Statistica Sinica **29** (2019), 2035-2053
doi:https://doi.org/10.5705/ss.202017.0126

MULTIPLY ROBUST NONPARAMETRIC MULTIPLE IMPUTATION FOR THE TREATMENT OF MISSING DATA

Sixia Chen and David Haziza

University of Oklahoma and Université de Montréal

Abstract: Imputation offers an effective solution to the problem of missing values. We propose a nonparametric multiple imputation procedure that uses multiple outcome regression models and/or multiple propensity score models. Our procedure leads to a multiply robust point estimator in the sense that it remains consistent if all models but one are misspecified. We obtain a variance estimator and establish the asymptotic properties of the proposed method. The results of a simulation study, that assesses the proposed method in terms of bias, efficiency, and coverage probability, support our findings.

Key words and phrases: Double robustness, missing data, multiple imputation, multiple robustness, variance estimation.

1. Introduction

The problem of missing data is common in clinical trials, social research, and surveys, among others. One solution is to exclude cases with missing values, a technique often referred to as a complete case analysis. However, this typically leads to biased estimators, unless the data are missing completely at random (Rubin (1976)). Another common approach to resolve the problem of missing values is imputation, whereby a missing value is replaced by one or more substitute values. For example, item nonresponse in surveys conducted by national statistical offices is usually addressed using some form of single imputation in order to produce public-use data. This allows secondary analysts to obtain point estimates using estimation procedures designed for complete data.

In this study, we consider multiple imputation, where we replace a missing value by $M > 1$ imputed values, leading to M completed data files. An estimate is obtained from each data file using a complete-data estimation procedure, and then the M estimates are pooled to obtain a final point estimate and a variance estimate; see, for example, Rubin (1987), Little and Rubin (2002), and van Buuren (2012) for a comprehensive discussion of multiple imputation. An attractive feature of multiple imputation is that variance estimates can be readily obtained using what is known as Rubin's rule; see Wang and Robins (1998) for a discussion of proper (type A) and improper (type B) multiple imputation procedures. Despite the possibility of Rubin's variance estimator being biased in the case of improper procedures, it is often used in practice owing to its simplicity. Rubin's variance formula has been shown to work reasonably well in a number of settings with $M = 5$, although Wang and Robins (1998) have shown that it leads to an underestimation of the true variance in the case of finite M.

Whether or not we can successfully reduce the nonresponse bias depends on the availability of powerful auxiliary variables for the sample units (respondents and nonrespondents) and the validity of the postulated outcome regression model, which is a set of assumptions about the distribution of the variable being imputed. Traditionally, multiple imputation procedures have been based on a single outcome regression model, making the resulting point estimators vulnerable to model misspecification. To overcome this problem, Long, Hsu and Li (2012) proposed a doubly robust nonparametric multiple imputation procedure for ignorable missing data. Their method uses two working models: an outcome regression model and a propensity score model, where the latter is a set of assumptions about the unknown nonresponse mechanism. An estimator is doubly robust if it remains consistent for the true parameter when either model is correctly specified; see, for example, Robins, Rotnitzky and Zhao (1994), Scharfstein, Rotnitzky and Robins (1999), Bang and Robins (2005), Haziza and Rao (2006), Tan (2006), Kang and Schafer (2007), Cao, Tsiatis and Davidian (2009), and Kim and Haziza (2014), among others. Although doubly robust methods offer some protection, the point and variance estimators may perform poorly in terms of bias and efficiency if both models are misspecified, as noted by Kang and Schafer (2007) and Chen and Haziza (2017), among others.

To provide additional protection, Han and Wang (2013) introduced the concept of multiple robustness; see also Han (2014a,b, 2016a,b), Chan and Yam (2014), Chen and Haziza (2017), and Duan and Yin (2017). Multiply robust procedures use multiple outcome regression models and/or multiple propensity score models. An estimation procedure is said to be multiply robust if it remains consistent if any one of these multiple models is correctly specified. Multiply robust procedures are attractive in many practical situations; see Chen and Haziza (2017) for a discussion. When all of the models are misspecified, the literature provides evidence that multiply robust estimation procedures outperform doubly robust procedures based on the same working models in terms of bias and efficiency; see, for example, Han (2014b) and Chen and Haziza (2017). In this study, we extend the procedure of Long, Hsu and Li (2012) by proposing a multiply robust nonparametric multiple imputation (MRM) procedure that can be implemented easily. As argued by Long, Hsu and Li (2012), the proposed imputation procedure is proper. Therefore, we advocate using Rubin's variance formula. The empirical results presented in Section 5 suggest that Rubin's variance estimator works well, at least in our experiments. Han (2014b) proposed a variance estimator for multiply robust procedures based on Taylor linearization procedures. However, Rubin's variance formula is simpler to implement in practice.

The remainder of this paper is organized as follows. In Section 2, we present the basic setup. The proposed method is presented in Section 3, and we we establish its asymptotic properties in Section 4. The results of a simulation study that assesses the performance of the method in terms of bias, efficiency, and coverage probability, are reported in Section 5. All technical details and proofs are relegated to the Appendix.

2. Basic Setup

We are interested in estimating the population mean, $\mu_0 = E(Y)$, of a study variable Y subject to missingness. We consider a random sample of size n, selected at random from the target population. Let X be a p-vector of fully observed variables, and let R be a response indicator such that $R = 1$ if Y is observed, and $R = 0$, otherwise. The observed data are the independent and identically distributed triplets $(Y_i R_i, X_i, R_i)$, for $i = 1, 2, \ldots, n$. Let s_r and s_m denote the sets of responding units and nonresponding units, respectively for the study variable Y.

The relationship between Y and X is described by the following outcome regression model:

$$Y_i = m(X_i; \beta_0) + \epsilon_i, \quad i = 1, 2 \ldots, n, \tag{2.1}$$

where $m(X_i; \cdot)$ is a given function with a parameter β evaluated at β_0 such that $E(\epsilon_i \mid X_i) = 0$, $var(\epsilon_i \mid X_i) = \sigma^2$ and $E(\epsilon_i \epsilon_j \mid X_i, X_j) = 0$, for $i \neq j$. We assume that Y is missing at random (Rubin (1976)); that is, the true propensity score $p(X_i)$ satisfies

$$\Pr(R_i = 1 | X_i, Y_i) = \Pr(R_i = 1 | X_i) \equiv p(X_i; \alpha_0), \tag{2.2}$$

where $p(X_i; \cdot)$ is a given function with a parameter α evaluated at α_0.

