



HAL
open science

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

Pascal Bessonneau, Gwennaëlle Brillhaut, Guillaume Chauvet, Cédric Garcia

► **To cite this version:**

Pascal Bessonneau, Gwennaëlle Brillhaut, Guillaume Chauvet, Cédric Garcia. With-replacement bootstrap variance estimation for household surveys: principles, examples and implementation. 2020. hal-02514797

HAL Id: hal-02514797

<https://hal.science/hal-02514797>

Preprint submitted on 22 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

With-replacement bootstrap variance estimation for household surveys

Principles, examples and implementation

Pascal Bessonneau¹, Gwennaëlle Brilhaut¹, Guillaume Chauvet², and Cédric Garcia³

¹Ined

²Ensaï (Irmair), Campus de Ker Lann, Bruz - France

³Université Gustave Eiffel

March 22, 2020

Abstract

Variance estimation is a challenging problem in surveys, for which bootstrap is an important tool. In this work, we are interested in the use of the with-replacement bootstrap for 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 also explain how the with-replacement bootstrap can be used for sampling, treatment of unit non-response and calibration. 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.

1 Introduction

Variance estimation is a challenging problem in surveys. In particular, the final weights used for estimation involve several statistical treatments (e.g., correction of unit non-response and calibration) which need to be accounted for. Bootstrap is an interesting tool, leading to the creation of so-called bootstrap weights released with the survey data set. These random 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 explicitly written.

There is an extensive literature on bootstrap in survey sampling, see for example Shao and Tu (1995, Chapter 6), Davison and Hinkley (1997, Section 3.7), Davison and Sardy (2007), Chauvet (2007) and Mashreghi et al. (2016) for detailed reviews. One of these techniques is the so-called rescaled bootstrap proposed by Rao and Wu (1988), which may be summarized as follows. First, initial bootstrap weights are obtained by using the classical i.i.d. bootstrap on the first-stage sample. 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.

Alternatively, the classical i.i.d. bootstrap may be applied on the first-stage sample, without further rescaling. This is sometimes referred to as the bootstrap of Primary Sampling Units (PSUs) or with-replacement bootstrap. It leads to conservative variance estimation, in the sense that the sampling variance is overestimated if the first-stage sampling design is more efficient than with-replacement sampling. This is the approach that we pursue here, our paper aiming at being 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 inside 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 the SAS program to perform bootstrap for the one-stage case, and apply it for illustration on a sample of households from the French Panel for Urban Policy. We present in Section 6 the SAS program to perform bootstrap for the two-stage case, and apply it for illustration on a sample of individuals from the French Panel for Urban Policy. The SAS programs are available as Supplementary material.

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_{hou} 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

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

2.1.1 Sampling design

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

$$d_k = \frac{1}{\pi_k}. \quad (2.2)$$

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

$$\hat{Y}_{hou} = \sum_{k \in S_{hou}} d_k y_k. \quad (2.3)$$

We consider as a benchmark variance estimator

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

with n_h the size of the sample S_{hou}^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 et al., 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 if the sampling rates inside strata are so, 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_{hou} is prone to unit non-response, which leads to the observation of a sub-sample of respondents $S_{r,hou}$ 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_{hou} may be partitioned into C RHGs denoted as $S_{1,hou}, \dots, S_{C,hou}$ such that the response probability p_k is constant inside a RHG.

We let p_c denote the common response probability inside the RHG $S_{c,hou}$. It is estimated by

$$\hat{p}_c = \frac{\sum_{k \in S_{c,hou}} \omega_k r_k}{\sum_{k \in S_{c,hou}} \omega_k}, \quad (2.5)$$

with ω_k some weight attached to the household k . The choice $\omega_k = 1$ leads to estimating p_c by the unweighted response rate inside the RHG. The choice $\omega_k = d_k$ 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

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

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

$$\hat{Y}_{r,hou} = \sum_{k \in S_{r,hou}} d_{rk} y_k. \quad (2.7)$$

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

$$\begin{aligned} v_{mult}(\hat{Y}_{r,hou}) &= \sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{k \in S_{hou}^h} \left(d_k u_{1k} - \frac{1}{n_h} \sum_{k' \in S_{hou}^h} d_{k'} u_{1k'} \right)^2, \quad (2.8) \\ \text{with } u_{1k} &= \omega_k \pi_k \bar{y}_{rc(k)} + \frac{r_k}{\hat{p}_{c(k)}} \{ y_k - \omega_k \pi_k \bar{y}_{rc(k)} \}, \\ \text{and } \bar{y}_{rc} &= \frac{\sum_{k \in S_{c,hou}} d_k r_k y_k}{\sum_{k \in S_{c,hou}} \omega_k r_k}. \end{aligned}$$

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

2.1.3 Calibration

Lastly, the weights adjusted for non-response are calibrated on some 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_{hou} the total on the population U_{hou} . For the sample $S_{r,hou}$, this leads to the calibrated weights

$$\begin{aligned} w_k &= d_{rk} \left(1 + x_k^\top \lambda_{hou} \right), \\ \text{with } \lambda_{hou} &= \left(\sum_{k \in S_{r,hou}} d_{rk} x_k x_k^\top \right)^{-1} \left(X_{hou} - \hat{X}_{r,hou} \right), \quad (2.9) \end{aligned}$$

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

$$\hat{Y}_{cal,hou} = \sum_{k \in S_{r,hou}} w_k y_k. \quad (2.10)$$

The sampling and estimation steps are summarized in Figure 1.

