SSC Annual Meeting, May 2023
Proceedings of the Survey Methods Section

# ROBUST IMPUTATION IN THE PRESENCE OF INFLUENTIAL UNITS IN SURVEYS

Jia Ning Zhang [1], David Haziza[2], and Sixia Chen [3]

## ABSTRACT

Item nonresponse in surveys is often treated by some form of single imputation. In some cases, one faces the problem of influential units in the sample. This problem is especially acute in business surveys that collect economic variables whose distributions are highly skewed. In the presence of influential units, the classical imputed estimators are approximately unbiased if the first moment of the imputation model is correctly specified but they may be very unstable. Therefore, it is desirable to develop robust imputation methods that produce biased but more stable imputed estimators, i.e., estimators whose mean square error is smaller than that of the corresponding non-robust counterparts. In this paper, we consider three robust imputed estimators that rely on an adaptative tuning constant. We present the results of a simulation that suggest that the proposed methods perform well in terms of efficiency for a wide class of distributions.

KEY WORDS: item nonresponse, imputed estimator, adaptative tuning constant, influential unit.

## RÉSUMÉ

La non-réponse partielle dans les enquêtes est souvent traitée par une forme ou une autre d'imputation simple. Dans certains cas, on est confronté au problème des unités d'échantillonnage influentes. Ce problème est particulièrement aigu dans les enquêtes auprès des entreprises qui collectent des variables économiques dont les distributions sont fortement asymétriques. En présence d'unités influentes, les estimateurs imputés classiques sont approximativement sans biais si le premier moment du modèle d'imputation est correctement spécifié, mais ils peuvent s'avérer très instables. Il est donc souhaitable de développer des méthodes d'imputation robustes qui produisent des estimateurs imputés biaisés mais plus stables, c'est-à-dire des estimateurs dont l'erreur quadratique moyenne est inférieure à celle des estimateurs non robustes. Dans cet article, nous examinons trois estimateurs d'imputation robustes qui reposent sur un seuil adaptatif. Nous présentons les résultats d'une simulation qui suggèrent que les méthodes proposées sont performantes en termes d'efficacité pour une large classe de distributions.

MOTS CLÉS : non-réponse partielle, estimateur imputé, seuil adaptatif, unité influente.

## 1 INTRODUCTION

Many organizations worldwide, including national statistical offices (e.g., Statistics Canada), market research firms, and polling organizations, conduct surveys to collect valuable information. The common issue of missing data due to nonresponse poses a challenge in surveys, and without appropriate statistical treatment, point estimates may be greatly affected by nonresponse bias. In this paper, we focus on the problem of item nonresponse, which is typically treated by some form of imputation. In addition to missing values, some surveys suffer from the presence of influential units in the set of responding units. The problem of influential units is especially acute in business surveys as the distribution of economic variables tends to be highly skewed. A unit is said to be influential if its inclusion or exclusion in the calculation of point estimates may have a drastic impact on their magnitude. At this stage, we distinguish influential units, which are accurately recorded values and may represent other similar units in the set of nonrespondents or in the non-sampled part of the population, from gross measurement errors, which are typically identified and corrected during the data-editing stage.

[1] Jia Ning Zhang, STEM Complex, 150 Louis-Pasteur Pvt Ottawa, ON, Canada, K1N 6N5, jzhan434@uottawa.ca
[2] David Haziza, STEM Complex, 150 Louis-Pasteur Pvt Ottawa, ON, Canada, K1N 6N5, dhaziza@uottawa.ca
[3] Sixia Chen, University of Oklahoma Health Sciences Center, 801 NE 13th St, Room 325, Oklahoma City, Oklahoma, USA, 73126-0901, Sixia-Chen@ouhsc.edu

If the first moment of the imputation model is correctly specified, the resulting imputed estimator of a population total is consistent for the true total. However, point estimators may be highly unstable when influential units belong to the set of observed data. It is therefore desirable to develop robust estimators that exhibit a smaller mean square error than their corresponding counterparts. This is achieved at the expense of introducing a bias. Therefore, the treatment of influential units involves a bias-variance trade-off.

To cope with the problem of influential units, it may be tempting to select one of the many robust estimation procedures (e.g., $M$-estimators) that have been developed in the context of infinite populations. However, as we illustrate empirically in Section 3, the blind application of these methods to survey data may lead to unsatisfactory results. Indeed, the classical robust methods use a fixed tuning constant, for instance $c = 1.345$ for $M$-estimators based on the Huber function; see, e.g., Andersen (2008). While this is appropriate when the goal is to describe the behavior of the inliers (i.e., the non-outliers), it may lead to deceiving results when the goal is to estimate a finite population total/mean. A more sensible approach is to use a robust procedure based on an adaptative tuning constant; i.e., a tuning constant that increases as the sample size and the population size increase to infinity. The reason for opting for an adaptive tuning constant instead of a fixed tuning constant lies in the fact that as the sample size increases, the variance of non-robust estimators decreases. In other words, non-robust estimators become more stable, reducing the need for addressing influential cases.

In this paper, we consider the problem of influential units in the case of deterministic linear regression imputation. We first discuss two naive methods for the treatment of influential values at the imputation stage. We illustrate empirically that both methods are generally unsatisfactory. In Section 4, we describe three robust imputed estimators that share a common feature: they are all based on an adaptative tuning constant. The first two rely on the concept of conditional bias, which serves as an appropriate measure of influence in a finite population setting (e.g., Beaumont et al., 2013). The third one uses an optimal tuning constant in the sense that it minimizes the estimated mean square error of the robust estimator. In Section 5, we present the results of a simulation study that assesses the performance of the proposed methods in terms of bias and efficiency for a wide class of distributions. We make some final remarks in Section 6.