If complete reliance is imposed on the outcome regression model (2.1), a natural estimator of μ_0 is the imputed estimator

$$\hat{\mu}_{RI} = \frac{1}{n}\sum_{i=1}^{n}\left\{R_iY_i + (1-R_i)m(X_i;\hat{\beta})\right\}, \tag{2.3}$$

where $\hat{\beta}$ is a consistent estimator of β_0 obtained by solving the following estimating equations:

$$S_m(\beta) = \frac{1}{n}\sum_{i=1}^{n}\phi_i R_i\left\{Y_i - m(X_i;\beta)\right\}\frac{\partial m(X_i;\beta)}{\partial\beta} = 0, \tag{2.4}$$

where ϕ_i is a coefficient attached to unit i. Often, the coefficient ϕ_i is set to one.

If complete reliance is imposed on the propensity score model (2.2), a natural estimator of μ_0 is the inverse probability weighted estimator

$$\hat{\mu}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{p(X_i;\hat{\alpha})}Y_i, \tag{2.5}$$

where $\hat{\alpha}$ is a consistent estimator of α_0 obtained by solving the following estimating equations:

$$S_p(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i - p(X_i;\alpha)}{p(X_i;\alpha)\left\{1-p(X_i;\alpha)\right\}}\frac{\partial p(X_i;\alpha)}{\partial\alpha} = 0. \tag{2.6}$$

Both (2.3) and (2.5) are based on a single working model. Estimators of the form

$$\hat{\mu}_{DR} = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{p(X_i;\hat{\alpha})}Y_i + \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{R_i}{p(X_i;\hat{\alpha})}\right)m(X_i;\hat{\beta}) \tag{2.7}$$

are doubly robust in the sense that they remain consistent for μ_0 if either (2.1) or (2.2) is correctly specified. Unlike (2.3) and (2.5), the estimator (2.7) uses two working models: an outcome regression model and a propensity score model. However, when both models are incorrectly specified, doubly robust estimators are inconsistent and tend to have poor numerical properties; see Han (2014b) and Chen and Haziza (2017).

Multiply robust procedures provide additional protection against model misspecification because they rely on multiple propensity score models and/or multiple outcome regression models see, for example, Han and Wang (2013), Han (2014a, 2016a,b), Chan and Yam (2014), and Chen and Haziza (2017). Let $\mathcal{C}_1 = \{p^j(X_i;\alpha^j); j = 1,\ldots,J\}$ be a class of J propensity score models, where $p^j(\cdot;\alpha^j)$ is a known function associated with the jth propensity score model, and let $\mathcal{C}_2 = \{m^k(x_i;\beta^k); k = 1,\ldots,K\}$ be a class of K outcome regression working models, where $m^k(\cdot;\beta^k)$ is a known function associated with the kth outcome regression model. The corresponding estimators $\hat{\alpha}^j$ and $\hat{\beta}^k$ may be obtained by solving the following estimating equations:

$$S_p^j(\alpha^j) = \frac{1}{n}\sum_{i=1}^{n} s_p^j(X_i, R_i; \alpha^j) = 0, \quad j = 1, \ldots J, \tag{2.8}$$

and

$$S_m^k(\beta^k) = \frac{1}{n}\sum_{i=1}^{n} s_m^k(X_i, R_i; \beta^k) = 0, \quad k = 1, \ldots K, \tag{2.9}$$

respectively, where

$$s_p^j(X_i, R_i; \alpha^j) = \frac{R_i - p^j(X_i; \alpha^j)}{p^j(X_i; \alpha^j)\{1 - p^j(X_i; \alpha^j)\}} \frac{\partial p^j(X_i; \alpha^j)}{\partial \alpha^j}$$

and

$$s_m^k(X_i, R_i; \beta^k) = R_i \left\{Y_i - m^k(X_i; \beta^k)\right\} \frac{\partial m^k(X_i; \beta^k)}{\partial \beta}.$$

A multiply robust estimator of μ_0 is

$$\hat{\mu}_{MR} = \sum_{i=1}^{n} w_i R_i Y_i, \tag{2.10}$$

where w_i is the weight attached to unit i that satisfies the calibration constraints

$$\sum_{i=1}^{n} w_i R_i = 1,$$

$$\sum_{i=1}^{n} w_i R_i L\left\{\frac{1}{p^j(X_i; \widehat{\alpha}^j)}\right\} = n^{-1}\sum_{i=1}^{n} L\left\{\frac{1}{p^j(X_i; \widehat{\alpha}^j)}\right\},$$

$$\sum_{i=1}^{n} w_i R_i m^k(X_i; \widehat{\beta}^k) = n^{-1}\sum_{i=1}^{n} m^k(X_i; \widehat{\beta}^k),$$

where the function $L(t)$ depends on the calibration method; see Chen and Haziza (2017) for a discussion on choosing the calibration method. The estimator (2.10) is multiply robust in the sense that it is consistent if all models in $\mathcal{C}_1$ or $\mathcal{C}_2$ but one are misspecified.

3. Proposed Method

The multiply robust procedure described above relies on an optimization problem with $K + J + 1$ constraints. To reduce the computational burden, we adopt the approach of Duan and Yin (2017), who proposed compressing the working models using weighted averages. Let

$$\widehat{U}_{pi} = (p^1(X_i; \widehat{\alpha}^1), \ldots, p^J(X_i; \widehat{\alpha}^J))^\top, \quad \widehat{U}_{mi} = (m^1(X_i; \widehat{\beta}^1), \ldots, m^K(X_i; \widehat{\beta}^K))^\top.$$

To summarize information provided by the working models, we regress R_i on $\widehat{U}_{pi}$ and Y_i on $\widehat{U}_{mi}$, which leads to the least square regression coefficients

$$\widehat{\eta}_p = \left(\sum_{i=1}^{n} \widehat{U}_{pi}\widehat{U}_{pi}^{\top}\right)^{-1} \left(\sum_{i=1}^{n} \widehat{U}_{pi}R_i\right) \tag{3.1}$$

and

$$\widehat{\eta}_m = \left(\sum_{i=1}^{n} R_i\widehat{U}_{mi}\widehat{U}_{mi}^{\top}\right)^{-1} \left(\sum_{i=1}^{n} R_i\widehat{U}_{mi}Y_i\right). \tag{3.2}$$