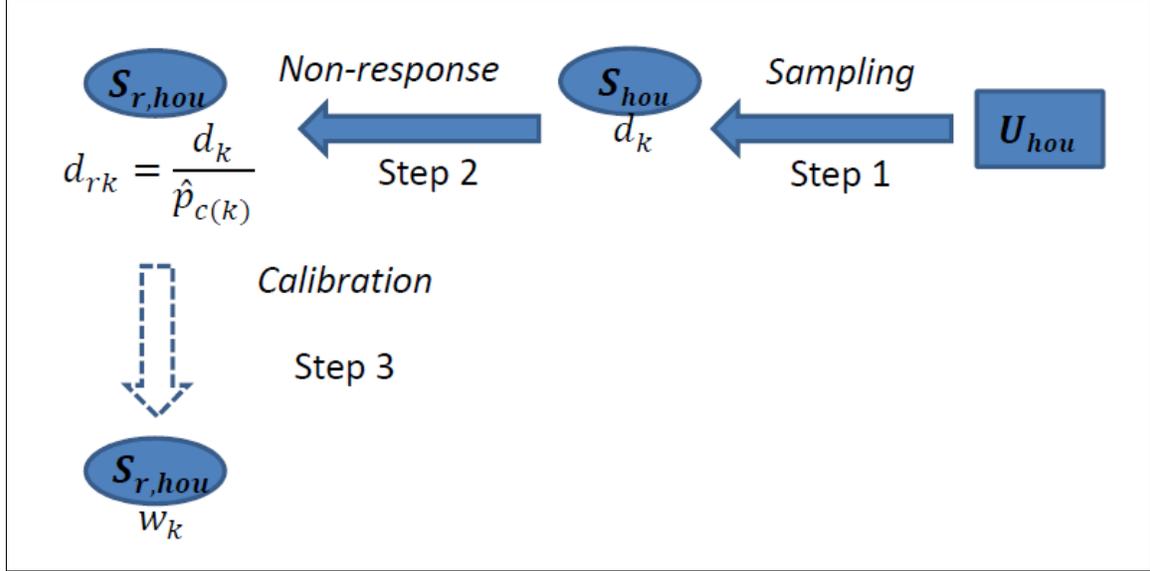


Figure 1: Sampling and estimation steps for a household sample

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

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

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

$$\text{and } \bar{e}_{rc} = \frac{\sum_{k \in S_{c,hou}} d_k r_k e_k}{\sum_{k \in S_{c,hou}} \omega_k r_k},$$

where we let

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

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

2.1.4 Computation of household weights on an example

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

households is selected.

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

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

resulting in the design weights

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

Among the 10 selected households, 7 only are surveyed due to non-response. It is accounted for by using the method of Response Homogeneity Groups (RHGs), and we suppose that there are two RHGs: 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 ($\omega_k = d_k$). This leads to

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

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

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

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

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

The sampling and estimation steps are summarized in Figure 2.

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_{hou} 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

$$Y_{ind} = \sum_{l \in U_{ind}} y_l. \quad (2.18)$$

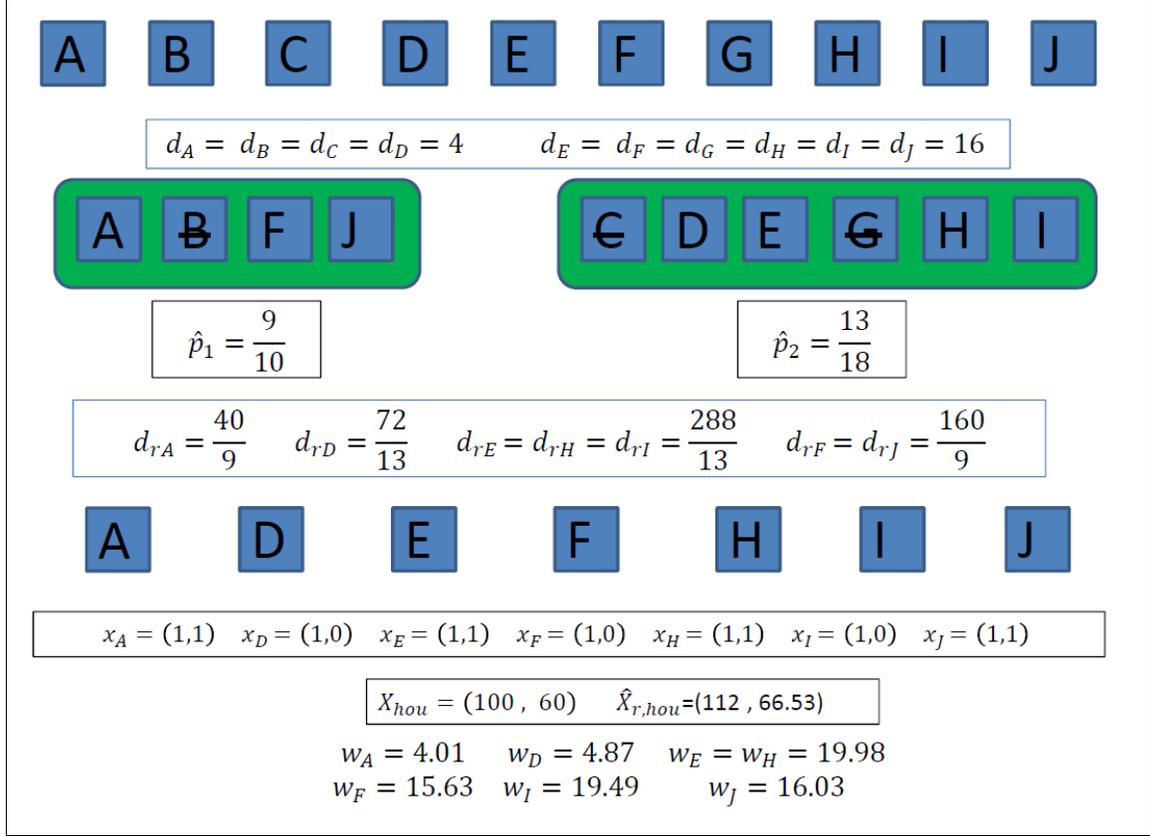


Figure 2: Estimation steps for the weighting of households

2.2.1 Sampling design

Inside any sampled household $k \in S_{hou}$, a subsample $S_{ind,k}$ 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

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

and the non-conditional design weight is

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

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

$$\hat{Y}_{ind} = \sum_{k \in S_{hou}} d_k \sum_{l \in S_{ind,k}} d_{l|k} y_l = \sum_{k \in S_{ind}} d_l y_l. \quad (2.21)$$

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

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

2.2.2 Treatment of non-response

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

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

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

$$S_{r,ind} = \bigcup_{k \in S_{r,hou}} S_{ind,k} \quad (2.24)$$

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

The individuals in $S_{r,ind}$ 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 $S_{rr,ind}$ 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 $S_{r,ind}$ may be partitioned into D RHGs denoted as $S_{r1,ind}, \dots, S_{rD,ind}$ such that the response probability p_l is constant inside a RHG.

We let p_d denote the common response probability inside the RHG $S_{rd,ind}$. It is estimated by

$$\hat{p}_d = \frac{\sum_{l \in S_{rd,ind}} \omega_l r_l}{\sum_{l \in S_{rd,ind}} \omega_l}, \quad (2.25)$$

with ω_l some weight attached to the individual l . The choice $\omega_l = 1$ leads to estimating p_d by the unweighted response rate inside the RHG. The choice $\omega_l = d_l$ leads to estimating p_d by the response rate inside the RHG, weighted by the individual sampling weights. The choice $\omega_l = d_{rl}$ 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

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

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

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

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 $S_{rr,ind}$, this leads to the calibrated weights

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

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

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

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

The sampling and estimation steps are summarized in Figure 3.

