Mathematical Modelling and Control

http://www.aimspress.com/journal/mmc

MMC, 2(1): 24–33
DOI:10.3934/mmc.2022004
Received: 25 December 2021
Revised: 06 February 2022
Accepted: 10 February 2022
Published: 08 March 2022

Research article

Multiple robust estimation of parameters in varying-coefficient partially linear model with response missing at random

Yaxin Zhao and Xiuli Wang*

School of Mathematics and Statistics, Shandong Normal University, Ji'nan 250358, China

* **Correspondence:** E-mail: wxlmath@163.com.

Abstract: In this paper, we consider the multiple robust estimation of the parameters in the varying-coefficient partially linear model with response missing at random. The multiple robust estimation method is proposed, and the multiple robustness of the proposed method is proved. Numerical simulations are conducted to investigate the finite sample performance of the proposed estimators compared with other competitors.

Keywords: multiple robust estimation; missing at random; lagrange multiplier method; estimation equation

1. Introduction

The model considered in this paper is a classical semi-parametric model, varying-coefficient partially linear model, and it has the following form

$$Y = X^T\theta(T) + Z^T\beta + \varepsilon, \qquad (1.1)$$

where Y is the response variable, X, Z and T are $q-$dimensional, $p-$dimensional and one-dimensional covariates, respectively. $\beta = (\beta_1, \cdots, \beta_p)^T$ is a $p-$dimensional unknown parameter vector, $\theta(\cdot) = (\theta_1(\cdot), \cdots, \theta_q(\cdot))^T$ is a $q-$dimensional unknown non-parametric function vector, ε is the random error and satisfies $E(\varepsilon|X, Z, T) = 0$. Model (1.1) has been well studied by many statisticians, see the literatures, for example, Fan and Huang [1], You and Zhou [2], Huang and Zhang [3], Zhao [4], Feng, Zhang and Lu [5] among others.

In practical applications, missing data problems are frequently encountered in almost all research areas, such as psychological sciences, medical studies, industrial and agricultural production. The complete-case (CC) method will lose the estimation efficiency due to the disregard of the information from the missing values, and may result in biased results if the data is not missing completely at random. For details, see Little and Rubin [6]. The inverse probability weighted (IPW) method is another frequently used method dated back to Horvitz and Thompson [7] that can be applied to the case of missing covariates. This method is to take the inverse of the selection probability as the weight to the fully observed data, and under missing at random (MAR) assumption this method is unbiased. It has attracted much attention in statistical analysis with missing data, but still doesn't make full use of the incomplete data. The imputation method is a popularly method to deal with missing responses in many studies which was introduced by Yates [8]. The concept of imputation is to fill in each missing data with a suitable value, and then use the observed value and the imputed value for statistical inference by the standard method. This method can improve the efficiency of the resulted estimators, see the literatures, for example, Cheng [9], Wang and Rao [10,11], Wang, Linton and Hardle [12], and so forth. In order to further improve the efficiency of estimation, Robins, Rotnitzky and Zhao [13] propose an augmented inverse probability weighted (AIPW) method. This method has the double robustness, that is, if the selection probability and the conditional expectation function are both correctly specified, the resulted estimator will reach the semi-parametric effective bound, and if either of the two assumed models is correctly specified, the estimator is consistent, see the details in Robins and Rotnitzky [14] and Scharfstein, Rotnitzky and Robins [15]. In subsequent ten years, the doubly robust estimation has been well studied, see for example, Kang and Schafer [16], Qin, Shao and Zhang [17], Cao, Tsiatis and Davidian [18], Han [19], and Rotnitzky et al. [20].

However, double robustness does not provide sufficient protection for estimation consistency, since it allows only one model for the selection probability and one for the conditional expectation function. It is often risky to assume that one of these two models is correctly specified with an unknown data generating process. Noticed this, Han and Wang [21] propose multiple robust estimator for the population mean when the response variable is subject to ignorable missingness. They suggest multiple models for both the selection probability function and the outcome regression model, and the resulted estimator is consistent if any of the multiple models is correctly specified, and attains the semi-parametric efficiency bound when one selection probability and one outcome regression model are correctly specified, without requiring knowledge of which models are correct. For the details please resort to Han and Wang [21]. Subsequently, Han [22] studies the multiple robust estimator for the linear regression model. He discusses the numerical implementation of the proposed method through a modified Newton-Raphson algorithm, derives the asymptotic distribution of the resulted estimator and provides some ways to improve the estimation efficiency. Later, Sun, Wang and Han [23] propose multiple robust kernel estimating equations (MRKEEs) for nonparametric regression, demonstrate its multiple robustness, and show that the resulted estimator achieves the optimal efficiency within the class of augmented inverse propensity weighted (AIPW) kernel estimators when including correctly specified models for both the missingness mechanism and the outcome regression. Please refer to Sun, Wang and Han [23] for more discussion. In addition, the multiple robust estimation with nonignorably missing data has been studied recently, and here we just list some literatures, see for example, Han [24] and Li, Yang and Han [25].