The prediction $\widehat{U}_{pi}^{\top}\widehat{\eta}_p$ converges to $p(X_i;\alpha_0)$ if one of the propensity score models is correctly specified, whereas the prediction $\widehat{U}_{mi}^{\top}\widehat{\eta}_m$ converges to $m(X_i;\beta_0)$ if one of the outcome regression models is correctly specified. Define

$$\widehat{p}_i = \widehat{U}_{pi} \times \frac{\widehat{\eta}_p^2}{\widehat{\eta}_p^{\top}\widehat{\eta}_p} \tag{3.3}$$

and

$$\widehat{m}_i = \widehat{U}_{mi} \times \frac{\widehat{\eta}_m^2}{\widehat{\eta}_m^{\top}\widehat{\eta}_m}, \tag{3.4}$$

with a^2 denoting the column vector $(a_1^2, \ldots, a_q^2)^{\top}$ for a given vector $a = (a_1, \ldots, a_q)^{\top}$. The scores $\widehat{p}_i$ and $\widehat{m}_i$ are standardized versions of $\widehat{U}_{pi}^{\top}\widehat{\eta}_p$ and $\widehat{U}_{mi}^{\top}\widehat{\eta}_m$, respectively, and thus enjoy the same statistical properties. The prediction $\widehat{p}_i$ compresses the J propensity score models into a single number, whereas the prediction $\widehat{m}_i$ compresses the K outcome regression models.

Let $Z_i = (Z_{1i}, Z_{2i})$, for $i = 1, \ldots, n$, with $Z_{1i} = \widehat{m}_i$, and $Z_{2i} = \widehat{p}_i$. Furthermore, let $S_i = (S_{1i}, S_{2i})$ be the vector of standardized scores, where $S_{1i} = \widehat{\sigma}_m^{-1}(\widehat{m}_i - \bar{\widehat{m}}_n)$ and $S_{2i} = \widehat{\sigma}_p^{-1}(\widehat{p}_i - \bar{\widehat{p}}_n)$, with $\bar{\widehat{m}}_n$ and $\bar{\widehat{p}}_n$ denoting the usual sample means of $\widehat{p}_i$ and $\widehat{m}_i$, respectively, and $\widehat{\sigma}_m$ and $\widehat{\sigma}_p$ denoting the corresponding sample standard deviations. Our proposed MRM procedure is implemented as follows:

(Step 1). Calculate S_i, for $i = 1, \ldots, n$, from the original sample.

(Step 2). To obtain the lth imputation, draw a bootstrap sample of size n, with replacement, from the original sample, fit the working models in $\mathcal{C}_1$ and $\mathcal{C}_2$ using the bootstrap sample, and calculate $S_i^{(l)}$, the version of S_i corresponding to the lth imputation.

(Step 3). Calculate the distance between each missing subject $i \in s_m$ and every response subject $j \in s_r^{(l)}$ as

$$d^{(l)}(i,j) = \left\{ \lambda \left(S_{1i} - S_{1j}^{(l)} \right)^2 + (1-\lambda) \left(S_{2i} - S_{2j}^{(l)} \right)^2 \right\}^{1/2}, \tag{3.5}$$

where $0 \le \lambda \le 1$ and $s_r^{(l)}$ denotes the set of respondents corresponding to the lth imputation. The value of λ reflects our confidence in a working model. A small value of λ places more weight on the outcome regression models, whereas a large value of λ places more weight on the propensity score models.

(Step 4). Define the nearest-H neighborhood $R_H^{(l)}(i)$ for each missing unit $i \in s_m$ as the H units in $s_r^{(l)}$ that have the smallest H distances (d) from unit i.

(Step 5). Randomly select one unit from $R_H^{(l)}(i)$, and use its value $Y_i^{*(l)}$ as the imputed value for unit i.

(Step 6). Repeat (Step 2) to (Step 5) L times to obtain L multiply imputed data sets with $\tilde{Y}_i^{(l)} = \left\{ R_i Y_i + (1-R_i) Y_i^{*(l)} \right\}$, for $i = 1, \ldots, n$, $l = 1, \ldots, L$.

A point estimator is obtained by pooling the L multiply imputed data sets, which leads to

$$\widehat{\mu}_{MRM} = \frac{1}{L} \sum_{l=1}^{L} \widehat{\mu}^{(l)}, \tag{3.6}$$

where $\widehat{\mu}^{(l)} = n^{-1} \sum_{i=1}^{n} \tilde{Y}_i^{(l)}$. Using Rubin's rule, the variance of (3.6) is readily estimated by

$$\widehat{V}_{MRM} = \bar{U}_L + \left(1 + \frac{1}{L} \right) B_L, \tag{3.7}$$

where $\bar{U}_L$ and B_L denote the within- and between-components given respectively by

$$\bar{U}_L = \frac{1}{L} \sum_{l=1}^{L} U^{(l)} \tag{3.8}$$

and

$$B_L = \frac{1}{L-1} \sum_{l=1}^{L} \left(\widehat{\mu}^{(l)} - \widehat{\mu}_{MRM} \right)^2, \tag{3.9}$$

with $U^{(l)}$ denoting the sampling variance of $\widehat{\mu}_{MRM}$ based on the lth imputed data set.

An $100(1-\gamma)$th confidence interval of μ_0 can be constructed as follows:

$$\left(\widehat{\mu}_{MRM} - z_{1-\gamma/2}\widehat{V}_{MRM}^{1/2}, \quad \widehat{\mu}_{MRM} + z_{1-\gamma/2}\widehat{V}_{MRM}^{1/2}\right),$$

where $z_{1-\gamma/2}$ is the $100(1-\gamma/2)$th percentile taken from a standard normal distribution.

Remark 1. The size of the neighborhood H in Step 4 of the proposed procedure is assumed to satisfy $H/n = o(1)$ and $\log(n)/H = o(1)$; see Condition C4 in Appendix A. Long, Hsu and Li (2012) empirically evaluated different choices of H for their doubly robust imputation procedure. They found that the bias of the point estimators increased as H increased, whereas their standard errors decreased slightly. In their empirical experiments, the point estimators exhibited slightly lower mean square errors with $H = 3$. In practice, choosing the optimal value of H is challenging. We suggest applying the proposed procedure with different values of H (say between 2 and 6), and then choosing the value of H corresponding to the smallest standard error computed using (3.7).

4. Asymptotic Properties

Here, we show that the estimator (3.6) is multiply robust and establish its asymptotic normality. The regularity conditions are described in Appendix A.

Theorem 1 establishes the multiply robustness property of (3.6).

Theorem 1. *Assume that the regularity conditions* (C1)–(C5) *in Appendix A hold. If* $0 < \lambda < 1$, *and at least one of the outcome regression models or the propensity score models is correctly specified, then* $\widehat{\mu}_{MRM}$ *is consistent for* μ_0.

An outline of the proof of Theorem 1 is presented in Appendix B.

