Journées de méthodologie statistique de l'Insee 2022

Estimation de la variance pour un plan produit

Jean RUBIN (*), Guillaume CHAUVET (**)

(*) Insee (ENSAI)

(**) ENSAI (Irmar), Rennes

jean.rubin@ensae.fr guillaume.chauvet@ensai.fr

Mots-clés. bootstrap, consistance, estimation de variance

Domaines. Théorie des sondages aval : estimation de variance

Abstract

We study a particular type of sampling designs called Cross-Classified Sampling (CCS), for an arbitrary number of dimensions. We propose a decomposition of the variance which allows to derive general asymptotic properties of these designs, and to construct simple and approximately unbiased variance estimators. Bootstrap methods are also studied, and shown to lead to approximately unbiased variance estimation, under mild assumptions. A simulation study supports our findings.

Résumé

Nous étudions une famille de plans de sondage appelés plans de sondage produit (CCS), pour un nombre quelconque de dimensions. Nous proposons une décomposition de la variance qui permet d'obtenir les propriétés asymptotiques de ces plans de sondages, et de construire des estimateurs de variance simples et approximativement sans biais. Des méthodes de bootstrap sont également étudiées. Nous montrons que sous des hypothèses raisonnables, elles conduisent à des estimateurs de variance approximativement sans biais. Nos résultats sont appuyés par une étude par simulations.

1 Introduction

Multistage sampling designs are commonly used in household and health surveys. In the case of the two-stage sampling, the population units are grouped into large blocks (e.g., municipalities or counties), called Primary Sampling Units (PSUs), which are sampled at the first stage. At the second stage, a list of population units is obtained inside the selected PSUs, and a sample of these units is selected. A detailed treatment of multistage sampling may be found in Cochran (1977), Särndal et al. (1992) and Fuller (2011). In some situations, a population unit $\mathbf{k}$ is more easily represented as a couple (k_1, k_2). For example, in the ELFE maternity survey (Juillard et al., 2017), a sample is obtained by selecting a sample of maternities (unit k_1), a sample of days (unit k_2), and by crossing the two samples. In this case, a population unit is therefore given by a day-maternity couple (k_1, k_2).

In such situations, Ohlsson (1996) introduced the cross-classified sampling (CCS) designs, under which independent samples S_1 and S_2 are selected in each dimension. By taking the cartesian product of S_1 and S_2, the final sample $S = S_1 \times S_2$ is naturally obtained. Other examples of use of cross-classified sampling designs include consumer price index surveys (Dalén and Ohlsson, 1995) and business surveys (Skinner, 2015). To produce reliable estimators with associated confidence intervals, some basic statistical properties are needed for cross-classified sampling designs, including the consistency and the asymptotic normality of Horvitz-Thompson estimators. It is also desirable to provide (at least approximately) unbiased variance estimators. Thereupon, it would be of interest to derive appropriate bootstrap methods for cross-classified sampling. With the notable exception of Skinner (2015), who proposed a bootstrap algorithm for with-replacement sampling designs in each dimension, this last topic remains poorly studied in the literature.

In this work, we extend cross-classified sampling to an arbitrary number of dimensions. Under some mild conditions, we prove the consistency and the asymptotic normality of the Horvitz-Thompson (HT)-estimator. Using the Hoeffding-Sobol decomposition (Hoeffding, 1948), we generalize the variance formula given in Ohlsson (1996) and prove that the variance of the HT-estimator can be decomposed into a sum of multiple terms with different orders of magnitude. By identifying the leading terms in this variance decomposition, we obtain simple, consistent variance estimators. The decomposition is also used to derive bootstrap methods suitable for cross-classified sampling.

The article is organized as follows. In section 2, we define our notations and state our main assumptions. In Section 3, we make use of the Hoeffding-Sobol decomposition to obtain a general variance decomposition and to prove the consistency and the asymptotic normality of the HT-estimator. We also obtain simple consistent variance estimators. Some illustrations of the use of these simplified estimators are presented in Section 4. A weighted bootstrap method is proposed and studied in Section 5. In Section 6, we consider the particular case of bootstrap variance estimation when simple random sampling is used in all dimensions. The results of a simulation study are given in Section 7. The proofs are given in Appendix.

2 Multi-dimensional cross-classified sampling

2.1 Notation

Suppose that we are interested in a finite population $\mathcal{U} = \prod_{d=1}^{D} \mathcal{U}_d$ which can be seen as the cartesian product of D finite populations of respective sizes $N_d,\ d = 1, \ldots, D$. The size of the product population $\mathcal{U}$ is therefore $N = \prod_{d=1}^{D} N_d$. For each unit $\mathbf{k} = (k_1, \ldots, k_D) \in \mathcal{U}$, a variable of interest y takes the value $y_{\mathbf{k}}$. We are interested in estimating the population total

$$Y = \sum_{\mathbf{k} \in \mathcal{U}} y_{\mathbf{k}}. \tag{2.1}$$

For computations considered in the following sections, it is convenient to introduce some notations for sub-totals. Let $I \subseteq \{1, \ldots, D\}$ denote a subset of dimensions, and let $\mathcal{U}_I = \prod_{d \in I} \mathcal{U}_d$ be the product population associated to these dimensions. For any $\mathbf{k}' \in \mathcal{U}_I$, we let

$$Y_{\mathbf{k}'} = \sum_{\substack{\mathbf{l} \in \mathcal{U} \\ \forall d \in I,\ l_d = k'_d}} y_{\mathbf{l}} \tag{2.2}$$

denote the sub-total of y when the set of coordinates in $\mathcal{U}_I$ remains fixed and equal to $\mathbf{k}'$.

For $d \in \{1, \ldots, D\}$, we let $p_d(\cdot)$ denote a sampling design used in the population $\mathcal{U}_d$. Under a D-dimensional cross-classified sampling design, D independent samples S_d are selected in the populations $\mathcal{U}_d,\ d = 1, \ldots, D$, and their cartesian product $S := \prod_{d=1}^{D} S_d$ is the overall sample. Therefore, the resulting sampling design $p(\cdot)$ is such that

$$\forall d \in \{1, \ldots, D\}, \forall s_d \in \mathcal{P}(\mathcal{U}_d),\ p\left(\prod_{d=1}^{D} s_d\right) = \prod_{d=1}^{D} p_d(s_d). \tag{2.3}$$

In the particular case when $D = 2$, we obtain the usual two-dimensional cross-classified design introduced by Ohlsson (1996). It is somewhat similar to a two-stage sampling design where each unit k_1 in $\mathcal{U}_1$ would be a PSU, while $\mathcal{U}_2$ would be the sub-population of SSUs for any k_1. The main difference is that, due to the independence in the selection of S_1 and S_2, the same subsample of SSUs is used inside any PSU: in other words, the cross-classified design does not verify the independence property (see Särndal et al., 1992). However, the invariance hypothesis is satisfied.

In the population $\mathcal{U}_d$, we let δ_k^d denote the sample membership indicator of the unit k in the sample S_d. We let π_k^d denote the probability that k is selected in S_d, $n_d = \sum_{k \in \mathcal{U}_d} \pi_k^d$ denote the expected size of the sample S_d, and π_{kl}^d denote the probability that units k and l are jointly selected in S_d. We will assume in the following that these probabilities are all strictly positive. Finally we also use the notation

$$\Delta_{kl}^d = Cov(\delta_k^d, \delta_l^d) = \pi_{kl}^d - \pi_k^d \pi_l^d \text{ for any } k, l \in \mathcal{U}_d. \tag{2.4}$$

2.2 Estimation

We consider weighted total estimators of the form:

$$\hat{Y} = \sum_{\mathbf{k} \in \mathcal{U}} y_{\mathbf{k}} \prod_{d=1}^{D} w_{k_d}^d(S_d), \tag{2.5}$$

where $\{w_k^d(S_d)\}_{k \in \mathcal{U}_d}$ is a set of estimation weights available for each dimension $d \in \{1, \ldots, D\}$. We suppose that $\{w_k^d(S_d)\}_{k \in \mathcal{U}_d}$ depends only on the sample S_d, and not on the samples selected in the other dimensions. When there is no risk of confusion, we simplify the notation as $w_k^d(S_d) \equiv w_k^d$. We also suppose that for any $d \in \{1, \ldots, D\}$ and $k \in \mathcal{U}_d$, $w_k^d = 0$ if $k \notin S_d$. This implies that each unit $\mathbf{k} \in \mathcal{U}$ has an associated weight $w_{\mathbf{k}} = \prod_{d=1}^{D} w_{k_d}^d$ that can be decomposed as a product of independent weights, leading to the following compact expression:

$$\hat{Y} = \sum_{\mathbf{k} \in \mathcal{U}} y_{\mathbf{k}} w_{\mathbf{k}}. \tag{2.6}$$

In the important particular case when $w_k^d = \{\pi_k^d\}^{-1}\delta_k^d$ for $k \in \mathcal{U}_d$, we obtain the Horvitz-Thompson (HT) estimator

$$\hat{Y}_\pi = \sum_{\mathbf{k}\in S} \frac{y_{\mathbf{k}}}{\prod_{d=1}^D \pi_{k_d}^d} = \sum_{\mathbf{k}\in\mathcal{U}} y_{\mathbf{k}} \prod_{d=1}^D \frac{\delta_{k_d}^d}{\pi_{k_d}^d}. \tag{2.7}$$

We may further introduce for $\mathbf{k} \in \mathcal{U}$ the quantities $\delta_{\mathbf{k}} = \prod_{d=1}^D \delta_{k_d}^d$ and $\pi_{\mathbf{k}} = \prod_{d=1}^D \pi_{k_d}^d$, leading to the compact expression

$$\hat{Y}_\pi = \sum_{\mathbf{k}\in S} \frac{y_{\mathbf{k}}}{\pi_{\mathbf{k}}} = \sum_{\mathbf{k}\in\mathcal{U}} y_{\mathbf{k}} \frac{\delta_{\mathbf{k}}}{\pi_{\mathbf{k}}}. \tag{2.8}$$

We will also use the notation $\pi_{\mathbf{kl}} = \prod_{d=1}^D \pi_{k_d l_d}^d$ for any $\mathbf{k}, \mathbf{l} \in \mathcal{U}$.

