Multiply robust imputation procedures for the treatment of item nonresponse in surveys

Sixia Chen[1] *(*), David Haziza*[2] *(**)*
(*)WESTAT
(**) Université de Montréal

Résumé

La non-réponse partielle est fréquemment traitée au moyen d'une imputation. Dans cet article, nous introduisons le concept multiple robustesse dans un contexte de populations finies. Ce concept est intimement lié à celui proposé par Han et Wang (2013). Le concept de multiple robustesse peut être vu comme une généralisation de celui de double robustesse. En pratique, on peut vouloir ajuster de multiples modèles de non-réponse et de multiples modèles d'imputation, chacun comportant différents prédicteurs et/ou différentes fonction de liens. Une procédure d'imputation est multiple robuste si l'estimateur résultant est convergent si tous les modèles sauf un sont mal spécifiés. Des estimateurs de variance possédant la propriété de multiple robustesse sont dévelopés. La généralisation au cas de l'imputation aléatoire et l'imputation fractionnelle est également discutée. Finalement, les résultats d'une étude par simulation, mesurant les propriétés des estimateurs ponctuels et de variance, sont présentés.

Mots-clés : non-réponse partielle, imputation multiple robuste, estimation de la variance.

[1]chensixiaster@gmail.com
[2]david.haziza@umontreal.ca

Abstract

Item nonresponse in surveys is often treated through some form of imputation. In this paper, we introduce the concept of multiply robust imputation procedures in the context of finite population sampling, which is closely related to the concept of multiple robustness proposed by Han and Wang (2013). Multiple robustness can be viewed as an extension of the concept of double robustness. In practice, multiple nonresponse models and multiple imputation models may be fitted, each involving different subsets of covariates and possibly different link functions. An imputation procedure is said to be multiply robust if the resulting estimator is consistent if all but one model are misspecified. A jackknife variance estimator is proposed and shown to be consistent provided that the sampling fraction is negligible. Multiply robust point and variance estimators are proposed even when the sampling fraction is not negligible. Extensions to random and fractional imputations, as well as estimation of distribution functions are discussed. Finally, the results of a simulation study, assessing the performance of the proposed point and variance estimators are presented.

Keywords: item nonresponse, multiply robust imputation, variance estimation.

1 Introduction

In surveys conducted by statistical agencies, item nonresponse is usually handled through some form of imputation. Most often, a single imputation procedure is used, whereby a missing value is replaced by a single imputed value. The latter is constructed using auxiliary information, which is a set of variables available for all the sample units (respondents and nonrespondents). The reason for using imputation is two-fold: (i) to reduce the nonresponse bias and (ii) to produce a complete rectangular data file, which allows the secondary analysts to obtain point estimates using complete data estimation procedures.

In the presence of missing data, there are two different ways to adjust for the nonresponse bias: (i) the nonresponse model approach, whereby one postulates a nonresponse model, which is a set of assumptions about the unknown nonresponse mechanism. (ii) the imputation model approach (also called the outcome regression approach) that requires the specification of the model describing the distribution of the study variable. In the last two decades, a number of procedures, called doubly robust (or doubly protected) procedures, have been proposed in the literature; e.g., Robins et al. (1994); Scharfstein et al. (1999); Tan (2006); Bang and Robins (2005); Kang and Schafer (2008); and Cao et al. (2009). In the context of survey data, doubly robust procedures have been discussed in Kott (1994), Kott (2006), Kim and Park (2006), Haziza and Rao (2006), Kott and Chang (2010), Haziza et al. (2014), and Kim and Haziza (2014), among others. An estimator is said to be doubly robust if it remains asymptotically unbiased and consistent if either model (nonresponse or imputation) is true. Thus, doubly robust procedures offer some protection if either the nonresponse model or the imputation model is misspecified. In this paper, we refer to an imputation procedure resulting in a doubly robust imputed estimator as a doubly robust imputation procedure. Double robustness is an attractive property that is closely related with the philosophy of model-assisted inference in survey sampling; see Särndal et al. (1992).

Recently, Han and Wang (2013) introduced the concept of multiple robustness in an infinite population set-up; see also Han (2014). Multiple robustness can be viewed as an extension of the concept of double robustness. In practice, multiple nonresponse models and multiple imputation models may be fitted, each involving different subsets of covariates and possibly different link functions. An estimation procedure is said to be multiply robust when it is consistent if any one of those multiple models, for either the response probability or the study variable, is correctly specified. Multiply robust procedures are attractive because they provide some protection if all but one model are misspecified. In practice, survey statisticians do not always perform all the necessary diagnostics to ensure that the link function is correctly specified or that appropriate interaction/curvature terms are included. In such cases, multiply robust procedures are attractive.

The paper is organized as follows. In Section 2, we present the basic theoretical set-up. In Section 3, motivated by Han and Wang (2013), we develop a multiply robust imputation procedure in the context of finite population sampling. We establish asymptotic properties of the resulting multiply robust imputed estimator in Section 4. Section 5 presents variance estimation for negligible sampling fractions. In Section 6, we propose multiply robust point and variance estimation procedures for non-negligible sampling fractions. Section 7 contains multiply robust random and fractional imputations. The results of a simulation study, assessing the performance of the proposed point and variance estimators are presented in Section 8. Section 9 concludes the article with some discussion.

2 Theoretical set-up

Consider a finite population U of size N. We are interested in estimating the population total of a study variable y, $Y = \sum_{i\in U} y_i$. We select a sample s, of size n, according to a sampling design $F(\mathbf{I})$, where $\mathbf{I} = (I_1, \ldots, I_N)^\top$ and I_i is a sample selection indicator associated with unit i such that $I_i = 1$ if unit $i \in s$ and $I_i = 0$, otherwise.

In the absence of nonresponse, a complete data estimator of Y is the expansion estimator given by

$$\widehat{Y}_\pi = \sum_{i\in s} w_i y_i,$$

where $w_i = 1/\pi_i$ denotes the design weight attached to unit i and π_i denotes its inclusion probability in the sample. The expansion estimator is design-unbiased and design-consistent for Y (e.g., Isaki and Fuller, 1982).

In the presence of nonresponse to the study variable y, an estimator of Y

$$\widehat{Y}_I = \sum_{i\in s} w_i r_i y_i + \sum_{i\in s} w_i (1 - r_i) y_i^*, \tag{1}$$

where y_i^* denotes the imputed value used to replace the missing value y_i and r_i is a response indicator attached to unit i, such that $r_i = 1$ if unit i is a respondent to the study variable y and $r_i = 0$, otherwise. Note that $\widehat{Y}_I$ is readily computed from an imputed data set with n rows, each row corresponding to a given sample unit, a column consisting of the design weights w_i and a column consisting of the $\widetilde{y}$-values, where $\widetilde{y}_i = r_i y_i + (1-r_i) y_i^*$. Throughout the paper, we assume that the MAR assumption (Rubin, 1976) holds:

$$\Pr(r_i = 1|\mathbf{x}_i, y_i) = \Pr(r_i = 1|\mathbf{x}_i) \equiv p_i.$$

The imputed values y_i^* are constructed on the basis of auxiliary information collected for both the respondents and the nonrespondents. Let $\mathbf{x}_i$ be a q-vector of auxiliary variables associated with unit i. We assume that the first component of $\mathbf{x}_i$ is 1 for all i. In order to construct the imputed values y_i^*, we postulate the following imputation model:

$$y_i = m(\mathbf{x}_i; \boldsymbol{\beta}) + \epsilon_i, \tag{2}$$

where $\boldsymbol{\beta}$ is a q vector of unknown coefficients. We assume that $E_m(\epsilon_i) = 0$, $E_m(\epsilon_i \epsilon_j) = 0, \quad i \neq j$ and $V(\epsilon_i) = \sigma^2$, where the subscript m refers to the imputation model (2).