## 2 THE SETUP

Let $U = \{1, 2, \ldots, i, \ldots, N\}$ denote a finite population of size $N$, and let $S$ be a random sample of size $n$ selected from $U$ according to a probability sampling design $p(S)$. We are interested in estimating the population total, $t_y = \sum_{i\in U} y_i$, of survey variable $y$. Let $I_i$ be a sample selection indicator attached to unit $i$, such that $I_i = 1$ if $i \in S$, and $I_i = 0$, otherwise. Let $\pi_i = p(I_i = 1)$ denote the first-order inclusion probability attached to unit $i$ and let $\pi_{ij} = p(I_i = 1, I_j = 1)$ denote the second-order inclusion probability for units $i$ and $j$, $i \neq j$. A full sample estimator of $t_y$ is the Horvitz–Thompson estimator (Horvitz and Thompson, 1952) given by

$$\widehat{t}_{y,\pi} = \sum_{i\in S} \frac{y_i}{\pi_i} = \sum_{i\in S} w_i y_i, \tag{1}$$

where $w_i = 1/\pi_i$ denotes the sampling (or basic) weight assigned to unit $i$. In practice, the $y$-variable may be prone to missing values. Let $r_i$ be a response indicator attached to unit $i$ such that $r_i = 1$ if $y_i$ is observed, and $r_i = 0$ if $y_i$ is missing. Let $S_r = \{i \in S; r_i = 1\}$ denote the set of responding units to the survey variable $y$, and let $S_m = S - S_r$ denote the set of nonresponding units. In this paper, we assume that the data are Missing At Random (Rubin, 1976):

$$p_i \equiv p(r_i = 1|\boldsymbol{v}_i, y_i) = p(r_i = 1|\boldsymbol{v}_i),$$

where $\boldsymbol{v}_i$ is a vector of fully observed variables associated with unit $i$. An estimator of $t_y$ after imputation, called an imputed estimator, is defined as

$$\widehat{t}_{y,I} = \sum_{i\in S} w_i r_i y_i + \sum_{i\in S} w_i (1 - r_i) y_i^*, \tag{2}$$

where $y_i^*$ denotes the imputed value for the missing $y_i$. In this paper, we consider deterministic linear regression imputation. The underlying imputation model is thus given by

$$y_i = \mathbf{v}_i^\top \boldsymbol{\beta} + \epsilon_i, \tag{3}$$

such that

$$\mathbb{E}(\epsilon_i \mid \mathbf{v}_i) = 0, \mathbb{E}(\epsilon_i\epsilon_j \mid \mathbf{v}_i, \mathbf{v}_j) = 0, i \neq j \text{ and } \mathbb{V}(\epsilon_i \mid \mathbf{v}_i) = \sigma^2\phi_i.$$

In (3), $\boldsymbol{\beta}$ is a vector of unknown parameters, $\epsilon_i$ is a random error associated with unit $i$, $\sigma^2$ is an unknown parameter, and $\phi_i$ is a known positive constant attached to unit $i$. An estimator of $\boldsymbol{\beta}$ based on the responding units is the weighted least squares estimator given by

$$\widehat{\mathbf{B}}_{\text{WLS}} = \left(\sum_{i\in S} w_i r_i \mathbf{v}_i \phi_i^{-1} \mathbf{v}_i^\top\right)^{-1} \sum_{i\in S} w_i r_i \mathbf{v}_i \phi_i^{-1} y_i. \tag{4}$$

The imputed values are given by $y_i^* = \mathbf{v}_i^\top \widehat{\mathbf{B}}_{\text{WLS}}$, $i \in S_m$. The resulting imputed estimator of $t_y$ is thus given by

$$\widehat{t}_{I,WLS} = \sum_{i\in S} w_i r_i y_i + \sum_{i\in S} w_i(1-r_i)\mathbf{v}_i^\top \widehat{\mathbf{B}}_{\text{WLS}}. \tag{5}$$

If the first moment of the imputation model (3) is correctly specified, we have $\mathbb{E}_m\mathbb{E}_p\mathbb{E}_q(\widehat{t}_{I,WLS} - t_y) = 0$, where $E_m(.)$, $E_p(.)$, and $E_q(.)$, denote the expectation with respect to the imputation model, the sampling design and the nonresponse mechanism, respectively. However, $\widehat{t}_{I,WLS}$ may be very inefficient in the presence of influential units.

## 3 NAIVE METHODS

A first approach for tackling the influential units is to replace the weighted least squares estimator $\widehat{\mathbf{B}}_{\text{WLS}}$ by a robust version $\widehat{\mathbf{B}}_{\text{R}}(c)$, where $c$ denotes a tuning constant, whose role is to adjust the resistance of the estimator. For instance, one may use an $M$-estimator based on the Huber function, in which case $\widehat{\mathbf{B}}_{\text{R}}(c)$ is solution of the following estimating equation:

$$\sum_{i\in S_r} w_i \psi_c\left(\frac{y_i - \mathbf{v}_i^\top\boldsymbol{\beta}}{\sqrt{\phi_i}\widehat{\sigma}}\right)\frac{\mathbf{v}_i}{\sqrt{\phi_i}} = \mathbf{0}, \tag{6}$$

where $\psi_c(\cdot)$ is the so-called Huber function such that $\psi_c(t) = t$ if $|t| \leq c$ and $\psi_c(t) = \text{sgn}(t)c$ if $|t| > c$. With the Huber function, the standard tuning constant is set to 1.345 as it produces a relative efficiency of approximately 95% if the data are normally distributed. The resulting imputed values are given by $y_i^* = \mathbf{v}_i^\top \widehat{\mathbf{B}}_{\text{R}}(c)$, $i \in S_m$. It follows that a robust estimator of $t_y$ is given by

$$\widehat{t}_{I,R}(c) = \sum_{i\in S} w_i r_i y_i + \sum_{i\in S} w_i(1-r_i)\boldsymbol{v}_i^\top \widehat{\boldsymbol{B}}_R(c). \tag{7}$$

Other $\psi$-functions can be used; e.g., Biweight and Andrew, etc. Also, alternative robust estimators may be used; e.g., S-estimators, MM-estimators and LTS estimators. The reader is referred to Andersen (2008) for more details on robust regression methods.

A second approach consists of identifying the influential units (using an outlier detection method), removing these units and obtaining the imputed values by fitting the customary linear regression model based on the remaining responding units. This leads to $y_i^* = \boldsymbol{v}_i^\top \widehat{\boldsymbol{B}}_{WLS}^*$, $i \in S_m$, where

$$\widehat{\boldsymbol{B}}_{WLS}^* = \left(\sum_{i\in S} w_i r_i a_i \mathbf{v}_i \phi_i^{-1} \mathbf{v}_i^\top\right)^{-1} \sum_{i\in S} w_i r_i a_i \mathbf{v}_i \phi_i^{-1} y_i, \tag{8}$$

with $a_i = 1$ if unit $i$ is not discarded and $a_i = 0$, if unit $i$ is discarded. The imputed estimator is then given by

$$\widehat{t}_{I,WLS}^* = \sum_{i\in S} w_i r_i y_i + \sum_{i\in S} w_i(1-r_i)\boldsymbol{v}_i^\top \widehat{\boldsymbol{B}}_{WLS}^*. \tag{9}$$