Let $\theta = (\alpha^1, \ldots, \alpha^J, \beta^1, \ldots, \beta^K)^\top$ and $\widehat{\theta} = (\widehat{\alpha}^1, \ldots, \widehat{\alpha}^J, \widehat{\beta}^1, \ldots, \widehat{\beta}^K)^\top$. Let $\theta^* = (\alpha^{*1}, \ldots, \alpha^{*J}, \beta^{*1}, \ldots, \beta^{*K})^\top$ denote the probability limit of $\widehat{\theta}$. Define

$$S_{mp}(\theta) = \left(S_p^1(\alpha^1), \ldots, S_p^J(\alpha^J), S_m^1(\beta^1), \ldots, S_m^K(\beta^K)\right)$$

and

$$s(X_i, R_i; \theta) = (s_p^1(X_i, R_i; \alpha^1), \ldots, s_p^J(X_i, R_i; \alpha^J), s_m^1(X_i, R_i; \beta^1), \ldots, s_m^K(X_i, R_i; \beta^K)).$$

Theorem 2 establishes the root-n consistency and asymptotic normality of (3.6).

Theorem 2. *Assume that the regularity conditions* (C1)-(C5) *in Appendix A hold. If* $0 < \lambda < 1$ *and at least one of the outcome regression models or the propensity score models is correctly specified, then*

$$\widehat{\mu}_{MRM} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{R_i}{\pi(Z_i^*)}Y_i + \left\{1 - \frac{R_i}{\pi(Z_i^*)}\right\}\mu(Z_i^*) + As(X_i, R_i; \theta^*)\right] + o_p(n^{-1/2}), \tag{4.1}$$

as $n \to \infty$ *and* $L \to \infty$, *where* Z_i^* *denotes* Z_i *evaluated at* θ^*, $\pi(Z_i^*) = E(R_i = 1|Z_i^*)$, $\mu(Z_i^*) = E(Y_i|Z_i^*)$, *and*

$$A = -E\left\{(1 - R_i)\frac{\partial\mu(Z_i^*)}{\partial\theta}\right\}\left\{E\left(\frac{\partial S_{mp}(\theta^*)}{\partial\theta}\right)\right\}^{-1}.$$

Furthermore, we have

$$n^{1/2}(\widehat{\mu}_{MRM} - \mu_0) \to^d N(0, \sigma^2_{MRM}),$$

where

$$\sigma^2_{MRM} = V\left\{\frac{R_i}{\pi(Z_i^*)}Y_i + \left\{1 - \frac{R_i}{\pi(Z_i^*)}\right\}\mu(Z_i^*) + As(X_i, R_i; \theta^*)\right\}. \tag{4.2}$$

An outline of the proof of Theorem 2 is presented in Appendix C.

Remark 2. Alternatively, the estimator $\widehat{\mu}_{MRM}$ in Theorem 2 can be written as

$$\widehat{\mu}_{MRM} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{R_i}{\pi(Z_i)}Y_i + \left\{1 - \frac{R_i}{\pi(Z_i)}\right\}\mu(Z_i)\right] + o_p(n^{-1/2}), \tag{4.3}$$

where $\pi(Z_i) = E(R_i = 1|Z_i)$ and $\mu(Z_i) = E(Y_i|Z_i)$. If one of propensity score models and one of outcome regression models are correctly specified, then according to (4.3), it can be shown that

$$\begin{aligned}\widehat{\mu}_{MRM} &= \frac{1}{n}\sum_{i=1}^{n} m(X_i;\beta_0) + \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{p(X_i;\alpha_0)}\{Y_i - m(X_i;\beta_0)\} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} R_i\left\{\frac{1}{\pi(Z_i)} - \frac{1}{p(X_i;\alpha_0)}\right\}\{Y_i - m(X_i;\beta_0)\} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{1 - \frac{R_i}{\pi(Z_i)}\right\}\{\mu(Z_i) - m(X_i;\beta_0)\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{R_i}{p(X_i;\alpha_0)}Y_i + \left\{1 - \frac{R_i}{p(X_i;\alpha_0)}\right\}m(X_i;\beta_0)\right] \\ &\quad + o_p(n^{-1/2}). \end{aligned} \tag{4.4}$$

Therefore, $\widehat{\mu}_{MRM}$ achieves a semiparametric lower bound in this scenario.

Using Theorem 2, we can construct a normal approximation-based confidence interval of μ_0. However, estimating the variance of $\widehat{\mu}_{MRM}$ given by (4.2) is generally tedious and requires a nonparametric method, such as kernel smoothing, to estimate the density. To overcome this problem, we recommend using Rubin's formula for variance estimation. According to Long, Hsu and Li (2012), the proposed imputation procedure is proper. The addition of a bootstrap step allows us to estimate the variance of the estimators in other settings as well (Heitjan and Little (1991); Rubin and Schenker (1991)). In our numerical experiments, $L = 5$ imputations led to good performance for finite samples; see also Rubin (1987).

5. Simulation Study

We performed a simulation study to evaluate the performance of the proposed estimators. We used the simulation setup of Kang and Schafer (2007), also considered by Chan and Yam (2014), Han (2014a), and Chen and Haziza (2017). The Monte Carlo size was set to $B = 1,000$ and the sample size was set to $n = 400$. We first generated a vector of covariates $X = (X_1, X_2, X_3, X_4)^\top$ from a standard multivariate normal distribution. Then, we generated a study variable Y from the linear regression model $Y = 210 + 27.4X_1 + 13.7(X_2 + X_3 + X_4) + \epsilon$, with the errors ϵ generated from a standard normal distribution. Given X and Y, the response indicators R were generated from a Bernoulli distribution with probability

$$p = \frac{\exp(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4)}{1 + \exp(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4)},$$

where $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = (0, -1, 0.5, -0.25, -0.1)$. The corresponding response rate was approximately 50%.

To evaluate the performance of the proposed procedure when some models are misspecified, we considered the following transformations of the X-variables: $V_1 = X_2\{1 + \exp(X_1)\}^{-1} + 10$ and $V_2 = (X_1X_3/25 + 0.6)^3$; see also Kang and Schafer (2007). The correct models were fitted using $X = (X_1, X_2, X_3, X_4)$ as predictors in both the outcome regression and the propensity score models. The incorrect models were fitted using $V = (V_1, V_2)$ as predictors in both models.

We were interested in estimating the population mean $\mu_0 = E(Y) = 210$. Because the proposed imputation procedure may be based on different combinations of the models, we use four digits between parentheses to distinguish between the estimators constructed from the different models. The first two digits correspond to the correct and incorrect propensity score models, respectively. The last two digits correspond to the correct and incorrect outcome regression models, respectively. For example, the estimator $\widehat{\mu}(1010)$ is based on correct nonresponse and outcome regression models, whereas $\widehat{\mu}(1111)$ denotes the estimator based on all models.