To the best of our knowledge, the multiple robust estimation for the parameters of the varying-coefficient partially linear model with response missing at random has not been studied. So in this paper, applying the idea of Han [22] and Sun, Wang and Han [23], we consider the multiple robust estimation method for the parameters of the varying-coefficient partially linear model with missing response, and the proposed method is demonstrated superior over the existing competitors via simulation studies.

This paper is organized as follows. The proposed estimation technique and its multiple robustness are presented in Section 2. Numerical simulation studies are conducted in Section 3 in order to examine the performance of the proposed method. The technical proofs are also provided in Section 4. Conclusions are summarized in Section 5.

2. The proposed estimator

Suppose the available incomplete data $\{(R_i, Y_i, X_i, Z_i, T_i), i = 1, 2, \ldots, n\}$ is a random sample from model (1.1), that is

$$Y_i = X_i^T \theta(T_i) + Z_i^T \beta + \varepsilon_i, \tag{2.1}$$

where R_i is an indicator variable, when Y_i can be observed, then $R_i = 1$, and when Y_i is missing with $R_i = 0$. The covariate X_i, Z_i and T_i are all observed. Following Han [22] and Sun, Wang and Han [23], we also suppose the auxiliary variables S_i relate to $(R_i, Y_i, X_i, Z_i, T_i)$ is available. Just as Han [22] points out that the auxiliary variables do not enter the regression model and are not of direct statistical interest, but they can reduce the impact of missing data on estimation and improve the estimation efficiency. Let $V_i = (X_i^T, Z_i^T)^T$ denote the covariates. The missing mechanism we assume in this paper is MAR mechanism that commonly used in practice. Specifically, given the covariates V_i, T_i and the available auxiliary variables S_i, the missing of Y_i is independent of Y_i, that is,

$$P\{R_i = 1 | Y_i, V_i, T_i, S_i\} = P\{R_i = 1 | V_i, T_i, S_i\} \hat{=} \pi(V_i, S_i). \tag{2.2}$$

Here we assume that $\pi(\cdot)$ is only related to V and S.

We first carry out the estimator of the varying coefficient functions $\theta(\cdot)$. For any t in a small neighborhood of t_0, using the local linear fitting for $\theta_j(t)$, $j = 1, 2, \ldots, q$, we have

$$\theta_j(t) \approx \theta_j(t_0) + \theta_j'(t_0)(t - t_0) = a_j + b_j(t - t_0).$$

Suppose the parameter β is known, and then minimizing the following objective function

$$\sum_{i=1}^{n} R_i\{Y_i - Z_i^T\beta - \sum_{j=1}^{q}(a_j + b_j(T_i - t_0))X_{ij}\}^2 K_h(T_i - t_0)$$

about (a_j, b_j), $j = 1, 2, \ldots, q$, we can obtain the estimator of $\theta(t)$ at t_0, where $K_h(\cdot) = h^{-1}k(\cdot/h)$, $k(\cdot)$ is a kernel function, and h is the bandwidth. Let

$$D_{t_0} = \begin{pmatrix} X_1^T & h^{-1}(T_1 - t_0)X_1^T \\ \vdots & \vdots \\ X_n^T & h^{-1}(T_n - t_0)X_n^T \end{pmatrix},$$

$$W_{t_0} = \mathrm{diag}(K_h(T_1 - t_0)R_1, K_h(T_2 - t_0)R_2, \cdots, K_h(T_n - t_0)R_n),$$

and

$$\begin{aligned} S(t_0) &= (I_q, 0_q)(D_{t_0}^T W_{t_0} D_{t_0})^{-1} D_{t_0}^T W_{t_0} \\ &= (S_1(t_0), S_2(t_0), \ldots, S_n(t_0)), \end{aligned}$$

then the estimator of the coefficient functions $\theta(t)$ at t_0 is given by

$$\tilde{\theta}(t_0) = \sum_{k=1}^{n} S_k(t_0)(Y_k - Z_k^T\beta). \tag{2.3}$$

Substituting (2.3) into (2.1), we obtain

$$\tilde{Y}_i = \tilde{Z}_i^T\beta + \varepsilon_i, \tag{2.4}$$

where $\tilde{Y}_i = Y_i - X_i^T\hat{g}(T_i)$, $\tilde{Z}_i = Z_i - \hat{\mu}^T(T_i)X_i$ with $\hat{g}(t) = \sum_{k=1}^{n} S_k(t)Y_k$ and $\hat{\mu}(t) = \sum_{k=1}^{n} S_k(t)Z_k^T$.