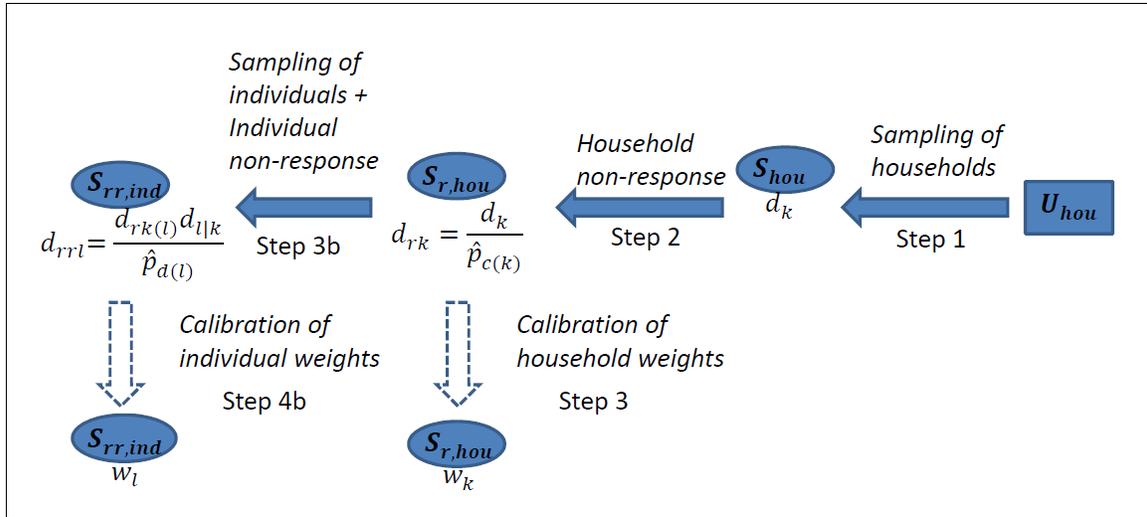


Figure 3: Sampling and estimation steps for a household sample with sub-sampling of individuals

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 $S_{r,hou} = \{A, D, E, F, H, I, J\}$. The set of all individuals inside the responding households is as follows (say):

$$\underbrace{(i_1, i_2, i_3)}_A \quad \underbrace{(i_4)}_D \quad \underbrace{(i_5, i_6)}_E \quad \underbrace{(i_7, i_8, i_9)}_F \quad \underbrace{(i_{10}, i_{11})}_H \quad \underbrace{(i_{12})}_I \quad \underbrace{(i_{13})}_J. \quad (2.30)$$

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

$$S_{r,ind} = \{i_1, i_4, i_6, i_8, i_{11}, i_{12}, i_{13}\}. \quad (2.31)$$

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

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

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 ($\omega_l = 1$). This leads to

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

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

$$d_{rr1} = \frac{160}{9}, \quad d_{rr6} = \frac{2304}{39}, \quad d_{rr8} = \frac{640}{9}, \quad d_{rr12} = \frac{864}{13}. \quad (2.34)$$

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

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

The sampling and estimation steps are summarized in Figure 4.

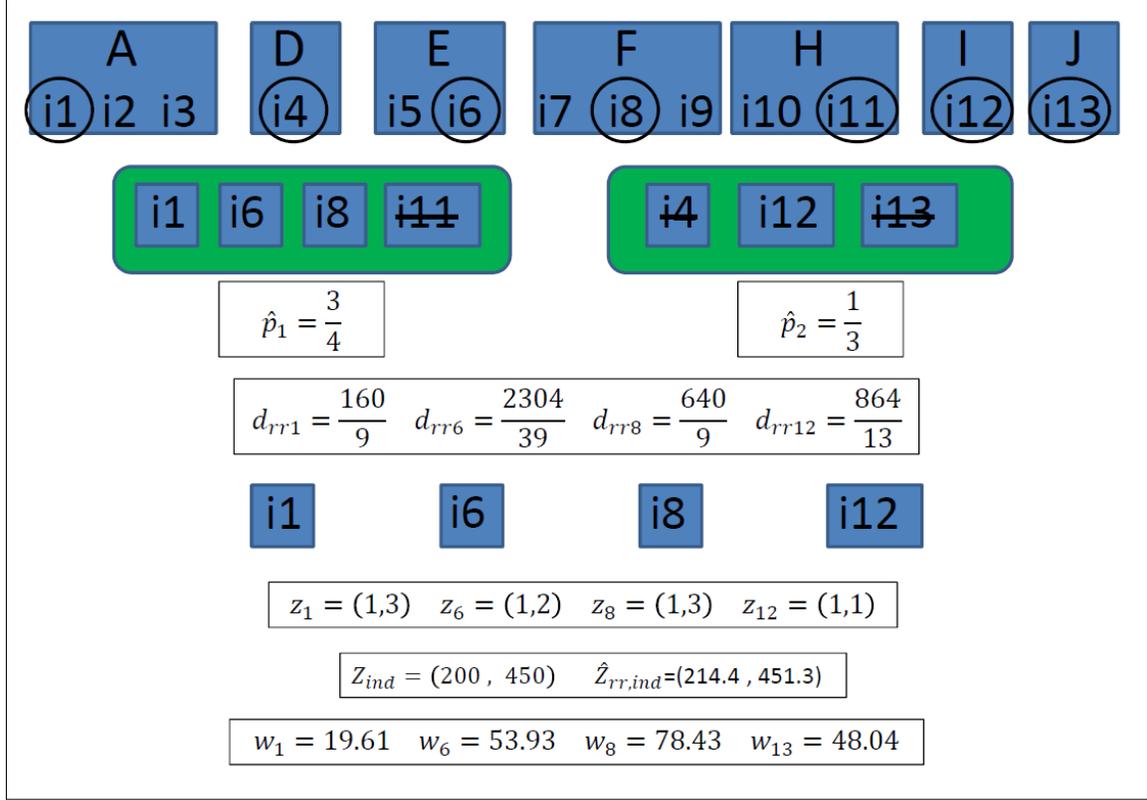


Figure 4: Estimation steps for the weighting of individuals

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_{hou}^h selected in the stratum U_{hou}^h a with-replacement resample S_{hou*}^h of $n_h - 1$ units, with equal probabilities. For any $k \in S_{hou}^h$, we take

$$W_k = \frac{n_h}{n_h - 1} \times \text{Number of times the household } k \text{ is selected in the resample } S_{hou*}^h. \quad (3.1)$$

The computation of the bootstrap weights accounting for the sampling design, for unit non-response and for the calibration is described in Algorithm 1. The steps refer to Figure 1.