Let s_r denote the set of respondents to the study variable y, of size n_r and $s_m = s - s_r$ denote the set of nonrespondents of size n_m such that $s = s_r \bigcup s_m$ and $n = n_r + n_m$. Deterministic imputation consists of replacing the missing value y_i by the imputed value y_i^* given by

$$y_i^* = m(\mathbf{x}_i; \widehat{\boldsymbol{\beta}}), \quad i \in s_m$$

where $\widehat{\boldsymbol{\beta}}$ is a solution of the estimating equation

$$\sum_{i \in s} \phi_i r_i \left\{ y_i - m\left(\mathbf{x}_i; \boldsymbol{\beta}\right) \right\} \frac{\partial m(\mathbf{x}_i; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0} \tag{3}$$

and ϕ_i in (3) is a coefficient attached to unit i. Regardless of the choice of ϕ_i, the imputed estimator (1) is consistent for Y if the imputation model (2) holds. The choice $\phi_i = w_i$ leads to survey weighted deterministic imputation, whereas the choice $\phi_i = 1$ leads to unweighted deterministic imputation. However, with these choices of ϕ, the imputed estimator (1) is generally biased if the imputation model (2) is misspecified. To cope with this problem, an alternative choice of ϕ_i can be obtained by postulating a nonresponse model:

$$p_i = p(\mathbf{x}_i; \boldsymbol{\alpha}), \tag{4}$$

where $\boldsymbol{\alpha}$ is a q-vector of unknown coefficients. Let $\widehat{p_i} = p(\mathbf{x}_i; \widehat{\boldsymbol{\alpha}})$ be an estimate of the response probability for unit i, where $\widehat{\boldsymbol{\alpha}}$ is a solution of the estimating equation

$$\sum_{i \in s} w_i \frac{r_i - p(\mathbf{x}_i; \boldsymbol{\alpha})}{p(\mathbf{x}_i; \boldsymbol{\alpha}) \left\{1 - p(\mathbf{x}_i; \boldsymbol{\alpha})\right\}} = \mathbf{0}.$$

The choice $\phi_i = w_i \widehat{p_i}^{-1} - 1$ ensures that the imputed estimator (1) is consistent for Y provided that the nonresponse model (4) is correctly specified, regardless of whether of not the imputation model (2) is correctly specified. Hence, with the choice $\phi_i = w_i \widehat{p_i}^{-1} - 1$, the resulting imputed estimator, denoted by $\widehat{Y}_{DR}$, is doubly robust; see, for example, Haziza and Rao (2006). As mentioned in Section 1, instead of fitting a single nonresponse model and a single imputation model, one may want to fit multiple nonresponse models and multiple imputation models. This is discussed in the next section.

3 The proposed method

In this section, we develop a multiply robust imputation procedure in the case of survey data. Let $\mathcal{C}_1 = \{p^j(\mathbf{x}_i; \boldsymbol{\alpha}^j); j = 1, \ldots, J\}$ denote the set consisting of J nonresponse models and $\mathcal{C}_2 = \{m^k(\mathbf{x}_i; \boldsymbol{\beta}^k); k = 1, \ldots, K\}$ be the set consisting of K imputation models.

The corresponding estimators $\widehat{\boldsymbol{\alpha}}^j$ and $\widehat{\boldsymbol{\beta}}^k$ are obtained by solving the following survey weighted estimating equations:

$$S_1^j(\boldsymbol{\alpha}^j) = \sum_{i\in s} w_i \frac{r_i - p^j(\mathbf{x}_i;\boldsymbol{\alpha}^j)}{p^j(\mathbf{x}_i;\boldsymbol{\alpha}^j)\left\{1 - p^j(\mathbf{x}_i;\boldsymbol{\alpha}^j)\right\}} = \mathbf{0} \tag{5}$$

and

$$S_2^k(\boldsymbol{\beta}^k) = \sum_{i\in s} w_i r_i \left\{y_i - m^k(\mathbf{x}_i;\boldsymbol{\beta}^k)\right\} \frac{\partial m^k(\mathbf{x}_i;\boldsymbol{\beta}^k)}{\partial \boldsymbol{\beta}} = \mathbf{0}, \tag{6}$$

respectively.

Our imputation procedure consists of two distinct steps: (i) In the first step, we obtain calibrated weights $\widetilde{w}_i$ as close as possible tothe initial weights w_i such that the following $J + K + 1$ calibration constraints are satisfied:

$$\sum_{i\in s_r} \widetilde{w}_i = \sum_{i\in s} w_i, \tag{7}$$

$$\frac{\sum_{i\in s_r} \widetilde{w}_i L\left\{1/p^j(\mathbf{x}_i;\widehat{\boldsymbol{\alpha}}^j)\right\}}{\sum_{i\in s_r}\widetilde{w}_i} = \frac{\sum_{i\in s} w_i L\left\{1/p^j(\mathbf{x}_i;\widehat{\boldsymbol{\alpha}}^j)\right\}}{\sum_{i\in s} w_i} \equiv \widehat{L}^j, \quad j = 1,\ldots,J, \tag{8}$$

and

$$\frac{\sum_{i\in s_r} \widetilde{w}_i m^k(\mathbf{x}_i;\widehat{\boldsymbol{\beta}}^k)}{\sum_{i\in s_r}\widetilde{w}_i} = \frac{\sum_{i\in s} w_i m^k(\mathbf{x}_i;\widehat{\boldsymbol{\beta}}^k)}{\sum_{i\in s} w_i} \equiv \widehat{m}^k, \quad k = 1,\ldots,K, \tag{9}$$

where $L(t)$ is the inverse function of $F(t)$, which is a calibration function defined below. The calibration constraints (7)-(9) are similar to those encountered in the context of model calibration for complete data (Wu and Sitter, 2001). More specifically, we seek calibrated weights $\widetilde{w}_i$ such that

$$\sum_{i\in s_r} G(\widetilde{w}_i/w_i)$$

is minimized subject to (7)-(9), where $G(\widetilde{w}_i/w_i)$ is such that $G(\widetilde{w}_i/w_i) \geq 0$, $G(1) = 0$, differentiable with respect to $\widetilde{w}_i$, strictly convex, with continuous derivatives $g(\widetilde{w}_i/w_i) = \partial G(\widetilde{w}_i/w_i)/\partial\widetilde{w}_i$ such that $g(1) = 0$; see Deville and Särndal (1992). Popular distance functions include (i) the generalized chi-square distance $G(\widetilde{w}_i/w_i) = 1/2\left\{(\widetilde{w}_i/w_i) - 1\right\}^2 w_i$; (ii) the pseudo-empirical likelihood distance $G(\widetilde{w}_i/w_i) = -w_i\log(\widetilde{w}_i/w_i) + \widetilde{w}_i - w_i$ and (iii) the Kullback-Leibler distance $G(\widetilde{w}_i/w_i) = \widetilde{w}_i\log(\widetilde{w}_i/w_i) - \widetilde{w}_i + w_i$. The weights $\widetilde{w}_i$ are given by

$$\widetilde{w}_i = w_i F(\widehat{\boldsymbol{\lambda}}_r^\top \mathbf{h}_i)$$

with $wF(.)$ denoting the inverse function of $g(.)$, $\widehat{\boldsymbol{\lambda}}_r^\top$ is a $J + K + 1$-vector of estimated coefficients and

$$\mathbf{h}_i = \left(1, \widehat{L}_i^1 - \widehat{L}^1, \ldots, \widehat{L}_i^J - \widehat{L}^J, \widehat{m}_i^1 - \widehat{m}^1, \ldots, \widehat{m}_i^K - \widehat{m}^K\right)^\top, \tag{10}$$

where $\widehat{L}_i^j \equiv L\left\{1/p^j(\mathbf{x}_i;\widehat{\boldsymbol{\alpha}}^j)\right\}$ and $\widehat{m}_i^k \equiv m^k(\mathbf{x}_i;\widehat{\boldsymbol{\beta}}^k)$. The function $F(.)$ is often referred to as the calibration function. In the case of the generalized chi-square distance, $\widehat{L}_i^j = 1/p^j(\mathbf{x}_i;\widehat{\boldsymbol{\alpha}}^j)$ and the weights $\widetilde{w}_i$ reduce to

$$\widetilde{w}_i = w_i(1 + \widehat{\boldsymbol{\lambda}}_r^\top \mathbf{h}_i), \tag{11}$$

where

$$\widehat{\boldsymbol{\lambda}}_r = \left(\sum_{i\in s_r} w_i \mathbf{h}_i \mathbf{h}_i^\top\right)^{-1} \left(\sum_{i\in s} w_i \mathbf{h}_i - \sum_{i\in s_r} w_i \mathbf{h}_i\right).$$

With the generalized chi-square distance, some weights $\widetilde{w}_i$ in (11) may be negative. Both the pseudo-empirical likelihood distance and the Kullback-Leibler distance ensures that $\widetilde{w}_i > 0$ for all i. For short, we write $\widehat{F}_i$ for $F(\widehat{\boldsymbol{\lambda}}_r^\top \mathbf{h}_i)$ in the sequel.