We computed the following estimators:

1. The complete data estimator $\widehat{\mu}_{COM} = n^{-1}\sum_{i=1}^{n} Y_i$, which assumes no missing values.

2. The unadjusted estimator $\widehat{\mu}_{RES} = n_r^{-1}\sum_{i=1}^{n} R_i Y_i$ based on the complete cases, where n_r denotes the number of responding units.

3. The customary multiply imputed estimators (MI) based on a single outcome regression model: $\widehat{\mu}_{MI}(0010)$ and $\widehat{\mu}_{MI}(0001)$, where the two estimators were obtained under the correct and incorrect outcome regression models, respectively, with normal errors. The imputations were performed using the R package 'mice'; see van Buuren et al. (2014).

4. The doubly robust nonparametric multiple imputation estimator (DRM) of Long, Hsu and Li (2012), based on a single propensity score model and a single outcome regression model: $\widehat{\mu}_{DRM}(1010)$, $\widehat{\mu}_{DRM}(1001)$, $\widehat{\mu}_{DRM}(0110)$, and $\widehat{\mu}_{DRM}(0101)$.

5. The proposed MRM estimator given by (3.6): $\widehat{\mu}_{MRM}(0111)$, $\widehat{\mu}_{MRM}(1011)$, $\widehat{\mu}_{MRM}(1101)$, $\widehat{\mu}_{MRM}(1110)$, and $\widehat{\mu}_{MRM}(1111)$.

For the DRM and the MRM estimators, we used $H = 3$ units in each neighborhood and $L = 5$ imputed values. We also evaluated the effect of λ in (3.5) by considering $\lambda = 0.2$, 0.5, and 0.8. For each estimator, we computed the Monte Carlo bias (Bias), standard error (SE), and relative root mean square error (RRMSE). The results, presented in Tables 1–3, correspond to values of λ equal to $0.2, 0.5$, and 0.8, respectively.

From Tables 1–3, the complete data estimator, $\widehat{\mu}_{COM}$, performed best in terms of RB, RSE, and RRMSE, as expected. On the other hand, the unadjusted estimator, $\widehat{\mu}_{RES}$, was biased in all scenarios, as expected. When either the propensity score model or the outcome regression model was correctly specified, the DRM estimators showed small values of RB, suggesting that they are doubly robust. However, when both models were misspecified, the estimator $\widehat{\mu}_{DRM}(0101)$ exhibited a large bias. For example, for $\lambda = 0.5$, the relative bias was approximately equal to -5.1%; see Table 2. The MRM estimators showed negligible bias in all scenarios. The estimators $\widehat{\mu}_{DRM}(1111)$ based on all four models were almost as efficient as the estimator $\widehat{\mu}_{DRM}(1010)$ based on two correctly specified models. This suggests that incorporating an additional propen-

Table 1. Relative bias (RB), relative standard error (RSE), and relative root mean squared error (RRMSE) of several estimators with $H = 3$, $L = 5$, and $\lambda = 0.2$.

Estimators	RB(%)	RSE(%)	RRMSE(%)
$\widehat{\mu}_{COM}$	0.000	0.871	0.871
$\widehat{\mu}_{RES}$	−4.761	1.220	4.915
$\widehat{\mu}_{MI}(0010)$	−0.002	0.871	0.871
$\widehat{\mu}_{MI}(0001)$	−5.771	1.178	5.890
$\widehat{\mu}_{DRM}(1010)$	−0.207	0.915	0.938
$\widehat{\mu}_{DRM}(1001)$	−0.370	1.093	1.154
$\widehat{\mu}_{DRM}(0110)$	−0.855	0.897	1.239
$\widehat{\mu}_{DRM}(0101)$	−5.096	1.172	5.229
$\widehat{\mu}_{MRM}(0111)$	−0.850	0.904	1.241
$\widehat{\mu}_{MRM}(1011)$	−0.195	0.924	0.944
$\widehat{\mu}_{MRM}(1101)$	−0.363	1.100	1.159
$\widehat{\mu}_{MRM}(1110)$	−0.194	0.924	0.944
$\widehat{\mu}_{MRM}(1111)$	−0.194	0.924	0.944

Table 2. Relative bias (RB), relative standard error (RSE), and relative root mean squared error (RRMSE) of several estimators with $H = 3$, $L = 5$, and $\lambda = 0.5$.

Estimators	RB(%)	RSE(%)	RRMSE(%)
$\widehat{\mu}_{COM}$	0.000	0.871	0.871
$\widehat{\mu}_{RES}$	−4.761	1.220	4.915
$\widehat{\mu}_{MI}(0010)$	−0.002	0.871	0.871
$\widehat{\mu}_{MI}(0001)$	−5.781	1.175	5.899
$\widehat{\mu}_{DRM}(1010)$	−0.191	0.907	0.927
$\widehat{\mu}_{DRM}(1001)$	−0.662	1.089	1.274
$\widehat{\mu}_{DRM}(0110)$	−0.506	0.899	1.032
$\widehat{\mu}_{DRM}(0101)$	−5.165	1.205	5.303
$\widehat{\mu}_{MRM}(0111)$	−0.505	0.888	1.022
$\widehat{\mu}_{MRM}(1011)$	−0.182	0.897	0.915
$\widehat{\mu}_{MRM}(1101)$	−0.622	1.077	1.244
$\widehat{\mu}_{MRM}(1110)$	−0.182	0.895	0.913
$\widehat{\mu}_{MRM}(1111)$	−0.182	0.892	0.911

sity score model and an additional outcome regression model does not have a significant effect on the efficiency of the resulting estimator.

We also computed the coverage rate (CR) and relative average length (RAL) of the confidence intervals based on the DRM and MRM estimators for $\lambda = 0.5$; see Table 4. The results for $\lambda = 0.2$ and $\lambda = 0.8$ were very similar, and thus are not presented here. Table 4 shows the CRs of the MRM estimators were close to the nominal rate of 95% and were slightly better than those obtained using the DRM estimators. In most cases, the RALs of the MRM estimators were smaller

Table 3. Relative bias (RB), relative standard error (RSE), and relative root mean squared error (RRMSE) of several estimators with $H = 3$, $L = 5$, and $\lambda = 0.8$.