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_{hou} 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., 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 . If we do not bootstrap these response probabilities and if we plug in equation (3.5) the original estimated probabilities \hat{p}_c , 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).

In the bootstrap process presented in Algorithm 1, the calibration step is performed on the true population total X_{hou} . Following the bootstrap principle which states that the sample S_{hou} is to the bootstrap sample S_{hou*} what the population U_{hou} is to the sample S_{hou} , it could seem more intuitive to rather calibrate on the estimated totals \hat{X}_{hou} obtained by plugging x_k into equation (2.3). Both approaches seem valid for variance estimation, but the calibration variables x_k may be prone to non-response on the sample S_{hou} , making the estimator \hat{X}_{hou} not possible to compute, while the total X_{hou} 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 $n = 9$ 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

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

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

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

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 $k \in S_{hou}$, the bootstrap sampling weight

$$d_{k*} = W_k d_k. \quad (3.2)$$

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

$$\hat{Y}_{hou*} = \sum_{k \in S_{hou}} d_{k*} y_k. \quad (3.3)$$

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

$$\hat{p}_{c*} = \frac{\sum_{k \in S_{c,hou}} W_k \omega_k r_k}{\sum_{k \in S_{c,hou}} W_k \omega_k}, \quad (3.4)$$

and we compute the bootstrap weights corrected for non-response

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

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

$$\hat{Y}_{r,hou*} = \sum_{k \in S_{r,hou}} d_{rk*} y_k. \quad (3.6)$$

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

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

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

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

$$\hat{Y}_{cal,hou*} = \sum_{k \in S_{r,hou}} w_{k*} y_k. \quad (3.8)$$

and to the bootstrap weights corrected for non-response

$$d_{rA^*} = \frac{40}{3} \quad d_{rD^*} = \frac{280}{39} \quad d_{rE^*} = d_{rH^*} = d_{rI^*} = \frac{1120}{39}. \quad (3.11)$$

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

$$w_{A^*} = 11.30 \quad w_{D^*} = 8.00 \quad w_{E^*} = w_{H^*} = 24.35 \quad w_{I^*} = 32.00 \quad (3.12)$$

The computation of the bootstrap weights is summarized in Figure 5.

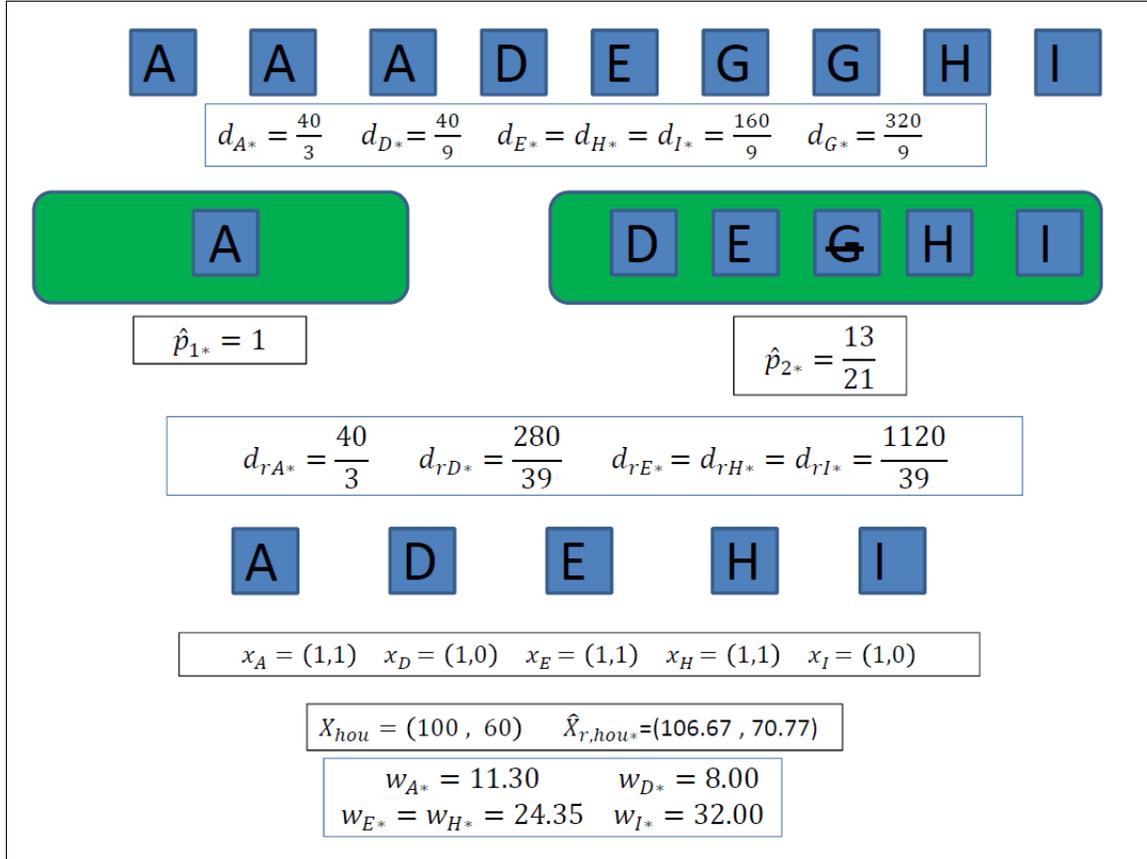


Figure 5: Computation of bootstrap household weights

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 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.

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 equations (2.29) and (2.30), this results in the bootstrap sample of individuals

$$S_{r,ind*} = \{i_1, i_4, i_6, i_{11}, i_{12}\}. \quad (3.19)$$

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

$$d_{r1*} = 40 \quad d_{r4*} = \frac{280}{39} \quad d_{r6*} = \frac{2240}{39} \quad d_{r11*} = \frac{2240}{39} \quad d_{r12*} = \frac{1120}{39}. \quad (3.20)$$

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 original bootstrap weights W_k . 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 ($W_A = 3$) 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 ($W_E = W_H = 1$). The computation is similar for the second RHG, and leads to

$$\begin{aligned} \hat{p}_{1*} &= \frac{W_A + W_E}{W_A + W_E + W_H} = \frac{4}{5}, \\ \hat{p}_{2*} &= \frac{W_I}{W_D + W_I} = \frac{1}{2}, \end{aligned} \quad (3.21)$$

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

$$d_{rr1*} = 50 \quad d_{r6*} = \frac{5600}{39} \quad d_{r12*} = \frac{2240}{39}. \quad (3.22)$$

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

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

The computation of bootstrap individual weights is summarized in Figure 6.