(ii) In the second step, the imputed values y_i^* are obtained by by fitting a weighted linear regression with y as the dependent variable, $\mathbf{h}$ given by (10) as the vector of independent variables and $w_i(\widehat{F}_i - 1)$ as the weights. This leads to

$$y_i^* = \mathbf{h}_i^\top \widehat{\boldsymbol{\gamma}}_p, \quad i \in s_m, \tag{12}$$

where

$$\widehat{\boldsymbol{\gamma}}_p = \left\{\sum_{i\in s} r_i w_i \left(\widehat{F}_i - 1\right) \mathbf{h}_i \mathbf{h}_i^\top\right\}^{-1} \left\{\sum_{i\in s} r_i w_i \left(\widehat{F}_i - 1\right) \mathbf{h}_i y_i\right\}. \tag{13}$$

The vector of estimated coefficients $\widehat{\boldsymbol{\gamma}}_p$ can be viewed as a weighted least square estimator. The resulting imputed estimator, denoted by $\widehat{Y}_{MR}$, is given by

$$\widehat{Y}_{MR} = \sum_{i\in s} r_i w_i y_i + \sum_{i\in s} (1 - r_i)\, w_i \mathbf{h}_i^\top \widehat{\boldsymbol{\gamma}}_p. \tag{14}$$

Because the first component of the vector $\mathbf{h}_i$ is equal to 1 for all i, the imputed estimator (14) can alternatively be written as

$$\widehat{Y}_{MR} = \sum_{i\in s_r} w_i \widehat{F}_i y_i + \left(\sum_{i\in s} w_i \mathbf{h}_i - \sum_{i\in s_r} w_i \widehat{F}_i \mathbf{h}_i\right)^\top \widehat{\boldsymbol{\gamma}}_p. \tag{15}$$

The form (14) is often referred to as the projection form. In the next section, we show that $\widehat{Y}_{MR}$ is multiply robust. For this reason, the imputation procedure (12) will be referred to as a multiply robust deterministic imputation procedure.

4 Asymptotic results

The following theorem establishes the consistency of $\widehat{Y}_{MR}$ when one of the imputation model is true or when the true model is a linear combination of the multiple imputation models.

Theorem 1. *If one of the imputation models is true or the true model is a linear combination of the multiple imputation models, then the proposed estimator $\widehat{Y}_{MR}$ in (14) is consistent.*

A sketch of the proof is presented in Appendix A. The following theorem establishes the consistency of $\widehat{Y}_{MR}$ when one of the nonresponse models is true.

Theorem 2. *If one of the nonresponse model is true, then the proposed estimator $\widehat{Y}_{MR}$ in (14) is consistent.*

A sketch of the proof is presented in Appendix B. Combining Theorem 1 and Theorem 2, we conclude that $\widehat{Y}_{MR}$ is multiply robust, in the sense that it is consistent if all but one of the model are misspecified. Finally, the next theorem presents an asymptotic expression of $\widehat{Y}_{MR}$.

Theorem 3. *Under the regularity conditions (C1)-(C4) in Appendix C, the proposed estimator (14) has the following expansion*

$$\widehat{Y}_{MR} = \sum_{i\in s} w_i \eta_{i,0} + O_p(Nn^{-1}),$$

where

$$\eta_{i,0} = r_i F_i y_i + C_1(1 - r_i F_i) h_{i,0} + C_2 S(\mathbf{x}_i; \boldsymbol{\theta}^*),$$

with

$$C_1 = 1 + E\left(\sum_{i\in s} w_i r_i \dot{F}_i \mathbf{h}_{i,0} e_{i,0}\right)\left\{E\left(\sum_{i\in s} w_i r_i \dot{F}_i \mathbf{h}_{i,0}\mathbf{h}_{i,0}^\top\right)\right\}^{-1}$$

and

$$C_2 = E\left\{\frac{1}{N}\left(\sum_{i\in s} w_i r_i \dot{F}_i \boldsymbol{\lambda}^* \dot{\mathbf{h}}_{i,0} e_{i,0} - \sum_{i\in s} w_i (r_i F_i - 1) r_i \dot{\mathbf{h}}_{i,0}\right)\right\},$$

where $S(\mathbf{x}_i; \boldsymbol{\theta}^*)$ is defined in Appendix C, $\mathbf{h}_{i,0} = \mathbf{h}(\mathbf{x}_i; \boldsymbol{\theta}^*)$, $\dot{\mathbf{h}}_{i,0} = \dot{\mathbf{h}}(\mathbf{x}_i; \boldsymbol{\theta}^*)$, $\dot{\mathbf{h}}(\mathbf{x}_i; \boldsymbol{\theta}^*) = \partial \mathbf{h}(\mathbf{x}_i; \boldsymbol{\theta})/\partial \boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, $e_{i,0} = y_i - \boldsymbol{\gamma}^{*\top} \mathbf{h}_{i,0}$ and $\dot{F}_i = \partial F(t)/\partial t$ evaluated at $t = \boldsymbol{\lambda}^* \mathbf{h}_{i,0}$.

The sketched proof of Theorem 3 is contained in Appendix C.

5 Variance estimation for negligible sampling fractions

In this section, we propose a variance estimator for the multiply robust estimator $\widehat{Y}_{MR}$ given by (14). We assume that the sampling fraction n/N is negligible. The case of non-negligible sampling fractions is discussed In Section 6. It is well known that treating the imputed values as observed values leads to an underestimation of the variance of the imputed estimators and ultimately to confidence intervals that are too narrow. In the case of negligible sampling fractions, several variance estimation methods, taking nonresponse and imputation into account, have been proposed in the literature including the adjusted jackknife variance estimator of Rao and Shao (1992) and the bootstrap variance estimator of Shao and Sitter (1996).

To motivate our variance estimation procedures, we use the reverse framework for variance estimation, which reverses the actual of sampling and nonresponse (Fay, 1991; Shao and Steel, 1999): instead of sampling the units first, we start by dividing the population U into a population of respondents U_r and a population of nonrespondents U_m according to the nonresponse mechanism and then select the sample s using the sampling design. The reverse decomposition may be used when the sampling design is independent of the non-response mechanism. This condition is similar to that of strong invariance in the context of two-phase sampling designs (Beaumont and Haziza, 2015).

In the case of deterministic imputation, the total variance of $\widehat{Y}_{MR}$ may be decomposed as

$$V_T = V_1 + V_2, \tag{16}$$

where

$$V_1 = EV_p\left(\widehat{Y}_{MR} - Y \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{r}\right)$$

and

$$V_2 = VE_p\left(\widehat{Y}_{MR} - Y \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{r}\right)$$

with $\mathbf{y}_U = (y_1, \ldots, y_N)^\top$, $\mathbf{x}_U = (x_1, \ldots, x_N)^\top$ and $\mathbf{r} = (r_1, \ldots, r_N)^\top$.

Under mild regularity conditions, the contribution of the term V_2 to the total variance, V_2/V_T is $O(n/N)$ (Shao and Steel, 1999). Thus, when the sampling fraction n/N is negligible, the contribution of the term V_2 to the total variance is negligible and may be omitted. It remains to estimate the term V_1 consistently, which requires obtaining a consistent estimator of $V_p\left(\widehat{Y}_{MR} - Y \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{r}\right)$. Conditionally on $\mathbf{y}_U$, $\mathbf{x}_U$ and $\mathbf{r}$, the estimator $\widehat{Y}_{MR}$ is expressed as a smooth function of estimated totals. Thus, the problem of estimating $V_p\left(\widehat{Y}_{MR} - Y \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{r}\right)$ reduces to the classical problem of estimating the sampling variance of a smooth function of estimated totals, conditionally on $\mathbf{y}_U$, $\mathbf{x}_U$ and $\mathbf{r}$. Therefore, any complete data variance estimation procedure may be used, including Taylor expansion procedures or resampling methods such as the jackknife, the random group method or the bootstrap; e.g., Wolter (2007).