For model (2.4), using the complete data, the CC estimator of β can be obtained by solving the following estimation equation

$$\sum_{i=1}^{n} R_i\hat{\xi}_i(\beta) = 0, \tag{2.5}$$

where

$$\begin{aligned} \hat{\xi}_i(\beta) &= \tilde{Z}_i(\tilde{Y}_i - \tilde{Z}_i^T\beta) \\ &= (Z_i - \hat{\mu}^T(T_i)X_i)[Y_i - X_i^T\hat{g}(T_i) - (Z_i - \hat{\mu}^T(T_i)X_i)^T\beta]. \end{aligned}$$

From Little and Rubin [6] we know that the CC estimator maybe biased unless the missing mechanism is missing completely at random. So following the works of Robins, Rotnitzky and Zhao [13], the doubly robust estimator $\hat{\beta}_{AIPW}$ of β can be defined by

$$\frac{1}{n}\sum_{i=1}^{n}\{\frac{R_i}{\hat{\pi}(V_i, S_i)}\hat{\xi}_i(\beta) - \frac{R_i - \hat{\pi}(V_i, S_i)}{\hat{\pi}(V_i, S_i)}\eta_i(\beta)\} = 0, \tag{2.6}$$

where $\hat{\pi}(V_i, S_i)$ is some estimated value of $\pi(V_i, S_i)$, $\eta_i(\beta) = E[\hat{\xi}_i(\beta)|V_i, T_i, S_i]$. $\hat{\beta}_{AIPW}$ has been improved in terms of consistency, but in practice it is still a great risk to assume that one of the two assumed models is correctly specified. So inspired by Han [22] and Sun, Wang and Han [23], next we shall give the multiple robust estimation for β.

Suppose there are J and K models used to estimate $\pi(V, S)$ and $E(Y|V, T, S)$. Let $\mathscr{P} = \{\pi^j(\alpha^j) : j = 1, \cdots, J\}$ and $\mathscr{F} = \{a^k(\gamma^k) : k = 1, \cdots, K\}$ denote the set of these two models respectively, where α^j and γ^k are the corresponding parameters. Let $\hat{\alpha}^j$, $\hat{\gamma}^k$ be the estimator of α^j, γ^k respectively. Usually, $\hat{\alpha}^j$ can be obtained by maximizing the binomial likelihood

$$\prod_{i=1}^{n}\{\pi_i^j(\alpha^j)\}^{R_i}\{1 - \pi_i^j(\alpha^j)\}^{1-R_i}.$$

According to the property of MAR assumption, it can be seen that Y and R are conditionally independent with respect to (V, T, S), that is, $E(Y|V, T, S) = E(Y|R = 1, V, T, S)$. Therefore, using the complete observation data to fit the model $a^k(\gamma^k)$, we can obtain $\hat{\gamma}^k$. Let $\hat{\beta}^k$ be the solution of

$$\frac{1}{n}\sum_{i=1}^{n}\{Z_i - \hat{\mu}^T(T_i)X_i\}\{R_iY_i + (1 - R_i)a_i^k(\hat{\gamma}^k) - X_i^T\tilde{\theta}(T_i) - Z_i^T\beta\} = 0. \tag{2.7}$$

Obviously, $\hat{\beta}^k$ is an estimated value of β.

Next, let $m = \sum_{i=1}^{n} R_i$ represents the number of the observable response variables. Without loss of generality, $R_1 = \cdots = R_m = 1$, $R_{m+1} = \cdots = R_n = 0$. Let $\omega(V, S) = \frac{1}{\pi(V,S)}$, similar to Han [22], the following formulas hold

$$E(\omega(V, S)[\pi^j(\alpha^j) - E\{\pi^j(\alpha^j)\}]|R = 1) = 0, \tag{2.8}$$

$$E(\omega(V, S)[U^k(\beta, \gamma^k) - E\{U^k(\beta, \gamma^k)\}]|R = 1) = 0, \tag{2.9}$$

where $j = 1, \cdots, J$, $k = 1, \cdots, K$, $U^k(\beta, \gamma^k) = \{Z - \mu^T(T)X\}\{a^k(\gamma^k) - X^T\theta(T) - Z^T\beta\}$. Therefore, the weights