2.3 Assumptions

In the paper, we consider the following assumptions:

H1. There exists some constant α such that

$$\frac{1}{N} \sum_{\mathbf{k}\in\mathcal{U}} y_{\mathbf{k}}^2 \leq \alpha. \tag{2.9}$$

H2. For any $d = 1, \ldots, D$, we have $n_d \to \infty$ and $N_d \to \infty$, and there exists some constant $f_d \in [0,1]$ such that

$$\frac{n_d}{N_d} \to f_d. \tag{2.10}$$

H3. For any $d = 1, \ldots, D$, there exists some constant $\lambda_d > 0$ such that

$$\forall k \in \mathcal{U}_d,\ \pi_k^d \geq \lambda_d \frac{n_d}{N_d}. \tag{2.11}$$

H4. For every $d = 1, \ldots, D$, there exists a constant γ_d such that

$$\forall k \neq l \in \mathcal{U}_d,\ |\Delta_{k,l}^d| \leq \gamma_d \frac{n_d}{N_d^2}. \tag{2.12}$$

The assumption (H1) is related to the variable of interest by assuming that it has a finite moment of order 2. The assumptions (H2)-(H4) are related to the sampling design. The asymptotic framework is defined in (H2). Note in particular that it is assumed that the sample size $n_d \to \infty$ in all dimensions, which is needed to obtain the consistency of the Horvitz-Thompson estimator, see Section 3.2. It is assumed in (H3) that in each dimension d, the first-order inclusion probabilities have a lower bound of order n_d/N_d. The assumption (H4) is related to the second-order inclusion probabilities. The quantity $|\Delta^d_{k,l}|$ may be thought of as a measure of dependency in the selection of the units $k, l \in \mathcal{U}_d$ in the sample S_d. These quantities are equal to 0 if the units are independently selected in S_d, which is known as Poisson sampling. This assumption is also respected for simple random sampling and rejective sampling (Hájek, 1964), for example. Overall, assumptions (H1)-(H4) are standard. In the particular case of two-dimensional cross-classified sampling, they reduce to assumptions (H1)-(H3) in Juillard et al. (2017).

3 Properties of total estimators

3.1 Hoeffding-Sobol variance decomposition of the CCS

From equation (2.5), it is possible to derive the following general variance formula:

$$V_p(\hat{Y}) = \sum_{\mathbf{k},\mathbf{l} \in \mathcal{U}} y_{\mathbf{k}} y_{\mathbf{l}} \left[\prod_{d=1}^{D} E_p(w^d_{k_d} w^d_{l_d}) - \prod_{d=1}^{D} E_p(w^d_{k_d}) E_p(w^d_{l_d}) \right]. \tag{3.1}$$

The proof is given in Appendix A.1. However, the naturally derived unbiased variance estimator from equation (3.1) appears to be unpractical in many sampling designs, as illustrated by Ohlsson (1996). In order to study the statistical properties of estimators and to propose simple variance estimators, we therefore follow an alternative variance decomposition generalizing the approach proposed by Ohlsson for the two-dimensional case, and write $\hat{Y}$ as a sum of uncorrelated components. This is summarized in our Proposition 1, whose proof is given in Appendix A.2. This type of decomposition based on the works of Hoeffding (1948) is now commonly called Hoeffding-Sobol decomposition or functional ANOVA decomposition.

Proposition 1. *We can write*

$$\hat{Y} = \sum_{I \subseteq \{1,\dots,D\}} \hat{Y}^I, \tag{3.2}$$

where for any subset $I \subseteq \{1,\dots,D\}$*:*

$$\hat{Y}^I = \sum_{I' \in \mathcal{P}(I)} (-1)^{|I|-|I'|} E_p(\hat{Y}|(S_d)_{d\in I'}) \tag{3.3}$$

$$= \sum_{\mathbf{k}\in\mathcal{U}} y_{\mathbf{k}} \prod_{d\in I} (w^d_{k_d} - E_p(w^d_{k_d})) \prod_{d\notin I} E_p(w^d_{k_d}). \tag{3.4}$$

Also, the components $\{\hat{Y}^I\}_{I\subseteq\{1,\dots,D\}}$ *are uncorrelated, and*

$$V_p(\hat{Y}) = \sum_{I \subseteq \{1,\dots,D\}} V_p(\hat{Y}^I), \tag{3.5}$$

where for any $I \neq \emptyset$*:*

$$V_p(\hat{Y}^I) = \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} y_{\mathbf{k}} y_{\mathbf{l}} \prod_{d\in I} Cov_p(w^d_{k_d}, w^d_{l_d}) \prod_{d\notin I} E_p(w^d_{k_d}) E_p(w^d_{l_d}). \tag{3.6}$$

From Proposition 1, we obtain an unbiased variance estimator for $\hat{Y}$, as summarized in Corollary 1. The proof is given in Appendix A.3.

Corollary 1. *Suppose that the set of weights is adapted to the sampling design, namely that for any* $d \in \{1,\dots,D\}$*:*

$$\forall k \in \mathcal{U}_d,\ E_p(w^d_k) = 1 \tag{3.7}$$

Also, suppose that for any $d \in \{1,\dots,D\}$ *and for any couple of units* $k,l \in \mathcal{U}_d$, $\widehat{Cov}^d(w^d_k, w^d_l)$ *is an unbiased estimator of* $Cov_p(w^d_k, w^d_l)$ *built from* S_d *and such that* $\widehat{Cov}^d(w^d_k, w^d_l) = 0$ *when* k *or* l *are not in* S_d*. Then* $V_p(\hat{Y})$ *may be unbiasedly estimated by*

$$\hat{V}_p(\hat{Y}) = \sum_{I \subseteq \{1,\dots,D\}} \hat{V}_p(\hat{Y}^I), \tag{3.8}$$

where

$$\hat{V}_p(\hat{Y}^I) = \sum_{\mathbf{k},\mathbf{l}\in S} y_{\mathbf{k}} y_{\mathbf{l}} \prod_{d\in I} \widehat{Cov}^d(w^d_{k_d}, w^d_{l_d}) \prod_{d\notin I} \frac{1}{\pi^d_{k_d,l_d}}. \tag{3.9}$$

3.2 Properties of the Horvitz-Thompson estimator

We now consider the HT-estimator, which is popular in practice. In this case, the general variance formula (3.1) simplifies to

$$V_p(\hat{Y}_\pi) = \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} \frac{y_\mathbf{k} y_\mathbf{l}}{\pi_\mathbf{k}\pi_\mathbf{l}}(\pi_{\mathbf{kl}} - \pi_\mathbf{k}\pi_\mathbf{l}). \tag{3.10}$$

By applying the results stated in Proposition 1 and Corollary 1, we obtain an alternative variance decomposition for $\hat{Y}_\pi$ and an unbiased variance estimator. This is summarized in Corollary 2. The proof is given in Appendix B.1.

Corollary 2. *The variance of the HT-estimator may be written as*

$$V_p(\hat{Y}_\pi) = \sum_{I\subseteq\{1,\ldots,D\}} V_p(\hat{Y}_\pi^I), \tag{3.11}$$

where for any $I \subseteq \{1,\ldots,D\}$

$$\hat{Y}_\pi^I = \sum_{\mathbf{k}\in\mathcal{U}} y_\mathbf{k} \prod_{d\in I}\left(\frac{\delta^d_{k_d}}{\pi^d_{k_d}} - 1\right). \tag{3.12}$$

Also, if $I \neq \varnothing$*, we have*

$$V_p(\hat{Y}_\pi^I) = \sum_{\mathbf{k}',\mathbf{l}'\in\mathcal{U}_I} \frac{Y_{\mathbf{k}'}}{\pi_{\mathbf{k}'}}\frac{Y_{\mathbf{l}'}}{\pi_{\mathbf{l}'}}\Delta_{\mathbf{k}'\mathbf{l}'}, \tag{3.13}$$

with $\Delta_{\mathbf{k}'\mathbf{l}'} = \prod_{d\in I}\Delta^d_{k'_d l'_d}$ *and* $\pi_{\mathbf{k}'} = \prod_{d\in I}\pi^d_{k'_d}$ *for any* $\mathbf{k}',\mathbf{l}' \in \mathcal{U}_I$*. This variance is unbiasedly estimated by*

$$\hat{V}_p(\hat{Y}_\pi^I) = \sum_{\mathbf{k},\mathbf{l}\in S} \frac{y_\mathbf{k} y_\mathbf{l}}{\pi_{\mathbf{kl}}} \prod_{d\in I}\frac{\Delta^d_{k_d,l_d}}{\pi^d_{k_d}\pi^d_{l_d}}. \tag{3.14}$$

The Hoeffding-Sobol decomposition therefore enables to write the HT-estimator as the sum of 2^D uncorrelated terms given in (3.12), leading to the unbiased variance estimator in (3.14). It is important to precisely state the orders of magnitude of the terms in the decomposition. This is established in Proposition 2, see Appendix B.2 for a proof.

Proposition 2. *Suppose that assumptions (H1), (H3) and (H4) hold. Then for every non-empty subset I of $\{1,\ldots,D\}$:*

$$V_p(\hat{Y}_\pi^I) = O\left(\frac{N^2}{\prod_{d\in I} n_d}\right). \tag{3.15}$$

Proposition 2 has some important consequences. Firstly, it states that the leading terms in the Hoeffding-Sobol variance decomposition correspond to the subsets I of size 1, and that the other terms are of negligible order under assumption (H2). It is therefore possible to obtain an approximately unbiased variance estimator by restricting the variance estimator in (3.14) to the D terms associated to the singletons, leading to the simplified variance estimator

$$\hat{V}^{SIMP}(\hat{Y}_\pi) = \sum_{\mathbf{k},\mathbf{l}\in S} \frac{y_{\mathbf{k}} y_{\mathbf{l}}}{\pi_{\mathbf{kl}}} \left(\sum_{d=1}^{D} \frac{\Delta^d_{k_d,l_d}}{\pi^d_{k_d}\pi^d_{l_d}}\right). \tag{3.16}$$

Secondly, it follows from Proposition 2 that assumption (H2) can not be dropped for the HT-estimator to be consistent. In other words, if the sample size n_d in some dimension d is bounded, then the variance of $N^{-1}\hat{Y}_\pi^{\{d\}}$ is not vanishing and the Horvitz-Thompson estimator is not consistent. The asymptotic properties of the HT-estimator are established in Proposition 3, see Appendix B.3 for a proof.