We start by a variance estimator based on a Taylor expansion procedure. According to Theorem 3, the term $V_p\left(\widehat{Y}_{MR} - Y \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{r}\right)$ can be estimated consistently by

$$\widehat{V}_1 = \sum_{i\in s}\sum_{j\in s} \frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}} \frac{\widehat{\eta}_i}{\pi_i}\frac{\widehat{\eta}_j}{\pi_j}, \tag{17}$$

where

$$\widehat{\eta}_i = r_i\widehat{F}_i y_i + \widehat{C}_1(1 - r_i\widehat{F}_i)h_i + \widehat{C}_2 S(\mathbf{x}_i; \widehat{\boldsymbol{\theta}}),$$

$$\widehat{C}_1 = 1 + \left(\sum_{i\in s} w_i r_i \widehat{\dot{F}}_i \mathbf{h}_i \widehat{e}_i\right)\left(\sum_{i\in s} w_i r_i \widehat{\dot{F}}_i \mathbf{h}_i \mathbf{h}_i^\top\right)^{-1}$$

and

$$\widehat{C}_2 = \frac{1}{\widehat{N}}\left\{\sum_{i\in s} w_i r_i \widehat{\dot{F}}_i \widehat{\boldsymbol{\lambda}}_r \dot{\mathbf{h}}(\mathbf{x}_i; \widehat{\boldsymbol{\theta}})\widehat{e}_i - \sum_{i\in s} w_i(r_i\widehat{F}_i - 1)r_i \widehat{\dot{\mathbf{h}}}_i\right\},$$

with $\widehat{N} = \sum_{i\in s} w_i$, $\dot{\mathbf{h}}_i = \dot{\mathbf{h}}(\mathbf{x}_i; \widehat{\boldsymbol{\theta}})$, $\dot{\mathbf{h}}(\mathbf{x}_i; \widehat{\boldsymbol{\theta}}) = \partial \mathbf{h}(\mathbf{x}_i; \boldsymbol{\theta})/\partial\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, $\widehat{e}_i = y_i - \widehat{\boldsymbol{\gamma}}_p^\top \mathbf{h}_i$ and $\widehat{\dot{F}}_i = \partial F(t)/\partial t$ evaluated at $t = \widehat{\boldsymbol{\lambda}}_r \mathbf{h}_i$.

Alternatively, to estimate the term $V_p\left(\widehat{Y}_{MR} - Y \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{r}\right)$, we consider a jackknife variance estimator that does not require adjusting the imputed values, unlike the Rao-Shao jackknife variance estimator. We illustrate the method in the case of the multiply robust estimator given by (14) and simple random sampling without replacement. Let $w_{i(j)}$ be the so-called jackknife weights given by

$$w_{i(j)} = \begin{cases} \frac{n}{n-1} w_i & \text{if } i \neq j \\ 0 & \text{if } i = j. \end{cases}$$

As an estimator of $V_p\left(\widehat{Y}_{MR} - Y \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{r}\right)$, we consider the jackknife variance estimator

$$\widehat{V}_J = \frac{n-1}{n} \sum_{i \in s} (\widehat{Y}_{MR(j)} - \widehat{Y}_{MR})^2, \tag{18}$$

where

$$\widehat{Y}_{MR(j)} = \sum_{i \in s} w_{i(j)} r_i y_i + \sum_{i \in s} w_{i(j)} (1 - r_i) \mathbf{h}_{i(j)}^\top \widehat{\boldsymbol{\gamma}}_{p(j)}$$

with $\widehat{\boldsymbol{\gamma}}_{p(j)}$ computed the same way as $\widehat{\boldsymbol{\gamma}}_p$ in (13) but with the jackknife weights $w_{i(j)}$ instead of the original weights w_i and with $\mathbf{h}_i$ replaced by $\mathbf{h}_{i(j)}$. Note that obtaining $\mathbf{h}_{i(j)}$ involves fitting the J nonresponse models and the K imputation models after deletion of unit j.

When the sampling fraction n/N is negligible, both variance estimators $\widehat{V}_1$ and $\widehat{V}_J$ in (17) and (18), respectively, are consistent estimators of $V_p\left(\widehat{Y}_{MR} - Y \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{r}\right)$, which implies that they are consistent for V_1 in (16). It is worth noting that the consistency property holds even if all the nonresponse models and all the imputation models are misspecified. Since the proposed variance estimator does not depend on the validity of the assumed models, it is also multiply robust. For unequal probability sampling designs, jackknife variance estimation can be performed through the use the complete data jackknife variance estimator of Berger (2007), provided that the sampling design belongs to the class of high entropy sampling designs that includes Conditional Poisson sampling (CPS) and the Rao-Sampford method as special cases.

Finally, a $(1 - \alpha\%)$ confidence interval for Y is

$$\widehat{Y}_{MR} \pm z_{\alpha/2} \sqrt{\widehat{V}}, \tag{19}$$

where $z_{\alpha/2}$ denotes the upper $(1-\alpha)/2$ critical value for the standard normal distribution and $\widehat{V}$ denotes either (17) or (18). The confidence interval (19) is multiply robust in the sense that its coverage probability is close to the nominal rate if all but one model are misspecified and the sampling fraction n/N is negligible.

6 Point and variance estimation for non-negligible sampling fractions

In Section 3, we introduced a multiply robust estimator, $\widehat{Y}_{MR}$, based on estimated coefficients $\widehat{\boldsymbol{\alpha}}^j$, $j = 1, \ldots, J$ and $\widehat{\boldsymbol{\beta}}^k$, $k = 1, \ldots, K$ that were obtained by solving the weighted score equations (5) and (6), respectively. In Section 5, the variance estimator based on a Taylor expansion procedure involved relatively messy calculations, which can be explained by the fact that it accounted for the variability associated with the $\widehat{\boldsymbol{\alpha}}^j$'s and the $\widehat{\boldsymbol{\beta}}^k$'s. For this reason, we considered a jackknife variance estimator that did not involve cumbersome derivation, but was highly computer intensive. In this section, we modify the procedure for estimating the $\boldsymbol{\alpha}^j$'s and the $\boldsymbol{\beta}^k$'s so that, at the variance estimation stage, we can safely ignore the variability associated with the estimation of these parameters. As a result, obtaining a variance estimator based on Taylor expansion procedures involves much simpler derivations.

Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\alpha}}^1, \ldots, \widehat{\boldsymbol{\alpha}}^J, \widehat{\boldsymbol{\beta}}^1, \ldots, \widehat{\boldsymbol{\beta}}^K)$ denote the $J+K$ vector of estimated coefficients. According to (B.2) in the proof of Theorem 3, we have

$$\begin{aligned}
\widehat{Y}_{MR}\left(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}_r\right) &= \widehat{Y}_\pi + \sum_{i \in s} w_i \left(r_i F_i - 1\right) e_{i,0} + \sum_{i \in s} w_i r_i \dot{F}_i \mathbf{h}_{i,0} e_{i,0} \left(\widehat{\boldsymbol{\lambda}}_r - \boldsymbol{\lambda}^*\right) \\
&+ \left\{\sum_{i \in s} w_i r_i \dot{F}_i \boldsymbol{\lambda}^* \dot{\mathbf{h}}_{i,0} e_{i,0} - \sum_{i \in s} w_i \left(r_i F_i - 1\right) \boldsymbol{\gamma}^* \dot{\mathbf{h}}_{i,0}\right\} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) \\
&+ O_p\left(n^{-1} N\right),
\end{aligned}$$

If the following conditions

$$\sum_{i \in s} w_i r_i \dot{F}_i \dot{\mathbf{h}}_{i,0} e_{i,0} = o_p(1), \quad \sum_{i \in s} w_i \left(r_i F_i - 1\right) \dot{\mathbf{h}}_{i,0} = o_p(1), \tag{20}$$

and

$$\sum_{i \in s} w_i r_i \dot{F}_i \mathbf{h}_{i,0} e_{i,0} = o_p(1), \tag{21}$$

are satisfied, it follows that

$$\widehat{Y}_{MR} = \widehat{Y}_\pi + \sum_{i \in s} w_i \left(r_i F_i - 1\right) e_{i,0} + o_p\left(n^{-1/2} N\right) \tag{22}$$

and the variability of $\widehat{\boldsymbol{\theta}}$ can be safely ignored.

In order to achieve (20) and (21), consider the following iterative procedures:

(Step1). Obtain the initial estimates $\widehat{\boldsymbol{\theta}}^{(0)}$ by solving the weighted score equations (5) and (6).

(Step2). Use $\widehat{\boldsymbol{\theta}}^{(t-1)}$ to calculate $\widehat{\boldsymbol{\lambda}}_r^{(t-1)}$ by minimizing proposed distance function subject to constraints (7)-(9).

(Step3). Calculate $\widehat{\boldsymbol{\theta}}^{(t)}$ by using the generalized method of moment (GMM) method which minimizes the distance $G(\boldsymbol{\theta}) \mathbf{D} G^\top(\boldsymbol{\theta})$ with

$$G(\boldsymbol{\theta}) = \left(A_1(\boldsymbol{\theta}), \quad A_2(\boldsymbol{\theta}), \quad A_3(\boldsymbol{\theta})\right),$$

where

$$A_1(\boldsymbol{\theta}) = \sum_{i \in s} w_i r_i \dot{F}_i^{(t-1)} \dot{\mathbf{h}}_i (y_i - \widehat{\boldsymbol{\gamma}}_p^{(t-1)} \mathbf{h}_i), \quad A_2(\boldsymbol{\theta}) = \sum_{i \in s} w_i \left(r_i F_i^{(t-1)} - 1\right) \dot{\mathbf{h}}_i,$$

$$A_3(\boldsymbol{\theta}) = \sum_{i \in s} w_i r_i \dot{F}_i^{(t-1)} \mathbf{h}_i (y_i - \widehat{\boldsymbol{\gamma}}_p^{(t-1)} \mathbf{h}_i),$$

and $\mathbf{D}$ is any positive definite matrix.