$\omega_i, i = 1, \cdots, m$ can be defined by

$$\omega_i \geq 0, i = 1, \cdots, m; \quad \sum_{i=1}^{m} \omega_i = 1,$$
$$\sum_{i=1}^{m} \omega_i \{\pi_i^j(\hat{\alpha}^j) - \nu^j(\hat{\alpha}^j)\} = 0, j = 1, \cdots, J,$$
$$\sum_{i=1}^{m} \omega_i \{\hat{U}_i^k(\hat{\beta}^k, \hat{\gamma}^k) - \eta^k(\hat{\beta}^k, \hat{\gamma}^k)\} = 0, k = 1, \cdots, K,$$

where

$$\nu^j(\hat{\alpha}^j) = \frac{1}{n} \sum_{i=1}^{n} \pi_i^j(\hat{\alpha}^j), j = 1, \cdots, J,$$

$$\eta^k(\hat{\beta}^k, \hat{\gamma}^k) = \frac{1}{n} \sum_{i=1}^{n} \hat{U}_i^k(\hat{\beta}^k, \hat{\gamma}^k), k = 1, \cdots, K,$$

$$\hat{U}_i^k(\hat{\beta}^k, \hat{\gamma}^k) = \{Z_i - \hat{\mu}^T(T_i)X_i\}\{a_i^k(\hat{\gamma}^k) - X_i^T \tilde{\theta}(T_i) - Z_i^T \hat{\beta}^k\}.$$

Based on the empirical likelihood method, under the above constraints, the Lagrange multiplier method is used to solve the maximum value problem of $\prod_{i=1}^{m} \omega_i$, and we use the solution as the weight $\omega_i (i = 1, \cdots, m)$ to estimate the parameter β. For ease of presentation, let $\hat{\alpha}^T = \{(\hat{\alpha}^1)^T, \cdots, (\hat{\alpha}^J)^T\}$, $\hat{\beta}^T = \{(\hat{\beta}^1)^T, \cdots, (\hat{\beta}^K)^T\}$, $\hat{\gamma}^T = \{(\hat{\gamma}^1)^T, \cdots, (\hat{\gamma}^K)^T\}$, and $\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})^T = [\pi_i^1(\hat{\alpha}^1) - \nu^1(\hat{\alpha}^1), \cdots, \pi_i^J(\hat{\alpha}^J) - \nu^J(\hat{\alpha}^J), \{\hat{U}_i^1(\hat{\beta}^1, \hat{\gamma}^1) - \eta^1(\hat{\beta}^1, \hat{\gamma}^1)\}^T, \cdots, \{\hat{U}_i^K(\hat{\beta}^K, \hat{\gamma}^K) - \eta^K(\hat{\beta}^K, \hat{\gamma}^K)\}^T]$. Based on the empirical likelihood theory, we have

$$\hat{\omega}_i = \frac{1}{m} \frac{1}{1 + \hat{\rho}^T \hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})}, \quad i = 1, \cdots, m, \tag{2.10}$$

where $\hat{\rho}^T = (\hat{\rho}_1, \cdots, \hat{\rho}_{J+pK})$ is the $(J + pK)$-dimension Lagrange multiplier, and is the solution of

$$\frac{1}{m} \sum_{i=1}^{m} \frac{\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})}{1 + \rho^T \hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})} = 0. \tag{2.11}$$

Due to the non-negativity of the weight $\hat{\omega}_i$, $\hat{\rho}$ satisfies

$$1 + \hat{\rho}^T \hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) > 0, \quad i = 1, \cdots, m. \tag{2.12}$$

So we can solve the equation

$$\sum_{i=1}^{m} \hat{\omega}_i \hat{\xi}_i(\beta) = 0 \tag{2.13}$$

to obtain the multiple robust estimator of the parameter β, denoted by $\hat{\beta}_{MR}$.

In calculation of the weight $\hat{\omega}_i$, the Lagrange multiplier $\hat{\rho}$ is essential. The calculation algorithm we used is similar to Han [22], for the details please refer to Han [22], here we omit.

The multiple robustness of $\hat{\beta}_{MR}$ is given by the following theorem.

Theorem 2.1. Suppose that the conditions C1–C5 in Section 4 hold, and if $\mathscr{P}$ contains a model that correctly specifies $\pi(V, S)$, or $\mathscr{F}$ contains a correctly specified model for $E(Y|V, T, S)$, then $\sum_{i=1}^{m} \hat{\omega}_i \hat{\xi}_i(\beta) \xrightarrow{P} 0$ with $n \to \infty$.

3. Simulation study

In this section, we conduct some numerical simulations to evaluate the feasibility of the above method and the finite sample performance of the proposed estimator $\hat{\beta}_{MR}$. Several indices of multiple robust estimates, inverse probability weighted estimates, and augmented inverse probability weighted estimates are compared and analyzed under different sample sizes.