To assess the performance of the above approaches, we conducted a limited simulation study. We repeated $R = 10,000$ iterations of the following process: (1) A population $U$ of size $N = 10,000$ was generated, consisting

of a survey variable $Y$ and one covariate $V$, using a mixture of normal distributions with a proportion of outliers equal to 5%: $Y_i = 0.95 \times \mathcal{N}(\mu_{1i}; \sigma^2_{1i}) + 0.05 \times \mathcal{N}(\mu_{2i}; \sigma^2_{2i})$. To that end, we generated the variable $V$ from a Gamma distribution with shape parameter equal to 1 and scale parameter equal to 10. For the distribution with asymmetric outliers, we set $\mu_{1i} = 1000 + 5v_i$, $\mu_{2i} = 9000 + 20v_i$, $\sigma^2_{1i} = 19600v_i$, and $\sigma^2_{2i} = 640000v_i$. For the distribution with symmetric outliers, we set $\mu_{1i} = 100 + 8v_i$, $\mu_{2i} = 100 + 8v_i$, $\sigma^2_{1i} = 64v_i$, and $\sigma^2_{2i} = 14400v_i$. (2) A sample $S$ of size $n = 100; 200; 500$, was selected from $U$ according to simple random sampling without replacement; (3) Nonresponse to the $Y$-variable was generated according to a uniform nonresponse mechanism such that $p_i = 0.5$ for all $i$; (4) In each sample, we computed three imputed estimators given by (5), (7) and (9). For the estimator (7), we used an $M$-estimator based on the Huber function with $c = 1.345$. To compute (9), we first needed to detect the outliers. To that end, we used two outlier detection methods: the method based on the Cook distance with threshold $c = 4/(n-3)$ and the method based on studentized residuals with $c = 2; 2.5; 3$.
As a measure of the bias of an estimator, we computed its Monte Carlo percent relative bias, defined as $RB_{MC} = \frac{1}{R}\sum_{r=1}^{R}\left\{\left(\widehat{t}_I^{(r)} - t_y\right)/t_y\right\} \times 100$, where $\widehat{t}_I$ is a generic notation used to denote an imputed estimator of $t_y$. As a measure of efficiency, we computed the Monte Carlo percent relative efficiency (RE), using the non-robust estimator $\widehat{t}_{I,WLS}$, as the reference: $RE = 100 \times \left\{\text{MSE}_{MC}(\widehat{t}_I)/\text{MSE}_{MC}(\widehat{t}_{I,WLS})\right\}$, where $\text{MSE}_{MC}(\widehat{t}_I) = \frac{1}{R}\sum_{r=1}^{R}\left(\widehat{t}_I^{(r)} - t_y\right)^2$. The results are shown in Table 1 for symmetric outliers, and in Table 2 for asymmetric outliers.

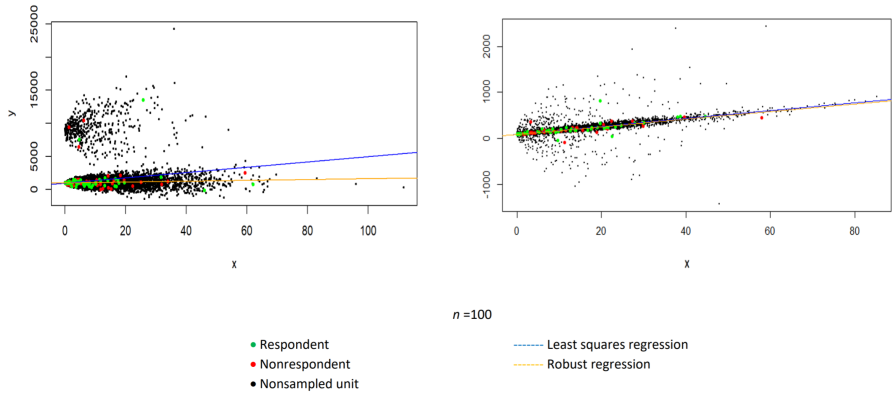

Figure 1: Data generated from a mixture distribution with asymmetric outliers (on the left) and symmetric outliers (on the right)

From Table 1, we note that both approaches behaved very well in terms of bias and efficiency. Indeed, all the estimators exhibited negligible values of RB and values of RE ranging from 53 to 57. However, in the case of asymmetric outliers (see Table 2), both approaches worked well in some scenarios but their performance deteriorated considerably for $n = 200$ and $n = 500$. For both sample sizes, both approaches led to significant bias and values of RE larger than 100, which is undesirable. For the approach based on robust regression, the bad performance of the imputed estimator can be explained by the fact that the tuning constant $c = 1.345$ was fixed and not adaptative. This approach is appropriate in the classical setup, whereby the interest lies in describing the behavior of the inliers. In survey sampling, the goal is different, as the interest lies in estimating the overall population total that consists of a mix of outliers and inliers. As a result, the tuning constant $c$ should be adaptative in the sense that $c$ should increase as $n$ increases. Finally, for the approach based on weighted least squares regression after removing outliers, the bad performance of the imputed estimator can be explained by the fact it relies on the assumption that the discarded respondent $y$-values are unique, i.e., they do not represent similar units in the non-responding set. In general, this assumption is not tenable.

| | WLS | Robust regression | | | WLS<br>(Exclude outliers) | | | |
|---|---|---|---|---|---|---|---|---|
| $n$ | | $c = 0.1$ | $c = 1.345$ | $c = 2.5$ | Studentized<br>$c = 2$ | Studentized<br>$c = 2.5$ | Studentized<br>$c = 3$ | Cook distance |
| 100 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | -0.0 | -0.0 |
| | (100) | (54) | (55) | (57) | (54) | (55) | (57) | (56) |
| 200 | -0.1 | -0.0 | -0.0 | -0.1 | -0.0 | -0.1 | -0.1 | -0.1 |
| | (100) | (54) | (55) | (57) | (53) | (55) | (57) | (55) |
| 500 | -0.0 | -0.0 | -0.0 | -0.0 | -0.0 | -0.0 | -0.0 | -0.0 |
| | (100) | (54) | (54) | (56) | (53) | (54) | (56) | (54) |

Table 1: Percent relative bias and relative efficiency of several estimators in the case of symmetric outliers

| | WLS | Robust regression | | | WLS<br>(Exclude outliers) | | | |
|---|---|---|---|---|---|---|---|---|
| $n$ | | $c = 0.1$ | $c = 1.345$ | $c = 2.5$ | Studentized<br>$c = 2$ | Studentized<br>$c = 2.5$ | Studentized<br>$c = 3$ | Cook distance |
| 100 | 0.0 | -13.0 | -12.7 | -11.9 | -10.7 | -9.6 | -8.6 | -8.6 |
| | (100) | (77) | (75) | (74) | (87) | (89) | (92) | (92) |
| 200 | -0.0 | -13.0 | -12.7 | -12.0 | -10.2 | -8.9 | -7.7 | -7.9 |
| | (100) | (122) | (118) | (112) | (117) | (114) | (111) | (114) |
| 500 | 0.0 | -13.1 | -12.8 | -12.1 | -9.7 | -8.1 | -6.6 | -6.9 |
| | (100) | (267) | (256) | (235) | (204) | (177) | (156) | (167) |

Table 2: Percent relative bias and relative efficiency of several estimators in the case of asymmetric outliers

# 4 PROPOSED APPROACHES

In this section, we describe three robust approaches that share a common feature: they all use an adaptative tuning constant. The first two approaches are based on the concept of conditional bias, which is a measure of influence. The last approach consists of determining the tuning constant that minimizes the estimated mean square error of the imputed estimator.