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

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

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

$$\hat{p}_{d*} = \frac{\sum_{l \in S_{rd,ind}} W_{k(l)} \omega_l r_l}{\sum_{l \in S_{rd,ind}} W_{k(l)} \omega_l}. \quad (3.14)$$

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

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

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

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

- 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

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

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

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

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

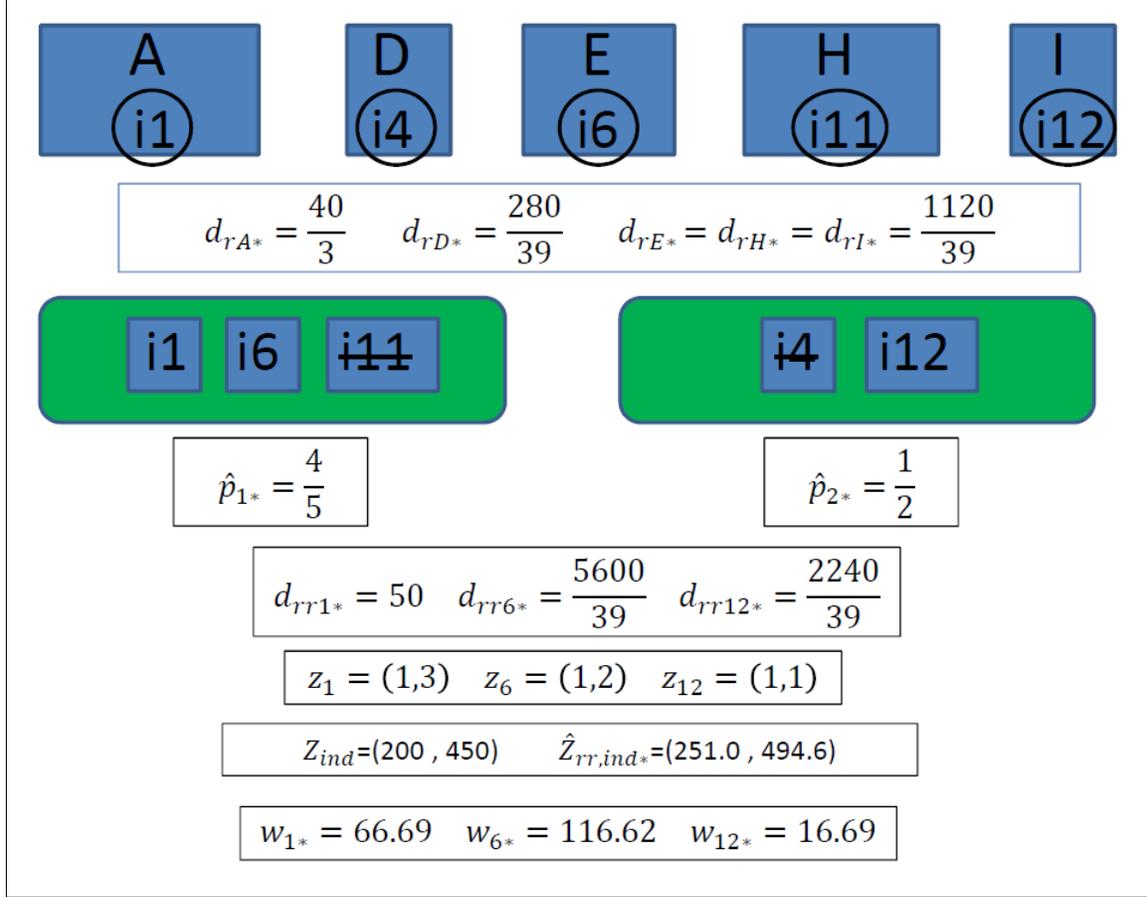


Figure 6: Computation of bootstrap household weights

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_{hou} . 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 $\theta_{hou} = f(Y_{hou})$ with $f : \mathbb{R}^q \rightarrow \mathbb{R}$ a known, smooth function. In case of full response, the substitution estimator of θ_{hou} is

$$\hat{\theta}_{hou} = f(\hat{Y}_{hou}), \quad (3.24)$$

see for example Deville (1999). In case of unit non-response at the household level, the estimator

of θ_{hou} corrected for unit non-response is

$$\hat{\theta}_{r,hou} = f(\hat{Y}_{r,hou}), \quad (3.25)$$

and the calibrated estimator of θ_{hou} is

$$\hat{\theta}_{cal,hou} = f(\hat{Y}_{cal,hou}). \quad (3.26)$$

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. A rigorous proof for the consistency of this bootstrap variance estimator is beyond the scope of this paper, but a heuristic justification is given in Appendix B.

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}_{hou*}^b , $\hat{Y}_{r,hou*}^b$ and $\hat{Y}_{cal,hou*}^b$ for the bootstrap estimators of totals computed on the b -th sample. Also, let us denote $\hat{\theta}_{hou*}^b$, $\hat{\theta}_{r,hou*}^b$ and $\hat{\theta}_{cal,hou*}^b$ for the associated bootstrap estimators of θ_{hou} .
2. The Bootstrap variance estimator for $\hat{\theta}_{hou}$ is

$$\hat{V}_{boot}(\hat{\theta}_{hou}) = \frac{1}{B-1} \sum_{b=1}^B \left\{ \hat{\theta}_{hou*}^b - \frac{1}{B} \sum_{b'=1}^B \hat{\theta}_{hou*}^{b'} \right\}^2, \quad (3.27)$$

and similarly for $\hat{\theta}_{r,hou}$ and $\hat{\theta}_{cal,hou}$.