We consider five mutually independent covariates, namely: $X \sim N(0, 1), T \sim U(0, 1), Z_1 \sim N(1, 5), Z_2 \sim B(0.5, 1), Z_3 \sim N(0, 1)$. The response variable is generated by the model $Y = X^T \theta(T) + Z^T \beta + \varepsilon$, where $\theta(t) = \sin(\pi t)$ and $\beta = (1, 1, 2)^T$. In addition, We consider three auxiliary variable, namely $S^{(1)} = 1 + Z^{(1)} - Z^{(2)} + \varepsilon_1$, $S^{(2)} = I\{S^{(1)} + 0.4\varepsilon_2 > 2.8\}$, $S^{(3)} = \exp[\{S^{(1)}/9\}^2] + \varepsilon_3$, where $I(\cdot)$ is an indicator function. $(\varepsilon, \varepsilon_1, \varepsilon_2, \varepsilon_3)^T \sim N(0, \Sigma)$. The diagonal elements of the matrix Σ are $1, 0.5, 1, 2$, the elements at positions $(1, 2)$ and $(2, 1)$ are 0.5, and the remaining elements are all 0. The probability of selection is $\text{logit}\{\pi(V, S)\} = 3.5 - 5S^{(2)}$, under which there are approximately 34% of the subjects with missing Y. The models for correctly estimating $\pi(V, S)$ and $E(Y|V, T, S)$ are $\text{logit}\{\pi^1(\alpha^1)\} = \alpha_1^1 + \alpha_2^1 S^{(2)}$ and $a^1(\gamma^1) = X^T \theta(T) + \gamma_1^1 Z_1 + \gamma_2^1 Z_2 + \gamma_3^1 Z_3 + \gamma_4^1 S^{(3)}$ respectively. In addition, we also use two incorrect models in the simulation process, namely $\text{logit}\{\pi^2(\alpha^2)\} = \alpha_1^2 + \alpha_2^2 Z_1 + \alpha_3^2 Z_2 + \alpha_4^2 Z_3$, $a^2(\gamma^2) = X^T(-4T^2 + 4T) + \gamma_1^2 Z_1 + \gamma_2^2 Z_2 + \gamma_3^2 Z_3 + \gamma_4^2 S^{(3)}$. For simplicity, we use the Rule of Thumb method to obtain the optimal bandwidth when estimating the nonparametric functions, that is, $h = 1.06 * \{min(qr, sig)\} * n^{-1/5}$, where sig is the standard deviation of covariate T, $qr = (Q_3 - Q_1)/1.34$, Q_1 and Q_3 are the first and third quartile, respectively. In simulation, we generate random samples with $n = 200$ and $n = 500$ respectively, and repeat the process 500 times to calculate the average biases, mean squared errors (MSEs), the root of mean squared errors (RMSEs) and median absolute error (MAEs).

In order to verify the superiority of the multiple robust estimation method, we give the calculated indices of the parameter β under different estimation methods, which are the inverse probability weighted estimates $\hat{\beta}_{IPW}$, and the augmented inverse probability weighted estimates $\hat{\beta}_{AIPW}$ and multiple robust estimates $\hat{\beta}_{MR}$. To distinguish all the estimators constructed based on different methods and models, each estimator is assigned a name with the form "Method-0000", where each digit of the four-digit number, from left to right, indicates whether $\pi^1(\alpha^1), \pi^2(\alpha^2), a^1(\gamma^1), a^2(\gamma^2)$ is used in the construction (1 means yes, 0 means no), respectively. The simulation results are reported in Table 1 and Table 2 with the sample size $n = 200$ and $n = 500$.

It can be seen from the two tables that regardless of the estimation method, the larger the sample size, the better the estimation effect. And when the models for estimating the selection probability and the conditional expectation are all specified correctly, the estimated results obtained by the multiple robust estimation method, the inverse probability weighted estimation method and the augmented inverse probability weighted estimation method are not much different, but the effect of multiple robust estimation is better in terms of MSE. When all the models for estimating the selection probability and the conditional expectation are specified incorrectly, the $AIPW - 0101$ has unsatisfactory effects, the resulted estimators have larger deviations, but our proposed $MRE - 0101$, despite using two incorrect models, can generate better estimators. The interesting observation that $\hat{\beta}_{MR}$ seems to still provide a reasonable (at least not too bad) estimate of β even if there is no model correctly specified is similar to Han [22]. In a word, it is obvious that our proposed multiple robust estimation method is better than the two competitors.

4. Proofs

Before we give the proof of Theorem 2.1, some notations and interpretations are presented firstly.

Let $\Phi(t) = E[RXZ^T|T = t]$, $\Psi(t) = E[RXX^T|T = t]$, then

$$\theta(T_i) = \{\Psi(T_i)\}^{-1}\{E[R_iX_iY_i|T_i] - \Phi(T_i)\beta\}. \quad (4.1)$$

Substituting (4.1) into (2.1), we obtain

$$\check{Y}_i = \check{Z}_i^T\beta + \varepsilon_i, \quad (4.2)$$

where $\check{Y}_i = Y_i - X_i^T g(T_i)$, $\check{Z}_i = Z_i - \mu^T(T_i)X_i$, with $g(T_i) = \{\Psi(T_i)\}^{-1}E[R_iX_iY_i|T_i]$, $\mu(T_i) = \{\Psi(T_i)\}^{-1}\Phi(T_i)$. From model (4.2), using the complete data, the CC estimator of β can be obtained by solving the following estimation equation

$$\sum_{i=1}^{n} R_i\xi_i(\beta) = 0,$$

where $\xi_i(\beta) = \check{Z}_i(\check{Y}_i - \check{Z}_i^T\beta) = (Z_i - \mu^T(T_i)X_i)[Y_i - X_i^T g(T_i) - (Z_i - \mu^T(T_i)X_i)^T\beta]$, and $E[\xi_i(\beta)] = 0$.