## 4.1 Conditional Bias

The conditional bias is an appropriate measure to quantify the influence (or impact) of a unit in a finite population setting. The conditional bias of the responding unit $i$ is defined as $B_i^I = \mathbb{E}_m\mathbb{E}_p\mathbb{E}_q\left(t_{I,WLS} - t_y | Y_i = y_i, I_i = 1, r_i = 1\right)$. It can be shown that $B_i^I$ can be approximated by

$$B_i^I \approx \sum_{j \in U}\left(\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}\right) y_j + w_i\mathbf{C}\mathbf{v}_i\phi_i^{-1}\left(y_i - \mathbf{v}_i^\top\boldsymbol{\beta}\right), \tag{10}$$

where $\mathbf{C} = \left\{\sum_{i\in U}(1-p_i)\mathbf{v}_i^\top\right\}\left\{\sum_{i\in U} p_i\mathbf{v}_i\phi_i^{-1}\mathbf{v}_i^\top\right\}^{-1}$. The first term on the right hand-side of (10) measures the influence of unit $i$ on the sampling error $\widehat{t}_{y,\pi} - t_y$, whereas the second term measures the influence of unit $i$ on the nonresponse error, $\widehat{t}_{I,WLS} - \widehat{t}_{y,\pi}$. Under simple linear regression imputation (i.e., $\mathbf{v}_i = (1, v_i)^\top$ and $\phi_i = 1$) and simple random sampling without replacement, an estimator of the conditional bias is given by

$$\widehat{B}_i^I \approx \left(\frac{N}{n} - 1\right)(y_i - \overline{y}_I) + \left(\frac{N}{n}\right)\frac{1}{\widehat{p}}\left\{(1-\widehat{p}) + \frac{(v_i - \overline{v}_r)(\overline{v} - \overline{v}_r)}{s_{vr}^2}\right\}\left(y_i - \widehat{B}_{0,WLS} - \widehat{B}_{1,WLS}v_i\right), \tag{11}$$

where $\overline{y}_I = \widehat{t}_{I,WLS}/N$, $\widehat{p} = n_r/n$, and $s_{vr}^2 = (n_r - 1)^{-1}\sum_{i\in S_r}(v_i - \overline{v}_r)^2$. Thus, the responding unit $i$ has a large influence if (1) the sampling fraction $n/N$ is small; (2) Its $y$-value is far from the overall estimated mean $\overline{y}_I$; (3) The response rate $\widehat{p}$ is low; (4) Its $v$-value is far from the mean of respondents $\overline{v}_r$ (which may indicate a high leverage point); (5) It has a large vertical residual: $y_i - \widehat{B}_{0,WLS} - \widehat{B}_{1,WLS}v_i$.

## 4.2 Three robust approaches

In this section, we consider three robust estimators of $t_y$ that all rely on an adaptative tuning constant. As a result, the three estimators converge to the non-robust estimator $\widehat{t}_{I,WLS}$ as the sample size and the population size increase.

Following Beaumont et al. (2013) and Chen et al. (2020), we first consider a robust version of $\widehat{t}_{I,WLS}$ (5) based on the concept of conditional bias:

$$\widehat{t}_{I,CB}(c) \quad = \quad \widehat{t}_{I,WLS} + \Delta(c), \tag{12}$$

where $c$ denotes a tuning constant. As in Beaumont et al. (2013) and Chen et al. (2020), we select the value of $c$ that minimizes $\max\limits_{i \in S_r} \left| \widehat{B}_i^R \right|$, where $\widehat{B}_i^R$ is the conditional bias of unit $i$ with respect to the robust estimator $\widehat{t}_{I,CB}(c)$. The resulting estimator is given by

$$\widehat{t}_{I,CB}(c_{opt}) \quad = \quad \widehat{t}_{I,WLS} - \frac{1}{2}\left[\min_{i \in S_r}\left\{\widehat{B}_i^I\right\} + \max_{i \in S_r}\left\{\widehat{B}_i^I\right\}\right], \tag{13}$$

where, if needed, the value $c_{opt}$ may be obtained by solving

$$\Delta(c) = -\frac{1}{2}\left[\min_{i \in S_r}\left\{\widehat{B}_i^I\right\} + \max_{i \in S_r}\left\{\widehat{B}_i^I\right\}\right].$$

A second robust estimator is obtained by estimating $\boldsymbol{\beta}$ in (3) using an $M$-estimator based on the Huber function with the following tuning constant:

$$c_{\text{new}} = 1.345\left\{1 + \left|\min_{i \in S_r}\left\{\widehat{B}_i^*\right\} + \max_{i \in S_r}\left\{\widehat{B}_i^*\right\}\right|/2\right\} + \frac{n}{N}\sqrt{n}, \tag{14}$$

where $\widehat{B}_i^*$ denotes the standardized version of $\widehat{B}_i^I$, obtained by the subtracting the average of the $\widehat{B}_i^I$'s and dividing by their standard deviation. A robust estimator of $t_y$ is thus given by

$$\widehat{t}_{I,R}(c_{\text{new}}) = \sum_{i \in S} w_i \mathbf{v}_i^\top \widehat{\boldsymbol{B}}_{\text{R}}(c_{\text{new}}), \tag{15}$$

which is written in the so-called projection form. Why not use

$$\widehat{t}_{I,R}(c_{\text{new}}) = \sum_{i \in S_r} w_i y_i + \sum_{i \in S_m} w_i \mathbf{v}_i^\top \widehat{\boldsymbol{B}}_{\text{R}}(c_{\text{new}}). \tag{16}$$

instead? The answer to this question is that, in (16), we are only "taking care" of the missing values and not the $y$-values observed for the respondents, some of which may be influential.