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 $\hat{\theta}_{hou}$ is

$$IC_{nor}(\theta_{hou}) = [\hat{\theta}_{hou} \pm u_{1-\alpha} \{ \hat{V}_{boot}(\hat{\theta}_{hou}) \}^{0.5}], \quad (3.28)$$

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 $\hat{\theta}_{hou}$, the percentile confidence interval

is obtained by using the distribution of $\hat{\theta}_{hou*}$ as an approximation of the distribution of $\hat{\theta}_{hou}$. It makes use of the ordered bootstrap estimates $\hat{\theta}_{hou*}^{(1)}, \dots, \hat{\theta}_{hou*}^{(B)}$ to form the confidence interval

$$IC_{per}(\theta_{hou}) = \left[\hat{\theta}_{hou*}^{(L)}, \hat{\theta}_{hou*}^{(U)} \right], \quad (3.29)$$

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 $(\hat{\theta}_{hou*} - \hat{\theta}_{hou})$ as an approximation of the distribution of $(\hat{\theta}_{hou} - \theta_{hou})$. It leads to the confidence interval

$$IC_{rev}(\theta_{hou}) = \left[2\hat{\theta}_{hou} - \hat{\theta}_{hou*}^{(U)}, 2\hat{\theta}_{hou} - \hat{\theta}_{hou*}^{(L)} \right]. \quad (3.30)$$

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.

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_{hou} containing $N_{hou} = 100,000$ households, with four auxiliary variables x_1, \dots, 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, \dots, y_3 according to the following models

$$\begin{aligned} y_{1k} &= 10 + x_{1k} + x_{2k} + \sigma_\epsilon \epsilon_k, \\ y_{2k} &= 10 + x_{1k} + x_{3k} + \sigma_\epsilon \epsilon_k, \\ y_{3k} &= 10 + x_{3k} + x_{4k} + \sigma_\epsilon \epsilon_k, \end{aligned} \quad (4.1)$$

where ϵ_k is generated according to a standard normal distribution. We set $\sigma_\epsilon = 10$, 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_{hou} 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_{ind} , we generate four auxiliary variables z_1, \dots, 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 $\sigma_\eta = 0.4$, 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_{hou} , we select a sample S_{hou} of $n_{hou} = 1,000$ 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 $S_{r,hou}$ of responding households. The estimated response probabilities \hat{p}_c are obtained from equation (2.5), with equal weight $\omega_k = 1$. Inside each $k \in S_{r,hou}$, one Kish individual is randomly selected with equal probabilities, which results in the sample of individuals $S_{r,ind}$. Inside $S_{r,ind}$, the non-response is generated according to the RHG individual model, resulting in a sample $S_{rr,ind}$ of responding individuals. The estimated response probabilities \hat{p}_d are obtained from equation (2.25), in three possible ways: equal weights $\omega_l = 1$, sampling weights $\omega_l = d_l$, or individuals weights corrected for the household non-response $\omega_l = d_{rl}$.

The sampling and non-response steps are repeated $C = 1,000$ times. On each sample S_{hou} , we compute the full-response estimator given in (2.3), and on each sample $S_{r,hou}$, we compute the estimator adjusted for non-response $\hat{Y}_{r,hou}$ given in (2.7) and the estimator $\hat{Y}_{cal,hou}$ given in (2.10) with the set of calibration variables $x_k = (1, x_{1k}, x_{2k})^\top$. On each sample $S_{rr,ind}$, we compute the estimator adjusted for non-response $\hat{Y}_{rr,ind}$ given in (2.27) and the estimator $\hat{Y}_{cal,ind}$ 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

coefficient of variation

$$CV(\hat{Y}) = 100 \times \frac{\sqrt{MSE(\hat{Y})}}{Y}, \quad (4.3)$$

with $MSE(\hat{Y})$ 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

$$RB\{v(\hat{Y})\} = 100 \times \frac{C^{-1} \sum_{c=1}^C v_c(\hat{Y}_c) - MSE(\hat{Y})}{MSE(\hat{Y})}, \quad (4.4)$$

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

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

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 1 for the estimation on the population of households. The coefficient of variation of the calibrated estimator $\hat{Y}_{cal,hou}$ 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}_{hou} , but almost no bias for the reweighted estimators $\hat{Y}_{r,hou}$ and $\hat{Y}_{cal,hou}$. 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 2. We observe that the relative bias of the bootstrap variance estimator is very small in all cases. The choice of the weights ω_k used in the estimation of the response probabilities seem to have no effect on the coefficient of variation of the estimators, but the use of the weights $\omega_l = d_{rl}$ adjusted for household non-response yields slightly more stable variance estimators for $\hat{Y}_{rr,ind}$. The coverage rates are approximately respected in all cases.

Table 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

		CV	RB	RS	Percentile bootstrap			Basic bootstrap			Normality-based		
					L	U	L+U	L	U	L+U	L	U	L+U
\hat{Y}_{hou}	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
$\hat{Y}_{r,hou}$	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
$\hat{Y}_{cal,hou}$	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

5 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 5.1. For clarity, a small example is presented in Section 5.2. An illustration for the Panel for Urban Policy is given in Section 5.3.

5.1 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.

Table 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

		CV	RB	RS	Percentile bootstrap			Basic bootstrap			Normality-based		
					L	U	L+U	L	U	L+U	L	U	L+U
Equal weights $\omega_l = 1$													
$\hat{Y}_{rr,ind}$	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
$\hat{Y}_{cal,ind}$	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 $\omega_l = d_l$													
$\hat{Y}_{rr,ind}$	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
$\hat{Y}_{cal,ind}$	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 $\omega_l = d_{rl}$													
$\hat{Y}_{rr,ind}$	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
$\hat{Y}_{cal,ind}$	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

- 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.