Suppose C to be a positive constant which can represent different values, and assume the following conditions C1–C5 hold.

C1 The bandwidth h satisfies $h = Cn^{-1/5}$, that is $h \to 0$ and $nh \to \infty$ as $n \to \infty$, where $C > 0$ is a given positive constant.

C2 The kernel function $K(\cdot)$ is a symmetric probability kernel function, and $\int t^2K(t)dt \neq 0$, $\int t^4K(t)dt < \infty$.

C3 For each $t \in (0, 1)$, $f(t), \Phi(t), \Psi(t)$ and $\theta(t)$ are twice continuous differentiable at point t , where $f(t)$ is the density function of the variable T.

C4 $\sup_{0\le t\le 1} E[\varepsilon_i^4|T_i = t] < \infty$, $\sup_{0\le t\le 1} E[X_{ir}^4|T_i = t] < \infty$, and they are continuous about t, where X_{ir} is the r-th component of X_i, $i = 1, \cdots, n$, $r = 1, \cdots, q$.

C5 For a given t, $\Psi(t)$ is a positive definite matrix.

Next, a Lemma is needed in proof of Theorem 2.1, and the proof can be found in Zhao [4].

Lemma 4.1. Suppose conditions C1–C5 hold, then we have

$$\sup_{0<t<1} \|\hat{\mu}(t) - \Psi(t)^{-1}\Phi(t)\| = O_p(C_n),$$

$$\sup_{0<t<1} \|\hat{g}(t) - \Psi(t)^{-1}\Phi(t)\beta - \theta(t)\| = O_p(C_n),$$

where $C_n = h^2 + (\frac{\log(1/h)}{nh})^{1/2}$.

Proof of Theorem 2.1: Assuming that $\mathscr{P}$ contains a model that correctly specifies $\pi(V, S)$, without loss of generality, let $\pi^1(\alpha^1)$ be the model, α_0^1 represents the truth

Table 1. The biases, MSEs, RMSEs and MAEs (multiplied by 10^2) of different estimators for parameter β when sample size $n = 200$.

Method	β_1				β_2				β_3			
	Bias	MSE	RMSE	MAE	Bias	MSE	RMSE	MAE	Bias	MSE	RMSE	MAE
IPW-1000	0.071	0.083	2.872	2.296	1.451	2.645	16.26	12.96	0.003	1.360	11.66	9.160
IPW-0100	0.329	0.233	4.826	3.677	0.021	4.591	21.43	16.04	2.113	3.915	19.79	14.65
AIPW-1010	0.030	0.078	2.799	2.211	1.324	2.654	16.29	12.88	0.073	1.402	11.84	9.207
AIPW-1001	0.113	0.074	2.724	2.206	0.644	2.629	16.22	13.18	0.181	1.546	12.44	9.692
AIPW-0110	0.065	0.069	2.625	2.125	0.387	2.283	15.11	12.17	0.232	1.551	12.45	9.778
AIPW-0101	-4.930	130.1	114.1	33.09	-5.591	212.9	145.9	51.96	-1.457	447.1	211.4	59.17
MR-1111	0.066	0.022	2.479	2.193	-0.921	2.341	15.46	12.00	-0.563	0.591	11.18	9.120
MR-1110	0.030	0.021	2.532	2.208	0.637	2.207	15.59	12.08	0.487	0.572	11.23	9.510
MR-1101	0.031	0.023	2.574	2.140	0.602	2.206	15.13	13.26	0.487	0.573	12.01	9.165
MR-1011	0.068	0.020	2.608	2.361	-0.927	2.334	16.35	12.31	-0.544	0.592	12.84	9.997
MR-1010	0.031	0.022	2.427	2.072	0.639	2.207	16.01	12.24	0.501	0.570	11.81	9.772
MR-1001	0.032	0.023	2.899	2.508	0.603	2.208	16.37	13.06	0.501	0.571	12.92	9.328
MR-0111	0.065	0.022	2.623	2.283	-0.909	2.341	15.24	13.52	-0.555	0.590	12.12	9.808
MR-0110	0.029	0.024	2.487	2.904	0.629	2.203	16.20	12.73	0.492	0.573	11.42	9.629
MR-0101	0.121	0.106	3.458	3.140	0.560	5.210	18.53	15.87	-1.064	1.371	16.54	12.61

Table 2. The biases, MSEs, RMSEs and MAEs (multiplied by 10^2) of different estimators for parameter β when sample size $n = 500$.

