} of individuals associated to the population U_{hh} of households considered in Section 2.1. If we let y_l denote the value taken by some variable of interest for the individual $l,$ the parameter of interest is

$$Y_{\text{ind}} = \sum_{l \in U_{\text{ind}}} y_l. \tag{2.18}$$

2.2.1 Sampling design

Within any sampled household $k \in S_{hh},$ a subsample $S_{\text{ind},k}$ of individuals is selected, and the sample S_{ind} is the union of these samples. We let $\pi_{l|k}$ denote the conditional inclusion probability of the individual l inside the household $k.$ The conditional design weight of l is

$$d_{l|k} = \frac{1}{\pi_{l|k}} \quad \text{for any} \quad l \in k, \tag{2.19}$$

and the non-conditional design weight is

$$d_l = d_{l|k} \times d_k \quad \text{for any} \quad l \in k. \tag{2.20}$$

In case of full response, the estimator of Y_{ind} is

$$\hat{Y}_{\text{ind}} = \sum_{k \in S_{hh}} d_k \sum_{l \in S_{\text{ind},k}} d_{l|k} y_l = \sum_{k \in S_{\text{ind}}} d_l y_l. \tag{2.21}$$

The benchmark variance estimator for $\hat{Y}_{\text{ind}}$ is obtained from (2.4), by replacing y_k with

$$\hat{y}_k = \sum_{l \in S_{\text{ind},k}} d_{l|k} y_l. \tag{2.22}$$

2.2.2 Treatment of non-response

The weights of individuals accounting for the non-response of households are

$$d_{rl} = d_{rk(l)} d_{l|k(l)} \quad \text{with } k(l) \text{ the household containing } l, \tag{2.23}$$

with d_{rk} the weight of household k corrected for unit non-response (see equation (2.6)), and $d_{l|k}$ the conditional sampling weight of individual l inside the household k (see equation (2.19)). We let

$$S_{r,\text{ind}} = \bigcup_{k \in S_{r,hh}} S_{\text{ind},k} \tag{2.24}$$

denote the set of all sampled individuals inside the responding households.

The individuals in $S_{r,\text{ind}}$ are themselves prone to non-response, though it is usually expected to be to a smaller extent. This leads to the observation of a sub-sample of respondents $S_{rr,\text{ind}}$ only. We let r_l denote the response indicator and p_l denote the response probability of the individual l. We suppose that the individuals respond independently of one another. Also, we suppose that this non-response is handled through the method of RHGs: the sample $S_{r,\text{ind}}$ may be partitioned into D RHGs denoted as $S_{r1,\text{ind}}, \ldots, S_{rD,\text{ind}}$ such that the response probability p_l is constant inside a RHG.

For $d = 1, \ldots, D,$ we let p_d denote the common response probability inside the RHG $S_{rd,\text{ind}}$. It is estimated by

$$\hat{p}_d = \frac{\sum_{l \in S_{rd,\text{ind}}} \theta_l r_l}{\sum_{l \in S_{rd,\text{ind}}} \theta_l}, \tag{2.25}$$

with θ_l some weight attached to the individual l. The choice $\theta_l = 1$ leads to estimating p_d by the unweighted response rate inside the RHG. The choice $\theta_l = d_l$ leads to estimating p_d by the response rate inside the RHG, weighted by the individual sampling weights. The choice $\theta_l = d_{rl}$ leads to estimating p_d by the response rate inside the RHG, weighted by the individual sampling weights corrected of household unit non-response. We compare these different choices in the simulation study performed in Section 4.

Accounting for the estimated response probabilities leads to the individual weights corrected for household/individual non-response

$$d_{rrl} = \frac{d_{rl}}{\hat{p}_{d(l)}} \quad \text{with } d(l) \text{ the household containing } l. \tag{2.26}$$

The estimator of Y_{ind} adjusted for household/individual non-response is

$$\hat{Y}_{rr,\text{ind}} = \sum_{l \in S_{rr,\text{ind}}} d_{rrl} y_l. \tag{2.27}$$

2.2.3 Calibration

We let z_l denote the vector of calibration variables at the individual level, and Z_{ind} denote the total on the population U_{ind}. For the sample $S_{rr,\text{ind}},$ this leads to the linear calibrated weights

$$w_l = d_{rrl}\left(1 + z_l^{\top}\lambda_{\text{ind}}\right),$$

with

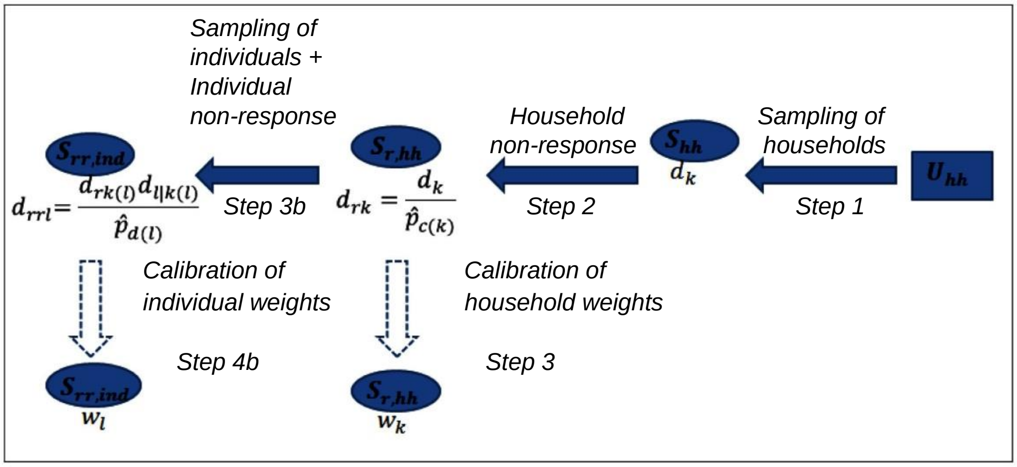
$$\lambda_{\text{ind}} = \left(\sum_{l \in S_{rr,\text{ind}}} d_{rrl} z_l z_l^{\top}\right)^{-1}\left(Z_{\text{ind}} - \hat{Z}_{rr,\text{ind}}\right), \tag{2.28}$$

and where $\hat{Z}_{rr,\text{ind}}$ is the estimator of Z_{ind}, obtained by plugging z_l into (2.27). The calibrated estimator is

$$\hat{Y}_{\text{cal,ind}} = \sum_{l \in S_{rr,\text{ind}}} w_l y_l. \tag{2.29}$$

The sampling and estimation steps are summarized in Figure 2.3.

Figure 2.3 Sampling and estimation steps for a household sample with sub-sampling of individuals.


Sampling of individuals + Individual non-response
Household non-response
Sampling of households
$S_{rr,ind}$
$d_{rrl} = \frac{d_{rk(l)} d_{l|k(l)}}{\hat{p}_{d(l)}}$
Step 3b
$S_{r,hh}$
$d_{rk} = \frac{d_k}{\hat{p}_{c(k)}}$
Step 2
S_{hh}
d_k
Step 1
U_{hh}
Calibration of individual weights
Step 4b
Calibration of household weights
Step 3
$S_{rr,ind}$
w_l
$S_{r,hh}$
w_k

2.2.4 Computation of individual weights on an example

We continue the example initiated in Section 2.1.4. Recall that the sample of responding households is $S_{r,hh} = \{A, D, E, F, H, I, J\}$. The set of all individuals inside the responding households is as follows (say):

$$\underbrace{(i_1, i_2, i_3)}_{A} \quad \underbrace{(i_4)}_{D} \quad \underbrace{(i_5, i_6)}_{E} \quad \underbrace{(i_7, i_8, i_9)}_{F} \quad \underbrace{(i_{10}, i_{11})}_{H} \quad \underbrace{(i_{12})}_{I} \quad \underbrace{(i_{13})}_{J}. \tag{2.30}$$

We suppose that the sampling design consists in selecting one individual exactly inside each household. The set $S_{r,\text{ind}}$ of all sampled individuals inside the responding households is

$$S_{r,\text{ind}} = \{i_1, i_4, i_6, i_8, i_{11}, i_{12}, i_{13}\}. \tag{2.31}$$

From equations (2.23) and (2.16), the individual weights corrected for household non-response are therefore

$$d_{r1} = \frac{40}{3},\ d_{r4} = \frac{72}{13},\ d_{r6} = \frac{576}{13},\ d_{r8} = \frac{160}{3},\ d_{r11} = \frac{576}{13},\ d_{r12} = \frac{288}{13},\quad d_{r13} = \frac{160}{9}. \tag{2.32}$$

Among these 7 selected individuals, 4 only are surveyed due to non-response, accounted for by using the method of Response Homogeneity Groups (RHGs). We suppose that there are two RHGs: the units i_1, i_6, i_8 and i_{11} in the first one, and the units i_4, i_{12} and i_{13} in the second one. The units i_4, i_{11} and i_{13} are non-respondents. Inside each RHG, we compute unweighted estimated response probabilities $(\theta_l = 1)$. This leads to

$$\begin{aligned} \hat{p}_1 &= \frac{\sum_{l \in S_{r1,\text{ind}}} r_l}{\sum_{l \in S_{r1,\text{ind}}} 1} = \frac{3}{4}, \\ \hat{p}_2 &= \frac{\sum_{l \in S_{r2,\text{ind}}} r_l}{\sum_{l \in S_{r2,\text{ind}}} 1} = \frac{1}{3}. \end{aligned} \tag{2.33}$$

The weights accounting for household/individual non-response are obtained for the respondents by dividing the weights in (2.32) by the estimated response probabilities. This leads to the weights