Proposition 3. *Suppose that assumptions (H1)-(H4) hold. Then*

$$V_p(N^{-1}\hat{Y}_\pi) = O\left(n_m^{-1}\right), \tag{3.17}$$

where $n_m = \min_{d=1,\ldots,D} n_d$. Additionally, suppose that:

H5. There exists some constant $C > 0$ such that

$$V_p(\hat{Y}_\pi) \geq C N^2 n_m^{-1}. \tag{3.18}$$

H6. For any $d = 1,\ldots,D$:

$$\frac{\hat{Y}_\pi^{\{d\}}}{\sqrt{V_p\left(\hat{Y}_\pi^{\{d\}}\right)}} \longrightarrow_{\mathcal{L}} \mathcal{N}(0,1), \tag{3.19}$$

where $\longrightarrow_{\mathcal{L}}$ stands for the convergence in distribution.

H7. For any $d = 1, \ldots, D$, there exists some constant $\gamma_d \geq 0$ such that

$$\frac{V_p(\hat{Y}_\pi^{\{d\}})}{V_p(\hat{Y}_\pi)} \longrightarrow (\gamma_d)^2. \tag{3.20}$$

Then

$$\frac{\hat{Y}_\pi - Y}{\sqrt{V_p(\hat{Y}_\pi)}} \longrightarrow_{\mathcal{L}} \mathcal{N}(0,1). \tag{3.21}$$

It is supposed in Assumption (H5) that the variance of the HT-estimator is non-vanishing, and has the usual order of magnitude $O(N^2 n_m^{-1})$. It is supposed in Assumption (H6) that for any dimension d, the HT-estimator is asymptotically normally distributed under the sampling design used.

3.3 Plug-in variance estimation

Another possible variance estimator can be obtained from the decomposition in equation (3.11). For any term $V_p(\hat{Y}_\pi^I)$ of the variance decomposition, a plug-in estimator based on the expression in (3.13)

$$\hat{V}^{PLUG}(\hat{Y}_\pi^I) = \sum_{\mathbf{k}',\mathbf{l}' \in S_I} \frac{\hat{Y}_{\mathbf{k}'}}{\pi_{\mathbf{k}'}} \frac{\hat{Y}_{\mathbf{l}'}}{\pi_{\mathbf{l}'}} \frac{\Delta_{\mathbf{k}',\mathbf{l}'}}{\pi_{\mathbf{k}',\mathbf{l}'}}, \tag{3.22}$$

is obtained by replacing for each $\mathbf{k}' \in S_I$ the partial sum $Y_{\mathbf{k}'}$ with the unbiased estimator

$$\hat{Y}_{\mathbf{k}'} = \sum_{\substack{\mathbf{l} \in S \\ \forall d \in I,\ l_d = k'_d}} \frac{y_{\mathbf{l}}}{\prod_{d \notin I} \pi_{l_d}}. \tag{3.23}$$

and by using a Horvitz-Thompson like variance estimator. This leads to the plug-in variance estimator

$$\hat{V}^{PLUG}(\hat{Y}_\pi) = \sum_{I \subseteq \{1,\ldots,D\}} \hat{V}^{PLUG}(\hat{Y}_\pi^I). \tag{3.24}$$

Following the result obtained in Proposition 2, this estimator may be further simplified by restricting the sum in (3.24) to the D terms associated to the singletons. This leads to the second simplified variance estimator

$$\hat{V}^{SIMP2}(\hat{Y}_\pi) = \sum_{d=1}^{D} \sum_{k_d,l_d \in S_d} \frac{\hat{Y}_{k_d}}{\pi_{k_d}^d} \frac{\hat{Y}_{l_d}}{\pi_{l_d}^d} \frac{\Delta_{k_d,l_d}^d}{\pi_{k_d,l_d}^d}. \tag{3.25}$$

By construction, this approximation only consider first order variance terms but intuitively it is simpler than $\hat{V}^{SIMP}$ in the sense that it removes higher order interactions inside these variance terms. More precisely, Proposition 4 shows that the bias of the plug-in estimators can be expressed in terms of the $V_p(\hat{Y}_\pi^I)$, see Appendix B.4 for a proof.

Proposition 4. *For every non-empty subset I of $\{1, \ldots, D\}$, we have*

$$E_p\left[\hat{V}^{PLUG}(\hat{Y}_\pi^I)\right] = \sum_{I' \supseteq I} V_p(\hat{Y}_\pi^{I'}) \tag{3.26}$$

It follows from Propositions 2 and 4 that for every non-empty subset $I \subseteq \{1, \ldots, D\}$, the estimator $\hat{V}^{PLUG}(\hat{Y}_\pi^I)$ is asymptotically unbiased, and likewise for $\hat{V}^{PLUG}(\hat{Y}_\pi)$. In particular, it results that $\hat{V}^{SIMP}(\hat{Y}_\pi)$ and $\hat{V}^{SIMP2}(\hat{Y}_\pi)$ are asymptotically similar.

4 Illustrations of the simplified estimations

In this section, we illustrate the proposed simplified variance estimators $\hat{V}^{SIMP}(\hat{Y}_\pi)$ and $\hat{V}^{SIMP2}(\hat{Y}_\pi)$ for specific sampling designs, in the two-dimensional situation. The case of simple random sampling in each dimension is considered in Section 4.1, and the case of Poisson sampling in each dimension is considered in Section 4.2.

4.1 Simple random sampling

We first consider the case when simple random sampling of size n_d is used in each dimension $d = 1, 2$. The sampling fraction in dimension d is denoted as $f_d = n_d/N_d$. The Horvitz-Thompson estimator can be written in the form

$$\hat{Y}_\pi = \frac{N_1 N_2}{n_1 n_2} \sum_{k_1 \in S_1} \sum_{k_2 \in S_2} y_{k_1 k_2} \tag{4.1}$$

By applying Corollary 2 (see also Ohlsson, 1996), the Hoeffding-Sobol variance decomposition is

$$V_p(\hat{Y}_\pi) = N_1^2 N_2^2 \left[(1-f_1)\frac{S_1^2}{n_1} + (1-f_2)\frac{S_2^2}{n_2} + (1-f_1)(1-f_2)\frac{S_{12}^2}{n_1 n_2}\right] \tag{4.2}$$

where

$$\begin{aligned}
S_1^2 &= \frac{1}{N_1-1}\sum_{k_1\in\mathcal{U}_1}(\overline{Y}_{k_1\bullet} - \overline{Y}_{\bullet\bullet})^2 \text{ with } \overline{Y}_{k_1\bullet} = \frac{1}{N_2}\sum_{k_2\in\mathcal{U}_2} y_{k_1k_2},\\
S_2^2 &= \frac{1}{N_2-1}\sum_{k_2\in\mathcal{U}_2}(\overline{Y}_{\bullet k_2} - \overline{Y}_{\bullet\bullet})^2 \text{ with } \overline{Y}_{\bullet k_2} = \frac{1}{N_1}\sum_{k_1\in\mathcal{U}_1} y_{k_1k_2},\\
S_{12}^2 &= \frac{1}{N_1-1}\frac{1}{N_2-1}\sum_{k_1\in\mathcal{U}_1}\sum_{k_2\in\mathcal{U}_2}(y_{k_1k_2} - \overline{Y}_{k_1\bullet} - \overline{Y}_{\bullet k_2} + \overline{Y}_{\bullet\bullet})^2,
\end{aligned} \tag{4.3}$$

and where $\overline{Y}_{\bullet\bullet} = N^{-1}Y$ is the population mean.

As mentionned by Skinner (2015), this variance is unbiasedly estimated by replacing each term with an unbiased counterpart. More specifically, we can use the following unbiased estimators:

$$\begin{aligned}
\hat{S}_1^2 &= \frac{1}{n_1-1}\sum_{k_1\in S_1}(\overline{y}_{k_1\bullet} - \overline{y}_{\bullet\bullet})^2 - (1-f_2)\frac{\hat{S}_{12}^2}{n_2},\\
\hat{S}_2^2 &= \frac{1}{n_2-1}\sum_{k_2\in S_2}(\overline{y}_{\bullet k_2} - \overline{y}_{\bullet\bullet})^2 - (1-f_1)\frac{\hat{S}_{12}^2}{n_1},\\
\hat{S}_{12}^2 &= \frac{1}{n_1-1}\frac{1}{n_2-1}\sum_{k_1\in S_1}\sum_{k_2\in S_2}(y_{k_1k_2} - \overline{y}_{k_1\bullet} - \overline{y}_{\bullet k_2} + \overline{y}_{\bullet\bullet})^2,
\end{aligned} \tag{4.4}$$

with

$$\overline{y}_{k_1\bullet} = \frac{1}{n_2}\sum_{k_2\in S_2} y_{k_1k_2} \quad \text{and} \quad \overline{y}_{\bullet k_2} = \frac{1}{n_1}\sum_{k_1\in S_1} y_{k_1k_2} \tag{4.5}$$

the unbiased estimators of the partial means $\overline{Y}_{k_1\bullet}$ and $\overline{Y}_{\bullet k_2}$, respectively, and with $\overline{y}_{\bullet\bullet}$ the sample mean. Note that both $\hat{S}_1^2$ and $\hat{S}_2^2$ include correction terms for unbiasedness, which are negligible if the sample sizes are large enough.

The first proposed variance estimator $\hat{V}^{SIMP}(\hat{Y}_\pi)$ given in (3.16) is obtained by removing the interaction term in (4.2), leading to

$$\hat{V}^{SIMP}(\hat{Y}_\pi) = N_1^2 N_2^2 \left[(1-f_1)\frac{\hat{S}_1^2}{n_1} + (1-f_2)\frac{\hat{S}_2^2}{n_2}\right]. \tag{4.6}$$

Note that this estimator still uses correction terms inside $\hat{S}_1^2$ and $\hat{S}_2^2$ and therefore requires the computation of $\hat{S}_{12}^2$. The second proposed variance estimator $\hat{V}^{SIMP2}(\hat{Y}_\pi)$ does not only remove the interaction term $\hat{S}_{12}^2$, but also these correction terms. This leads to

$$\hat{V}^{SIMP2}(\hat{Y}_\pi) = N_1^2 N_2^2 \left[(1-f_1)\frac{\hat{S}_1^{2,PLUG}}{n_1} + (1-f_2)\frac{\hat{S}_2^{2,PLUG}}{n_2}\right] \tag{4.7}$$

where

$$\begin{aligned} \hat{S}_1^{2,PLUG} &= \frac{1}{n_1-1}\sum_{k_1\in S_1}(\overline{y}_{k_1\bullet} - \overline{y}_{\bullet\bullet})^2, \\ \hat{S}_2^{2,PLUG} &= \frac{1}{n_2-1}\sum_{k_2\in S_2}(\overline{y}_{\bullet k_2} - \overline{y}_{\bullet\bullet})^2, \end{aligned} \tag{4.8}$$

are the plug-in estimators of S_1^2 and S_2^2.