5.2 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 dcmn 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);

```

5.3 Application to the French panel for urban policy

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

$$\theta_{g,hou} = \frac{\sum_{k \in U_{hou}} 1(y_k = g)}{N_{hou}}, \quad (5.1)$$

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

$$\hat{\theta}_{gr,hou} = \frac{\sum_{k \in S_{r,hou}} d_{rk} 1(y_k = g)}{\sum_{k \in S_{r,hou}} d_{rk}}, \quad (5.2)$$

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

$$\hat{\theta}_{gcal,hou} = \frac{\sum_{k \in S_{r,hou}} w_k 1(y_k = g)}{\sum_{k \in S_{r,hou}} w_k}, \quad (5.3)$$

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 3. The three confidence intervals are very similar in all cases.

6 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 6.1. For clarity, a small example is presented in Section 6.2. An illustration for the Panel for Urban Policy is given in Section 6.3.

6.1 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:

Table 3: 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	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	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	Yes	No	Yes	No	No opinion	Yes
Estim.	0.471	0.495	0.034	0.470	0.496	0.034	0.470	0.496	0.470	0.496	0.034	0.470
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]	[0.443,0.496]	[0.469,0.523]	[0.443,0.496]	[0.469,0.523]	[0.024,0.045]	[0.443,0.496]
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]	[0.440,0.495]	[0.470,0.524]	[0.440,0.495]	[0.470,0.524]	[0.025,0.045]	[0.440,0.495]
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]	[0.444,0.499]	[0.468,0.522]	[0.444,0.499]	[0.468,0.522]	[0.024,0.044]	[0.444,0.499]
	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	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]				

- **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.
- **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 household 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.
- PIKSACT=: 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.

6.2 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 x0 x1;
cards;
A 1 4 1 aa 4.44 1 1
B 1 4 0 aa . . .
C 1 4 0 bb . . .
D 1 4 1 bb 5.54 1 0
E 1 16 1 bb 22.15 1 1
F 1 16 1 aa 17.78 1 0
G 1 16 0 bb . . .
H 1 16 1 bb 22.15 1 1
I 1 16 1 bb 22.15 1 0
J 1 16 1 aa 17.78 1 1
;run;

data echind;
input idm$ idi$ piksaci drmen dr1_ind rind ghrind$ phat_ind dr2_ind xi1 xi2;
cards;
A i01 0.34 04.44 13.06 1 g1 0.67 19.59 1 3
A i02 0.33 04.44 13.45 0 g2 . . .
A i03 0.33 04.44 13.45 1 g1 0.67 20.18 1 3
D i04 1.00 05.54 05.54 0 g2 . . .
E i05 0.33 22.15 67.12 1 g2 0.50 134.24 1 3
E i06 0.34 22.15 65.15 0 g2 . . .
E i07 0.33 22.15 67.12 1 g1 0.67 100.68 1 3
F i08 0.33 17.78 53.88 1 g2 0.50 107.76 1 3
F i09 0.33 17.78 53.88 0 g2 . . .
F i10 0.34 17.78 52.29 0 g1 . . .
H i11 0.50 22.15 44.30 1 g2 0.50 88.60 1 2
H i12 0.50 22.15 44.30 1 g1 0.67 66.45 1 2
I i13 1.00 22.17 22.17 0 g1 . . .
J i14 1.00 17.78 17.78 1 g2 0.50 35.56 1 1
;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,echmen=echmen,echind=echind,
    itboot=1000,
    idmen=idm,stmen=stmen,dmen=dmen,drmen=drmen,rmen=rmen,ghrmen=ghrmen,wghrmen=1,
    xmenquant=x0,xmenquali=x1,
    id_ind=idi,r_ind=rind,dr_ind=dr2_ind,piksaci=piksaci,ghr_ind=ghrind,wghr_ind=0,
    xindquant=xi1 xi2,xindquali=,
    sort_men=sort_men_test,sort_rmen=sort_rmen_test,
    sort_rind=sort_rind_test);
```

6.3 Application to the French panel for urban policy

We continue the illustration initiated in Section 6.3. 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

$$\theta_{g,ind} = \frac{\sum_{l \in U_{ind}} 1(y_l = g)}{N_{ind}}, \quad (6.1)$$

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

$$\hat{\theta}_{grr,ind} = \frac{\sum_{l \in S_{rr,ind}} d_{rrl} 1(y_l = g)}{\sum_{l \in S_{rr,ind}} d_{rrl}}, \quad (6.2)$$

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

$$\hat{\theta}_{gcal,ind} = \frac{\sum_{l \in S_{rr,ind}} w_l 1(y_l = g)}{\sum_{l \in S_{rr,ind}} w_l}, \quad (6.3)$$

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 4. The three confidence intervals are very similar in all cases.

7 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 as Supplementary Material.

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. While the bootstrap variance estimator is known to be conservative in this case, the coverage properties of confidence intervals in such context remain unclear. This is an interesting matter for further research.

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.

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

Table 4: 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	
Estim.	One	None	One	None
Norm. CI	0.304	0.372	0.305	0.372
Perc. CI	[0.286,0.323]	[0.352,0.392]	[0.285,0.325]	[0.350,0.394]
Basic CI	[0.287,0.323]	[0.351,0.393]	[0.284,0.325]	[0.351,0.393]
	[0.286,0.322]	[0.351,0.393]	[0.285,0.325]	[0.352,0.393]
	Complementary full medical cover			
	Estimator adj. for non-response		Calibration estimator	
Estim.	Yes	No	Yes	No
Norm. CI	0.122	0.626	0.122	0.627
Perc. CI	[0.106,0.137]	[0.603,0.650]	[0.105,0.138]	[0.604,0.650]
Basic CI	[0.105,0.137]	[0.603,0.651]	[0.105,0.138]	[0.604,0.650]
	[0.106,0.138]	[0.602,0.649]	[0.105,0.138]	[0.605,0.651]

- Chauvet, G. (2007). *Méthodes de Bootstrap en population finie*. PhD thesis, Rennes 2 University.
- 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. and Sardy, S. (2007). Resampling variance estimation in surveys with missing data. *J. Off. Stat.*, 23:371–386.
- Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*, volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*.
- Deville, J. C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey methodology*, 25(2):193–204.
- Haziza, D. and Beaumont, J.-F. (2007). On the construction of imputation classes in surveys. *Int. Stat. Rev.*, 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.
- Kim, J. K. and Kim, J. J. (2007). Nonresponse weighting adjustment using estimated response probability. *Can. J. Stat.*, 35(4):501–514.
- 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.
- Mashreghi, Z., Haziza, D., and Léger, C. (2016). A survey of bootstrap methods in finite population sampling. *Statist. Surv.*, 10:1–52.
- Rao, J. N. K. and Wu, C.-F. J. (1988). Resampling inference with complex survey data. *J. Am. Stat. Assoc.*, 83(401):231–241.
- Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model assisted survey sampling*. Springer Series in Statistics.
- Shao, J. and Tu, D. S. (1995). *The jackknife and bootstrap*. Springer Series in Statistics.
- Tillé, Y. (2011). *Sampling algorithms*. Springer.

A Benchmark variance estimators for the sample of individuals

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

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

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

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

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

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

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

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

$$\text{and } \bar{y}_{rc} = \frac{\sum_{k \in S_{c,hou}} d_k r_k \hat{y}_{r,k}}{\sum_{k \in S_{c,hou}} \omega_k r_k},$$

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

and for the second linearized variable

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

$$\text{with } \bar{y}_{rrd} = \frac{\sum_{l \in S_{rd,ind}} d_{rl} r_l y_l}{\sum_{l \in S_{rd,ind}} \omega_l r_l}. \quad (\text{A.5})$$

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

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

B Justification for the bootstrap variance estimator

We provide some heuristic justification for the estimation of the total Y_{hou} . The case of a complex parameter can be handled similarly, by means of the linearization technique (Deville, 1999). For the estimation of Y_{hou} , we consider the case of the calibration estimator $\hat{Y}_{cal,hou}$ only, since the case of \hat{Y}_{hou} and $\hat{Y}_{r,hou}$ can be handled similarly.