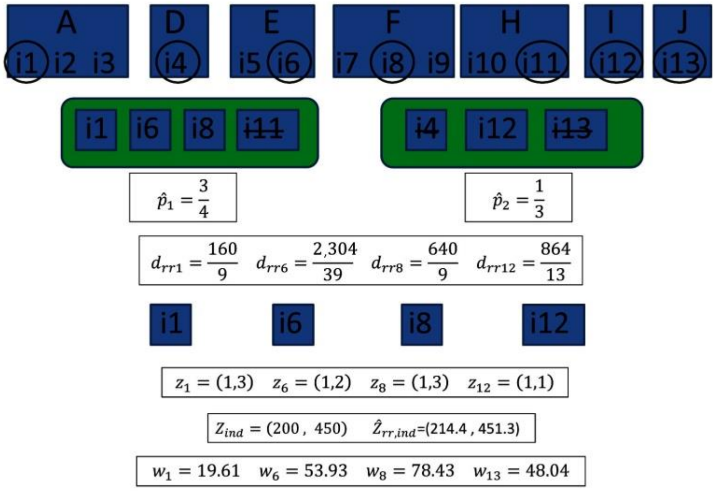
$$d_{rr1} = \frac{160}{9}, \quad d_{rr6} = \frac{2{,}304}{39}, \quad d_{rr8} = \frac{640}{9}, \quad d_{rr12} = \frac{864}{13}. \tag{2.34}$$

Finally, the weights are calibrated to match the population size $N_{\text{ind}} = 200$ and an auxiliary total $Z_{1,\text{ind}} = 450.$ Note that, using the sample of respondents, we obtain $\hat{N}_{r,\text{ind}} = 214.4$ and $\hat{Z}_{1r,\text{ind}} = 451.3$. The calibrated weights are

$$w_1 = 19.61, \quad w_6 = 53.93, \quad w_8 = 78.43, \quad w_{13} = 48.04. \tag{2.35}$$

The sampling and estimation steps are summarized in Figure 2.4.

Figure 2.4 Estimation steps for the weighting of individuals.


A
D
E
F
H
I
J
i1 i2 i3
i4
i5 i6
i7 i8 i9
i10 i11
i12
i13
i1 i6 i8 ~~i11~~
~~i4~~ i12 ~~i13~~
$\hat{p}_1 = \frac{3}{4}$
$\hat{p}_2 = \frac{1}{3}$
$d_{rr1} = \frac{160}{9}$ $d_{rr6} = \frac{2,304}{39}$ $d_{rr8} = \frac{640}{9}$ $d_{rr12} = \frac{864}{13}$
i1
i6
i8
i12
$z_1 = (1,3)$ $z_6 = (1,2)$ $z_8 = (1,3)$ $z_{12} = (1,1)$
$Z_{ind} = (200 , 450)$ $\hat{Z}_{rr,ind}$=(214.4 , 451.3)
$w_1 = 19.61$ $w_6 = 53.93$ $w_8 = 78.43$ $w_{13} = 48.04$

3. Bootstrap variance estimation

We begin in Section 3.1 with the description of the basic step of the bootstrap method when a sample of households only is selected. An illustration is given in Section 3.2 on the example initiated in

Section 2.1.4. The bootstrap method when individuals are sampled inside the households is described in Section 3.3, and an illustration is given in Section 3.4. In Section 3.5, we explain how the basic step of the proposed bootstrap method is used to perform variance estimation and to produce confidence intervals.

3.1 Basic step of the bootstrap for households

Using the with-replacement bootstrap, we first draw inside the original sample S_{hh}^h selected in the stratum U_{hh}^h a with-replacement resample S_{hh*}^h of $n_h - 1$ households, with equal probabilities. Note that the resampling is performed on the sampling unit (a household) rather than on the final unit of observation (an individual), which is key to correctly capture the sampling variance. In particular, this bootstrap method enables to capture the variance due to the second-stage sampling (selection of individuals) without resampling the final units in the bootstrap process. For any $k \in S_{hh}^h$, we define the reweighting adjustment factor

$$G_k = \frac{n_h}{n_h - 1} \times m_k, \tag{3.1}$$

with m_k the number of times the household k is selected in the resample S_{hh*}^h, a.k.a. the multiplicity. Note that some unit $k \in S_{hh}^h$ may not appear in the resample, in which case this unit has multiplicity zero; see Section 3.2 for an example. The reweighting adjustment factors G_k are used to obtain the bootstrap weights accounting for the sampling design, for unit non-response and for the calibration, as described in Algorithm 1. The steps refer to Figure 2.1. The resampling presented in Algorithm 1 is then repeated B times independently for variance estimation and/or to produce a confidence interval, see Algorithm 3 in Section 3.5.

Algorithm 1. Computation of bootstrap household weights accounting for non-response and calibration

- Step 1: we account for the sampling of households by computing, for any $k \in S_{hh}$, the bootstrap sampling weight

$$d_{k*} = G_k d_k. \tag{3.2}$$

The bootstrap version of the full-response estimator given in (2.3) is

$$\hat{Y}_{hh*} = \sum_{k \in S_{hh}} d_{k*} y_k. \tag{3.3}$$

- Step 2: we account for household unit non-response by computing the bootstrap estimated probabilities inside the RHGs

$$\hat{p}_{c*} = \frac{\sum_{k \in S_{c,hh}} G_k \theta_k r_k}{\sum_{k \in S_{c,hh}} G_k \theta_k}, \tag{3.4}$$

and we compute the bootstrap weights corrected for non-response

$$d_{rk*} = \frac{d_{k*}}{\hat{p}_{c(k)*}}, \tag{3.5}$$

with $c(k)$ the RHG containing the household k. The bootstrap version of the estimator corrected for unit non-response given in (2.7) is

$$\hat{Y}_{r,hh*} = \sum_{k \in S_{r,hh}} d_{rk*} y_k. \tag{3.6}$$

- Step 3: we account for the calibration by calibrating the weights d_{rk*} on the totals X_{hh}. This leads to the bootstrap calibrated weights

  $$w_{k*} = d_{rk*}\left(1 + x_k^\top \lambda_{hh*}\right), \tag{3.7}$$

  with

  $$\lambda_{hh*} = \left(\sum_{k \in S_{r,hh}} d_{rk*} x_k x_k^\top\right)^{-1} \left(X_{hh} - \hat{X}_{r,hh*}\right)$$

  and

  $$\hat{X}_{r,hh*} = \sum_{k \in S_{r,hh}} d_{rk*} x_k.$$

  The bootstrap version of the calibrated estimator given in (2.10) is

  $$\hat{Y}_{\text{cal},hh*} = \sum_{k \in S_{r,hh}} w_{k*} y_k. \tag{3.8}$$

The treatment of unit non-response in the bootstrap process deserves some explanations. Firstly, our approach is conditional on the response indicators r_k. Contrarily to the sample membership indicators which are bootstrapped at Step 1 of Algorithm 1, the response indicators remain fixed in the bootstrap process. This is due to the fact that we aim at reproducing a variance estimator which considers the sample S_{hh} as selected with replacement, and that in such case bootstrapping the r_k's is not needed. Secondly, accounting for unit non-response at Step 2 of Algorithm 1 is performed conditionally on the RHGs: we do not bootstrap the process leading to the building of the RHGs (e.g., Girard, 2009; Haziza and Beaumont, 2017). Finally, bootstrapping the response probabilities as described in equation (3.4) accounts for the estimation of the response probabilities p_c. In other words, we use within each resample the same RHGs identified on the basis of the sample, but the non-response adjustments inside the RHGs are based on a resample's content. This is illustrated in the example developed in Section 3.2. If we do not bootstrap the response probabilities and directly plug in equation (3.5) the original estimated probabilities $\hat{p}_c$, then the response probabilities are treated as if they were known, which usually results in an overestimation of the variance (Beaumont, 2005; Kim and Kim, 2007).

Now, we discuss bootstrap variance estimation for calibrated estimators, as considered in Step 3 of Algorithm 1 where the calibration step is performed on the true population total X_{hh}. Following the bootstrap principle which states that the sample S_{hh} is to the bootstrap sample S_{hh*} what the population U_{hh} is to the sample S_{hh}, it could seem more intuitive to rather calibrate on the estimated totals $\hat{X}_{hh}$ obtained by plugging x_k into equation (2.3). Both approaches seem valid for bootstrap variance estimation for the calibrated estimator $\hat{Y}_{\text{cal},hh}$, but the calibration variables x_k may be prone to non-response on the sample S_{hh}, making the estimator $\hat{X}_{hh}$ not possible to compute, while the total X_{hh} is known from an external source.

3.2 An example of computation of bootstrap household weights

We continue with the example initiated in Section 2.1.4. The bootstrap is performed by first selecting a resample of $n_{hh} - 1 = 9$ households, with replacement and with equal probabilities, among the original sampled households. In this example, we suppose that the household A is selected three times, that the household G is selected twice, and that the households D, E, H and I are selected once. Making use of equation (3.2), this leads to the bootstrap sampling weights

$$d_{A*} = \frac{40}{3} \quad d_{D*} = \frac{40}{9} \quad d_{E*} = d_{H*} = d_{I*} = \frac{160}{9} \quad d_{G*} = \frac{320}{9}. \tag{3.9}$$