The rationale behind the choice of $c_{\text{new}}$ is as follows. First, consider the case of a negligible sampling fraction $n/N$. In this case, the second term on the right-hand side of (14) is negligible. Now, suppose that the distribution has symmetric outliers and the weights $w_i$ are constant. In this case, we expect $\left|\min\limits_{i \in S_r}\left\{\widehat{B}_i^*\right\} + \max\limits_{i \in S_r}\left\{\widehat{B}_i^*\right\}\right|/2$ to be close to 0. Thus, $c_{\text{new}}$ will essentially be equal to 1.345, which is a desirable feature (see Section 3). If the distribution has asymmetric outliers (say to the right), the term $\left|\min\limits_{i \in S_r}\left\{\widehat{B}_i^*\right\} + \max\limits_{i \in S_r}\left\{\widehat{B}_i^*\right\}\right|/2$ will be larger than 0, which implies that $c_{\text{new}}$ will be larger than 1.345. Second, as the sample size $n$ gets larger, the second term on the right-hand side of (14) increases and $\widehat{\boldsymbol{B}}_{\text{R}}(c_{\text{new}})$ gets increasingly closer to $\widehat{\boldsymbol{B}}_{\text{WLS}}$.

For the third proposal, we determine the optimal tuning constant $c^*$ that minimizes the estimated mean square error of the robust estimator, which leads to

$$\widehat{t}_{I,R}(c^*) = \sum_{i \in S} w_i \mathbf{v}_i^\top \widehat{\boldsymbol{B}}_{\text{R}}(c^*). \tag{17}$$

The estimated mean squared error of $\widehat{t}_{I,R}(c) = \sum_{i \in S} w_i \mathbf{v}_i^\top \widehat{\boldsymbol{B}}_{\mathrm{R}}(c)$ is given by

$$\widehat{MSE}(\widehat{t}(\widehat{\mathbf{B}}_R, \sigma, c)) \quad = \quad \max\left\{0, \left(\widehat{t}(\widehat{\mathbf{B}}_R, \sigma, c) - \widehat{t}_{I,WLS}\right)^2 - \widehat{V}\left(\widehat{t}(\widehat{\mathbf{B}}_R, \sigma, c) - \widehat{t}_{I,WLS}\right)\right\} + \widehat{V}\left\{\widehat{t}(\widehat{\mathbf{B}}_R, \sigma, c)\right\}. \quad (18)$$

In practice, this involves computing the estimated mean square error of $\widehat{t}_{I,R}(c)$ for a grid of $c$-values and selecting the one that minimizes (18). The explicit expressions (and their derivations) for the terms on the right-hand side of (18), are given in the Appendix.

## 5 SIMULATION STUDY

We conducted a simulation study to assess the performance of the proposed estimators. We repeated $R = 10,000$ iterations of the following process: (i) A finite population $U$, of size $N = 5,000$ was generated, with a survey variable $Y$ and a single covariate $V$. To that end, we generated $V$ from a Gamma distribution with shape parameter equal to 5 and scale parameter equal to 10. Given the $v$-values, the $y$-values were generated according to the following model:

$$y_i \mid v_i \sim \mathcal{D}(\mu_i; \sigma^2),$$

where $\mu_i = \beta_0 + \beta_1 v_i$. We used the following distributions $\mathcal{D}$: Normal, Lognormal, Pareto, Frechet, Student and Double exponential (Laplace). The values of $\beta_0$, $\beta_1$ and $\sigma$ were set to 50, 12, and 600, respectively. This led to identical first two moments for all the distributions. We also considered mixture distributions with approximately 1% and 3% of outliers: $Y_i = \alpha_i \times \mathcal{D}(\mu_{1i}; \sigma_{1i}^2) + (1 - \alpha_i) \times \mathcal{D}(\mu_{2i}; \sigma_{2i}^2)$, where $P(\alpha_i = 1) = 0.99$ or $0.97$. For mixtures of normal distributions, we set $\mu_{1i} = 150 + 20v_i$, $\mu_{2i} = 2000 + 85v_i$, $\sigma_1 = 400$ and $\sigma_2 = 2000$. For mixtures of log-normal distributions, we set $\mu_{1i} = 150 + 8v_i$, $\mu_{2i} = 1200 + 60v_i$, $\sigma_1 = 150$ and $\sigma_2 = 1500$; (ii) From $U$ generated in Step (i), we selected a sample of size $n \in \{50, 100, 200\}$, according to simple random sampling without replacement; (iii) In each sample, nonresponse to the $y$-variable was generated with probability

$$p_i = 0.1 + 0.9 \frac{\exp(4 - 0.09v_i)}{1 + \exp(4 - 0.09v_i)}.$$

This led to a response rate approximately equal to 50%; (iv) In each sample, we computed the non-robust estimator (5), the naive estimator (7) with $c = 1.345$, the three proposed estimators given by (13), (15) and (17), as well as the (unfeasible) robust estimator $\widehat{t}_{I,R}(\widetilde{c})$ based on a tuning constant $\widetilde{c}$ that minimizes its Monte Carlo mean square error. The latter can be viewed as a "gold standard".