(Step4). Repeat (Step2) and (Step3) by using the updated estimates until convergence.

Turning to variance estimation, the total variance of $\widehat{Y}_{MR}$ with respect to the nonresponse model (NM) approach can be expressed as

$$V_T^{NM} = V_1^{NM} + V_2^{NM},$$

where

$$V_1^{NM} = E\left\{V\left(\widehat{Y}_{MR} \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{r}\right) \mid \mathbf{y}_U, \mathbf{x}_U\right\}$$

and

$$V_2^{NM} = V\left\{E\left(\widehat{Y}_{MR} \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{r}\right) \mid \mathbf{y}_U, \mathbf{x}_U\right\}.$$

Under the imputation model (IM) approach, the total variance of $\widehat{Y}_{MR}$ can be written as

$$V_T^{IM} = V_1^{IM} + V_2^{IM},$$

where

$$V_1^{IM} = E\left\{V\left(\widehat{Y}_{MR} - Y \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{r}\right) \mid \mathbf{x}_U, \mathbf{r}\right\}$$

and

$$V_2^{IM} = V\left\{E\left(\widehat{Y}_{MR} - Y \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{r}\right) \mid \mathbf{x}_U, \mathbf{r}\right\}.$$

According to (22), we can estimate V_1^{NM} and V_1^{IM} using any complete data variance estimation procedure such as resampling method or Taylor linearization procedures. Using a first-order Taylor expansion, an asymptotically unbiased variance estimator of either V_1^{NM} or V_1^{IM} is given by

$$\widehat{V}_1 = \sum_{i\in s}\sum_{j\in s} \frac{(\pi_{ij} - \pi_i\pi_j)}{\pi_{ij}} \frac{\widehat{\eta}_i}{\pi_i} \frac{\widehat{\eta}_j}{\pi_j},$$

where $\widehat{\eta}_i = r_i F\left(\widehat{\boldsymbol{\lambda}}_r^\top \mathbf{h}_i\right) y_i + \widehat{\mathbf{A}}_p \widehat{\mathbf{B}}_p^{-1}\left\{1 - r_i F\left(\widehat{\boldsymbol{\lambda}}_r^\top \mathbf{h}_i\right)\right\}\mathbf{h}_i$, with $\widehat{\mathbf{A}}_p = \sum_{i\in s_r} w_i \left\{F\left(\widehat{\boldsymbol{\lambda}}_r^\top \mathbf{h}_i\right) - 1\right\} y_i \mathbf{h}_i$ and $\widehat{\mathbf{B}}_p = \sum_{i\in s_r} w_i \left\{F\left(\widehat{\boldsymbol{\lambda}}_r^\top \mathbf{h}_i\right) - 1\right\}\mathbf{h}_i\mathbf{h}_i^\top$. The estimator $\widehat{V}_1$ is multiply robust as its validity does not depend on the validity of the assumed nonresponse or imputation models.

Now, an estimator V_2^{NM} is given by

$$\widehat{V}_2 = \sum_{i\in s} w_i r_i F\left(\widehat{\boldsymbol{\lambda}}_r^\top \mathbf{h}_i\right)\left\{F\left(\widehat{\boldsymbol{\lambda}}_r^\top \mathbf{h}_i\right) - 1\right\}\widehat{e}_i^2,$$

where $\widehat{e}_i = y_i - \widehat{\mathbf{A}}_p \widehat{\mathbf{B}}_p^{-1}\mathbf{h}_i$. The estimator $\widehat{V}_2$ is consistent if all but one nonresponse models are misspecified. However, it is biased with respect to the imputation model. In fact, it can be shown that the bias of $\widehat{V}_2$ with respect to the true imputation model is given by

$$\begin{aligned} B(\widehat{V}_2) &= E(\widehat{V}_2) - V_2^{IM} \\ &= \sum_{i\in U}\left(r_i p_i^{-1} - 1\right) V(y_i|\mathbf{x}_i). \end{aligned}$$

Following Kim and Haziza (2014), we propose the following bias-corrected multiply robust variance estimator of the variance of $\widehat{Y}_{MR}$:

$$\widehat{V}_T = \widehat{V}_1 + \widehat{V}_2 - \widehat{B}(\widehat{V}_2),$$

where

$$\widehat{B}(\widehat{V}_2) = \sum_{i \in s} w_i \left\{ r_i F\left(\widehat{\boldsymbol{\lambda}}_r^\top \mathbf{h}_i\right) - 1 \right\} \widehat{V}(y_i|\mathbf{x}_i)$$

and $\widehat{V}(y_i|\mathbf{x}_i)$ is a multiply robust estimator of $V(y_i|\mathbf{x}_i)$. If one of the nonresponse models is correctly specified, then $E\left\{\widehat{B}(\widehat{V}_2)\right\} \approx 0$ as $F\left(\widehat{\boldsymbol{\lambda}}_r^\top \mathbf{h}_i\right) \approx p_i^{-1}$. To obtain a multiply robust estimator of $V(y_i|\mathbf{x}_i)$, suppose that, for each of the K imputation models we have $V(y_i \mid \mathbf{x}_i) = \psi(\mathbf{x}_i; \boldsymbol{\delta}_0^k) \quad k = 1, \ldots, K$ for some $\boldsymbol{\delta}_0^k$. Suppose that a consistent estimator $\widehat{\boldsymbol{\delta}}$ is available. Then, one can perform the estimation process similar to that described in Section 3, with y_i^2 as the dependent variable to obtain a multiply robust estimator $\widehat{E}(y_i^2|\mathbf{x}_i)$ of $E(y_i^2|\mathbf{x}_i)$ first, and estimate $V(y_i|\mathbf{x}_i)$ by using $\widehat{V}(y_i|\mathbf{x}_i) = \widehat{E}(y_i^2|\mathbf{x}_i) - (\mathbf{h}_i^\top \widehat{\boldsymbol{\gamma}}_p)^2$.

7 Multiply robust random and fractional imputation

In this section, we consider two random counterparts of the multiply robust imputation procedure (12): (i) multiply robust random imputation and (ii) multiply robust fractional imputation.

7.1 Random imputation

First, we consider a random imputation procedure that consists of replacing the missing y_i by

$$y_{ij}^* = \mathbf{h}_i^\top \widehat{\boldsymbol{\gamma}}_p + e_j^*, \tag{23}$$

where $e_j^* = y_j - \mathbf{h}_j^\top \widehat{\boldsymbol{\gamma}}_p,\ j \in s_r$, is selected at random from the set of residuals observed among the responding units with probability

$$w_{ij}^* = \frac{w_j(\widehat{F}_j - 1)}{\sum_{k \in s} r_k w_k(\widehat{F}_k - 1)}.$$

The resulting imputed estimator obtained by using (23) in (1) is denoted by $\widehat{Y}_{MRR}$. The latter is multiply robust as $E\left(e_j^* \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{I}, \mathbf{r}\right) = 0$. Its variance can be expressed as

$$V(\widehat{Y}_{MRR}) \;=\; V(\widehat{Y}_{MR}) + E\left\{V\left(\widehat{Y}_{MRR} - \widehat{Y}_{MR} \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{I}, \mathbf{r}\right)\right\}, \tag{24}$$

where the first term on the right hand-side of (24) denotes the variance of $\widehat{Y}_{MR}$ under the multiply robust deterministic imputation procedure (12), whereas the second term denotes the imputation variance arising from the random selection of the residuals e_j^*. Therefore, a multiply robust variance estimator of $V(\widehat{Y}_{RI})$ is given by

$$\widehat{V}(\widehat{Y}_{MRR}) \;=\; \widehat{V}(\widehat{Y}_{MR}) + \widehat{V}\left(\widehat{Y}_{MRR} - \widehat{Y}_{MR} \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{I}, \mathbf{r}\right),$$

where $\widehat{V}(\widehat{Y}_{MR})$ is an estimator of the variance of $\widehat{Y}_{MR}$ under the multiply robust deterministic imputation procedure (12) (see Sections 5 and 6) and

$$\widehat{V}\left(\widehat{Y}_{MRR} - \widehat{Y}_{MR} \mid \mathbf{y}_U, \mathbf{x}_U, \mathbf{I}, \mathbf{r}\right) = \sum_{i \in s} w_i^2 (1 - r_i) \sum_{j \in s_r} w_{ij}^* \left(y_{ij}^* - \sum_{j \in s_r} w_{ij}^* y_{ij}^*\right)^2.$$