The bootstrap sampling weights are corrected for non-response in the same way than in the original correction of non-response: using the same RHGs, and weighted estimated probabilities. In this case, the first RHG contains only the unit A which is a respondent, so that $\hat{p}_{1*} = 1$. The second RHG contains D, E, G (non-respondent), H and I. This leads to

$$\hat{p}_{2*} = \frac{d_{D*} + d_{E*} + d_{H*} + d_{I*}}{d_{D*} + d_{E*} + d_{G*} + d_{H*} + d_{I*}} = \frac{13}{21}, \tag{3.10}$$

and to the bootstrap weights corrected for non-response

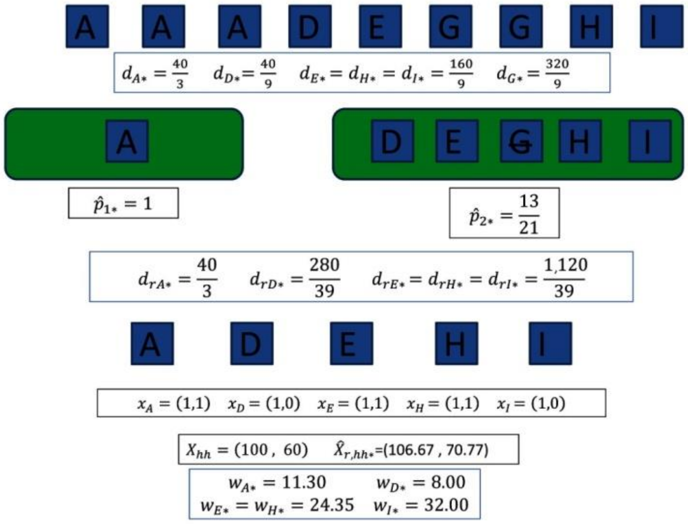
$$d_{rA*} = \frac{40}{3} \quad d_{rD*} = \frac{280}{39} \quad d_{rE*} = d_{rH*} = d_{rI*} = \frac{1{,}120}{39}. \tag{3.11}$$

Finally, the weights are calibrated to match the population size $N_{hh} = 100$ and the auxiliary total $X_{1,hh} = 60$. This leads to the bootstrap calibrated weights

$$w_{A*} = 11.30 \quad w_{D*} = 8.00 \quad w_{E*} = w_{H*} = 24.35 \quad w_{I*} = 32.00. \tag{3.12}$$

The computation of the bootstrap weights is summarized in Figure 3.1.

Figure 3.1 Computation of bootstrap household weights.


A
A
A
D
E
G
G
H
I
$d_{A*} = \frac{40}{3}$ $d_{D*} = \frac{40}{9}$ $d_{E*} = d_{H*} = d_{I*} = \frac{160}{9}$ $d_{G*} = \frac{320}{9}$
A
D
E
G
H
I
$\hat{p}_{1*} = 1$
$\hat{p}_{2*} = \frac{13}{21}$
$d_{rA*} = \frac{40}{3}$ $d_{rD*} = \frac{280}{39}$ $d_{rE*} = d_{rH*} = d_{rI*} = \frac{1,120}{39}$
A
D
E
H
I
$x_A = (1,1)$ $x_D = (1,0)$ $x_E = (1,1)$ $x_H = (1,1)$ $x_I = (1,0)$
$X_{hh} = (100, 60)$ $\hat{X}_{r,hh*}$=(106.67 , 70.77)
$w_{A*} = 11.30$ $w_{D*} = 8.00$
$w_{E*} = w_{H*} = 24.35$ $w_{I*} = 32.00$

3.3 Computation of bootstrap weights for individuals

The computation of the bootstrap weights accounting for the sampling design, for household/individual non-response and for calibration is described in Algorithm 2. The steps refer to Figure 2.3. In addition to the bootstrap steps in Algorithm 1, note that Algorithm 2 involves bootstrapping the computation of response individual probabilities only. Note that the sub-sampling of individuals inside households does not need to be bootstrapped, as discussed in Section 3.1.

Algorithm 2. Computation of bootstrap individual weights accounting for non-response of households, for non-response of individuals and for calibration

- Perform Steps 1 and 2 of Algorithm 1. The bootstrap weights of households corrected for non-response are d_{rk}^*, as given in equation (3.5).
- Step 3b: we first account for the sampling of individuals by computing the bootstrap individual weights corrected for household unit non-response

$$d_{rl*} = d_{rk(l)*} d_{l|k(l)} \quad \text{with } k(l) \text{ the household containing } l. \tag{3.13}$$

We then account for individual unit non-response. We compute the bootstrap estimated probabilities inside the RHGs

$$\hat{p}_{d*} = \frac{\sum_{l \in S_{rd,\text{ind}}} G_{k(l)} \theta_l r_l}{\sum_{l \in S_{rd,\text{ind}}} G_{k(l)} \theta_l}. \tag{3.14}$$

We compute the bootstrap weights of individuals corrected for household/individual non-response, namely

$$d_{rrl*} = \frac{d_{rl*}}{\hat{p}_{d(l)*}}, \tag{3.15}$$

with $d(l)$ the RHG containing the individual l. The bootstrap version of the estimator corrected for unit non-response given in (2.27) is

$$\hat{Y}_{rr,\text{ind}*} = \sum_{l \in S_{rr,\text{ind}}} d_{rrl*} y_l. \tag{3.16}$$

- Step 4b: we account for the calibration by calibrating the weights d_{rrl*} on the totals Z_{ind}. This leads to the bootstrap calibrated weights

$$w_{l*} = d_{rrl*}\left(1 + z_l^\top \lambda_{\text{ind}*}\right), \tag{3.17}$$

with

$$\lambda_{\text{ind}*} = \left(\sum_{k \in S_{rr,\text{ind}}} d_{rrl*} z_l z_l^\top\right)^{-1} \left(Z_{\text{ind}} - \hat{Z}_{rr,\text{ind}*}\right)$$

and

$$\hat{Z}_{rr,\text{ind}*} = \sum_{l \in S_{rr,\text{ind}}} d_{rrl*} z_l.$$

The bootstrap version of the calibrated estimator given in (2.29) is

$$\hat{Y}_{\text{cal},\text{ind}*} = \sum_{l \in S_{rr,\text{ind}}} w_{l*} y_l. \tag{3.18}$$

3.4 An example of computation of bootstrap individual weights

We continue with the example in Section 3.2. The bootstrap sample of households is constituted of A (three times), G (two times), and D, E, H and I (one time). Due to household non-response, we observe A, D, E, H and I only. From (2.30), this results in the bootstrap sample of individuals

$$S_{r,\text{ind}*} = \{i_1, i_4, i_6, i_{11}, i_{12}\}. \tag{3.19}$$

The bootstrap weights of households corrected for unit non-response are given in equation (3.11). From equation (3.13), the bootstrap weights of individuals adjusted for household non-response are

$$d_{r1*} = 40 \quad d_{r4*} = \frac{280}{39} \quad d_{r6*} = \frac{2{,}240}{39} \quad d_{r11*} = \frac{2{,}240}{39} \quad d_{r12*} = \frac{1{,}120}{39}. \tag{3.20}$$

These bootstrap weights are corrected for individual non-response in the same way than in the original correction of individual non-response: using the same RHGs and unweighted estimated probabilities. However, we need to account in these probabilities for the multiplicity m_k and the reweighting adjustment factor G_k, see equation (3.1). In our case, the first RHG contains the individuals i_1, i_6 and i_{11}, and i_{11} is a non-respondent. The individual i_1 belongs to the household A, which has been selected three times $(m_A = 3)$ in the bootstrap sample. The individual i_6 belongs to the household E, and the individual i_{11} belongs to the household H, which have both been selected one time in the bootstrap sample $(m_E = m_H = 1)$. The computation is similar for the second RHG, and leads to

$$\begin{aligned} \hat{p}_{1*} &= \frac{G_A + G_E}{G_A + G_E + G_H} = \frac{4}{5}, \\ \hat{p}_{2*} &= \frac{G_I}{G_D + G_I} = \frac{1}{2}, \end{aligned} \tag{3.21}$$

and to the bootstrap individuals weights corrected for household/individual non-response

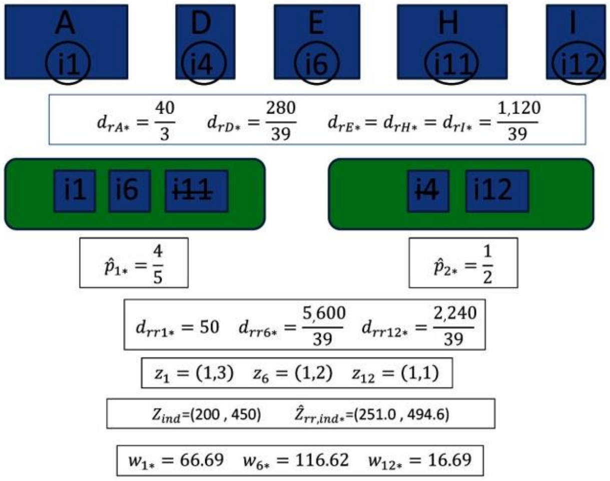
$$d_{rr1*} = 50 \quad d_{r6*} = \frac{5{,}600}{39} \quad d_{r12*} = \frac{2{,}240}{39}. \tag{3.22}$$

Finally, the weights are calibrated to match the population size $N_{\text{ind}} = 200$ and the auxiliary total $Z_{1,\text{ind}} = 450.$ This leads to the bootstrap calibrated weights

$$w_{1*} = 66.69 \quad w_{6*} = 116.62 \quad w_{12*} = 16.69. \tag{3.23}$$

The computation of bootstrap individual weights is summarized in Figure 3.2.

Figure 3.2 Computation of bootstrap individual weights.


A
i1
D
i4
E
i6
H
i11
I
i12
$d_{rA*} = \frac{40}{3}$ $d_{rD*} = \frac{280}{39}$ $d_{rE*} = d_{rH*} = d_{rI*} = \frac{1{,}120}{39}$
i1 i6 ~~i11~~
~~i4~~ i12
$\hat{p}_{1*} = \frac{4}{5}$
$\hat{p}_{2*} = \frac{1}{2}$
$d_{rr1*} = 50$ $d_{rr6*} = \frac{5{,}600}{39}$ $d_{rr12*} = \frac{2{,}240}{39}$
$z_1 = (1,3)$ $z_6 = (1,2)$ $z_{12} = (1,1)$
Z_{ind}=(200 , 450) $\hat{Z}_{rr,ind*}$=(251.0 , 494.6)
$w_{1*} = 66.69$ $w_{6*} = 116.62$ $w_{12*} = 16.69$

3.5 Bootstrap variance estimation and confidence intervals

In this section, we are interested in parameters which may be written as smooth functions of totals. We explain how the basic step of the proposed bootstrap method is used to perform variance estimation and to produce confidence intervals. For brevity, we focus on parameters defined over the population of households U_{hh}. The treatment for parameters of interest in the population of individuals U_{ind} is similar.