4.2 Poisson sampling

We consider the case when Poisson sampling is used in each dimension $d = 1, 2$. The HT-estimator may be written as

$$\hat{Y}_\pi = \sum_{k_1\in S_1}\sum_{k_2\in S_2}\frac{y_{k_1k_2}}{\pi_{k_1}^{(1)}\pi_{k_2}^{(2)}}. \tag{4.9}$$

By applying Corollary 2 (see also Ohlsson, 1996), the Hoeffding-Sobol variance decomposition is

$$\begin{aligned} V_p(\hat{Y}_\pi) &= \sum_{k_1\in\mathcal{U}_1}\left(\frac{1-\pi_{k_1}^{(1)}}{\pi_{k_1}^{(1)}}\right)Y_{k_1\bullet}^2 + \sum_{k_2\in\mathcal{U}_2}\left(\frac{1-\pi_{k_2}^{(2)}}{\pi_{k_2}^{(2)}}\right)Y_{\bullet k_2}^2 \\ &+ \sum_{k_1\in\mathcal{U}_1}\sum_{k_2\in\mathcal{U}_2}\left(\frac{1-\pi_{k_1}^{(1)}}{\pi_{k_1}^{(1)}}\right)\left(\frac{1-\pi_{k_2}^{(2)}}{\pi_{k_2}^{(2)}}\right)y_{k_1k_2}^2, \end{aligned} \tag{4.10}$$

with $Y_{k_1\bullet} = \sum_{k_2\in\mathcal{U}_2} y_{k_1k_2}$ and $Y_{\bullet k_2} = \sum_{k_1\in\mathcal{U}_1} y_{k_1k_2}$ the partial sums.

The variance is unbiasedly estimated by replacing each term in (4.11) with an unbiased counterpart. This leads to the variance estimator

$$\begin{aligned} \hat{V}_p(\hat{Y}_\pi) &= \sum_{k_1\in S_1} \frac{1}{\pi^{(1)}_{k_1}} \left(\frac{1-\pi^{(1)}_{k_1}}{\pi^{(1)}_{k_1}}\right) \left(\hat{Y}^2_{k_1\bullet} - \sum_{k_2\in S_2} \frac{1}{\pi^{(2)}_{k_2}} \left(\frac{1-\pi^{(2)}_{k_2}}{\pi^{(2)}_{k_2}}\right) y^2_{k_1k_2}\right) \\ &+ \sum_{k_2\in S_2} \frac{1}{\pi^{(2)}_{k_2}} \left(\frac{1-\pi^{(2)}_{k_2}}{\pi^{(2)}_{k_2}}\right) \left(\hat{Y}^2_{\bullet k_2} - \sum_{k_1\in S_1} \frac{1}{\pi^{(1)}_{k_1}} \left(\frac{1-\pi^{(1)}_{k_1}}{\pi^{(1)}_{k_1}}\right) y^2_{k_1k_2}\right) \\ &+ \sum_{k_1\in S_1}\sum_{k_2\in S_2} \left(\frac{1-\pi^{(1)}_{k_1}}{\pi^{(1)}_{k_1}}\right) \left(\frac{1-\pi^{(2)}_{k_2}}{\pi^{(2)}_{k_2}}\right) \frac{y^2_{k_1k_2}}{\pi^{(1)}_{k_1}\pi^{(2)}_{k_2}}, \end{aligned} \tag{4.11}$$

with $\hat{Y}_{k_1\bullet} = \sum_{k_2\in S_2} y_{k_1k_2}/\pi^2_{k_2}$ and $\hat{Y}_{\bullet k_2} = \sum_{k_1\in S_1} y_{k_1k_2}/\pi^1_{k_1}$ the unbiased estimators of the partial sums. Note that the two first terms include correction terms for unbiasedness, which can be shown to be negligible if the sample sizes are large enough.

The first proposed variance estimator $\hat{V}^{SIMP}(\hat{Y}_\pi)$ given in (3.16) is obtained by removing the third interaction term in (4.11). This leads to

$$\begin{aligned} \hat{V}^{SIMP}(\hat{Y}_\pi) &= \sum_{k_1\in S_1} \frac{1}{\pi^{(1)}_{k_1}} \left(\frac{1-\pi^{(1)}_{k_1}}{\pi^{(1)}_{k_1}}\right) \left(\hat{Y}^2_{k_1\bullet} - \sum_{k_2\in S_2} \frac{1}{\pi^{(2)}_{k_2}} \left(\frac{1-\pi^{(2)}_{k_2}}{\pi^{(2)}_{k_2}}\right) y^2_{k_1k_2}\right) \\ &+ \sum_{k_2\in S_2} \frac{1}{\pi^{(2)}_{k_2}} \left(\frac{1-\pi^{(2)}_{k_2}}{\pi^{(2)}_{k_2}}\right) \left(\hat{Y}^2_{\bullet k_2} - \sum_{k_1\in S_1} \frac{1}{\pi^{(1)}_{k_1}} \left(\frac{1-\pi^{(1)}_{k_1}}{\pi^{(1)}_{k_1}}\right) y^2_{k_1k_2}\right) \end{aligned} \tag{4.12}$$

The second proposed variance estimator $\hat{V}^{SIMP2}(\hat{Y}_\pi)$ also removes the correction terms, leading to

$$\hat{V}^{SIMP2}(\hat{Y}_\pi) = \sum_{k_1\in S_1} \left(\frac{1-\pi^{(1)}_{k_1}}{\pi^{(1)}_{k_1}}\right) \frac{\hat{Y}^2_{k_1\bullet}}{\pi^{(1)}_{k_1}} + \sum_{k_2\in S_2} \left(\frac{1-\pi^{(2)}_{k_2}}{\pi^{(2)}_{k_2}}\right) \frac{\hat{Y}^2_{\bullet k_2}}{\pi^{(2)}_{k_2}} \tag{4.13}$$

5 Weighted bootstrap method for CCS

We now construct a weighted bootstrap method under the assumption that for each $d \in \{1,\ldots,D\}$, a weighted bootstrap method adapted to the weights $(w^d_{k_d})_{k_d\in\mathcal{U}_d}$ is available. More precisely we consider for each $d \in \{1,\ldots,D\}$ a set of weights $(w^{d*}_{k_d})_{k_d\in S_d}$ satisfying the first and second order moment constraints from Beaumont and Patak (2012):

$$\forall k_d \in S_d,\ E_*(w_{k_d}^{d*}) = w_{k_d}^d, \tag{5.1}$$

$$\forall k_d, l_d \in S_d,\ Cov_*(w_{k_d}^{d*}, w_{l_d}^{d*}) = \widehat{Cov}^d(w_{k_d}^d, w_{l_d}^d). \tag{5.2}$$

A natural bootstrap estimator for CCS is:

$$\hat{Y}^* = \sum_{\mathbf{k}\in S} y_{\mathbf{k}} \prod_{d=1}^{D} w_{k_d}^{d*} \tag{5.3}$$

where the weights are simulated independently in each dimension $d \in \{1, \ldots, D\}$. It is then possible to consider the Hoeffding-Sobol decomposition of $\hat{Y}^*$, leading to the following expression for each non-empty subset $I \subseteq \{1, \ldots, D\}$:

$$\hat{Y}^{I*} = \sum_{\mathbf{k}\in S} y_{\mathbf{k}} \prod_{d\in I} (w_{k_d}^{d*} - w_{k_d}^d) \prod_{d\notin I} w_{k_d}^d \tag{5.4}$$

We note $V_*(\cdot)$ for the bootstrap variance, conditionally on the original sample S. By applying formula (3.6), we obtain:

$$\begin{aligned} V_*(\hat{Y}^{I*}) &= \sum_{\mathbf{k},\mathbf{l}\in S} y_{\mathbf{k}} y_{\mathbf{l}} \prod_{d\in I} Cov_*(w_{k_d}^{d*}, w_{l_d}^{d*}) \prod_{d\notin I} E_*(w_{k_d}^{d*}) E_*(w_{l_d}^{d*}) \\ &= \sum_{\mathbf{k},\mathbf{l}\in S} y_{\mathbf{k}} y_{\mathbf{l}} \prod_{d\in I} \widehat{Cov}_p(w_{k_d}^d, w_{l_d}^d) \prod_{d\notin I} w_{k_d}^d w_{l_d}^d. \end{aligned} \tag{5.5}$$

where the last line in (5.5) follows from equations (5.1) and (5.2). It is important to observe that in the case of HT-estimation, $V_*(\hat{Y}_\pi^{I*})$ is equal to the plug-in estimator $\hat{V}^{PLUG}(\hat{Y}_\pi^I)$ given in (3.22). Therefore, $V^*(\hat{Y}_\pi^*)$ is equal to $\hat{V}^{PLUG}(\hat{Y}_\pi)$ as defined in (3.24), and is also asymptotically unbiased by Proposition 4.

6 Case study: simple random sampling

In this section, we consider the particular important case of CCS in dimension $D = 2$, when simple random sampling is used in each dimension. The application of the pseudo-population bootstrap is studied in Section 6.1, and the application of the rescaled bootstrap is studied in Section 6.2.