Estimators	RB(%)	RSE(%)	RRMSE(%)
$\widehat{\mu}_{COM}$	0.000	0.871	0.871
$\widehat{\mu}_{RES}$	−4.761	1.220	4.915
$\widehat{\mu}_{MI}(0010)$	−0.002	0.872	0.872
$\widehat{\mu}_{MI}(0001)$	−5.766	1.180	5.886
$\widehat{\mu}_{DRM}(1010)$	−0.180	0.898	0.916
$\widehat{\mu}_{DRM}(1001)$	−1.049	1.063	1.493
$\widehat{\mu}_{DRM}(0110)$	−0.311	0.897	0.949
$\widehat{\mu}_{DRM}(0101)$	−5.179	1.177	5.311
$\widehat{\mu}_{MRM}(0111)$	−0.305	0.888	0.939
$\widehat{\mu}_{MRM}(1011)$	−0.169	0.894	0.910
$\widehat{\mu}_{MRM}(1101)$	−1.040	1.085	1.503
$\widehat{\mu}_{MRM}(1110)$	−0.170	0.892	0.908
$\widehat{\mu}_{MRM}(1111)$	−0.171	0.892	0.909

Table 4. Coverage rate (CR) and relative average length (RAL) of confidence intervals with $H = 3$, $L = 5$, and $\lambda = 0.5$.

Estimators	CR(%)	RAL(%)
$\widehat{\mu}_{MI}(0010)$	95.3	3.38
$\widehat{\mu}_{MI}(0001)$	0.2	4.52
$\widehat{\mu}_{DRM}(1010)$	95.7	3.80
$\widehat{\mu}_{DRM}(1001)$	91.6	4.76
$\widehat{\mu}_{DRM}(0110)$	92.4	3.77
$\widehat{\mu}_{DRM}(0101)$	1.7	4.74
$\widehat{\mu}_{MRM}(0111)$	92.9	3.76
$\widehat{\mu}_{MRM}(1011)$	95.5	3.79
$\widehat{\mu}_{MRM}(1101)$	92.7	4.83
$\widehat{\mu}_{MRM}(1110)$	95.4	3.79
$\widehat{\mu}_{MRM}(1111)$	95.8	3.79

than those of the DRM estimators. When both models were incorrect, the DRM method showed a very low CR of 11.7%.

Acknowledgement

The work of the first author was partially supported through National Institutes of Health, National Institute of General Medical Sciences [Grant 2U54GM 104938-06, PI Judith James]. The work of the second author was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Canadian Statistical Sciences Institute. The authors wish to thank the

Editor, an Associate Editor and two referees for their constructive comments.

Appendix

A: Regularity conditions

Let $\widehat{Z}_i = (\widehat{Z}_{1i}, \widehat{Z}_{2i})$, where $\widehat{Z}_1 = \widehat{m}_i$ and $\widehat{Z}_2 = \widehat{p}_i$. Let β^{*K}, α^{*j}, η^*_m, and η^*_p be the probability limits of $\widehat{\beta}^k$, $\widehat{\alpha}^j$, $\widehat{\eta}_m$, and $\widehat{\eta}_p$, respectively. Denote $Z^*_i = (Z^*_{1i}, Z^*_{2i})$ as

$$Z^*_{1i} = \widehat{U}^*_{mi} \times \frac{\eta^{*2}_m}{\eta^{*\top}_p \eta^*_m}, \quad Z^*_{2i} = \widehat{U}^*_{pi} \times \frac{\eta^{*2}_p}{\eta^{*\top}_p \eta^*_p},$$

where

$\widehat{U}^*_{mi} = \left(m^1(X_i;\beta^{*1}),\ldots,m^K(X_i;\beta^{*K})\right)$ and $\widehat{U}^*_{pi} = \left(p^1(X_i;\alpha^{*1}),\ldots,p^J(X_i;\alpha^{*J})\right)$.

Finally, let $f(Z^*)$ be the density function of Z^*.

We assume the following regularity conditions needed to prove Theorems 1–3. Conditions (C1) and (C2) apply to each propensity score model, for $j = 1,\ldots,J$, and each outcome regression model, for $k = 1,\ldots,K$, respectively.

(C1). $\widehat{\alpha}^j$ is the unique solution of $S^j_p(\alpha^j) = 0$ and $\widehat{\beta}^k$ is the unique solution of $S^k_m(\beta^k) = 0$, where $S^j_p(\alpha^j)$ and $S^k_m(\beta^k)$ are defined in Section 3.

(C2). $S^j_p(\alpha^j)$ converges almost surely to $S^{*j}_p(\alpha^j) = E\{S^j_p(\alpha^j)\}$, uniformly in α^j, and $S^{*j}_p(\alpha^j) = 0$ has a unique solution α^{*j}. In addition, $S^k_m(\beta^k)$ converges almost surely to $S^{*k}_m(\beta^k) = E\left\{S^k_m(\beta^k)\right\}$, uniformly in β^k, and $S^{*k}_m(\beta^k) = 0$ has a unique solution β^{*k}.

(C3). $E(Y^2) < \infty$ and $E\left\{\mu^2(Z^*)\right\} < \infty$, where $\mu(Z^*) = E(Y|Z^*)$.

(C4). $H/n = o(1)$ and $\log(n)/H = o(1)$.

(C5). $f(Z^*)$ and $\pi(Z^*)$ are continuous and bounded away from zero in the compact support of Z^*.

Conditions (C1) and (C2) ensure the consistency of $\widehat{\alpha}^j$ and $\widehat{\beta}^k$. They are satisfied for most regression models, such as linear and generalized linear models as well as other estimating equations. Condition (C3) is used to derive the asymptotic expansion of $\widehat{\mu}_{MRM}$ and its asymptotic normality. Condition (C4) is used to control the asymptotic order of H. Condition (C5) is used frequently in nonparametric statistics to avoid extreme values of the density and the propensity score.