Suppose that y_k is a q-vector of interest variables, and that we are interested in some parameter $\beta_{hh} = f(Y_{hh})$ with $f: \mathbb{R}^q \rightarrow \mathbb{R}$ a known, smooth function. In case of full response, the substitution estimator of β_{hh} is

$$\hat{\beta}_{hh} = f(\hat{Y}_{hh}), \tag{3.24}$$

see for example Deville (1999). In case of unit non-response at the household level, the estimator of β_{hh} corrected for unit non-response is

$$\hat{\beta}_{r,hh} = f(\hat{Y}_{r,hh}), \tag{3.25}$$

and the calibrated estimator of β_{hh} is

$$\hat{\beta}_{\text{cal},hh} = f(\hat{Y}_{\text{cal},hh}). \tag{3.26}$$

In each case, a bootstrap variance estimator is obtained by applying a large number of times (say B) the basic step of the bootstrap method in Algorithm 1, and then by computing the dispersion of the bootstrap estimators. This is summarized in Algorithm 3.

Algorithm 3. Bootstrap variance estimation for an estimation over the population of households

1. Repeat B times the bootstrap procedure described in Algorithm 1. Let us denote $\hat{Y}^b_{hh*}$, $\hat{Y}^b_{r,hh*}$ and $\hat{Y}^b_{\text{cal},hh*}$ for the bootstrap estimators of totals computed on the b^{th} sample. Also, let us denote $\hat{\beta}^b_{hh*}$, $\hat{\beta}^b_{r,hh*}$ and $\hat{\beta}^b_{\text{cal},hh*}$ for the associated bootstrap estimators of β_{hh}.
2. The Bootstrap variance estimator for $\hat{\beta}_{hh}$ is

$$\hat{V}_{\text{boot}}(\hat{\beta}_{hh}) = \frac{1}{B-1} \sum_{b=1}^{B} \left\{ \hat{\beta}^b_{hh*} - \frac{1}{B} \sum_{b'=1}^{B} \hat{\beta}^{b'}_{hh*} \right\}^2, \tag{3.27}$$

and similarly for $\hat{\beta}_{r,hh}$ and $\hat{\beta}_{\text{cal},hh}$.

The bootstrap variance estimator may be used to compute a normality-based confidence interval with targeted level $1-2\alpha$. For example, the confidence interval when using the full-response estimator $\hat{\beta}_{hh}$ is

$$\text{IC}_{\text{nor}}(\beta_{hh}) = \left[\hat{\beta}_{hh} \pm u_{1-\alpha} \left\{ \hat{V}_{\text{boot}}(\hat{\beta}_{hh}) \right\}^{0.5} \right], \tag{3.28}$$

with $u_{1-\alpha}$ the quantile of order $1-\alpha$ of the standard normal distribution. This confidence interval is expected to be conservative, since the proposed bootstrap method is conservative too.

We also consider the percentile and the reverse percentile (a.k.a. basic) bootstrap confidence intervals. They can be directly computed from the bootstrap weights and are therefore attractive from a data user's perspective, unlike more computationally intensive methods like the t-bootstrap (e.g. Davison and Hinkley, 1997; Shao and Tu, 1995). For $\hat{\beta}_{hh},$ the percentile confidence interval is obtained by using the distribution of $\hat{\beta}_{hh*}$ as an approximation of the distribution of $\hat{\beta}_{hh}.$ It makes use of the ordered bootstrap estimates $\hat{\beta}_{hh*}^{(1)}, \ldots, \hat{\beta}_{hh*}^{(B)}$ to form the confidence interval

$$\mathrm{IC}_{\text{per}}(\beta_{hh}) = \left[\hat{\beta}_{hh*}^{(L)}, \hat{\beta}_{hh*}^{(U)}\right], \tag{3.29}$$

with targeted level $1 - 2\alpha,$ where $L = \alpha B$ and $U = (1 - \alpha) B.$ The reverse percentile confidence interval is obtained by viewing the distribution of $(\hat{\beta}_{hh*} - \hat{\beta}_{hh})$ as an approximation of the distribution of $(\hat{\beta}_{hh} - \beta_{hh}).$ It leads to the confidence interval

$$\mathrm{IC}_{\text{rev}}(\beta_{hh}) = \left[2\hat{\beta}_{hh} - \hat{\beta}_{hh*}^{(U)}, 2\hat{\beta}_{hh} - \hat{\beta}_{hh*}^{(L)}\right]. \tag{3.30}$$

The properties of the bootstrap variance estimator and of the three confidence intervals are evaluated in the simulation study performed in Section 4 for the estimation of a total.

Choosing the number B of resamples is an important practical problem. Girard (2009) suggests considering several possible resample sizes (e.g., by increasing B with an increment of 100), and plotting the bootstrap variance estimators in function of $B.$ The value for which this variance estimator starts to stabilize is then retained. This is a simple method, but which may require some compromise solution if different variables of interest lead to different stabilizing values. Beaumont and Patak (2012) suggest choosing B such that with a high probability, the length of the bootstrap confidence interval given in (3.28) is close to the length of the confidence interval obtained with an analytical variance estimator. Under the assumption that conditionally on the original sample, the normalized bootstrap estimator of the total is normally distributed, they establish that the value B may be determined from the distribution of a chi-square variable (Beaumont and Patak, 2012, equation 10). Interestingly, the value obtained does not depend on the variable of interest. Based on these results, they suggest using a value B no smaller than 750, and a larger value if the normality assumption of the bootstrap estimator may fail. We used $B =$ 1,000 in the simulation study presented in the following section. For surveys that are to serve multiple analytical needs – ranging from simple to complex population parameters and various domain sizes – selecting no fewer than 1,000 replicates is the norm given the computing resources available nowadays.

4. Simulation study

In order to evaluate the proposed bootstrap method, we conducted a simulation study on an artificial population. We first generate a population U_{hh} containing $N_{hh} =$ 100,000 households, with four auxiliary variables $x_1, \ldots, x_4$ generated from a gamma distribution with shape and scale parameters 2 and 5. Inside the population, we generate three variables of interest $y_1, \ldots, y_3$ according to the following models

$$\begin{aligned} y_{1k} &= 10 + x_{1k} + x_{2k} + \sigma_\varepsilon \varepsilon_k, \\ y_{2k} &= 10 + x_{1k} + x_{3k} + \sigma_\varepsilon \varepsilon_k, \\ y_{3k} &= 10 + x_{3k} + x_{4k} + \sigma_\varepsilon \varepsilon_k, \end{aligned} \tag{4.1}$$

where ε_k is generated according to a standard normal distribution. We set $\sigma_\varepsilon = 10,$ which results in a coefficient of determination of approximately 0.50 for each model. The auxiliary variables $1, x_{1k}, x_{2k}$ are used as calibration variables at the household level in this simulation study. The three variables of interest therefore correspond to cases when the calibration model is well specified $(y_1),$ partly well specified $(y_2),$ or poorly specified $(y_3).$ The population U_{hh} is randomly split into five response homogeneity groups (RHG) of equal sizes. The response probability p_c inside the RHG c is equal to 0.5 for the first group, 0.6 for the second group, ..., and 0.9 for the fifth group, resulting in an average response rate of 70% for the households.

Inside each household $k,$ we generate N_k individuals, where $N_k - 1$ is generated according to a Poisson distribution with parameter 1, which results in an average number of 2 individuals per household. Inside the corresponding population $U_{\text{ind}},$ we generate four auxiliary variables $z_1, \ldots, z_4$ with shape and scale parameters 2 and 0.5. Also, we generate three variables of interest y_4, y_5, y_6 according to the following models

$$\begin{aligned} y_{4l} &= 5 + 0.5z_{1l} + 0.5z_{2l} + \sigma_\eta \eta_l, \\ y_{5l} &= 5 + 0.5z_{1l} + 0.5z_{3l} + \sigma_\eta \eta_l, \\ y_{6l} &= 5 + 0.5z_{3l} + 0.5z_{4l} + \sigma_\eta \eta_l, \end{aligned} \tag{4.2}$$

where η_l is generated according to a standard normal distribution. We set σ_η = 0.4, which results in a coefficient of determination of approximately 0.6 for each model. The auxiliary variables $1, z_{1l}, z_{2l}$ are used as calibration variables at the individual level in this simulation study. The three variables of interest therefore correspond to a case when the calibration model is well specified $(y_4),$ partly well specified $(y_5),$ or poorly specified $(y_6).$

The population U_{ind} is split into five RHGs as follows. The individuals which are alone in their household form a separate RHG, with a response probability of 1. The rationale behind this choice is that in such case, the individual is somewhat equivalent to his/her household, and that the non-response is modeled at the household level. Among the rest of the individuals living in a household k with $N_k = 2$ individuals or more, the variables z_1 and z_2 are used to form four RHGs of approximately equal size. The response probability p_d ranges from 0.80 to 0.95 in these four remaining RHGs. This results in an overall response rate of approximately 90% for the individuals.