| $n$ | $\widehat{t}_{I,WLS}$ | | | $\widehat{t}_{I,R}(1.345)$ | | | $\widehat{t}_{I,CB}(c_{opt})$ | | | $\widehat{t}_{I,R}(c_{\mathrm{new}})$ | | | $\widehat{t}_{I,R}(c^*)$ | | | $\widehat{t}_{I,R}(\widetilde{c})$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 50 | 100 | 200 | 50 | 100 | 200 | 50 | 100 | 200 | 50 | 100 | 200 | 50 | 100 | 200 | 50 | 100 | 200 |
| **Normal** | 0.2 | -0.2 | 0.1 | 0.2 | -0.2 | 0.1 | -0.1 | -0.4 | -0.1 | 0.4 | 0.0 | 0.1 | 0.1 | -0.3 | 0.0 | 0.2 | -0.2 | 0.1 |
| | (100) | (100) | (100) | (103) | (103) | (102) | (101) | (100) | (100) | (102) | (101) | (100) | (106) | (105) | (103) | (100) | (100) | (100) |
| **Student** | 0.1 | 0.0 | -0.1 | 0.0 | -0.1 | -0.1 | -0.1 | -0.2 | -0.2 | 0.4 | 0.3 | 0.1 | -0.0 | -0.1 | -0.1 | 0.1 | 0.0 | 0.0 |
| | (100) | (100) | (100) | (69) | (68) | (66) | (83) | (82) | (83) | (75) | (73) | (74) | (68) | (65) | (64) | (60) | (57) | (55) |
| **Laplace** | -0.2 | 0.0 | 0.1 | -0.1 | 0.0 | 0.1 | -0.4 | -0.2 | -0.1 | 0.2 | 0.3 | 0.2 | -0.3 | -0.1 | 0.1 | -0.2 | 0.0 | 0.2 |
| | (100) | (100) | (100) | (85) | (82) | (81) | (93) | (93) | (95) | (90) | (90) | (92) | (82) | (80) | (80) | (76) | (68) | (65) |
| **Pareto** | 0.2 | -0.1 | 0.1 | -7.9 | -8.6 | -8.7 | -3.4 | -3.1 | -2.3 | -6.3 | -5.5 | -4.1 | -6.9 | -5.4 | -4.0 | -7.1 | -4.8 | -3.1 |
| | (100) | (100) | (100) | (50) | (58) | (93) | (70) | (64) | (69) | (50) | (45) | (54) | (63) | (52) | (58) | (50) | (45) | (53) |
| **Frechet** | -0.3 | 0.1 | 0.0 | -7.8 | -8.2 | -8.5 | -3.8 | -2.9 | -2.2 | -6.6 | -5.4 | -4.1 | -7.8 | -5.4 | -4.0 | -6.5 | -4.7 | -3.0 |
| | (100) | (100) | (100) | (55) | (69) | (102) | (70) | (73) | (76) | (52) | (60) | (67) | (59) | (67) | (70) | (51) | (58) | (63) |
| **Lognormal** | -0.2 | 0.2 | 0.0 | -8.4 | -8.9 | -9.3 | -3.8 | -2.7 | -1.9 | -6.6 | -5.1 | -3.4 | -7.8 | -5.0 | -3.3 | -6.9 | -3.9 | -2.4 |
| | (100) | (100) | (100) | (79) | (95) | (134) | (88) | (90) | (94) | (83) | (88) | (94) | (90) | (92) | (96) | (83) | (86) | (91) |
| **Weibull** | -0.3 | 0.1 | 0.1 | -7.0 | -7.4 | -7.7 | -3.3 | -2.0 | -1.2 | -5.9 | -4.6 | -3.0 | -6.6 | -3.8 | -2.2 | -2.8 | -1.3 | -0.5 |
| | (100) | (100) | (100) | (92) | (111) | (141) | (95) | (98) | (98) | (91) | (103) | (102) | (95) | (101) | (100) | (91) | (98) | (97) |
| **Mixture of Normals** | 0.2 | -0.1 | 0.0 | -1.9 | -2.2 | -2.1 | -1.3 | -1.5 | -1.3 | -2.4 | -2.6 | -2.2 | -3.0 | -3.2 | -2.9 | -3.3 | -3.4 | -2.7 |
| **(1 %)** | (100) | (100) | (100) | (50) | (51) | (53) | (74) | (71) | (73) | (53) | (53) | (58) | (58) | (57) | (60) | (46) | (50) | (55) |
| **Mixture of Normals** | 0.0 | -0.1 | 0.1 | -5.8 | -5.9 | -5.9 | -3.2 | -3.1 | -2.5 | -7.2 | -7.0 | -5.9 | -7.7 | -7.0 | -5.6 | -10.0 | -7.6 | -5.4 |
| **(3%)** | (100) | (100) | (100) | (38) | (45) | (58) | (72) | (73) | (79) | (50) | (55) | (70) | (66) | (73) | (83) | (41) | (54) | (67) |
| **Mixture of Lognormals** | 0.0 | 0.0 | -0.1 | -3.9 | -4.1 | -4.2 | -2.4 | -2.4 | -2.2 | -5.2 | -4.9 | -4.5 | -5.9 | -5.6 | -4.9 | -5.6 | -5.0 | 4.1 |
| **(1%)** | (100) | (100) | (100) | (30) | (33) | (43) | (61) | (60) | (64) | (29) | (32) | (42) | (43) | (39) | (48) | (26) | (30) | (41) |
| **Mixture of Lognormals** | -0.1 | 0.3 | 0.0 | -9.4 | -9.5 | -9.6 | -5.0 | -4.4 | -3.9 | -12.6 | -12.2 | -11.1 | -11.5 | -10.9 | -10.1 | -15.0 | -12.6 | -9.4 |
| **(3%)** | (100) | (100) | (100) | (27) | (36) | (60) | (67) | (69) | (76) | (38) | (48) | (74) | (67) | (72) | (78) | (33) | (50) | (71) |

Table 3: Monte Carlo percent relative bias and Monte Carlo relative efficiency (in parentheses) of several estimators

Table 3 displays the Monte Carlo percent relative bias and the Monte Carlo percent relative efficiency for the six estimators listed above. In the case of the symmetric distributions (Normal, Student and Laplace), all the estimators exhibited a negligible bias in all the scenarios. For the normal distribution, all the robust estimators suffer from a slight loss of efficiency with values ranging from 100 to 106, which is a desirable feature. For the $t$-distribution and the Laplace distribution, the robust estimators were much more efficient than $\widehat{t}_{I,WLS}$. The estimator $\widehat{t}_{I,R}(c^*)$ was the best but, as expected, incurred some loss of efficiency with respect to the gold standard estimator $\widehat{t}_{I,R}(\widetilde{c})$. The estimator $\widehat{t}_{I,CB}(c_{opt})$ was outperformed by the other robust estimators.

In the case of asymmetric distributions (Pareto, Frechet, Lognormal, and Weibull), all the robust estimators exhibited some bias, as expected. In virtually all the scenarios, the estimator $\widehat{t}_{I,CB}(c_{opt})$ was less biased than its competitors. The naive estimator $\widehat{t}_{I,R}(1.345)$ performed well in some scenarios but performed poorly in others, especially for larger sample sizes. For instance, for the lognormal distribution, the estimator $\widehat{t}_{I,R}(1.345)$ exhibited a value of RE equal to 134% for $n = 200$. For highly skewed distributions such as Pareto and Frechet, the proposed robust estimators showed substantial improvement in terms of relative efficiency with respect to $\widehat{t}_{I,WLS}$. In particular, $\widehat{t}_{I,R}(c_{\text{new}})$ was the best estimator with a value of RE close to that of the gold standard $\widehat{t}_{I,R}(\widetilde{c})$. For the Lognormal distribution, all the proposed estimators were more efficient than the non-robust estimator for all the sample sizes. The robust estimator $\widehat{t}_{I,R}(c_{\text{new}})$ was the best with a value of RE close to that of the gold standard estimator $\widehat{t}_{I,R}(\widetilde{c})$. For the Weibull distribution, all the estimators showed a value of RE close to that of $\widehat{t}_{I,WLS}$.