B: An outline of the proof of Theorem 1

Let $\widehat{\alpha} = (\widehat{\alpha}^1, \ldots, \widehat{\alpha}^J)$, $\widehat{\beta} = (\widehat{\beta}^1, \ldots, \widehat{\beta}^K)$, $\alpha^* = (\alpha^{*1}, \ldots, \alpha^{*J})$, and $\beta^* = (\beta^{*1}, \ldots, \beta^{*K})$. According to (C1), (C2), and Section 5.2 of Van der Vaart (1998), it can be shown that $\widehat{\alpha} \to^p \alpha^*$ and $\widehat{\beta} \to^p \beta^*$, even when the working models are incorrect. Next, we show the multiply robustness property of $\widehat{\mu}_{MRM}$. If one of the outcome regression models is correct, say $m^1(X_i; \beta^1)$, then we have $\beta^{*1} = \beta_0$ and $\eta_m^* = (1, 0, \ldots, 0)^\top$, which implies that $Z_{1i}^* = m(X_i; \beta_0)$. Then, it follows that

$$\begin{aligned} E(Y|R, Z^*) &= E\left\{E(Y|R, Z^*, X)|R, Z^*\right\} \\ &= E\left\{E(Y|Z^*, X)|R, Z^*\right\} \\ &= E(Z_1^*|R, Z^*) = Z_1^* = E(Y|Z^*). \end{aligned} \tag{B.1}$$

If one of the propensity score models is correctly specified, say $p^1(X_i; \alpha^1)$, then $\alpha^{*1} = \alpha_0$, $\eta_p^* = (1, 0, \ldots, 0)^\top$, and $Z_{2i}^* = p(X_i; \alpha_0)$. Therefore,

$$E(Y|R, Z^*) = E(Y|Z^*), \tag{B.2}$$

because Y is independent of R, given Z_2^*. According to Devroye and Wagner (1977) and Silverman (1978), it can be shown that

$$\sum_{j \in R_H(i)} \frac{1}{H} Y_j \to^p E(Y_i|Z_i^*, R_i = 1), \tag{B.3}$$

$$\frac{1}{H} \sum_{j \in R_H(i)} (Y_j - \bar{Y}_{R_H(i)})^2 \to^p V(Y_i|Z_i^*, R_i = 1), \tag{B.4}$$

uniformly for any $i \in S$ as $n \to \infty$, $L \to \infty$, and $H \to \infty$ and conditions (C4) and (C5). In addition, according to Chebyshev's inequality, (B.3) and (B.4), we have

$$\begin{aligned} &\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n}(1 - R_i)\frac{1}{L}\sum_{l=1}^{L} Y_i^{*(l)} - E\left\{(1 - R_i)E(Y_i^{*(l)}|Y, X, R)\right\}\right| > \epsilon\right) \\ &\le \epsilon^{-2} V\left\{\frac{1}{n}\sum_{i=1}^{n}(1 - R_i)\frac{1}{L}\sum_{l=1}^{L} Y_i^{*(l)}\right\} \\ &= \epsilon^{-2}\left[V\left\{\frac{1}{n}\sum_{i=1}^{n}(1 - R_i)E(Y_i^{*(l)}|Y, X, R)\right\}\right. \\ &\qquad \left. + E\left\{\frac{1}{n^2}\sum_{i=1}^{n}(1 - R_i)\frac{1}{L}V(Y_i^{*(l)}|Y, X, R)\right\}\right] \end{aligned}$$

$$
\begin{aligned}
&= \epsilon^{-2}\Bigg[V\Bigg\{\frac{1}{n}\sum_{i=1}^{n}(1-R_i)\sum_{j\in R_H(i)}\frac{1}{H}Y_j\Bigg\} \\
&\quad + E\Bigg\{\frac{1}{n^2}\sum_{i=1}^{n}(1-R_i)\frac{1}{L}\frac{1}{H}\sum_{j\in R_H(i)}(Y_j-\bar{Y}_{R_H(i)})^2\Bigg\}\Bigg] \\
&= \epsilon^{-2}\Bigg[V\left\{\frac{1}{n}\sum_{i=1}^{n}(1-R_i)E(Y_i|Z_i^*,R_i=1)\right\} \\
&\quad + E\Bigg\{\frac{1}{n^2}\sum_{i=1}^{n}(1-R_i)\frac{1}{L}V(Y_i|Z_i^*,R_i=1)\Bigg\}\Bigg] \\
&\quad + o(n^{-1}) + o(n^{-1}L^{-1}) \\
&= O(n^{-1}) + O(n^{-1}L^{-1}). \qquad \text{(B.5)}
\end{aligned}
$$

According to (B.5), we have

$$
\frac{1}{n}\sum_{i=1}^{n}(1-R_i)\frac{1}{L}\sum_{l=1}^{L}Y_i^{*(l)} \to^p E\left\{(1-R_i)E(Y_i^{*(l)}|Y,X,R)\right\} \qquad \text{(B.6)}
$$

as $n\to\infty$, $L\to\infty$, and $H\to\infty$. Therefore, if at least one the models is correctly specified from (B.3) and (B.6), we have

$$
\begin{aligned}
\widehat{\mu}_{MRM} &= \frac{1}{n}\sum_{i=1}^{n}\left\{R_iY_i + (1-R_i)\frac{1}{L}\sum_{l=1}^{L}Y_i^{*(l)}\right\} \\
&\to^p E(R_iY_i) + E\left\{(1-R_i)E(Y_i^{*(l)}|Y,X,R)\right\} \\
&= E(R_iY_i) + E\left\{(1-R_i)E\left(\sum_{j\in R_H^{(l)}(i)}\frac{1}{H}Y_j^{*(l)}|X,Y,R\right)\right\} \\
&= E(R_iY_i) + E\left\{(1-R_i)\sum_{j\in R_H(i)}\frac{1}{H}Y_j\right\} \\
&\to E(R_iY_i) + E\left\{(1-R_i)E(Y_i|Z_i^*,R_i=1)\right\} \\
&= E(R_iY_i) + E\left\{(1-R_i)Y_i\right\} \\
&= E(Y) \qquad \text{(B.7)}
\end{aligned}
$$

as $n\to\infty$, $L\to\infty$, and $H\to\infty$, where we used the consistent probability weights argument defined in Stone (1977) in the derivations. Hence, Theorem 1 is proved.

C: An outline of the proof of Theorem 2

The proposed estimator can be written as $\widehat{\mu}_{MRM} = \widehat{\mu} + \widehat{\mu}_{MRM} - \widehat{\mu}$, where

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\left\{R_iY_i + (1-R_i)\sum_{j\in R_H(i)}\frac{1}{H}Y_j\right\} \tag{C.1}$$

and

$$\widehat{\mu}_{MRM} - \widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n}(1-R_i)\left(\frac{1}{L}\sum_{l=1}^{L}Y_i^{*(l)} - \frac{1}{H}\sum_{j\in R_H(i)}Y_j\right). \tag{C.2}$$

By using a first-order Taylor expansion of $\widehat{\mu}$ at $\theta = \theta^*$, we have

$$\widehat{\mu} = \widehat{\mu}^* + E\left(\frac{\partial\widehat{\mu}^*}{\partial\theta}\right)\left(\widehat{\theta} - \theta^*\right) + o_p(n^{-1/2}), \tag{C.3}$$

where $\partial\widehat{\mu}^*/\partial\theta$ is $\partial\widehat{\mu}/\partial\theta$ evaluated at θ^*. Because $\widehat{\theta}$ is the solution to the estimating equation $S_{mp}(\theta) = 0$, it can be shown that

$$\widehat{\theta} - \theta^* = -\left[E\left\{\frac{\partial S_{mp}(\theta^*)}{\partial\theta}\right\}\right]^{-1} S_{mp}(\theta^*) + o_p(n^{-1/2}). \tag{C.4}$$