Inside the population $U_{hh},$ we select a sample S_{hh} of n_{hh} = 1,000 households by simple random sampling without replacement. Note that the sampling rate is small (1%), so that simple random sampling with/without replacement are not much different, and the bias of the bootstrap variance estimators is expected to be small under this set-up. The non-response is generated according to the RHG household model, which results in a sample $S_{r,hh}$ of responding households. The estimated response probabilities $\hat{p}_c$ are obtained from equation (2.5), with equal weight $\theta_k = 1.$ Inside each $k \in S_{r,hh},$ one Kish individual is randomly selected with equal probabilities, which results in the sample of individuals $S_{r,\text{ind}}.$ Inside $S_{r,\text{ind}},$ the non-response is generated according to the RHG individual model, resulting in a sample $S_{rr,\text{ind}}$ of responding individuals. The estimated response probabilities $\hat{p}_d$ are obtained from equation (2.25), in three possible ways: equal weights $\theta_l = 1,$ sampling weights $\theta_l = d_l,$ or individuals weights corrected for the household non-response $\theta_l = d_{rl}.$

The sampling and non-response steps are repeated $R =$ 1,000 times. On each sample $S_{hh},$ we compute the full-response estimator given in (2.3), and on each sample $S_{r,hh},$ we compute the estimator adjusted for non-response $\hat{Y}_{r,hh}$ given in (2.7) and the estimator $\hat{Y}_{\text{cal},hh}$ given in (2.10) with the set of calibration variables $x_k = (1, x_{1k}, x_{2k})^\top.$ On each sample $S_{rr,\text{ind}},$ we compute the estimator adjusted for non-response $\hat{Y}_{rr,\text{ind}}$ given in (2.27) and the estimator $\hat{Y}_{\text{cal,ind}}$ given in (2.29) with the set of calibration variables $z_l = (1, z_{1l}, z_{2l})^\top.$ For these five estimators, we compute the normalized root mean square error

$$\text{NRMSE}(\hat{Y}) = 100 \times \frac{\sqrt{\text{MSE}(\hat{Y})}}{Y}, \tag{4.3}$$

with $\text{MSE}(\hat{Y})$ a simulation-based approximation of the mean square error of $\hat{Y},$ obtained from an independent run of 10,000 simulations.

For these five estimators, we also compute the bootstrap variance estimators obtained by applying Algorithm 3 with $B =$ 1,000. So as to measure the bias of a variance estimator $v(\hat{Y}),$ we use the Monte Carlo Percent Relative Bias

$$\text{RB}\left\{v(\hat{Y})\right\} = 100 \times \frac{R^{-1}\sum_{c=1}^{R} v_c(\hat{Y}_c) - \text{MSE}(\hat{Y})}{\text{MSE}(\hat{Y})}, \tag{4.4}$$

where $v_c(\hat{Y}_c)$ stands for the variance estimator in the c^{th} sample. As a measure of stability of $v(\hat{Y}),$ we use the Relative Stability

$$\text{RS}\left\{v(\hat{Y})\right\} = 100 \times \frac{\left[R^{-1}\sum_{c=1}^{R} \left\{v_c(\hat{Y}_c) - \text{MSE}(\hat{Y})\right\}^2\right]^{1/2}}{\text{MSE}(\hat{Y})}. \tag{4.5}$$

Also, we compute the coverage rates of the confidence interval associated to the percentile Bootstrap, to the basic bootstrap and to the normality-based confidence interval, with nominal one-tailed error rate of 2.5% in each tail.

The results are presented in Table 4.1 for the estimation on the population of households. The normalized root mean square error of the calibrated estimator $\hat{Y}_{\text{cal},hh}$ is smaller when the calibration variables are explanatory for the variable of interest, as expected. We observe a slight positive bias of the bootstrap variance estimator for the full-response estimator $\hat{Y}_{hh},$ but almost no bias for the reweighted estimators $\hat{Y}_{r,hh}$ and $\hat{Y}_{\text{cal},hh}.$ The bootstrap variance estimator is slightly less stable with the reweighted estimators, which is likely due to the additional variability associated to the correction of unit non-response. Concerning the confidence intervals, we note that the coverage rates are well respected in all cases and for the three studied methods.

We now turn to the result on the population of individuals, which are presented in Table 4.2. We observe that the relative bias of the bootstrap variance estimator is very small in all cases. The choice of the weights θ_k used in the estimation of the response probabilities seem to have no effect on the normalized root mean square error of the estimators, but the use of the weights $\theta_l = d_{rl}$ adjusted for household non-response yields slightly more stable variance estimators for $\hat{Y}_{rr,\text{ind}}$. The coverage rates are approximately respected in all cases.

Table 4.1
Coefficient of variation of the estimator of the total, Relative Bias and Relative Stability of the Bootstrap variance estimator, and Nominal One-Tailed Error Rates of the percentile bootstrap and of the basic bootstrap for 3 variables on the population of households

					Percentile bootstrap			Basic bootstrap			Normality-based		
		NRMSE	RB	RS	L	U	L+U	L	U	L+U	L	U	L+U
$\hat{Y}_{hh}$	y_1	1.47	2.48	7.2	2.2	3.1	5.3	2.1	3.3	5.4	2.2	3.2	5.4
	y_2	1.48	0.73	6.6	2.6	3.3	5.9	2.7	3.4	6.1	2.6	3.2	5.8
	y_3	1.48	1.11	6.6	2.6	2.7	5.3	2.7	3.0	5.7	2.4	2.7	5.1
$\hat{Y}_{r,hh}$	y_1	1.82	0.42	8.7	2.4	2.4	4.8	2.3	2.7	5.0	2.3	2.6	4.9
	y_2	1.83	-0.76	8.2	2.7	2.8	5.5	2.5	3.0	5.5	2.2	2.7	4.9
	y_3	1.82	0.72	8.4	2.8	2.1	4.9	2.8	2.2	5.0	2.8	1.9	4.7
$\hat{Y}_{\text{cal},hh}$	y_1	1.29	1.27	8.3	2.4	2.7	5.1	2.8	2.8	5.6	2.8	2.7	5.5
	y_2	1.58	-0.55	8.2	2.5	3.5	6.0	2.8	3.9	6.7	2.8	3.6	6.4
	y_3	1.82	0.49	8.4	2.9	1.8	4.7	3.0	2.2	5.2	2.9	2.0	4.9

Table 4.2
Coefficient of variation of the estimator of the total, Relative Bias and Relative Stability of the Bootstrap variance estimator, and Nominal One-Tailed Error Rates of the percentile bootstrap and of the basic bootstrap for 3 variables on the population of individuals

					Percentile bootstrap			Basic bootstrap			Normality-based		
		NRMSE	RB	RS	L	U	L+U	L	U	L+U	L	U	L+U
		Equal weights $\theta_l = 1$											
$\hat{Y}_{rr,\text{ind}}$	y_4	2.01	0.31	9.6	2.0	3.2	5.2	1.9	3.3	5.2	1.9	3.0	4.9
	y_5	2.02	-0.17	9.6	2.4	3.4	5.8	2.2	3.7	5.9	2.3	3.5	5.8
	y_6	2.02	-0.24	9.6	2.2	3.3	5.5	2.0	3.7	5.7	2.0	3.2	5.2
$\hat{Y}_{\text{cal,ind}}$	y_4	0.29	1.72	10.8	2.1	2.4	4.5	2.1	2.3	4.4	2.1	2.2	4.3
	y_5	0.39	1.04	11.3	2.3	2.5	4.8	2.3	2.5	4.8	2.2	2.4	4.6
	y_6	0.47	1.90	11.2	2.8	2.1	4.9	2.2	2.5	4.7	2.3	2.0	4.3
		Sampling weights $\theta_l = d_l$											
$\hat{Y}_{rr,\text{ind}}$	y_4	2.00	-0.08	9.5	1.8	3.8	5.6	1.7	3.8	5.5	1.7	3.4	5.1
	y_5	2.00	0.14	9.4	1.9	3.3	5.2	2.2	3.5	5.7	1.8	3.5	5.3
	y_6	1.99	0.61	9.3	1.7	3.2	4.9	1.7	3.4	5.1	1.7	3.2	4.9
$\hat{Y}_{\text{cal,ind}}$	y_4	0.29	-0.57	10.3	2.9	2.4	5.3	3.3	2.2	5.5	3.0	2.3	5.3
	y_5	0.39	0.40	11.6	2.4	3.2	5.6	2.7	3.3	6.0	2.3	3.2	5.5
	y_6	0.47	-0.05	11.2	2.3	2.2	4.5	1.8	2.3	4.1	1.8	2.3	4.1
		Weights adjusted for household non-response $\theta_l = d_{rl}$											
$\hat{Y}_{rr,\text{ind}}$	y_4	1.99	-0.71	8.9	2.5	2.3	4.8	2.6	2.7	5.3	2.5	2.4	4.9
	y_5	1.99	-0.82	8.9	3.1	2.2	5.3	2.9	2.5	5.4	2.5	2.2	4.7
	y_6	1.99	-0.26	9.1	3.1	2.3	5.4	3.0	3.0	6.0	2.9	2.5	5.4
$\hat{Y}_{\text{cal,ind}}$	y_4	0.29	1.70	10.6	2.7	3.4	6.1	2.6	3.3	5.9	2.5	3.3	5.8
	y_5	0.39	1.38	11.3	2.1	2.7	4.8	2.2	3.0	5.2	1.7	3.0	4.7
	y_6	0.47	0.61	10.9	2.5	2.8	5.3	2.3	3.0	5.3	2.3	2.8	5.1

5. Application to the French panel for urban policy

In this section, we present an illustration of the proposed methodology on a French panel for urban policy. The sampling design and the estimation steps for the sample of households are briefly described in Section 5.1, and three possible bootstrap confidence intervals are computed. The SAS macro developed to implement the proposed methodology for one-stage sampling is given in Appendix B, along with a small example. The additional sampling and estimation steps for the sample of individuals are described in Section 5.2, and three possible bootstrap confidence intervals are computed. The SAS macro developed to implement the proposed methodology for two-stage sampling is given in Appendix C, along with a small example.

5.1 Sample of households

The Panel for Urban Policy (PUP) is a survey in four waves, conducted between 2011 and 2014 by the French General Secretariat of the Inter-ministerial Committee for Cities (SGCIV). The survey aims at collecting information about security, employment, precariousness, schooling and health, for people living in the Sensitive Urban Zones (ZUS). We are only interested in the 2011 wave of the survey. A sample of households is selected, and all the individuals living in the selected households are theoretically surveyed.