6.1 Pseudo-Population Bootstrap method

In this section, we propose a pseudo-population bootstrap method for simple random sampling without replacement in each dimension, based on the method proposed by Gross (1980). To simplify the presentation, suppose that both N_2/n_2 and N_1/n_1 are integers. Once S_1 and S_2 have been sampled, the idea is to construct a pseudo-population $\mathcal{U}_d^*$ by reproducing N_d/n_d times each unit in S_d for each $d = 1, 2$. Simple random sampling without replacement is then applied in each pseudo-population to obtain the bootstrap sample S_d^*. The bootstrap estimator is thus given by

$$\hat{Y}_\pi^* = \frac{N_1 N_2}{n_1 n_2} \sum_{k_1 \in S_1^*} \sum_{k_2 \in S_2^*} y_{k_1 k_2}. \tag{6.1}$$

We can easily check that $E_*(\hat{Y}_\pi^*) = \hat{Y}_\pi$, i.e. this is a conditionally unbiased estimator for $\hat{Y}_\pi$, with $E_*(\cdot)$ the expectation under the bootstrap procedure, conditionall on the original sample S. Furthermore, the bootstrap variance is

$$V_*(\hat{Y}_\pi^*) = N_1^2 N_2^2 \left[(1-f_1)\frac{S_1^{2*}}{n_1} + (1-f_2)\frac{S_2^{2*}}{n_2} + (1-f_1)(1-f_2)\frac{S_{12}^{2*}}{n_1 n_2} \right] \tag{6.2}$$

with

$$\begin{aligned} S_1^{2*} &= \frac{N_1(n_1-1)}{n_1(N_1-1)} \left[\hat{S}_1^2 + (1-f_2)\frac{\hat{S}_{12}^2}{n_2} \right], \\ S_2^{2*} &= \frac{N_2(n_2-1)}{n_2(N_2-1)} \left[\hat{S}_2^2 + (1-f_1)\frac{\hat{S}_{12}^2}{n_1} \right], \\ S_{12}^{2*} &= \frac{N_1(n_1-1)}{n_1(N_1-1)} \frac{N_2(n_2-1)}{n_2(N_2-1)} \hat{S}_{12}^2. \end{aligned} \tag{6.3}$$

Under our asymptotic framework where $n_1, n_2 \to \infty$, the bootstrap variance is therefore asymptotically identical to the unbiased variance estimator of $V_p(\hat{Y}_\pi)$, see equation (4.4).

6.2 Rescaled bootstrap method

In this section, we propose a rescaled bootstrap method for simple random sampling without replacement in each dimension, based on the method proposed by Rao et al. (1992). The Horvitz-Thompson estimator uses weights of the form $w_{k_1k_2} = w_{k_1}^{(1)} w_{k_2}^{(2)}$ for each unit $(k_1, k_2) \in \mathcal{U}$, where $w_{k_1}^{(1)} = 1/f_1$ and $w_{k_2}^{(2)} = 1/f_2$. The method by Rao and Wu consists in constructing bootstrap weights $w_{k_d}^{d*}$ for $d = 1, 2$ by multiplying the original weights by an adjustment factor $a_{k_d}^{d*}$, leading to the bootstrap weights $w_{k_d}^{d*} = a_{k_d}^{d*} w_{k_d}^{d}$. The method consists in sampling so-called multiplicities $(m_{k_d}^*)_{k_d \in S_d}$ from a multinomial distribution $\mathcal{M}\left(n_d^*; \frac{1}{n_d}, \ldots, \frac{1}{n_d}\right)$, leading to the adjustment factors

$$a_{k_d}^{d*} = 1 + \sqrt{\frac{n_d^*(1-f_d)}{n_d - 1}} \left(\frac{n_d m_{k_d}^{d*}}{n_d^*} - 1 \right) \text{ for any } k_d \in S_d, \tag{6.4}$$

Note that this method requires to choose the resampling size n_d^*. A customary choice is $n_d^* = n_d - 1$.

The proposed rescaled bootstrap for CCS consists in building bootstrap weights $w_{k_1k_2}^* = w_{k_1}^{(1*)} w_{k_2}^{(2*)}$, by using the rescaled bootstrap in each dimension. Some straightforward algebra leads to

$$V_*(\hat{Y}_\pi^*) = N_1^2 N_2^2 \left[(1-f_1)\frac{S_1^{2*}}{n_1} + (1-f_2)\frac{S_2^{2*}}{n_2} + (1-f_1)(1-f_2)\frac{S_{12}^{2*}}{n_1 n_2} \right], \tag{6.5}$$

with

$$\begin{aligned} S_1^{2*} &= \hat{S}_1^2 + (1-f_2)\frac{\hat{S}_{12}^2}{n_2}, \\ S_2^{2*} &= \hat{S}_2^2 + (1-f_1)\frac{\hat{S}_{12}^2}{n_1}, \\ S_{12}^{2*} &= \hat{S}_{12}^2. \end{aligned} \tag{6.6}$$

The fact that the method by Rao and Wu satisfies the second order moment condition in each dimension does not imply that the bootstrap variance estimator for the CCS is unbiased. Anyway, it guarantees that the leading variance terms are unbiasedly estimated.

7 Simulations

In this section, we carry out a simulation study under the $D = 2$ dimensional case, under a set-up inspired from Juillard et al. (2017). We are interested in evaluating the proposed variance estimators for the estimation of a total and a ratio. The basic generation model that we use is

$$y_{k_1k_2} = \mu + \sigma_1 U_{k_1} + \sigma_2 V_{k_2} + \sigma_{12} W_{k_1k_2}, \tag{7.1}$$

where the $U_{k_1}, V_{k_2}, W_{k_1k_2}$ are independent standard normal variables.

7.1 Variance estimation for a total

The values of the variable of interest in the first population are generated according to the model 7.1. We use $N_1 = N_2 = 1000$, $\mu = 200$, and $\sigma_1 = \sigma_2 = 5$. We let σ_{12} vary in $\{5, 10, 50\}$ to evaluate the influence of the variance due to interactions.

Given the population values, we repeat $T = 10000$ times the sample selection by means of CCS with simple random sampling in each dimension, with $n_1, n_2 \in \{5, 10, 100, 500\}$. We are interested in estimating the total Y, and the estimator for the t^{th} sample is denoted as $\hat{Y}^{(t)}$. We consider the *Unbiased* variance estimator, the simplified variance estimator $\hat{V}^{SIMP}(\hat{Y}_\pi)$ (*Simp*) given in equation (4.6), and the simplified variance estimator $\hat{V}^{SIMP2}(\hat{Y}_\pi)$ (*Simp2*) given in equation (4.7). We also compute the bootstrap variance estimator $\hat{V}^{(t)}_{\text{Gross}}$ associated to the pseudo-population bootstrap (*Gross*), $\hat{V}^{(t)}_{\text{RaoWu}}$ associated to the rescaled bootstrap (*Rao Wu*), and $\hat{V}^{(t)}_{\text{Skinner}}$ associated to a bootstrap procedure for with-replacement sampling in each dimension proposed by Skinner (2015) (*Skinner*). In each case, we use $B = 1000$ bootstrap resamples, and the bootstrap variance estimator $\hat{V}^{(t)}$ is computed according to the formula

$$\hat{V}^{(t)} = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{Y}^{(b*)(t)} - \frac{1}{B} \sum_{b=1}^{B} \hat{Y}^{(b*)(t)} \right)^2 . \tag{7.2}$$

To evaluate the proposed variance estimators, we consider the Monte-Carlo relative bias

$$\mathrm{RB}(\hat{V}) = 100 \times \frac{\frac{1}{T}\sum_{t=1}^{T} \hat{V}^{(t)} - V}{V}, \tag{7.3}$$

where V is an approximation of the true variance of $\hat{Y}$, computed from an independent run of $100,000$ simulations. We also compute the Monte-Carlo relative stability

$$\mathrm{RS}(\hat{V}) = 100 \times \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{V}^{(t)} - V)^2}}{V}. \tag{7.4}$$

Finally, we compute confidence intervals for the bootstrap methods according to the reverse percentile method. More precisely, assume that the bootstrap estimates are reordered as $\hat{Y}^{(1*)(t)} \leq \hat{Y}^{(2*)(t)} \leq \cdots \leq \hat{Y}^{(B*)(t)}$. Then the reverse percentile confidence interval is obtained by considering $(\hat{Y}^* - \hat{Y})$ as an approximation of the distribution of $\hat{Y} - Y$:

$$\mathrm{CI}_{\alpha}^{REV,(t)}(Y) = \left[2\hat{Y}^{(t)} - \hat{Y}^{(U*)(t)}, 2\hat{Y}^{(t)} - \hat{Y}^{(L*)(t)}\right]. \tag{7.5}$$

For $\mathrm{CI}_{\alpha}^{REV,(t)}(Y)$ to have the level α, the values of L and U are given by $L = \lfloor \frac{1-\alpha}{2} B \rfloor$ and $U = \lfloor \frac{1+\alpha}{2} B \rfloor$.

7.2 Variance estimation for a ratio

In the second population, the values for three variables of interest are generated accrding to the model

$$z_{k_1k_2} = \sigma_1 U_{k_1} + \sigma_2 V_{k_2} + \sigma_{12} W_{k_1k_2}, \tag{7.6}$$

$$x_{k_1k_2} = \mu_x + \alpha z_{k_1,k_2} + (1-\alpha)(\sigma_1' U_{k_1}' + \sigma_2' V_{k_2}' + \sigma_{12}' W_{k_1k_2}') \tag{7.7}$$

$$y_{k_1k_2} = \mu_y + \beta z_{k_1,k_2} + (1-\beta)(\sigma_1'' U_{k_1}'' + \sigma_2'' V_{k_2}'' + \sigma_{12}'' W_{k_1k_2}''). \tag{7.8}$$

The variables $U_{k_1}, U_{k_1}', U_{k_1}''$, $V_{k_2}, V_{k_2}', V_{k_2}''$, and $W_{k_1k_2}, W_{k_1k_2}', W_{k_1k_2}''$, are independent standard normal variables. The parameter α in equation (7.7) is used to control the correlation between variables x and z. Similarly, the parameter β in equation (7.7) is used to control the correlation between variables y and z. Thus, high values of α and β will lead to high correlations of the variables x and y. We used $\mu_x = 100$, $\mu_y = 300$, and $\sigma_1 = \sigma_2 = \sigma_{12} = \sigma_1' = \sigma_1'' = 5, \sigma_2' = \sigma_2'' = 10, \sigma_{12}' = \sigma_{12}'' = 15$. Finally we considered $\alpha \in \{0.1, 0.5, 0.8\}$ and $\beta = 0.5$.