Finally, in the case of the mixture distributions, the robust estimators exhibited substantial improvement over $\widehat{t}_{I,WLS}$. Both $\widehat{t}_{I,R}(1.345)$ and $\widehat{t}_{I,R}(c_{\text{new}})$ performed well, and outperformed $\widehat{t}_{I,CB}(c_{opt})$ by a significant margin. Again, in terms of efficiency, the robust estimator $\widehat{t}_{I,R}(c_{\text{new}})$ showed values of RE comparable to those obtained with the gold standard $\widehat{t}_{I,R}(\widetilde{c})$.

## 6 CONCLUSION

In this paper, we considered the problem of robust imputation in the presence of influential units. We proposed two new robust estimators that were shown to perform well for a wide class of distributions. Overall, among the three robust estimators, $\widehat{t}_{I,R}(c^*)$ had the best performance in terms of relative efficiency for symmetric outliers, whereas the estimator $\widehat{t}_{I,R}(c_{\text{new}})$ was generally the best for asymmetric distributions

We considered the case of linear regression imputation. The extension to imputation procedures based on generalized linear models and non-parametric methods is a topic of future research. Estimating the mean square error of the proposed robust estimators is a challenging problem and is currently under investigation.

## APPENDIX

### A Mean Square Error Estimation: derivation with known constant and known standard deviation

Consider the proposed robust imputed estimator

$$\widehat{t}(\widehat{\mathbf{B}}_R, \sigma, c) = \sum_{i \in S} w_i \mathbf{v}_i^\top \widehat{\mathbf{B}}_R, \tag{19}$$

where $\widehat{\mathbf{B}}_R$ is the solution of the following robust estimating equation

$$\widehat{U}(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i \in S_r} \psi_c \left( \frac{y_i - \mathbf{v}_i^\top \boldsymbol{\beta}}{\sigma \phi_i^{1/2}} \right) \frac{w_i \mathbf{v}_i}{\phi_i^{1/2}} = 0, \tag{20}$$

and $\psi_c(t)$ is the Huber function such that $\psi_c(t) = cI(t \geq c) + tI(-c \leq t \leq c) + (-c)I(t \leq -c)$. Suppose the probability limit of $\widehat{\mathbf{B}}_R$ is $\boldsymbol{\beta}^*$. Using a first-order Taylor expansion, we have

$$\begin{aligned} 0 &= \widehat{U}(\widehat{\mathbf{B}}_R) \\ &= \widehat{U}(\boldsymbol{\beta}^*) + \frac{\partial \mathbb{E}(\widehat{U}(\boldsymbol{\beta}^*))}{\partial \boldsymbol{\beta}} \left( \widehat{\mathbf{B}}_R - \boldsymbol{\beta}^* \right) + o_p(n^{-1/2}). \end{aligned} \tag{21}$$

In addition, it can be shown that

$$\frac{\partial \mathbb{E}(\widehat{U}(\boldsymbol{\beta}^*))}{\partial \boldsymbol{\beta}} = M_1 + M_2 + M_3, \tag{22}$$

where

$$M_1 = -\mathbb{E}\left\{ p(\mathbf{v}_i) c \frac{\mathbf{v}_i \mathbf{v}_i^\top}{\phi_i^{1/2}} f_{y|\mathbf{v}}(\mathbf{v}_i^\top \boldsymbol{\beta}^* + c\sigma\phi_i^{1/2}) \right\}, \tag{23}$$

$$\begin{aligned} M_2 &= \mathbb{E}\left\{ p(\mathbf{v}_i) \frac{\mathbf{v}_i \mathbf{v}_i^\top}{\phi_i^{1/2}} c f_{y|\mathbf{v}}(\mathbf{v}_i^\top \boldsymbol{\beta}^* + c\sigma\phi_i^{1/2}) \right\} + \mathbb{E}\left\{ p(\mathbf{v}_i) \frac{\mathbf{v}_i \mathbf{v}_i^\top}{\phi_i^{1/2}} c f_{y|\mathbf{v}}(\mathbf{v}_i^\top \boldsymbol{\beta}^* - c\sigma\phi_i^{1/2}) \right\} \\ &- \mathbb{E}\left\{ p(\mathbf{v}_i) \frac{\mathbf{v}_i \mathbf{v}_i^\top}{\sigma\phi_i} I(-c \leq \frac{y_i - \mathbf{v}_i^\top \boldsymbol{\beta}^*}{\sigma\phi_i^{1/2}} \leq c) \right\}, \end{aligned} \tag{24}$$

and

$$M_3 = -\mathbb{E}\left\{ p(\mathbf{v}_i) \frac{\mathbf{v}_i \mathbf{v}_i^\top}{\phi_i^{1/2}} c f_{y|\mathbf{v}}(\mathbf{v}_i^\top \boldsymbol{\beta}^* - c\sigma\phi_i^{1/2}) \right\}. \tag{25}$$

According to (22)-(25), we have

$$\frac{\partial \mathbb{E}(\widehat{U}(\boldsymbol{\beta}^*))}{\partial \boldsymbol{\beta}} = -\mathbb{E}\left\{ p(\mathbf{v}_i) \frac{\mathbf{v}_i \mathbf{v}_i^\top}{\sigma\phi_i} I(-c \leq \frac{y_i - \mathbf{v}_i^\top \boldsymbol{\beta}^*}{\sigma\phi_i^{1/2}} \leq c) \right\}. \tag{26}$$

According to (21), we have

$$\widehat{\mathbf{B}}_R - \boldsymbol{\beta}^* = -\left\{ \frac{\partial \mathbb{E}(\widehat{U}(\boldsymbol{\beta}^*))}{\partial \boldsymbol{\beta}} \right\}^{-1} \widehat{U}(\boldsymbol{\beta}^*) + o_p(n^{-1/2}). \tag{27}$$