7.2 Fractional imputation

We now turn to fractional imputation that was considered in Kim and Fuller (2004) and Fuller and Kim (2005). The imputed estimator (1) under fractional imputation can be written as

$$\widehat{Y}_{MRF} = \sum_{i \in s} w_i r_i y_i + \sum_{i \in s} w_i (1 - r_i) \sum_{j \in s_r} w_{ij}^* y_{ij}^*, \tag{25}$$

where y_{ij}^* is given by (23) and w_{ij}^* denotes the fraction of the original weight of recipient i assigned to the value from donor j. We have $w_{jj}^* = 1$ and $\sum_{j \in s_r} w_{ij}^* = 1$. It follows that the estimator (25) can be rewritten as

$$\widehat{Y}_{MRF} = \sum_{i \in s} w_i \mathbf{h}_i^\top \widehat{\boldsymbol{\gamma}}_p + \sum_{i \in s} r_i \left\{ w_i + \sum_{j \in s} (1 - r_j) w_j w_{ij}^* \right\} e_i^*. \tag{26}$$

Comparing (26) with (15), we have

$$w_i + \sum_{j \in s} (1 - r_j) w_j w_{ij}^* = w_i \widehat{F}_i. \tag{27}$$

Because $\sum_{j \in s} (1 - r_j) w_j = \sum_{j \in s} r_j w_j (\widehat{F}_j - 1)$, the fractional weights satisfying (27) are given by

$$w_{ij}^* = \frac{w_i (\widehat{F}_i - 1)}{\sum_{i \in s} r_i w_i (\widehat{F}_i - 1)}. \tag{28}$$

Using the fractional weights (28) ensures that the imputation variance is eliminated. The estimator $\widehat{Y}_{MRF}$ is then said to be fully efficient, a term coined by Kim and Fuller (2004). Its asymptotic properties are thus identical to those of (14). As a result, the estimator $\widehat{Y}_{MRF}$ is multiply robust.

8 Simulation study

We performed a simulation study to assess the performance of the multiply robust estimator $\widehat{Y}_{MR}$ in terms of bias and efficiency. In addition, we assessed the performance of the proposed jackknife variance estimator presented in Section 5 in terms of relative bias and coverage probability.

8.1 Monte Carlo properties of point estimators

We generated $B = 2,000$ finite populations of size $N = 10,000$, each one consisting of three variables: two auxiliary variables x and z and a study variable y. First, the x-values were generated independently from a uniform distribution with parameters -2.5 and 2.5. The z-values were generated according to $z_i = 0.5\chi_i + 1$, where χ_i was generated from a chi-square distribution with one degree of freedom. Given the x-values, the y-values were generated according to

$$y_i = m(x_i) + \epsilon_i, \quad i = 1, \ldots, N,$$

where the errors ϵ_i were generated from a normal distribution with mean equal to 0 and variance equal to $4x^2 + 2$. We used two choices for $m(x)$: (i) (IM1). $m(x) = 1 + 2x + 3x^2$;

and (ii) (IM2). $m(x) = 1 + 2x + 3\exp(x)$.

In each population, samples of expected size $n = 300$ were selected according to Poisson sampling with inclusion probability proportional-to-size. That is, $\pi_i = 300 z_i / \sum_{i \in U} z_i$.

In each sample, nonresponse to the study variable y was generated according to two distinct nonresponse mechanisms. More specifically, the response indicators r_i were generated independently from a Bernoulli distribution with probability p_i, where p_i was assigned as follows: (i) (NM1) $p_i = \{1 + \exp(0.8 + 0.5x_i - 0.3x_i^2)\}^{-1}$; and (NM2) $p_i = 1 - \exp\left[-\exp\{0.5 + 0.5x_i - 0.3\exp(x_i)\}\right]$. The overall response rate was set to 0.5.

This led to four types of scenarios: in the first one (NM1-IM1), the true imputation model is given by IM1 and nonresponse was generated according NM1. In the second (NM2-IM1), the true imputation model is given by IM1 and nonresponse was generated according NM2. In the third (NM1-IM2), the true imputation model is given by IM2 and nonresponse was generated according NM2. Finally, in the fourth (NM2-IM2), the true imputation model is given by IM2 and nonresponse was generated according NM2.

We were interested in estimating the finite population mean, $\overline{Y} = N^{-1} \sum_{i \in U} y_i$. We computed 3 estimators of $\overline{Y}$:

1. The complete data estimator (COM): $\widehat{\overline{Y}}_{COM} = \sum_{i \in s} w_i y_i / \sum_{i \in s} w_i$, which assumes no missing values.

2. Four doubly robust estimators (DR) of the form $\widehat{\overline{Y}}_{DR} = \widehat{Y}_{DR} / \sum_{i \in s} w_i$, where $\widehat{Y}_{DR}$ is described in Section 2: $\widehat{\overline{Y}}_{DR}(1010)$, $\widehat{\overline{Y}}_{DR}(1001)$, $\widehat{\overline{Y}}_{DR}(0110)$ and $\widehat{\overline{Y}}_{DR}(0101)$. The four digits between parentheses indicate which models are used in the estimation. The first two digits correspond to the nonresponse models NM1 and NM2, respectively, whereas the last two digits correspond to the imputation models IM1 and IM2, respectively. For example, the estimator $\widehat{\overline{Y}}_{DR}(1010)$ corresponds to the doubly robust estimator $\widehat{\overline{Y}}_{DR}$, for which the imputed values (3) were obtained by fitting the nonresponse model NM1 and the imputation model IM1.

3. Five multiply robust estimators of the form $\widehat{\overline{Y}}_{MR} = \widehat{Y}_{MR} / \sum_{i \in s} w_i$ based on the chi-squared distance function, where $\widehat{Y}_{MR}$ is given by (14): $\widehat{\overline{Y}}_{MR}(1110)$, $\widehat{\overline{Y}}_{MR}(1101)$, $\widehat{\overline{Y}}_{MR}(1011)$, $\widehat{\overline{Y}}_{MR}(011$ and $\widehat{\overline{Y}}_{MR}(1111)$. Once again, the four digits between parentheses indicate which models were used in the construction of the imputed values (12). For example, the estimator $\widehat{\overline{Y}}_{MR}(1111)$ denotes the multiply robust estimator based on the four models NM1, NM2, IM1 and IM2.

As a measure of bias of an estimator $\widehat{\theta}$ of a parameter θ, we computed the Monte Carlo bias given by

$$B_{MC}(\widehat{\theta}) = E_{MC}(\widehat{\theta}) - \theta,$$

where $E_{MC}(\widehat{\theta}) = R^{-1} \sum_{r=1}^{R} \widehat{\theta}_{(r)}$ with $\widehat{\theta}_{(r)}$ denoting the estimator $\widehat{\theta}$ in the r-th iteration. As a measure of efficiency, we computed the Monte Carlo standard error (SE) of $\widehat{\theta}$:

$$SE_{MC}(\widehat{\theta}) = \left[\frac{1}{R} \sum_{r=1}^{R} \left\{ \widehat{\theta}_{(r)} - E_{MC}(\widehat{\theta}) \right\}^2 \right]^{1/2}.$$

Finally, we computed the Monte Carlo root mean square error (RMSE) of $\widehat{\theta}$:

$$RMSE_{MC}(\widehat{\theta}) = \left\{ \frac{1}{R} \sum_{r=1}^{R} \left(\widehat{\theta}_{(r)} - \theta \right)^2 \right\}^{1/2}.$$

Table 1 shows the Monte Carlo bias, the standard deviation and the root mean squared error of ten estimators.

From Table 1, we note that the complete data estimator $\widehat{\overline{Y}}_{COM}$ showed negligible bias in all the scenarios, as expected. Its efficiency was greater than that of all the other estimators, which is not surprising as $\widehat{\overline{Y}}_{COM}$ does not suffer from variability due to nonresponse. Turning to the doubly robust estimators, their bias was small when either the nonresponse model or the imputation was correctly specified. For example, in the scenario (NM1-IM1) all the doubly robust estimators having either the first digit or the third digit equal to 1 showed negligible bias, as expected. However, when both models were specified incorrectly (e.g., $\widehat{\overline{Y}}_{DR}(0101)$) the relative bias was large, with a value equal to 0.1582. In terms of efficiency, all the doubly robust estimators showed very similar performances in all the scenarios. Finally, the multiply robust estimators showed negligible bias in all the scenarios. In terms of efficiency, they performed as well as the doubly robust estimators, illustrating that the multiply robust estimators do not seem to suffer from instability.