Method	β_1				β_2				β_3			
	Bias	MSE	RMSE	MAE	Bias	MSE	RMSE	MAE	Bias	MSE	RMSE	MAE
IPW-1000	-0.071	0.030	1.732	1.401	0.202	1.023	10.11	8.104	-0.243	0.649	8.054	6.333
IPW-0100	0.208	0.135	3.677	2.869	0.034	4.570	21.38	15.17	0.264	3.559	18.86	13.43
AIPW-1010	0.014	0.027	1.649	1.334	-0.426	1.013	10.06	7.950	0.126	0.611	7.815	5.072
AIPW-1001	-0.009	0.029	1.692	1.334	-0.173	0.964	9.820	7.813	0.364	0.556	7.457	5.692
AIPW-0110	0.017	0.031	1.762	1.380	-0.026	0.870	9.328	7.529	-0.945	0.567	7.527	6.027
AIPW-0101	3.621	165.3	128.6	38.69	-1.019	294.4	171.6	63.55	15.32	123.2	351.0	69.06
MR-1111	0.040	0.018	1.665	1.308	0.257	0.950	9.379	7.060	-0.248	0.387	6.919	4.147
MR-1110	-0.031	0.022	1.689	1.347	0.219	0.913	9.828	7.301	0.114	0.404	6.484	4.991
MR-1101	0.039	0.021	1.669	1.386	-0.460	1.081	10.33	7.582	0.228	0.396	6.873	4.136
MR-1011	0.057	0.020	1.670	1.302	0.265	0.893	9.785	8.108	0.220	0.382	6.622	4.618
MR-1010	0.027	0.015	1.537	1.256	-0.071	0.860	9.603	7.096	0.265	0.359	6.140	5.054
MR-1001	0.044	0.026	1.714	1.350	0.441	1.039	10.36	8.224	0.326	0.462	7.427	6.043
MR-0111	-0.028	0.031	1.746	1.371	0.329	1.014	9.751	7.301	0.164	0.406	7.520	5.150
MR-0110	0.034	0.029	1.657	1.395	0.480	1.076	9.560	8.061	0.326	0.441	7.293	5.875
MR-0101	1.027	0.103	2.853	2.331	0.537	4.031	11.39	10.57	0.931	1.230	11.36	9.480

value of α^1, that is $\pi^1(\alpha_0^1) = \pi(V,S)$. Next, we combine the theory of empirical likelihood to prove that $\hat{\beta}_{MR}$ is a consistent estimator of β.

Referring to the method in Han [22] to establish the relationship between the weight $\hat{\omega}_i$ and the empirical likelihood on the biased sample. Let p_i represent the conditional empirical probability on the biased sample $(Y_i, X_i, Z_i, T_i, S_i), R_i = 1, i = 1, \cdots, m$, based on (2.8), (2.9) and $\omega(V,S) = \frac{1}{\pi^1(\alpha_0^1)}$, a more reasonable value of p_i can be given by the following constrained optimization problem:

$$\max_{p_1,\cdots,p_m} \prod_{i=1}^m p_i; \quad p_i \geq 0, i = 1, \cdots, m; \quad \sum_{i=1}^m p_i = 1,$$

$$\sum_{i=1}^m p_i\{\pi_j^i(\hat{\alpha}^j) - \nu^j(\hat{\alpha}^j)\}/\pi_i^1(\hat{\alpha}^1) = 0, j = 1, \cdots, J,$$

$$\sum_{i=1}^m p_i\{\hat{U}_i^k(\hat{\beta}^k, \hat{\gamma}^k) - \eta^k(\hat{\beta}^k, \hat{\gamma}^k)\}/\pi_i^1(\hat{\alpha}^1) = 0, k = 1, \cdots, K.$$

Using the Lagrange multiplier method again, we get

$$\hat{p}_i = \frac{1}{m}\frac{1}{1 + \hat{\lambda}^T\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})/\pi_i^1(\hat{\alpha}^1)}, i = 1, \cdots, m,$$

where $\hat{\lambda}^T = (\hat{\lambda}_1, \cdots, \hat{\lambda}_{J+pK})$ is the $(J + pK)$-dimensional Lagrange multiplier, and satisfies

$$\frac{1}{m}\sum_{i=1}^m \frac{\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})/\pi_i^1(\hat{\alpha}^1)}{1 + \lambda^T\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})/\pi_i^1(\hat{\alpha}^1)} = 0.$$