We follow the same approach as in Kim and Kim (2007). We first write

$$\begin{aligned}\hat{Y}_{cal,hou} &= \hat{Y}_{r,hou} + \hat{B}_{r,hou}^\top (X_{hou} - \hat{X}_{r,hou}) \\ &= \hat{Y}_{r,hou} + B_{hou}^\top (X_{hou} - \hat{X}_{r,hou}) + O_p(Nn^{-1}) \\ &= B_{hou}^\top X_{hou} + \hat{E}_{r,hou} + O_p(Nn^{-1}),\end{aligned}\tag{B.1}$$

with

$$E_k = y_k - B_{hou}^\top x_k \text{ and } B_{hou} = \left(\sum_{k \in U_{hou}} x_k x_k^\top \right) \sum_{k \in U_{hou}} x_k y_k.$$

From the writing

$$\frac{1}{\hat{p}_c} - \frac{1}{p_c} = \frac{\sum_{k \in S_{c,hou}} \omega_k (1 - r_k/p_c)}{\sum_{k \in S_{c,hou}} \omega_k r_k} = \frac{1}{p_c} \frac{\sum_{k \in S_{c,hou}} \omega_k (1 - r_k/p_c)}{\sum_{k \in S_{c,hou}} \omega_k} + O_p(n^{-1}),\tag{B.2}$$

we obtain

$$\begin{aligned}\hat{E}_{r,hou} &= \sum_{k \in S_{r,hou}} \frac{d_k r_k E_k}{\hat{p}_{c(k)}} \\ &= \sum_{k \in S_{r,hou}} \frac{d_k r_k E_k}{p_{c(k)}} + \sum_{c=1}^C \sum_{k \in S_{c,hou}} \omega_k \left(1 - \frac{r_k}{p_{c(k)}} \right) \bar{E}_{c,hou} + O_p(Nn^{-1}) \\ &= \sum_{k \in S_{hou}} d_k \left[\omega_k \pi_k \bar{E}_{c(k),hou} + \frac{r_k}{p_{c(k)}} (E_k - \omega_k \pi_k \bar{E}_{c(k),hou}) \right] + O_p(Nn^{-1}),\end{aligned}\tag{B.3}$$

with $\bar{E}_{c,hou} = (\sum_{k \in U_{c,hou}} E_k) / (\sum_{k \in U_{c,hou}} \omega_k \pi_k)$. From (B.1) and (B.3), we have

$$\begin{aligned}\hat{Y}_{cal,hou} &= B_{hou}^\top X_{hou} + \sum_{k \in S_{hou}} d_k A_k + O_p(Nn^{-1}),\end{aligned}\tag{B.4}$$

with $A_k = \omega_k \pi_k \bar{E}_{c(k),hou} + \frac{r_k}{p_{c(k)}} (E_k - \omega_k \pi_k \bar{E}_{c(k),hou})$.

From (B.4), the asymptotic variance of $\hat{Y}_{cal,hou}$ is that of $\sum_{k \in S_{hou}} d_k A_k$. If the sampling design used

inside the strata is more efficient than multinomial sampling, a conservative variance estimator is

$$v_{HH}(\hat{Y}_{cal,hou}) = \sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{k \in S_{hou}^h} \left(d_k a_k - \frac{1}{n_h} \sum_{k' \in S_{hou}^h} d_{k'} a_{k'} \right)^2, \quad (\text{B.5})$$

$$\text{with } a_k = \omega_k \pi_k \bar{e}_{rc(k),hou} + \frac{r_k}{\hat{p}_{c(k)}} (e_k - \omega_k \pi_k \bar{e}_{rc(k),hou}), \quad (\text{B.6})$$

$$\text{with } \bar{e}_{rc,hou} = \frac{\sum_{k \in S_{c,hou}} d_k r_k e_k}{\sum_{k \in S_{c,hou}} \omega_k r_k}, \quad (\text{B.7})$$

and where e_k is given in (2.12).

We can obtain a similar expansion for the bootstrap calibrated estimator $\hat{Y}_{cal,hou*}$, conditionally on the sample S_{hou} and on the sample of respondents $S_{r,hou}$. We first obtain

$$\hat{Y}_{cal,hou*} = \hat{B}_{r,hou}^\top X_{hou} + \hat{e}_{r,hou*} + O_{p*}(Nn^{-1}) \quad (\text{B.8})$$

and

$$\frac{1}{\hat{p}_{c*}} - \frac{1}{\hat{p}_c} = \frac{1}{\hat{p}_c} \frac{\sum_{k \in S_{c,hou}} \omega_k W_k (1 - r_k / \hat{p}_c)}{\sum_{k \in S_{c,hou}} \omega_k} + O_{p*}(n^{-1}), \quad (\text{B.9})$$

where O_{p*} stands for the order of magnitude in probability, conditionally on S_{hou} and $S_{r,hou}$. This leads to

$$\hat{e}_{r,hou*} = \sum_{k \in S_{hou}} d_k W_k \left[\omega_k \pi_k \bar{e}_{rc(k),hou} + \frac{r_k}{\hat{p}_{c(k)}} (e_k - \omega_k \pi_k \bar{e}_{rc(k),hou}) \right] + O_{p*}(Nn^{-1}). \quad (\text{B.10})$$

From (B.8) and (B.10), we obtain

$$\hat{Y}_{cal,hou*} = \hat{B}_{r,hou}^\top X_{hou} + \sum_{k \in S_{hou}} d_k W_k a_k + O_{p*}(Nn^{-1}). \quad (\text{B.11})$$

Therefore, the variance of $\hat{Y}_{cal,hou*}$ conditionally on S_{hou} and $S_{r,hou}$ (say, V_*) is approximately

$$V_*(\hat{Y}_{cal,hou*}) \doteq V_* \left(\sum_{k \in S_{hou}} W_k d_k a_k \right) \quad (\text{B.12})$$

$$= v_{mult}(\hat{Y}_{cal,hou}), \quad (\text{B.13})$$

where $v_{mult}(\hat{Y}_{cal,hou})$ is given in equation (2.11).