A similar procedure is then done as the previous subsection in order to compute relative biases of the different methods, but this time applied to the substitution estimator $\hat{R} = \hat{Y}/\hat{X}$ of the ratio $R = Y/X$. For each sample $t \in \{1, \ldots, T\}$, it is possible to produce bootstrap estimations $\hat{R}^{(b*)(t)} = \hat{Y}^{(b*)(t)}/\hat{X}^{(b*)(t)}$ from the bootstrap estimations $(\hat{Y}^{(b*)(t)}, \hat{X}^{(b*)(t)})$ of $(\hat{Y}^{(t)}, \hat{X}^{(t)})$. For the non-bootstrap methods, a linearization technique will be applied, replacing the variable $(y_{k_1k_2}, x_{k_1k_2})$ by its associated linearization $\nu_{k_1k_2} = (y_{k_1k_2} - \hat{R}x_{k_1k_2})/\hat{X}$.

7.3 Results

The simulation results for the estimation of a total are gathered in Table 7.3. We can observe the influence of the value of σ_{12} in the quality of the proposed estimations. A higher value of σ_{12} against σ_1 and σ_2 tends to increase the importance of S_{12}^2 against S_1^2 and S_2^2 and thus leads to a higher relative bias of the proposed simplified variance estimators. On the other hand, we observe that the relative biases of the proposed variance estimators tend towards 0 as the sample sizes increase, as expected. Overall, the Rao-Wu bootstrap method and the Skinner bootstrap method Skinner (2015) have quite similar results for low sampling rates, but the relative bias of Skinner's method increases with the sample sizes. This is due to the fact taht this method relies on the assumption of negligible sampling fractions, an assumption which is not needed for the proposed variance estimators. The simulation results also confirm that, in accordance with our theoretical results, the Rao-Wu bootstrap variance estimator tends to be conservative, which is not the case for the Gross method.

The simulation results for the estimation of a ratio are given in Table 7.3. The conclusions are very similar. We observe that a high value of α leads to better results, which can be explained by the fact that increasing α diminishes the overall interaction noise, by increasing the part generated with a value of $\sigma_{12} = 5$ against the part generated with a value of $\sigma_{12} = 15$.

A Proofs for Section 3.1

A.1 Proof of equation (3.1)

From equation (2.6), a straightforward computation leads to the result:

$$V_p(\hat{Y}) = \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} Cov_p\left(y_{\mathbf{k}}\prod_{d=1}^{D} w^d_{k_d}, y_{\mathbf{l}}\prod_{d=1}^{D} w^d_{l_d}\right) \tag{A.1}$$

$$= \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} y_{\mathbf{k}}y_{\mathbf{l}}\left[E_p\left(\prod_{d=1}^{D} w^d_{k_d}w^d_{l_d}\right) - E_p\left(\prod_{d=1}^{D} w^d_{k_d}\right)E_p\left(\prod_{d=1}^{D} w^d_{l_d}\right)\right] \tag{A.2}$$

$$= \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} y_{\mathbf{k}}y_{\mathbf{l}}\left[\prod_{d=1}^{D} E_p(w^d_{k_d}w^d_{l_d}) - \prod_{d=1}^{D} E_p(w^d_{k_d})E_p(w^d_{l_d})\right] \tag{A.3}$$

where the last line follows from the independence of the weights in each dimension.

A.2 Proof of Proposition 1

Equations (3.2) as well as (3.5) in the first part of the proposition are general properties of the Hoeffding-Sobol decomposition applied to $\hat{Y}$ seen as a function of the independent variables $(S_d)_{d=1}^{D}$, the general formula corresponding to (3.3). It remains to check if the equation (3.4) is true in our case by computing for each $I \subseteq \{1,\ldots,D\}$ the associated $\hat{Y}^I$ starting from the equation (3.3)

$$\hat{Y}^I = \sum_{I'\in\mathcal{P}(I)} (-1)^{|I|-|I'|}E_p(\hat{Y}|(S_d)_{d\in I'}) \tag{A.4}$$

$$= \sum_{I'\in\mathcal{P}(I)} (-1)^{|I|-|I'|}\sum_{\mathbf{k}\in\mathcal{U}} y_{\mathbf{k}}\prod_{d=1}^{D}\left[(w^d_{k_d} - E_p(w^d_{k_d}))\mathbf{1}_{(d\in I')} + E_p(w^d_{k_d})\right] \tag{A.5}$$

$$= \sum_{\mathbf{k}\in\mathcal{U}} y_{\mathbf{k}}\sum_{I'\in\mathcal{P}(I)} (-1)^{|I|-|I'|}\sum_{I''\subseteq\{1,\ldots,D\}}\prod_{d\in I''}(w^d_{k_d} - E_p(w^d_{k_d}))\mathbf{1}_{(d\in I')}\prod_{d\notin I''}E_p(w^d_{k_d}). \tag{A.6}$$

Table 1: Relative biases, relative stabilities and coverage probabilities of multiple estimators of the variance of $\hat{Y}_\pi$ in percentage

			RB						RS						$CI_{95\%}$		
σ_{12}	n_1	n_2	Gross	RaoWu	Skinner	Unbiased	Simp	Simp2	Gross	RaoWu	Skinner	Unbiased	Simp	Simp2	Gross	RaoWu	Skinner
5	5	5	−6	16	21	2	−12	10	46	57	59	57	56	53	91.1	94.2	94.2
	10	10	−1	8	11	−2	−4	6	32	35	37	35	36	34	93.4	94.5	94.9
	10	100	−5	2	5	0.6	0.1	2	41	46	47	45	45	45	91.6	93.0	93.0
	100	100	−0.3	0.6	12	−0.1	−0.8	0.2	10	11	17	10	10	10	95.1	94.9	96.0
	500	500	−0.8	−0.6	99	−0.8	−0.8	−0.7	5	6	99	3	3	3	95.1	94.8	99.5
10	5	5	23	59	60	0.9	−30	30	57	88	90	66	73	71	94.7	96.9	96.8
	10	10	17	33	35	−0.1	−17	16	40	51	53	39	43	42	95.6	96.9	97.1
	10	100	−4	6	9	−0.8	−4	3	41	45	46	44	44	44	92.3	93.3	94.0
	100	100	2	2	14	−0.9	−3	0.6	11	11	19	10	10	10	95.2	95.0	96.2
	500	500	0.5	0.3	101	−0.1	−0.1	0.1	6	5	102	3	3	3	95.1	94.7	99.4
50	5	5	109	180	182	−0.8	−90	89	135	206	208	100	145	131	98.6	99.5	99.5
	10	10	128	161	165	−1	−82	81	140	173	177	60	105	101	99.5	99.8	99.7
	10	100	69	83	95	−1	−42	41	80	95	106	44	62	60	98.3	98.9	99.0
	100	100	56	58	79	0.5	−28	29	58	60	80	13	31	32	98.4	98.5	99.0
	500	500	8	8	125	0.0	−4	4	10	10	126	3	5	6	95.9	95.7	99.6

Table 2: Relative biases, relative stabilities and coverage probabilities of multiple estimators of the variance of $\hat{R}$ in percentage

			RB						RS						$CI_{95\%}$		
α	n_1	n_2	Gross	RaoWu	Skinner	Unbiased	Simp	Simp2	Gross	RaoWu	Skinner	Unbiased	Simp	Simp2	Gross	RaoWu	Skinner
0.1	5	5	25	63	63	6	−24	28	73	108	112	78	80	82	93.9	96.5	96.7
	10	10	15	32	30	0.5	−18	18	46	59	55	43	46	50	95.3	96.5	96.2
	10	100	3	11	17	−0.7	−3	6	33	38	41	35	37	36	94.3	94.7	95.1
	100	100	2	3	15	−0.2	−2	0.9	13	13	20	11	12	11	94.7	94.4	96.7
	500	500	0.5	0.7	101	0.3	0.1	0.3	6	6	102	4	4	4	94.5	95.1	99.2
0.5	5	5	17	53	51	0.6	−24	22	61	88	85	65	71	71	94.2	95.4	96.4
	10	10	14	28	31	2	−13	14	43	51	56	44	44	44	95.4	96.0	94.8
	10	100	−0.1	7	13	−0.6	−4	6	36	39	41	38	39	39	92.4	94.4	94.6
	100	100	1	3	14	−0.2	−1	0.6	12	12	19	11	11	11	94.1	94.7	95.7
	500	500	0.4	0.4	102	0.4	−0.2	0.6	6	6	102	4	4	4	94.5	94.4	99.5
0.8	5	5	1	30	31	0.5	−14	15	47	68	68	57	60	60	92.4	94.8	95.2
	10	10	5	18	18	1	−8	8	35	42	43	38	38	38	94.4	95.2	95.4
	10	100	−5	4	7	0.4	−1	2	37	41	42	41	41	41	92.2	93.1	93.2
	100	100	−0.1	0.8	13	−0.4	−2	0.1	11	11	17	9	9	10	94.4	95.3	95.7
	500	500	−0.4	0.1	100	−0.2	−0.4	−0.1	6	6	100	3	3	3	94.8	94.8	99.4

By observing that for any $I'' \subseteq \{1,\ldots,D\}$, $\prod_{d\in I''} \mathbf{1}_{(d\in I')} = \mathbf{1}_{(I''\subset I')}$ and that $\sum_{I'\in\mathcal{P}(I)}(-1)^{|I|-|I'|}\mathbf{1}_{(I''\subseteq I')} = \mathbf{1}_{(I''=I)}$, one can conclude by rearranging the sum that

$$\hat{Y}^I = \sum_{\mathbf{k}\in\mathcal{U}} y_{\mathbf{k}} \sum_{I''\subseteq\{1,\ldots,D\}} \prod_{d\in I''} (w^d_{k_d} - E_p(w^d_{k_d})) \prod_{d\notin I''} E_p(w^d_{k_d}) \sum_{I'\in\mathcal{P}(I)} (-1)^{|I|-|I'|}\mathbf{1}_{(I''\subseteq I')} \tag{A.7}$$

$$= \sum_{\mathbf{k}\in\mathcal{U}} y_{\mathbf{k}} \prod_{d\in I}(w^d_{k_d} - E_p(w^d_{k_d})) \prod_{d\notin I} E_p(w^d_{k_d}), \tag{A.8}$$

from which we can deduce the value of $V_p(\hat{Y}^I)$ for $I \neq \varnothing$:

$$V_p(\hat{Y}^I) = \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} y_{\mathbf{k}}y_{\mathbf{l}} Cov_p\left(\prod_{d\in I}(w^d_{k_d} - E_p(w^d_{k_d})), \prod_{d\in I}(w^d_{l_d} - E_p(w^d_{l_d}))\right) \prod_{d\notin I} E_p(w^d_{k_d})E_p(w^d_{l_d}) \tag{A.9}$$

$$= \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} y_{\mathbf{k}}y_{\mathbf{l}} E_p\left(\prod_{d\in I}(w^d_{k_d} - E_p(w^d_{k_d}))(w^d_{l_d} - E_p(w^d_{l_d}))\right) \prod_{d\notin I} E_p(w^d_{k_d})E_p(w^d_{l_d}) \tag{A.10}$$

$$= \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} y_{\mathbf{k}}y_{\mathbf{l}} \prod_{d\in I} Cov_p(w^d_{k_d}, w^d_{l_d}) \prod_{d\notin I} E_p(w^d_{k_d})E_p(w^d_{l_d}). \tag{A.11}$$

where the last line follows from the independence of the weights in each dimension.