The sample of households is obtained by two-stage sampling, see for example Chauvet (2015); Chauvet and Vallée (2018). Firstly, the population of districts is partitioned into 4 strata, and a global sample of $n_I = 40$ districts is selected by means of probability proportional to size sampling inside strata. A sample of households is then selected at the second-stage inside each selected district by means of simple random sampling, in such a way that the final inclusion probabilities of households are approximately equal inside strata (self-weighted sampling design). For the purpose of illustration, the two-stage selection of the households is not considered here, and the sample of households is viewed as directly selected by means of stratified simple random sampling.

The sample contains 2,971 households, but due to unit non-response only 1,256 households are observed. Non-response is accounted for by using Response Homogeneity Groups, defined with respect to five auxiliary variables: housing construction period, type of dwelling (apartment/house), number of rooms, low-income housing (yes/no), region. By using a logistic regression and the score method (e.g. Haziza and Beaumont, 2007), we obtain 8 response homogeneity groups. The five auxiliary variables used in the definition of the RHGs are also used for calibration.

We are interested in four categorical variables related to security, town planning and residential mobility. The variable y_1 gives the perceived reputation of the district (good, fair, poor, no opinion). The variable y_2 indicates if a member of the household has witnessed trafficking (never, rarely, sometimes, no opinion). The variable y_3 indicates if some significant roadworks have been done in the neighborhood in the twelve last months (yes, no, no opinion). The variable y_4 indicates if the household intends to leave the district during the next twelve months (certainly/probably, certainly not, probably not, no opinion). For each category g of each variable $y,$ we are interested in the proportion

$$\beta_{g,hh} = \frac{\sum_{k \in U_{hh}} 1(y_k = g)}{N_{hh}}, \tag{5.1}$$

with N_{hh} the total number of households. The estimator of β_g adjusted for non-response is

$$\hat{\beta}_{\text{gr},hh} = \frac{\sum_{k \in S_{r,hh}} d_{rk} 1(y_k = g)}{\sum_{k \in S_{r,hh}} d_{rk}}, \tag{5.2}$$

see equation (2.7). The calibrated estimator of β_g is

$$\hat{\beta}_{\text{gcal},hh} = \frac{\sum_{k \in S_{r,hh}} w_k 1(y_k = g)}{\sum_{k \in S_{r,hh}} w_k}, \tag{5.3}$$

see equation (2.10).

For each proportion, we give the normality-based confidence interval making use of the bootstrap variance estimator, the percentile bootstrap and the basic bootstrap confidence intervals, see Section 3.5. We use the with-replacement Bootstrap presented in Algorithm 1 with $B = 1{,}000$ resamples. The results with a nominal one-tailed error rate of 2.5% are presented in Table 5.1. The three confidence intervals are very similar in all cases.

Table 5.1
Estimation of the marginal proportions with three confidence intervals for four variables on interest

	Perceived reputation of district status							
	Estimator adj. for non-response				Calibration estimator			
	Good	Fair	Poor	No opinion	Good	Fair	Poor	No opinion
Estim.	0.217	0.225	0.531	0.027	0.217	0.224	0.532	0.027
Norm. CI	[0.194,0.241]	[0.201,0.249]	[0.503,0.559]	[0.018,0.036]	[0.193,0.240]	[0.200,0.248]	[0.504,0.560]	[0.018,0.036]
Perc. CI	[0.195,0.241]	[0.201,0.251]	[0.504,0.558]	[0.019,0.036]	[0.193,0.240]	[0.201,0.251]	[0.505,0.560]	[0.019,0.036]
Basic CI	[0.193,0.240]	[0.200,0.249]	[0.503,0.557]	[0.018,0.035]	[0.193,0.240]	[0.198,0.248]	[0.504,0.559]	[0.018,0.035]
	Witnessed trafficking							
	Estimator adj. for non-response				Calibration estimator			
	Never	Rarely	Sometimes	No opinion	Never	Rarely	Sometimes	No opinion
Estim.	0.599	0.065	0.155	0.181	0.606	0.065	0.156	0.173
Norm. CI	[0.571,0.627]	[0.050,0.079]	[0.135,0.175]	[0.161,0.201]	[0.581,0.632]	[0.050,0.079]	[0.135,0.176]	[0.159,0.188]
Perc. CI	[0.572,0.628]	[0.050,0.080]	[0.134,0.175]	[0.161,0.201]	[0.582,0.633]	[0.051,0.080]	[0.134,0.175]	[0.160,0.188]
Basic CI	[0.570,0.626]	[0.049,0.078]	[0.136,0.176]	[0.161,0.201]	[0.579,0.630]	[0.049,0.078]	[0.136,0.177]	[0.159,0.187]
	Roadworks in neighborhood							
	Estimator adj. for non-response				Calibration estimator			
	Yes	No	No opinion		Yes	No	No opinion	
Estim.	0.471	0.495	0.034		0.470	0.496	0.034	
Norm. CI	[0.444,0.498]	[0.468,0.523]	[0.024,0.044]		[0.443,0.496]	[0.469,0.523]	[0.024,0.045]	
Perc. CI	[0.442,0.496]	[0.469,0.524]	[0.025,0.045]		[0.440,0.495]	[0.470,0.524]	[0.025,0.045]	
Basic CI	[0.445,0.500]	[0.466,0.522]	[0.023,0.043]		[0.444,0.499]	[0.468,0.522]	[0.024,0.044]	
	Intention to leave the district							
	Estimator adj. for non-response				Calibration estimator			
	Cert./Prob.	Prob. not	Cert. not	No opinion	Cert./Prob.	Prob. not	Cert. not	No opinion
Estim.	0.286	0.130	0.548	0.036	0.287	0.131	0.546	0.036
Norm. CI	[0.260,0.312]	[0.111,0.149]	[0.520,0.576]	[0.025,0.047]	[0.261,0.313]	[0.112,0.150]	[0.518,0.573]	[0.025,0.047]
Perc. CI	[0.260,0.313]	[0.111,0.149]	[0.521,0.576]	[0.026,0.047]	[0.261,0.313]	[0.113,0.151]	[0.520,0.574]	[0.026,0.048]
Basic CI	[0.259,0.312]	[0.111,0.149]	[0.520,0.575]	[0.025,0.046]	[0.261,0.313]	[0.111,0.149]	[0.517,0.572]	[0.025,0.047]

5.2 Sample of individuals

The sample of responding households contains 3,098 individuals who are theoretically surveyed, but due to unit non-response we observe a subset of 2,804 individual respondents only. Non-response is accounted for by using Response Homogeneity Groups, defined with respect to eight auxiliary variables: three at the individual level (sex, age, nationality), and five at the dwelling level (housing construction period, type of dwelling, number of rooms, low-income housing or not, region). By using a logistic regression and the score method, we obtain 8 response homogeneity groups. The three individual auxiliary variables used in the definition of the RHGs are also used for calibration.

We are interested in three variables of interest. The variable y_5 is quantitative, and gives the number of children. The variable y_6 indicates whether the individual has one or several jobs (one, several, none, no answer). The variable y_7 indicates whether the individual benefits from a complementary full medical cover (yes, no, no answer). For the variable $y_5,$ we compute the estimator of the total adjusted for non-reponse and the calibrated estimator given in equations (2.27) and (2.29), respectively. For the two other variables of interest and for each category $g,$ we are interested in the proportion

$$\beta_{g,\text{ind}} = \frac{\sum_{l \in U_{\text{ind}}} 1(y_k = g)}{N_{\text{ind}}}, \tag{5.4}$$

with N_{ind} the total number of individuals. The estimator of $\beta_{g,\text{ind}}$ adjusted for non-response is

$$\hat{\beta}_{\text{grr,ind}} = \frac{\sum_{l \in S_{rr,\text{ind}}} d_{rrl} 1(y_l = g)}{\sum_{l \in S_{rr,\text{ind}}} d_{rrl}}, \tag{5.5}$$

see equation (2.27). The calibrated estimator of $\beta_{g,\text{ind}}$ is

$$\hat{\beta}_{\text{gcal,ind}} = \frac{\sum_{l \in S_{rr,\text{ind}}} w_l 1(y_l = g)}{\sum_{l \in S_{rr,\text{ind}}} w_l}, \tag{5.6}$$

see equation (2.29).

For each parameter, we give the normality-based confidence interval making use of the bootstrap variance estimator, the percentile bootstrap and the basic bootstrap confidence intervals. We use the with-replacement Bootstrap presented in Algorithm 2 with $B = 1{,}000$ resamples. The results with a nominal one-tailed error rate of 2.5% are presented in Table 5.2. The three confidence intervals are very similar in all cases.