Therefore, we have

$$\begin{aligned} \widehat{t}(\widehat{\mathbf{B}}_R, \sigma, c) &= \sum_{i \in S} w_i \mathbf{v}_i^\top \widehat{\mathbf{B}}_R \\ &= \sum_{i \in S} w_i \mathbf{v}_i^\top (\boldsymbol{\beta}^* + \widehat{\mathbf{B}}_R - \boldsymbol{\beta}^*) \\ &= \sum_{i \in S} w_i \mathbf{v}_i^\top \boldsymbol{\beta}^* - \sum_{i \in S} w_i \mathbf{v}_i^\top \left\{ \frac{\partial \mathbb{E}(\widehat{U}(\boldsymbol{\beta}^*))}{\partial \boldsymbol{\beta}} \right\}^{-1} \widehat{U}(\boldsymbol{\beta}^*) + o_p(Nn^{-1/2}) \\ &= \sum_{i \in S} w_i \eta_i + o_p(Nn^{-1/2}), \end{aligned} \tag{28}$$

where

$$\eta_i = \mathbf{v}_i^\top \boldsymbol{\beta}^* - \left( \frac{1}{N} \sum_{i \in S} w_i \mathbf{v}_i^\top \right) \left\{ \frac{\partial \mathbb{E}(\widehat{U}(\boldsymbol{\beta}^*))}{\partial \boldsymbol{\beta}} \right\}^{-1} r_i \psi_c \left( \frac{y_i - \mathbf{v}_i^\top \boldsymbol{\beta}^*}{\sigma\phi_i^{1/2}} \right) \frac{\mathbf{v}_i}{\phi_i^{1/2}}. \tag{29}$$

Hence, the mean squared error of $\widehat{t}(\widehat{\mathbf{B}}_R, \sigma, c)$ can be written as

$$\begin{aligned} MSE(\widehat{t}(\widehat{\mathbf{B}}_R, \sigma, c)) &= \left\{ \mathbb{E}(\widehat{t}(\widehat{\mathbf{B}}_R, \sigma, c)) - t_y \right\}^2 + \mathbb{V}\left\{ \widehat{t}(\widehat{\mathbf{B}}_R, \sigma, c) \right\} \\ &= \left( \sum_{i=1}^N \eta_i - t_y \right)^2 + \sum_{i=1}^N \sum_{j=1}^N (\pi_{ij} - \pi_i \pi_j) \frac{\eta_i}{\pi_i} \frac{\eta_j}{\pi_j} \\ &+ o(N^2/n). \end{aligned} \tag{30}$$

Therefore, the estimated mean squared error can be written as

$$\begin{aligned}\widehat{MSE}(\widehat{t}(\widehat{\mathbf{B}}_R,\sigma,c)) &= \max\left\{\left(\widehat{t}(\widehat{\mathbf{B}}_R,\sigma,c)-\widehat{t}_{I,WLS}\right)^2-\widehat{\mathbb{V}}\left(\widehat{t}(\widehat{\mathbf{B}}_R,\sigma,c)-\widehat{t}_{I,WLS}\right),0\right\}\\ &+ \widehat{\mathbb{V}}\left\{\widehat{t}(\widehat{\mathbf{B}}_R,\sigma,c)\right\}.\end{aligned} \tag{31}$$

In addition, it can be shown that

$$\begin{aligned}\widehat{t}_{I,WLS} &= \sum_{i\in S_r} w_i y_i + \sum_{i\in S_m} w_i \mathbf{v}_i^\top \widehat{\mathbf{B}}_{WLS}\\ &= \sum_{i\in S_r} w_i y_i + \sum_{i\in S_m} w_i \mathbf{v}_i^\top \boldsymbol{\beta}^*_{WLS} + \sum_{i\in S_m} w_i \mathbf{v}_i^\top\left(\widehat{\mathbf{B}}_{WLS}-\boldsymbol{\beta}^*_{WLS}\right)\\ &= \sum_{i\in S} w_i \tau_i,\end{aligned} \tag{32}$$

where

$$\tau_i = r_i y_i + (1-r_i)\mathbf{v}_i^\top\boldsymbol{\beta}^*_{WLS} + A r_i \mathbf{v}_i \phi_i^{-1}\left(y_i - \mathbf{v}_i^\top\boldsymbol{\beta}^*_{WLS}\right), \tag{33}$$

and $A = \sum_{i\in S_m} w_i \mathbf{v}_i^\top \left(\sum_{i\in S_r} w_i \mathbf{v}_i \phi_i^{-1}\mathbf{v}_i^\top\right)^{-1}$. Therefore, we have

$$\widehat{\mathbb{V}}\left(\widehat{t}(\widehat{\mathbf{B}}_R,\sigma,c)-\widehat{t}_{I,WLS}\right) = \sum_{i\in S}\sum_{j\in S}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}}\frac{\widehat{\eta}_i-\widehat{\tau}_i}{\pi_i}\frac{\widehat{\eta}_j-\widehat{\tau}_j}{\pi_j}, \tag{34}$$

where

$$\eta_i = \mathbf{v}_i^\top\widehat{\mathbf{B}}_R - \left(\sum_{i\in S} w_i \mathbf{v}_i^\top\right)\left\{N\frac{\partial\mathbb{E}(U(\boldsymbol{\beta}^*))}{\partial\boldsymbol{\beta}}\right\}^{-1} r_i \psi_c\left(\frac{y_i-\mathbf{v}_i^\top\widehat{\mathbf{B}}_R}{\sigma\phi_i^{1/2}}\right)\frac{\mathbf{v}_i}{\phi_i^{1/2}}, \tag{35}$$

$$\frac{\partial\mathbb{E}(\widehat{U}(\boldsymbol{\beta}^*))}{\partial\boldsymbol{\beta}} = -\frac{1}{N}\sum_{i\in S_r} w_i \frac{\mathbf{v}_i\mathbf{v}_i^\top}{\widehat{\sigma}\phi_i} I(-c \le \frac{y_i-\mathbf{v}_i^\top\widehat{\mathbf{B}}_R}{\widehat{\sigma}\phi_i^{1/2}} \le c), \tag{36}$$

and

$$\widehat{\tau}_i = r_i y_i + (1-r_i)\mathbf{v}_i^\top\widehat{\mathbf{B}}_{WLS} + A r_i \mathbf{v}_i \phi_i^{-1}\left(y_i - \mathbf{v}_i^\top\widehat{\mathbf{B}}_{WLS}\right). \tag{37}$$

In addition, we have

$$\widehat{\mathbb{V}}\left\{\widehat{t}(\widehat{\mathbf{B}}_R,\sigma,c)\right\} = \sum_{i\in S}\sum_{j\in S}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}}\frac{\widehat{\eta}_i}{\pi_i}\frac{\widehat{\eta}_j}{\pi_j}, \tag{38}$$

so, $\widehat{MSE}(\widehat{t}(\widehat{\mathbf{B}}_R,\sigma,c))$ can be obtained according to (31)-(38).

## REFERENCES

Andersen, R. (2008). *Modern methods for robust regression* (No. 152). Sage.

Beaumont, J.-F., Haziza, D., and Ruiz-Gazen, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika*, **100** 555–569.

Chen, S., Haziza, D., and Michal, V. (2020). Efficient multiply robust imputation in the presence of influential units in surveys. To appear in the *Canadian Journal of Statistics*.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association* **47**, 663–685.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.