A.3 Proof of Corollary 1

It suffices to show that for every non-empty subset $I \subseteq \{1,\ldots,D\}$, $V_p(\hat{Y}^I)$ is unbiasedly estimated by $\hat{V}_p(\hat{Y}^I)$. It is possible to rewrite the sum over $\mathcal{U}$ by introducing the sample membership indicators. We then obtain

$$E_p(\hat{V}_p(\hat{Y}^I)) = \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} y_{\mathbf{k}}y_{\mathbf{l}} E_p\left[\prod_{d\in I} \widehat{Cov}^d(w^d_{k_d}, w^d_{l_d}) \prod_{d\notin I} \frac{\delta^d_{k_d}\delta^d_{l_d}}{\pi^d_{k_d,l_d}}\right] \tag{A.12}$$

$$= \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} y_{\mathbf{k}}y_{\mathbf{l}} \prod_{d\in I} E_p\left[\widehat{Cov}^d(w^d_{k_d}, w^d_{l_d})\right] \prod_{d\notin I} \frac{E_p(\delta^d_{k_d}\delta^d_{l_d})}{\pi^d_{k_d,l_d}} \tag{A.13}$$

$$= \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} y_{\mathbf{k}}y_{\mathbf{l}} \prod_{d\in I} Cov_p(w^d_{k_d}, w^d_{l_d}) \tag{A.14}$$

$$= V_p(\hat{Y}^I). \tag{A.15}$$

where the second line follows from the fact that for a fixed $\mathbf{k}$ and $\mathbf{l} \in \mathcal{U}$, each $\widehat{Cov}^d(w^d_{k_d}, w^d_{l_d})$ and $\delta^d_{k_d}\delta^d_{l_d}$ are built from the sample S_d only and are thus all independent from each other as long as they correspond to different dimensions.

B Proofs for Section 3.2

B.1 Proof of Corollary 2

Equation (3.12) is a direct consequence of equation (3.4), applied for $w^d_{k_d} = \delta^d_{k_d}/\pi^d_{k_d}$. Similarly, equation (3.13) follows from equation (3.6). Finally, equation (3.14) follows by applying equation (3.9) with

$$\widehat{Cov}^d(w^d_k, w^d_l) = \frac{\Delta^d_{k_d l_d}}{\pi^d_{k_d,l_d}} \frac{\delta^d_{k_d}\delta^d_{l_d}}{\pi^d_{k_d}\pi^d_{l_d}}. \tag{B.1}$$

B.2 Proof of Proposition 2

Let us introduce the partition $\mathcal{U}_I \times \mathcal{U}_I = \bigcup_{I' \in \mathcal{P}(I)} P_{I'}$ for a given non-empty $I \subseteq \{1, \dots, D\}$, where we defined for each $I' \in \mathcal{P}(I)$, the set $P_{I'}$ by

$$P_{I'} = \{\mathbf{k}', \mathbf{l}' \in \mathcal{U}_I,\ \forall d \in I, (k'_d \neq l'_d \Leftrightarrow d \in I')\} \tag{B.2}$$

furthermore, note that we can identify every element $\mathbf{k} \in \mathcal{U}$ with a couple $(\mathbf{k}_1, \mathbf{k}_2) \in \mathcal{U}_I \times \mathcal{U}_{I^c}$ where we introduced $\mathcal{U}_{I^c} = \prod_{d \notin I} \mathcal{U}_d$. More precisely, the coordinates of $\mathbf{k}_1$ are given by the coordinates of $\mathbf{k}$ associated to the dimensions in I and the coordinates of $\mathbf{k}_2$ are given by the coordinates of $\mathbf{k}$ associated to the dimensions that are not in I. We will therefore denote by $y_{\mathbf{k}_1\mathbf{k}_2}$ the quantity $y_{\mathbf{k}}$. This splitting allows us for example to rewrite the subtotal $Y_{\mathbf{k}'}$ in the form $Y_{\mathbf{k}'} = \sum_{\mathbf{l}' \in \mathcal{U}_{I^c}} y_{\mathbf{k}'\mathbf{l}'}$. Now let us fix some subset $I' \in \mathcal{P}(I)$. We can verify using the Cauchy-Schwarz inequality that

$$\sum_{\mathbf{k}' \in \mathcal{U}_I} Y^2_{\mathbf{k}'} = \sum_{\mathbf{k}' \in \mathcal{U}_I} \left(\sum_{\mathbf{l}' \in \mathcal{U}_{I^c}} y_{\mathbf{k}'\mathbf{l}'} \right)^2 \leq \left(\prod_{d \notin I} N_d \right) \sum_{\mathbf{k}' \in \mathcal{U}_I} \sum_{\mathbf{l}' \in \mathcal{U}_{I^c}} y^2_{\mathbf{k}'\mathbf{l}'} \leq \alpha \frac{N^2}{\prod_{d \in I} N_d} \tag{B.3}$$

by using the assumption (H1).

Now we can similarly identify an element $\mathbf{k}' \in \mathcal{U}_I$ with a couple $(\mathbf{k}'_1, \mathbf{k}'_2) \in \mathcal{U}_{I'} \times \mathcal{U}_{I \setminus I'}$. We will therefore denote by $Y_{\mathbf{k}'_1\mathbf{k}'_2}$ the subtotal $Y_{\mathbf{k}'}$. Thus, using again the Cauchy-Scharz inequality, it is now easy to verify that

$$\sum_{(\mathbf{k}',\mathbf{l}')\in P_{I'}} Y_{\mathbf{k}'}Y_{\mathbf{l}'} \leq \sum_{\substack{\mathbf{k}'',\mathbf{l}''\in\mathcal{U}_{I'}\\ \mathbf{m}''\in\mathcal{U}_{I\setminus I'}}} Y_{\mathbf{k}''\mathbf{m}''}Y_{\mathbf{l}''\mathbf{m}''} \tag{B.4}$$

$$= \sum_{\mathbf{m}''\in\mathcal{U}_{I\setminus I'}} \left(\sum_{\mathbf{k}''\in\mathcal{U}_{I'}} Y_{\mathbf{k}''\mathbf{m}''}\right)^2 \tag{B.5}$$

$$\leq \left(\prod_{d\in I'} N_d\right) \sum_{\mathbf{m}''\in\mathcal{U}_{I\setminus I'}} \sum_{\mathbf{k}''\in\mathcal{U}_{I'}} Y^2_{\mathbf{k}''\mathbf{m}''} \tag{B.6}$$

$$\leq \alpha \frac{N^2}{\prod_{d\in I\setminus I'} N_d}. \tag{B.7}$$

Furthermore we know by construction of $P_{I'}$ that for any $(\mathbf{k}',\mathbf{l}')\in P_{I'}$, $k'_d \neq l'_d$ if and only if $d\in I'$ and thus using assumption (H3)-(H4) we get that

$$\forall(\mathbf{k}',\mathbf{l}')\in P_{I'},\ \forall d\in I',\ \left|\frac{\Delta^d_{k'_d,l'_d}}{\pi^d_{k'_d}\pi^d_{l'_d}}\right| \leq \frac{\gamma_d}{\lambda_d^2}\frac{1}{n_d}. \tag{B.8}$$

On the other hand, when $d\notin I'$, we have that $k'_d = l'_d$ and thus

$$\forall(\mathbf{k}',\mathbf{l}')\in P_{I'},\ \forall d\notin I',\ \left|\frac{\Delta^d_{k'_d,l'_d}}{\pi^d_{k'_d}\pi^d_{l'_d}}\right| = \frac{1-\pi^d_{k'_d}}{\pi^d_{k'_d}} \leq \frac{1}{\lambda_d}\frac{N_d}{n_d}. \tag{B.9}$$

From there, we can build an upper bound of the sum

$$\sum_{(\mathbf{k}',\mathbf{l}')\in P_{I'}} Y_{\mathbf{k}'}Y_{\mathbf{l}'}\prod_{d\in I}\frac{\Delta^d_{k'_d l'_d}}{\pi^d_{k'_d}\pi^d_{l'_d}} = \sum_{(\mathbf{k}',\mathbf{l}')\in P_{I'}} Y_{\mathbf{k}'}Y_{\mathbf{l}'}\prod_{d\in I'}\frac{\Delta^d_{k'_d l'_d}}{\pi^d_{k'_d}\pi^d_{l'_d}}\prod_{d\in I\setminus I'}\frac{\Delta^d_{k'_d,l'_d}}{\pi^d_{k'_d}\pi^d_{l'_d}} \tag{B.10}$$

$$= O\left(\frac{N^2}{\prod_{d\in I\setminus I'} N_d}\prod_{d\in I'}\frac{1}{n_d}\prod_{d\in I\setminus I'}\frac{N_d}{n_d}\right) \tag{B.11}$$

$$= O\left(\frac{N^2}{\prod_{d\in I} n_d}\right). \tag{B.12}$$

Now, we can finally conclude that

$$V_p(\hat{Y}^I) = \sum_{I'\in\mathcal{P}(I)}\left(\sum_{(\mathbf{k}',\mathbf{l}')\in P_{I'}} Y_{\mathbf{k}'}Y_{\mathbf{l}'}\prod_{d\in I}\frac{\Delta^d_{k'_d l'_d}}{\pi^d_{k'_d}\pi^d_{l'_d}}\right) = O\left(\frac{N^2}{\prod_{d\in I} n_d}\right). \tag{B.13}$$

B.3 Proof of Proposition 3

We can write

$$\hat{Y}_\pi - Y = \sum_{d=1}^{D} \hat{Y}_\pi^{\{d\}} + \Delta, \text{where } \Delta = \sum_{\substack{I \subset \{1,\ldots,D\} \\ Card(I) \geq 2}} \hat{Y}_\pi^{I}.$$

We obtain

$$V_p(\hat{Y}_\pi) = \sum_{d=1}^{D} V_p(\hat{Y}_\pi^{\{d\}}) + V_p(\Delta). \tag{B.14}$$

It follows from Proposition 2 that

$$V_p\left(N^{-1}\hat{Y}_\pi^{\{d\}}\right) = O(n_m^{-1}) \quad \text{and} \quad V_p(N^{-1}\Delta) = o(n_m^{-1}). \tag{B.15}$$

Therefore, we obtain (3.17). Note that from equations (B.14) and (B.15), the constants in assumption (H7) are such that $\sum_{d=1}^{D}(\gamma_d)^2 = 1$.