8.2 Monte Carlo properties of variance estimators

We also assessed the performance of the proposed jackknife variance estimator (see Section 5) in terms of relative bias, coverage probability of normal confidence intervals, and average length of the confidence interval. For simplicity, we confine our discussion to the case of the first scenario, (NM1-IM1).

As a measure of bias of $\widehat{V}_J$, we used the Monte Carlo percent relative bias

$$RB_{MC}(\widehat{V}_J) = 100 \times \frac{E_{MC}(\widehat{V}_J) - V_{MC}(\widehat{\overline{Y}}_{MR})}{V_{MC}(\widehat{\overline{Y}}_{MR})},$$

where $E_{MC}(\widehat{V}_J) = R^{-1} \sum_{r=1}^{R} \widehat{V}_{J(r)}$ and $\widehat{V}_{J(r)}$ denotes the estimator $\widehat{V}_J$ in the r-th sample and

$$V_{MC}(\widehat{\overline{Y}}_{MR}) = \frac{1}{R} \sum_{r=1}^{R} \left\{ \widehat{\overline{Y}}_{MR} - E_{MC}(\widehat{\overline{Y}}_{MR}) \right\}^2.$$

Finally, we computed the coverage probability of 95% normal confidence intervals. That is, in the r-th sample, we computed the confidence interval

$$\widehat{\overline{Y}}_{MR(r)} \pm 1.96 \sqrt{\widehat{V}_{J(r)}}.$$

The Monte Carlo coverage probability was defined as the proportion of the confidence intervals covering the true total $\overline{Y}$ among the $2,000$ selected samples. The Monte Carlo average length of confidence interval was defined as the average of the length $\mathcal{L} = 2\sqrt{\widehat{V}_J(r)}$

Table 1: Bias, Standard error (SE) and Root mean squared error (RMSE) of different estimators under four model setups.

Estimators	Bias	SE	RMSE	Bias	SE	RMSE
	Scenario: (NM1-IM1)			Scenario:(NM1-IM2)		
$\widehat{\overline{Y}}_{COM}$	-0.0026	0.4347	0.4348	0.0022	0.7540	0.7540
$\widehat{\overline{Y}}_{DR}(1010)$	0.0027	0.4960	0.4960	-0.0057	0.8044	0.8044
$\widehat{\overline{Y}}_{DR}(1001)$	0.0165	0.5158	0.5161	0.0074	0.7946	0.7946
$\widehat{\overline{Y}}_{DR}(0110)$	0.0042	0.5016	0.5016	-0.1150	0.8025	0.8107
$\widehat{\overline{Y}}_{DR}(0101)$	0.1582	0.5293	0.5525	0.0089	0.7995	0.7995
$\widehat{\overline{Y}}_{MR}(1110)$	0.0032	0.4963	0.4963	-0.0002	0.7973	0.7973
$\widehat{\overline{Y}}_{MR}(1101)$	0.0147	0.5028	0.5030	0.0072	0.7948	0.7948
$\widehat{\overline{Y}}_{MR}(1011)$	0.0025	0.4966	0.4966	0.0073	0.7948	0.7948
$\widehat{\overline{Y}}_{MR}(0111)$	0.0032	0.4977	0.4978	0.0079	0.7955	0.7955
$\widehat{\overline{Y}}_{MR}(1111)$	0.0026	0.4967	0.4967	0.0073	0.7945	0.7945
	Scenario: (NM2-IM1)			Scenario: (NM2-IM2)		
$\widehat{\overline{Y}}_{COM}$	-0.0026	0.4347	0.4348	0.0022	0.7540	0.7540
$\widehat{\overline{Y}}_{DR}(1010)$	-0.0128	0.4959	0.4961	-0.1998	0.7981	0.8228
$\widehat{\overline{Y}}_{DR}(1001)$	0.1990	0.5196	0.5564	-0.0079	0.7958	0.7959
$\widehat{\overline{Y}}_{DR}(0110)$	-0.0155	0.5054	0.5057	-0.0109	0.8152	0.8153
$\widehat{\overline{Y}}_{DR}(0101)$	-0.0155	0.5337	0.5339	-0.0106	0.8034	0.8035
$\widehat{\overline{Y}}_{MR}(1110)$	-0.0141	0.5091	0.5093	0.0318	0.8114	0.8121
$\widehat{\overline{Y}}_{MR}(1101)$	-0.0833	0.5550	0.5613	-0.0100	0.8048	0.8049
$\widehat{\overline{Y}}_{MR}(1011)$	-0.0145	0.5021	0.5023	-0.0098	0.7998	0.7999
$\widehat{\overline{Y}}_{MR}(0111)$	-0.0136	0.5075	0.5077	-0.0104	0.8027	0.8027
$\widehat{\overline{Y}}_{MR}(1111)$	-0.0140	0.5167	0.5169	-0.0099	0.8113	0.8114

across the selected samples.

The results are presented in Table 2. We note that the variance estimator $\widehat{V}_J$ performed very well in terms of relative bias in all the scenarios, with absolute relative bias smaller than 2.2%. Also, the Monte Carlo coverage rates ranging from 93.8% to 94.3% were all close to the nominal values. The results suggest that the proposed confidence interval is multiply robust in the sense that the coverage probability is close to the nominal rate if all but one model are misspecified.

9 Concluding remarks

In this paper, we have extended the concept of multiply robustness of Han and Wang (2013) to survey sampling setups. More specifically, we proposed a novel determinist and

Table 2: Coverage rate (CR), Average length (AL) and Relative bias (RB) of variance estimators of different estimators under the scenario (NM1-IM1).

Estimators	CR	AL	RB
	(NM1-IM1)		
$\widehat{\widehat{Y}}_{MR}(1110)$	93.8%	1.912	-1.35%
$\widehat{\widehat{Y}}_{MR}(1101)$	94.3%	1.954	0.55%
$\widehat{\widehat{Y}}_{MR}(1011)$	94.0%	1.906	-2.11%
$\widehat{\widehat{Y}}_{MR}(0111)$	94.1%	1.920	-1.05%
$\widehat{\widehat{Y}}_{MR}(1111)$	94.3%	1.918	-0.73%

random imputation procedures in the context of complex survey data. In an empirical study, the proposed methods performed well in terms of bias and efficiency. We have also proposed multiply robust variance estimators that performed well in terms of bias and coverage probability. Our methods are easy to implement in practice and have an advantage in terms of protection compared to existing imputation procedures.

In this paper, we focussed on the estimation of a population total. Multiply robust procedures for estimating finite population distribution functions and quantiles are currently under investigation. Topics of interest that will be investigated in the future include the use of multiply robust procedures for the treatment of undercoverage errors, which is an important topic in statistical agencies. Finally, in the context of classical statistics, it would be of interest to develop multiply robust versions of multiple imputation procedures.

References

Bang, H. and Robins, J. (2005) Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962–973.

Beaumont, J.-F. and Haziza, D. (2015) A note on the concept of invariance in two-phase sampling designs. *In revision for Survey Methodology.*

Berger, Y. G. (2007) A jackknife variance estimator for unistage stratified samples with unequal probabilities. *Biometrika* **94**, 953–964.

Cao, W., Tsiatis, A. A. and Davidian, M. (2009) Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika* **96**, 723–734.

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

Fay, R. E. (1991) A Design-Based Perspective on Missing Data Variance. *Proceedings of the 1991 Annual Research Conference, US Bureau of the Census*, 429–440.

Han, P. (2014) A Further Study of the Multiply Robust Estimator in Missing Data Analysis. *Journal of Statistical Planning and Inference* **148**, 101–110.

Han, P. and Wang, L. (2013) Estimation with Missing Data: Beyond Double Robustness. *Biometrika* **100**, 417–430.

Fuller, W. A. and Kim, J. K. (2005) Hot deck imputation for the response model, *Survey Methodology* **31**, 139–149.

Haziza, D. and Rao, J. N. K. (2006) A nonresponse model approach to inference under imputation for missing survey data. *Survey Methodology* **32**, 53–64.

Isaki, C. T. and Fuller, W. A. (1982) Survey design under the regression superpopulation model. *Journal of the American Statistical Association* **77**, 89–96.

Kang, J. D. Y. and Schafer, J. L. (2008) Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* **22**, 523–539.

Kim, J. K. and Fuller, W. A. (2004) Fractional hot-deck imputation. *Biometrika* **91**, 559–578.

Kim, J. K. and Haziza, D. (2014) Doubly robust inference with missing data in survey sampling. *Statistica Sinica* **24**, 375–394.