Table 5.2
Estimation of the marginal proportions with three confidence intervals for four variables on interest

	Number of children							
	Estimator adj. for non-response				Calibration estimator			
Estim. ($\times 10^6$)	4.40				4.39			
Norm. CI	[4.15,4.64]				[4.21,4.58]			
Perc. CI	[4.16,4.65]				[4.21,4.58]			
Basic CI	[4.14,4.63]				[4.20,4.57]			
	Does the individual have several jobs?							
	Estimator adj. for non-response				Calibration estimator			
	One	Several	None	No answer	One	Several	None	No answer
Estim.	0.304	0.016	0.372	0.308	0.305	0.016	0.372	0.307
Norm. CI	[0.286,0.323]	[0.011,0.021]	[0.352,0.392]	[0.290,0.326]	[0.285,0.325]	[0.011,0.021]	[0.350,0.394]	[0.283,0.332]
Perc. CI	[0.287,0.323]	[0.011,0.021]	[0.351,0.393]	[0.289,0.326]	[0.284,0.325]	[0.011,0.020]	[0.351,0.393]	[0.284,0.333]
Basic CI	[0.286,0.322]	[0.011,0.020]	[0.351,0.393]	[0.289, 0.326]	[0.285,0.325]	[0.011,0.020]	[0.352,0.393]	[0.282,0.330]
	Complementary full medical cover							
	Estimator adj. for non-response				Calibration estimator			
	Yes	No	No answer		Yes	No	No answer	
Estim.	0.122	0.626	0.252		0.122	0.627	0.251	
Norm. CI	[0.106,0.137]	[0.603,0.650]	[0.234,0.270]		[0.105,0.138]	[0.604,0.650]	[0.227,0.275]	
Perc. CI	[0.105,0.137]	[0.603,0.651]	[0.235,0.269]		[0.105,0.138]	[0.604,0.650]	[0.230,0.276]	
Basic CI	[0.106,0.138]	[0.602,0.649]	[0.235,0.269]		[0.105,0.138]	[0.605,0.651]	[0.227,0.273]	

6. Conclusion and future work

In this paper, we have explained how the with-replacement bootstrap may be applied to household surveys, in order to account for the whole variability of the sampling process including sampling and non-response, and to a posteriori adjustments like calibration. The methods have been illustrated on a toy example for clarity of exposition, evaluated via a simulation study and applied to a French panel for urban policy. To make the implementation of the method easier for users, we have developed two SAS macros which are available upon request to the corresponding author.

The results in the simulation study show that both the bootstrap variance estimators and three bootstrap confidence intervals work well in case of a small sampling fraction. If the sampling fraction is larger, the bootstrap variance estimator is known to be conservative, and the normality-based confidence interval is therefore expected to be conservative as well. However, the coverage properties of the two other confidence intervals in such context remain unclear. This is an interesting matter for further research.

In this paper, we focused on applying the bootstrap for variance estimation, after the statistical adjustments (treatment of unit non-response and calibration) have been performed by the survey methodologist. Bootstrap may also be used a priori, as a diagnosis tool to evaluate the relevance of possible statistical adjustments. For example, it may be tempting to use a large number of Response Homogeneity Groups (RHGs) to correct unit non-response, so as to reduce the non-response bias. However, this may result in an increased variability of the reweighted estimators. Bootstrap may be used to evaluate several possible sets of RHGs, for example by producing histograms of the bootstrap non-response adjustments and/or of the bootstrap estimators corrected for unit non-response, to give some insight on the stability of estimation with a possible set of RHGs. This is helpful in finding a bias/variance trade-off. This approach in mentioned in Girard (2009), and is an important matter for further work.

We have considered the situation when the survey is performed at one time only. If we wish to perform longitudinal estimations, units are typically followed over time. If we are also interested in cross-sectional estimations at several times, additional samples are selected at posterior waves and mixed with the original sample. Bootstrap variance estimation in the context of longitudinal surveys is a very important matter for further investigation.

Acknowledgements

The authors are grateful to the associate editor and two referees for numerous suggestions on a previous version of the paper, which have helped us in significantly improving the article.

Appendix

A. Benchmark variance estimators for the sample of individuals

We first consider the estimator $\hat{Y}_{\text{ind}}$ in equation (2.21), that we use in case of full response. The benchmark variance estimator is

$$v_{\text{mult}}(\hat{Y}_{\text{ind}}) = \sum_{h=1}^{H} \frac{n_h}{n_h - 1} \sum_{k \in S_{hh}^h} \left(d_k \hat{y}_k - \frac{1}{n_h} \sum_{k' \in S_{hh}^h} d_{k'} \hat{y}_{k'} \right)^2, \tag{A.1}$$

with

$$\hat{y}_k = \sum_{l \in S_{\text{ind},k}} d_{l|k} y_l.$$

We now consider the estimator $\hat{Y}_{rr,\text{ind}}$ given in equation (2.27), which is adjusted for the non-response of both households and individuals. The benchmark variance estimator is

$$v_{\text{mult}}(\hat{Y}_{\text{ind}}) = \sum_{h=1}^{H} \frac{n_h}{n_h - 1} \sum_{k \in S_{hh}^h} \left(d_k v_{1k} - \frac{1}{n_h} \sum_{k' \in S_{hh}^h} d_{k'} v_{1k'} \right)^2, \tag{A.2}$$

where

$$v_{1k} = \hat{u}_{1k} + u_{3k},$$

where the first linearized variable $\hat{u}_{1k}$ is similar to that given in equation (2.8), while the second linearized variable u_{3k} accounts for the estimation of the individual response probabilities. We have for the first linearized variable

$$\hat{u}_{1k} = \theta_k \pi_k \bar{\hat{y}}_{rc(k)} + \frac{r_k}{\hat{p}_{c(k)}} \left\{ \hat{y}_{r,k} - \theta_k \pi_k \bar{\hat{y}}_{rc(k)} \right\},$$

and

$$\bar{\hat{y}}_{rc} = \frac{\sum_{k \in S_{c,hh}} d_k r_k \hat{y}_{r,k}}{\sum_{k \in S_{c,hh}} \theta_k r_k},$$

and

$$\hat{y}_{r,k} = \sum_{l \in S_{\text{ind},k}} \frac{d_{l|k} r_l}{\hat{p}_l} y_l, \tag{A.3}$$

and for the second linearized variable

$$u_{3k} = \frac{r_k}{d_k} \sum_{l \in S_{\text{ind},k}} \theta_l \left(1 - \frac{r_l}{\hat{p}_l}\right) \bar{y}_{rrd(l)}, \tag{A.4}$$

with

$$\bar{y}_{rrd} = \frac{\sum_{l \in S_{rd,\text{ind}}} d_{rl} r_l y_l}{\sum_{l \in S_{rd,\text{ind}}} \theta_l r_l}. \tag{A.5}$$

We now consider the calibrated estimator $\hat{Y}_{\text{cal,ind}}$ given in equation (2.29). The benchmark variance estimator is the same than given in equation (A.2) for $\hat{Y}_{rr,\text{ind}}$, by replacing the variable y_l with the estimated regression residuals of the variable of interest on the calibration variables, namely

$$e_l = y_l - \hat{B}_{rr,\text{ind}}^{\top} z_l \quad \text{with} \quad \hat{B}_{rr,\text{ind}} = \left(\sum_{l \in S_{rr,\text{ind}}} d_{rrl} z_l z_l^{\top}\right)^{-1} \sum_{l \in S_{rr,\text{ind}}} d_{rrl} z_l y_l. \tag{A.6}$$

B. SAS Program for one-stage sampling

In this section, we present the SAS macro developed to implement the proposed methodology for a sampling of households only (one-stage sampling). The parametrization of the SAS program for computing bootstrap weights is presented in Section B.1. For clarity, a small example is presented in Section B.2.

B.1 Program for computing bootstrap weights

The parameters related to the database are:

- `BASE`: library containing the SAS table with the list of sampled units. The default value is `BASE=WORK`.
- `ECHMEN`: SAS table containing the list of sampled units in the population. The non-respondents need also to be included in this table.

The parameters related to the bootstrap are:

- `ITBOOT`: number of bootstrap iterations. The default value is `ITBOOT=1000`.

The parameters related to the variables needed in the SAS table are:

- IDMEN: list of variables identifying the statistical unit. They need to be character variables.
- STMEN: list of variables of stratification used for the sample selection.
- DMEN: sampling weight.
- RMEN: response indicator (1 for a respondent, 0 for a non-respondent).
- DRMEN: sampling weight, corrected for non-response. The values are only needed for the respondents.
- DCMEN: calibrated weight. The values are only needed for the respondents.
- GHRMEN: list of variables identifying the response homogeneity groups.
- WGHRMEN: weighting used in the computation of the response probabilities inside RHGs.
  - With WGHRMEN=0, the response rates are not weighted. This is the default value.
  - With WGHRMEN=1, the response rates are weighted by the design weights.
- XMENQUANT: list of quantitative variables used in the calibration. The values are only needed for the respondents.
- XMENQUALI: list of qualitative variables used in the calibration. The values are only needed for the respondents.

The parameters related to the output are:

- SORT_MEN: SAS table containing the bootstrap sampling weights WB_D1,...,WB_D&ITBOOT for the whole sample.
- SORT_RMEN: SAS table containing the bootstrap weights WB_N1,...,WB_N&ITBOOT corrected for non-response, and the bootstrap weights WB_C1,...,WB_C&ITBOOT corrected for non-response and calibration, for the sub-sample of respondents.

B.2 A small example

We consider the example treated in Section 2.1.4. The sample is as follows:

```
data ech;
input idm$ stmen$ dmen rmen ghrmen$ drmen dcmen x0 x1;
cards;
A  1  4   1  aa  4.44   4.01   1  1
B  1  4   0  aa  .      .      .  .
C  1  4   0  bb  .      .      .  .
D  1  4   1  bb  5.54   4.87   1  0
E  1  16  1  bb  22.15  19.98  1  1
F  1  16  1  aa  17.78  15.63  1  0
G  1  16  0  bb  .      .      .  .
H  1  16  1  bb  22.15  19.98  1  1
I  1  16  1  bb  22.15  19.49  1  0
J  1  16  1  aa  17.78  16.03  1  1
;run;
```

We can obtain $B = 1{,}000$ bootstrap weights as follows. Since WGHRMEN=1, it is supposed that when unit non-response has been originally corrected by the method of RHGs, the response rates inside RHGs were weighted by the sampling weights.

```
%BOOTUP_1DEG(BASE=work,ECHMEN=ech,
      ITBOOT=1000,
      IDMEN=idm,STMEN=stmen,DMEN=dmen,
      RMEN=rmen,DRMEN=drmen,DCMEN=dcmen,GHRMEN=ghrmen,WGHRMEN=1,
      XMENQUANT=x0 x1,XMENQUALI=,
      SORT_MEN=ech_boot,SORT_RMEN=echr_boot);
```

C. SAS Program for two-stage sampling

In this section, we present the SAS macro developed to implement the proposed methodology for a sampling of households and a sub-sampling of individuals (two-stage sampling). The parametrization of the SAS program for computing bootstrap weights is presented in Section C.1. For clarity, a small example is presented in Section C.2.