From equation (B.14), we also obtain

$$\frac{\hat{Y}_\pi - Y}{\sqrt{V_p(\hat{Y}_\pi)}} = \underbrace{\sum_{d=1}^{D} \gamma_d \frac{\hat{Y}_\pi^{\{d\}}}{\sqrt{V_p(\hat{Y}_\pi^{\{d\}})}}}_{\Delta_1} + \underbrace{\sum_{d=1}^{D} \left(\sqrt{\frac{V_p(\hat{Y}_\pi^{\{d\}})}{V_p(\hat{Y}_\pi)}} - \gamma_d\right) \frac{\hat{Y}_\pi^{\{d\}}}{\sqrt{V_p(\hat{Y}_\pi^{\{d\}})}}}_{\Delta_2} + \underbrace{\frac{\Delta}{\sqrt{V_p(\hat{Y}_\pi)}}}_{\Delta_3}. \tag{B.16}$$

By using assumption (H5) and the right-hand side of (B.14), we have $\Delta_3 \longrightarrow_{Pr} 0$, where $\longrightarrow_{Pr}$ stands for the convergence in probability. Also, by using Assumptions (H6) and (H7) and the Slutsky theorem, we obtain that $\Delta_2 \longrightarrow_{Pr} 0$. Finally, since the variables $\hat{Y}_\pi^{\{d\}}$ are independent, we obtain from assumption (H6) by standard arguments that $\Delta_1 \longrightarrow \mathcal{N}(0,1)$, which completes the proof.

B.4 Proof of Proposition 4

Let us fix a non-empty subset $I \subseteq \{1, \ldots, D\}$. We can write using the definitions of $\hat{V}^{PLUG}(\hat{Y}_\pi^I)$

$$E_p\left[\hat{V}^{PLUG}(\hat{Y}_\pi^I)\right] = \sum_{\mathbf{k}',\mathbf{l}' \in \mathcal{U}_I} \frac{E_p(\hat{Y}_{\mathbf{k}'}\hat{Y}_{\mathbf{l}'}\delta_{\mathbf{k}'}\delta_{\mathbf{l}'})}{\pi_{\mathbf{k}'}\pi_{\mathbf{l}'}} \frac{\Delta_{\mathbf{k}'\mathbf{l}'}}{\pi_{\mathbf{k}'\mathbf{l}'}} \tag{B.17}$$

$$= \sum_{\mathbf{k}',\mathbf{l}' \in \mathcal{U}_I} \frac{E_p(\hat{Y}_{\mathbf{k}'}\hat{Y}_{\mathbf{l}'})E_p(\delta_{\mathbf{k}'}\delta_{\mathbf{l}'})}{\pi_{\mathbf{k}'}\pi_{\mathbf{l}'}} \frac{\Delta_{\mathbf{k}'\mathbf{l}'}}{\pi_{\mathbf{k}'\mathbf{l}'}} \tag{B.18}$$

$$= \sum_{\mathbf{k}',\mathbf{l}' \in \mathcal{U}_I} \frac{E_p(\hat{Y}_{\mathbf{k}'}\hat{Y}_{\mathbf{l}'})}{\pi_{\mathbf{k}'}\pi_{\mathbf{l}'}} \Delta_{\mathbf{k}'\mathbf{l}'} \tag{B.19}$$

Where the line (B.18) follows from the fact that $\delta_{\mathbf{k}'}$ and $\delta_{\mathbf{l}'}$ are functions of $(S_d)_{d\in I}$ whereas $\hat{Y}_{\mathbf{k}'}$ and $\hat{Y}_{\mathbf{l}'}$ are functions of $(S_d)_{d\notin I}$, and are therefore independent. We can then similarly expand the sum by replacing $\hat{Y}_{\mathbf{k}'}$ and $\hat{Y}_{\mathbf{l}'}$ by their definitions in (3.23):

$$E_p\left[\hat{V}^{PLUG}(\hat{Y}_\pi^I)\right] = \sum_{\mathbf{k}',\mathbf{l}' \in \mathcal{U}_I} \sum_{\substack{\mathbf{k},\mathbf{l} \in \mathcal{U} \\ \forall d\in I, k_d=k'_d, l_d=l'_d}} \frac{y_\mathbf{k}}{\pi_\mathbf{k}} \frac{y_\mathbf{l}}{\pi_\mathbf{l}} \prod_{d\in I} \Delta^d_{k_d l_d} E_p\left(\prod_{d\notin I} \delta^d_{k_d}\delta^d_{l_d}\right) \tag{B.20}$$

$$= \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} \frac{y_\mathbf{k}}{\pi_\mathbf{k}} \frac{y_\mathbf{l}}{\pi_\mathbf{l}} \prod_{d\in I} \Delta^d_{k_d l_d} \prod_{d\notin I} \pi^d_{k_d l_d}. \tag{B.21}$$

Now we can furthermore observe that for any $\mathbf{k}, \mathbf{l} \in \mathcal{U}$, we have the following identity

$$\prod_{d\notin I} \pi^d_{k_d l_d} = \prod_{d\notin I} (\Delta^d_{k_d l_d} + \pi^d_{k_d}\pi^d_{l_d}) = \sum_{I'\subseteq I^c} \prod_{d\in I'} \Delta^d_{k_d l_d} \prod_{d\in I^c\setminus I'} \pi^d_{k_d}\pi^d_{l_d} \tag{B.22}$$

which can then be substituted in the previous result to give

$$E_p\left[\hat{V}^{PLUG}(\hat{Y}_\pi^I)\right] = \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} \frac{y_\mathbf{k}}{\pi_\mathbf{k}} \frac{y_\mathbf{l}}{\pi_\mathbf{l}} \prod_{d\in I} \Delta^d_{k_d l_d} \sum_{I'\subseteq I^c} \prod_{d\in I'} \Delta^d_{k_d l_d} \prod_{d\in I^c\setminus I'} \pi^d_{k_d}\pi^d_{l_d} \tag{B.23}$$

$$= \sum_{I'\subseteq I^c} \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} \frac{y_\mathbf{k}}{\prod_{d\in I\cup I'} \pi^d_{k_d}} \frac{y_\mathbf{l}}{\prod_{d\in I\cup I'} \pi^d_{k_d}} \prod_{d\in I\cup I'} \Delta^d_{k_d l_d} \tag{B.24}$$

Finally we can note that $\{I \cup I',\ I' \subseteq I^c\} = \big\{I',\ I \subseteq I' \subseteq \{1,\dots,D\}\big\}$. From this observation, we can make a change of variable and conclude

$$E_p\left[\hat{V}^{PLUG}(\hat{Y}_\pi^I)\right] = \sum_{I'\supseteq I} \sum_{\mathbf{k},\mathbf{l}\in\mathcal{U}} \frac{y_\mathbf{k}}{\prod_{d\in I'} \pi^d_{k_d}} \frac{y_\mathbf{l}}{\prod_{d\in I'} \pi^d_{k_d}} \prod_{d\in I'} \Delta^d_{k_d l_d} \tag{B.25}$$

$$= \sum_{I'\supseteq I} \sum_{\substack{\mathbf{k}'',\mathbf{l}''\in\mathcal{U}_{I'} \\ \mathbf{k},\mathbf{l}\in\mathcal{U} \\ \forall d\in I', k_d = k''_d, l_d = l''_d}} \frac{y_\mathbf{k}}{\prod_{d\in I'} \pi^d_{k_d}} \frac{y_\mathbf{l}}{\prod_{d\in I'} \pi^d_{k_d}} \prod_{d\in I'} \Delta^d_{k_d l_d} \tag{B.26}$$

$$= \sum_{I'\supseteq I} \sum_{\mathbf{k}'',\mathbf{l}''\in\mathcal{U}_{I'}} \frac{Y_{\mathbf{k}''}}{\pi_{\mathbf{k}''}} \frac{Y_{\mathbf{l}''}}{\pi_{\mathbf{l}''}} \Delta_{\mathbf{k}''\mathbf{l}''} \tag{B.27}$$

$$= \sum_{I'\supseteq I} V_p(\hat{Y}_\pi^{I'}). \tag{B.28}$$

References

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

Cochran, W. G. (1977). *Sampling techniques.* John Wiley & Sons, New York-London-Sydney, third edition. Wiley Series in Probability and Mathematical Statistics.

Dalén, J. and Ohlsson, E. (1995). Variance estimation in the swedish consumer price index. *Journal of Business & Economic Statistics*, 13(3):347–356.

Fuller, W. A. (2011). *Sampling statistics*, volume 560. John Wiley & Sons.

Gross, S. (1980). Median estimation in sample surveys. *Proceedings of the Section on Survey Research Methods*, pages 181–184.

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

Hoeffding, W. (1948). A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19(3):293 – 325.

Juillard, H., Chauvet, G., and Ruiz-Gazen, A. (2017). Estimation under cross-classified sampling with application to a childhood survey. *Journal of the American Statistical Association*, 112(518):850–858.

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

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

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

Skinner, C. (2015). Cross-classified sampling: some estimation theory. *Statistics & Probability Letters*, 104:163–168.