Kim, J. K. and Park, H. A. (2006) Imputation using response probability. *Canadian Journal of Statistics* **34**, 171–182.

Kott, P. S. (1994) A note on handling nonresponse in sample surveys. *Journal of the American Statistical Association* **89**, 693–696.

Kott, P. S. (2006) Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology* **32**, 133–142.

Kott, P. S. and Chang, T. (2010) Using calibration weighting to adjust for nonignorable unit nonresponse. *Journal of the American Statistical Association* **105**, 1265–1275.

Rao, J. N. K. and Shao, J. (1992) Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* **79**, 811–822.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994) Estimation of regression coefficient when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.

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

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

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999) Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion and rejoinder). *Journal of the American Statistical Association* **94**, 1096–1146.

Shao, J. and Sitter, R. R. (1996) Bootstrap for imputed survey data. *Journal of the American Statistical Association* **93**, 819–831.

Shao, J. and Steel, P. (1999) Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association* **94**, 254–265.

Tan, Z. (2006) A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association* **101**, 1619–1637.

Wolter, K. M. (2007) *Introduction to variance estimation.* Springer-Verlag.

Wu, C. and Sitter, R. R (2001) A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association* **96**, 185–93.

Appendix

A: Proof of Theorem 1

Assume $m(\mathbf{x}_i;\boldsymbol{\beta}) = a_0 + \sum_{k=1}^{K} a_k m^k(\mathbf{x}_i;\boldsymbol{\beta}^k)$, then we have $\widehat{\boldsymbol{\gamma}}_p \to^p \boldsymbol{\gamma}_p^*$, where

$$\boldsymbol{\gamma}_p^* = (a_0^*, 0, \ldots, 0, a_1, a_2, \ldots, a_K),$$

with $a_0^* = a_0 + \sum_{k=1}^{K} E\left\{m^k(\mathbf{x};\boldsymbol{\beta}^k)\right\}$. Therefore,

$$\begin{aligned}
\widehat{Y}_{MR}/Y &= \left\{\sum_{i\in s} r_i w_i y_i + \sum_{i\in s}(1-r_i)\, w_i \mathbf{h}_i^{\top}\widehat{\boldsymbol{\gamma}}_p\right\}/Y \\
&\to^p \left\{\sum_{i\in s} r_i w_i y_i + \sum_{i\in s}(1-r_i)\, w_i m(\mathbf{x}_i;\boldsymbol{\beta})\right\}/Y \\
&\to^p 1,
\end{aligned}$$

as $N \to \infty$.

B: Proof of Theorem 2

Without loss of generality, we assume the model $p^1(\mathbf{x}_i;\boldsymbol{\alpha}^1)$ is correctly specified. Because the solution of minimizing the distance function subject to constraints (7)-(9) is unique, then we have $\widehat{\boldsymbol{\gamma}}_p \to^p \boldsymbol{\gamma}_p^*$, where

$$\boldsymbol{\lambda}_r^* = (\boldsymbol{\lambda}_0^*, 1, 0, \ldots, 0),$$

where $\boldsymbol{\lambda}_0^* = E\left[L\left\{1/p^1(\mathbf{x}_i;\boldsymbol{\alpha}^1)\right\}\right]$. Then we have

$$\begin{aligned}
F(\widehat{\boldsymbol{\lambda}}_r^{\top}\mathbf{h}_i) &= F\left\{\widehat{\lambda}_0 + \widehat{\lambda}_1(\widehat{L}_i^1 - \widehat{L}^1) + \ldots + \widehat{\lambda}_{J+K}(m_i^K - \widehat{m}^K)\right\} \\
&\to^p F\left\{F^{-1}(1/p_i^1)\right\} \\
&= 1/p_i^1.
\end{aligned}$$

Hence,

$$\begin{aligned}
\widehat{Y}_{MR}/Y &= \left\{\sum_{i\in s_r} w_i \widehat{F}_i y_i + \left(\sum_{i\in s} w_i \mathbf{h}_i - \sum_{i\in s_r} w_i \widehat{F}_i \mathbf{h}_i\right)^{\top}\widehat{\boldsymbol{\gamma}}_p\right\}/Y \\
&\to^p \left\{\sum_{i\in s_r} w_i/p_i^1 y_i + \left(\sum_{i\in s} w_i \mathbf{h}_i - \sum_{i\in s_r} w_i/p_i^1 \mathbf{h}_i\right)^{\top}\widehat{\boldsymbol{\gamma}}_p\right\}/Y \\
&\to^p 1.
\end{aligned}$$

C: Proof of Theorem 3

We first assume the following regularity conditions:

(C1). Assume $(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*, \boldsymbol{\gamma}^*)$ are the unique probability limits of $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}_r, \widehat{\boldsymbol{\gamma}}_p)$ and $\widehat{\boldsymbol{\theta}}$ has the following influence function:

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = \frac{1}{N}\sum_{i\in s} w_i S(\mathbf{x}_i; \boldsymbol{\theta}^*) + o_p(n^{-1/2}),$$

where

$$\boldsymbol{\theta}^* = (\boldsymbol{\alpha}^{1*}, \ldots, \boldsymbol{\alpha}^{J*}, \boldsymbol{\beta}^{1*}, \ldots, \boldsymbol{\beta}^{K*}), \quad \widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\alpha}}^1, \ldots, \widehat{\boldsymbol{\alpha}}^J, \widehat{\boldsymbol{\beta}}^1, \ldots, \widehat{\boldsymbol{\beta}}^K).$$

(C2). Assume $F(t)$ is the calibration function defined in Section 3 and it has continuous first derivative. We also assume $\mathbf{h}(\mathbf{x}; \boldsymbol{\theta})$ has continuous first derivative with respect to $\boldsymbol{\theta}$.

(C3). Assume the following moments are bounded: $E(Y^2)$, $E(F^2)$, $E(\mathbf{h}^2)$, $E(S^2)$, $E(\dot{F})$ and $E(\dot{\mathbf{h}})$.

(C4). Assume the true response probability $0 < a < p(\mathbf{x}; \boldsymbol{\alpha}) < 1$ almost surely.

Define

$$U(\widehat{\boldsymbol{\lambda}}_r) = \sum_{i\in s_r} w_i F(\widehat{\boldsymbol{\lambda}}_r^\top \mathbf{h}_i)\mathbf{h}_i - \sum_{i\in s} w_i \mathbf{h}_i,$$

then by using Taylor linearization and conditions (C1) and (C2), it can be shown that

$$\widehat{\boldsymbol{\lambda}}_r^\top - \boldsymbol{\lambda}^* = -\left\{E\left(\frac{\partial U}{\partial \widehat{\boldsymbol{\lambda}}_r}\right)\right\}^{-1} U(\boldsymbol{\lambda}^*) + o_p(n^{-1/2}). \quad \text{(B.1)}$$

By using Taylor linearization, condition (C1) and (B.1) and because $\sum_{i\in s} w_i \left(r_i \widehat{F}_i - 1\right) \mathbf{h}_i^\top = 0$, we have

$$\begin{aligned}
\widehat{Y}_{MR}\left(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\lambda}}_r\right) &= \widehat{Y}_\pi + \sum_{i\in s} w_i \left(r_i \widehat{F}_i - 1\right) \mathbf{h}_i^\top \left(\boldsymbol{\gamma}^* - \widehat{\boldsymbol{\gamma}}_p\right) + \sum_{i\in s} w_i \left(r_i \widehat{F}_i - 1\right) \widehat{e}_i \\
&= \widehat{Y}_\pi + \sum_{i\in s} w_i \left(r_i \widehat{F}_i - 1\right) \widehat{e}_i \\
&= \widehat{Y}_\pi + \sum_{i\in s} w_i \left(r_i F_i - 1\right) e_{i,0} + \sum_{i\in s} w_i r_i \dot{F}_i \mathbf{h}_{i,0} e_{i,0} \left(\widehat{\boldsymbol{\lambda}}_r - \boldsymbol{\lambda}^*\right) \\
&+ \left\{\sum_{i\in s} w_i r_i \dot{F}_i \boldsymbol{\lambda}^* \dot{\mathbf{h}}_{i,0} e_{i,0} - \sum_{i\in s} w_i \left(r_i F_i - 1\right) \boldsymbol{\gamma}^* \dot{\mathbf{h}}_{i,0}\right\} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) \\
&+ O_p\left(n^{-1}N\right) \\
&= \sum_{i\in s} w_i \eta_{i,0} + O_p\left(n^{-1}N\right), \quad \text{(B.2)}
\end{aligned}$$

where $\widehat{Y}_\pi = \sum_{i\in s} w_i y_i$ and $\eta_{i,0}$ is defined in Theorem 3.