C.1 Program for computing bootstrap weights

The SAS macro %BOOTUP_2DEG enables to compute bootstrap weights for a household survey with sub-sampling of individuals, and to account for correction of unit non-response via Response Homogeneity groups, and for the calibration of weights, both for households and individuals.

The parameters with equality sign are mandatory. All identifying variables must be of character type.

The parameters related to the database are:

- BASE: library containing the SAS tables ECHMEN and ECHIND. The default value is BASE=WORK.
- BASESOR: library containing the output. The default value is BASESOR=WORK.
- ECHMEN=: SAS table containing the list of sampled households in the population. The household non-respondents need also to be included in this table.
- ECHIND=: SAS table containing the list of sampled individuals inside all the responding households. The individual non-respondents need also to be included in this table.

The parameters related to the bootstrap are:

- ITBOOT: number of bootstrap iterations. The default value is ITBOOT=1000.

The parameters related to the variables needed in the household SAS table ECHMEN are:

- IDMEN=: list of variables identifying the household. This variable is required in both ECHMEN and ECHIND.
- STMEN: list of variables of stratification used for the sample selection.

- DMEN: sampling weight of the household.
- RMEN: response indicator of the household (1 for a respondent, 0 for a non-respondent).
- DRMEN: sampling weight of the household, corrected for non-response. The values are only needed for the respondents.
- DCMEN: calibrated weight. The values are only needed for the respondents.
- GHRMEN: list of variables identifying the response homogeneity groups for households.
- WGHRMEN: weighting used in the computation of the response probabilities inside RHGs:
  - With WGHRMEN=0, the response rates are not weighted. This is the default value.
  - With WGHRMEN=1, the response rates are weighted by the design weights DMEN.
- XMENQUANT: list of quantitative variables used in the calibration. The values are only needed for the respondents.
- XMENQUALI: list of qualitative variables used in the calibration. The values are only needed for the respondents.

The parameters related to the variables needed in the individual SAS table ECHIND are:

- ID_IND=: list of variables identifying the individual (character variable).
- R_IND: response indicator of the individual (1 for a respondent, 0 for a non-respondent).
- DR_IND: weight of the individual, corrected for both household and individual unit non-response. The values are only needed for the respondents.
- DC_IND: calibrated weight. The values are only needed for the respondents.
- PIKSACI=: conditional inclusion probability of the individual inside its household.
- GHR_IND: list of variables identifying the response homogeneity groups.
- WGHR_IND: weighting used in the computation of the response probabilities inside RHGs:
  - With WGHR_IND=0, the response rates are not weighted. This is the default value.
  - With WGHR_IND=1, the response rates are weighted by the design weights of individuals.
  - With WGHR_IND=2, the response rates are weighted by the weights of individuals, adjusted for household unit non-response.
- XINDQUANT: list of quantitative variables used in the calibration. The values are only needed for the respondents.
- XINDQUALI: list of qualitative variables used in the calibration. The values are only needed for the respondents.

The parameters related to the output are:

- SORT_MEN: SAS table containing all the sampled households, and the bootstrap sampling weights WB_D1,...,WB_D&ITBOOT for the whole sample.
- SORT_RMEN: SAS table containing all the responding households, and the bootstrap weights
  - WB_N1,...,WB_N&ITBOOT corrected for non-response,
  - WB_C1,...,WB_C&ITBOOT corrected for non-response and calibration.
- SORT_RIND: SAS table containing all the responding individuals inside the responding households, and the bootstrap weights

- `WB_N1,...,WB_N&ITBOOT` corrected for household non-response,
- `WB_NN1,...,WB_NN&ITBOOT` corrected for both household non-response and individual non-response,
- `WB_C1,...,WB_C&ITBOOT` corrected for non-response and calibration.

C.2 A small example

We consider the example treated in Section 2.2.4. The sample of households and the sample of individuals are as follows:

```
data echmen;
input idm$ stmen$ dmen rmen ghrmen$ drmen dcmen x0 x1;
cards;
A  1  4   1  aa  4.44   4.01   1  1
B  1  4   0  aa  .      .      .  .
C  1  4   0  bb  .      .      .  .
D  1  4   1  bb  5.54   4.87   1  0
E  1  16  1  bb  22.15  19.98  1  1
F  1  16  1  aa  17.78  15.63  1  0
G  1  16  0  bb  .      .      .  .
H  1  16  1  bb  22.15  19.98  1  1
I  1  16  1  bb  22.15  19.49  1  0
J  1  16  1  aa  17.78  16.03  1  1
;run;
```

```
data echind;
input idm$ idi$ piksaci dr1_ind rind ghrind$ phat_ind dr2_ind xi1 xi2 dc_ind;
cards;
A   i01   0.34   13.06   1   g1   0.75   17.41   1   3   19.61
D   i04   1.00   5.54    0   g2   0.33   .       .   .   .
E   i06   0.34   65.15   1   g1   0.75   86.86   1   2   53.93
F   i08   0.33   53.88   1   g1   0.75   71.84   1   3   78.43
H   i11   0.50   44.30   0   g1   0.75   .       .   .   .
I   i13   1.00   22.15   1   g2   0.33   67.12   1   1   48.04
J   i14   1.00   17.78   0   g2   0.33   .       .   .   .
;run;
```

We can obtain $B = 1{,}000$ bootstrap weights as follows. Since `WGHRMEN=1`, it is supposed that when unit non-response of households has been originally corrected by the method of RHGs, the response rates inside RHGs were weighted by the sampling weights. Since `WGHR_IND=0`, it is supposed that when unit non-response of individuals has been originally corrected by the method of RHGs, the response rates inside RHGs were unweighted.

```
%bootup_2deg(BASE=work,BASESOR=work,ECHMEN=echmen,ECHIND=echind,
     ITBOOT=1000,
     IDMEN=idm,STMEN=stmen,DMEN=dmen,RMEN=rmen,DRMEN=drmen,GHRMEN=ghrm
     en,WGHRMEN=0,
     DCMEN=dcmen,XMENQUANT=x0 x1,XMENQUALI=,
     ID_IND=idi,R_IND=rind,DR_IND=dr2_ind,PIKSACI=piksaci,GHR_IND=ghri
     nd,WGHR_IND=0,
     DC_IND=dc_ind,XINDQUANT=xi1 xi2,XINDQUALI=,
     SORT_MEN=sort_men,SORT_RMEN=sort_rmen,
     SORT_RIND=sort_rind);
```

References

Beaumont, J.-F. (2005). Calibrated imputation in surveys under a quasi-model-assisted approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(3), 445-458.

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.

Brick, J.M. (2013). Unit non-response and weighted adjustments: A critical review. *Journal of Official Statistics*, 29(3), 329-353.

Chauvet, G. (2007). *Méthodes de Bootstrap en population finie*. PhD thesis, University of Rennes 2.

Chauvet, G. (2015). Coupling methods for multistage sampling. *The Annals of Statistics*, 43(6), 2484-2506.

Chauvet, G., and Vallée, A.-A. (2018). Inference for two-stage sampling designs with application to a panel for urban policy. *arXiv preprint arXiv:1808.09758*.

Davison, A.C., and Hinkley, D.V. (1997). Bootstrap methods and their application, volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*.

Davison, A.C., and Sardy, S. (2007). Resamping variance estimation in surveys with missing data. *Journal of Official Statistics*, 23(3), 371-386.

Deville, J.-C. (1999). Variance estimation for complex statistics and estimators: Linearization and residual techniques. *Survey Methodology*, 25, 2, 193-203. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1999002/article/4882-eng.pdf.

Girard, C. (2009). The Rao-Wu rescaling bootstrap: from theory to practice. In *Proceedings of the Federal Committee on Statistical Methodology Research Conference*, pages 2-4. Citeseer.

Haziza, D., and Beaumont, J.-F. (2007). On the construction of imputation classes in surveys. *International Statistical Review*, 75(1), 25-43.

Haziza, D., and Beaumont, J.-F. (2017). Construction of weights in surveys: A review. *Statistical Science*, 32(2), 206-226.

Juillard, H., and Chauvet, G. (2018). Variance estimation under monotone non-response for a panel survey. *Survey Methodology*, 44, 2, 269-289. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2018002/article/54952-eng.pdf.

Kim, J.K., and Kim, J.J. (2007). Nonresponse weighting adjustment using estimated response probability. *The Canadian Journal of Statistics/La revue canadienne de statistique*, 35(4), 501-514.

Kim, J.K., Navarro, A. and Fuller, W.A. (2006). Replication variance estimation for two-phase stratified sampling. *Journal of the American Statistical Association*, 101(473), 312-320.

Kott, P.S. (2012). Why one should incorporate the design weights when adjusting for unit nonresponse using response homogeneity groups. *Survey Methodology*, 38, 1, 95-99. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2012001/article/11689-eng.pdf.

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

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.

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.

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18, 2, 209-217. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1992002/article/14486-eng.pdf.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer Series in Statistics.

Shao, J. (1994). *L*-Statistics in complex survey problems. *The Annals of Statistics*, 22(2), 946-967.

Shao, J., and Rao, J. (1993). Standard errors for low income proportions estimated from stratified multi-stage samples. *Sankhyā: The Indian Journal of Statistics, Series B*, 393-414.

Shao, J., and Tu, D.S. (1995). *The Jackknife and Bootstrap*. Springer Series in Statistics.

Tillé, Y. (2011). *Sampling Algorithms*. Springer.

Yeo, D., Mantel, H. and Liu, T.-P. (1999). Bootstrap variance estimation for the national population health survey. In *Proceedings of the Survey Research Methods Section*, American Statistical Association. Citeseer.