Substituting (C.4) into (C.3) leads to

$$\widehat{\mu} = \widehat{\mu}^* + AS_{mp}(\theta^*), \tag{C.5}$$

where A is defined in Theorem 2 of Section 4. By using a similar argument to that used by Long, Hsu and Li (2012), under the regularity conditions (C1)–(C5) in Appendix A, it can be shown that

$$\widehat{\mu}^* - \mu_0 = T_1 + T_2 + T_3 + o_p(n^{-1/2}), \tag{C.6}$$

with

$$T_1 = \frac{1}{n}\sum_{i=1}^{n}\left\{\mu(Z_i^*) - \mu_0\right\}, \quad T_2 = \frac{1}{n}\sum_{i=1}^{n}R_i\left\{Y_i - \mu(Z_i^*)\right\}$$

and

$$T_3 = \frac{1}{n}\sum_{i=1}^{n}\frac{1-\pi(Z_i^*)}{\pi(Z_i^*)}R_i\left\{Y_i - \mu(Z_i^*)\right\}.$$

Because $\widehat{\mu}_{MRM} - \widehat{\mu}$ is asymptotically independent of $\widehat{\mu}$, we have

$$\begin{aligned}
V\left(\widehat{\mu}_{MRM} - \widehat{\mu}\right) &= V\left\{E\left(\widehat{\mu}_{MRM} - \widehat{\mu}|X,Y,R\right)\right\} + E\left\{V\left(\widehat{\mu}_{MRM} - \widehat{\mu}|X,Y,R\right)\right\} \\
&= E\left\{V\left(\widehat{\mu}_{MRM} - \widehat{\mu}|X,Y,R\right)\right\} \\
&= \frac{1}{nL}E\left\{\frac{1}{n}\sum_{i=1}^{n}(1-R_i)V\left(Y_i^{*(l)}|Y,X,R\right)\right\}
\end{aligned}$$

$$= \frac{1}{nL} E \left\{ \frac{1}{n} \sum_{i=1}^{n} (1 - R_i) \frac{1}{H} \sum_{j \in R_H(i)} \left(Y_j - \bar{Y}_{R_H(i)} \right)^2 \right\}$$
$$= O\left(\frac{1}{nL} \right), \qquad \text{(C.7)}$$

where $\bar{Y}_{R_H(i)} = H^{-1} \sum_{j \in R_H(i)} Y_j$. From the asymptotic independence between $\widehat{\mu}_{MRM} - \widehat{\mu}$ and $\widehat{\mu}$, T_1 and T_2, T_1 and T_3, (C.1),(C.2),(C.5)–(C.7), and the regularity condition (C3), we have

$$\widehat{\mu}_{MRM} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{R_i}{\pi(Z_i^*)} Y_i + \left\{ 1 - \frac{R_i}{\pi(Z_i^*)} \right\} \mu(Z_i^*) + As(X_i, R_i; \theta^*) \right] + o_p(n^{-1/2})$$

as $n \to \infty$ and $L \to \infty$. By the central limit theorem, we have

$$n^{1/2} \left(\widehat{\mu}_{MRM} - \mu_0 \right) \to^d N(0, \sigma^2_{MRM}),$$

with σ^2_{MRM} defined in Theorem 2 of Section 4.

References

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

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.

Chan, K. C. G. and Yam, S. C. P. (2014). Oracle, multiple robust and multipurpose calibration in missing response problem. *Statistical Science* **29**, 380–386.

Chen, S. and Haziza, D. (2017). Multiply robust imputation procedures for the treatment of item nonresponse in surveys. *Biometrika***102**, 439–453.

Devroye, L. P. and Wagner, T. J. (1977). The strong uniform consistency of nearest neighbor density estimates. *The Annals of Statistics* **5**, 536–540.

Duan, X. and Yin, G. (2017). Ensemble approaches to estimation of the population mean with missing responses. *Scandinavian Journal of Statistics* **44**, 899–917.

Han, P. (2014a). A further study of the multiply robust estimator in missing data analysis. *Journal of Statistical Planning and Inference* **148**, 101–110.

Han, P. (2014b). Multiply robust estimation in regression analysis with missing data. *Journal of the American Statistical Association* **109**, 1159–1173.

Han, P. and Wang, L. (2013). Estimation with missing data: Beyond double robustness. *Biometrika* **100**, 417–430.

Han, P. (2016a). Combining inverse probability weighting and multiple imputation to improve robustness of estimation. *Scandinavian Journal of Statistics* **43**, 246–260.

Han, P. (2016b). Intrinsic efficiency and multiple robustness in longitudinal studies with dropout. *Biometrika* **103**, 683–700.

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.

Heitjan, D. F. and Little, R. J. A. (1991). Multiple imputation for the fata accident reporting system. *Applied Statistics* **40**, 13–29.

Kang, J. D. Y. and Schafer, J. L. (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data (with discussion). *Statistical Science* **22**, 523–539.

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

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. 2nd edition. Wiley, New York.

Long, Q., Hsu, C.-H. and Li, Y. (2012). Doubly robust nonparametric multiple imputation for ignorable missing data. *Statistica Sinica* **22**, 149–172.

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.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York.

Rubin, D. B. and Schenker, N. (1991). Mutiple imputation in health-care databases: An overview and some applications. *Statistics in Medecine* **10**, 585–598.

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.

Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics* **6**, 177–184.

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics* **6**, 595–645.

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

van Buuren, S. (2012). *Flexible Multiple Imputation of Missing Data*. CRC Press.

van Buuren, S., Groothuis-Oudshoorn, K., Robitzsch, A., Vink, G., Doove, L., and Jolani, S. (2014). Package 'mice'. Retrieved from: `cran.rproject.org/web/packages/mice/mice.pdf`

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York.

Wang, N. and Robins, J. M. (1998). Large-sample theory for parametric multiple imputation procedures. *Biometrika* **85**, 935–948.

Department of Biostatistics and Epidemiology, University of Oklahoma, Oklahoma City, OK 73126-0901, U.S.A.

E-mail: Sixia-Chen@ouhsc.edu

Department of mathematics and statistics, Université de Montréal, Montreal, QC H3C 3J7, Canada

E-mail: haziza@dms.umontreal.ca

(Received March 2017; accepted February 2018)