Due to the non-negativity of $\hat{p}_i$, $\hat{\lambda}$ satisfies $1 + \hat{\lambda}^T\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})/\pi_i^1(\hat{\alpha}^1) > 0, i = 1, \cdots, m$. Since

$$\begin{aligned}
&\frac{1}{m}\sum_{i=1}^m \frac{\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})/\pi_i^1(\hat{\alpha}^1)}{1 + \lambda^T\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})/\pi_i^1(\hat{\alpha}^1)} \\
&= \frac{1}{\nu^1(\hat{\alpha}^1)}\frac{1}{m}\sum_{i=1}^m \frac{\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})}{1 + \frac{\pi_i^1(\hat{\alpha}^1) - \nu^1(\hat{\alpha}^1)}{\nu^1(\hat{\alpha}^1)} + \{\frac{\lambda}{\nu^1(\hat{\alpha}^1)}\}^T\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})} \\
&= \frac{1}{\nu^1(\hat{\alpha}^1)}\frac{1}{m}\sum_{i=1}^m \frac{\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})}{1 + \{\frac{\lambda_1 + 1}{\nu^1(\hat{\alpha}^1)}, \frac{\lambda_2}{\nu^1(\hat{\alpha}^1)}, \cdots, \frac{\lambda_{J+pK}}{\nu^1(\hat{\alpha}^1)}\}\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})},
\end{aligned}$$

then the solution of (2.11), $\hat{\rho}$, can be written as $\hat{\rho}_1 = (\hat{\lambda}_1 + 1)/\nu^1(\hat{\alpha}^1)$ and $\hat{\rho}_l = \hat{\lambda}_l/\nu^1(\hat{\alpha}^1), l = 2, \cdots, J + pK$. Therefore

$$\hat{\omega}_i = \frac{1}{m}\frac{\nu^1(\hat{\alpha}^1)/\pi_i^1(\hat{\alpha}^1)}{1 + \hat{\lambda}^T\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})/\pi_i^1(\hat{\alpha}^1)} = \frac{\hat{p}_i\nu^1(\hat{\alpha}^1)}{\pi_i^1(\hat{\alpha}^1)}.$$

Just like White [24], let α_*^j, β_*^k and γ_*^k are the minimum points of the corresponding Kullback-Leibler distance respectively, then we have $\hat{\alpha}^j \xrightarrow{P} \alpha_*^j, \hat{\beta}^k \xrightarrow{P} \beta_*^k, \hat{\gamma}^k \xrightarrow{P} \gamma_*^k$, and $n^{1/2}(\hat{\alpha}^j - \alpha_*^j)$, $n^{1/2}(\hat{\beta}^k - \beta_*^k)$ and $n^{1/2}(\hat{\gamma}^k - \gamma_*^k)$ are bounded by probability. At the same time, $\nu^j(\hat{\alpha}^j) \xrightarrow{P} \nu_*^j$, $\eta^k(\hat{\beta}^k, \hat{\lambda}^k) \xrightarrow{P} \mu_*^k$, where $\nu_*^j = E[\pi^j(\alpha_*^j)], \mu_*^k = E[U^k(\beta_*^k, \gamma_*^k)]$. Generally speaking, when the model $\pi^j(\alpha^j)$ for $\pi(V,S)$ is correctly specified, we have $\pi^j(\alpha_*^j) = \pi(V,S)$, and when the model $a^k(\gamma^k)$ for $E(Y|V,T,S)$ is correctly specified, we have $a^k(\gamma_*^k) = E(Y|V,T,S)$. Let $\alpha_*^T = \{(\alpha_*^1)^T, \cdots, (\alpha_*^J)^T\}$, $\beta_*^T = \{(\beta_*^1)^T, \cdots, (\beta_*^K)^T\}$, $\gamma_*^T = \{(\gamma_*^1)^T, \cdots, (\gamma_*^K)^T\}$, and suppose $\hat{\rho} \xrightarrow{P} \rho_*$.

Based on the empirical likelihood theory, it can be known that $\hat{\lambda} \xrightarrow{P} 0$. According to the appendix in Han [22], $\hat{\lambda} = O_p(n^{-1/2})$ holds. Since the model $\pi^1(\alpha^1)$ is correct, then we have $\frac{m}{n} \xrightarrow{P} \nu_*^1$, and

$$\begin{aligned}
\sum_{i=1}^m \hat{\omega}_i\hat{\xi}_i(\beta) &= \frac{1}{m}\sum_{i=1}^n \frac{R_i\nu^1(\hat{\alpha}^1)/\pi_i^1(\hat{\alpha}^1)}{1 + \hat{\lambda}^T\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})/\pi_i^1(\hat{\alpha}^1)}\hat{\xi}_i(\beta) \\
&= \frac{\nu^1(\hat{\alpha}^1)}{m}\sum_{i=1}^n \frac{R_i/\pi_i^1(\hat{\alpha}^1)}{1 + \hat{\lambda}^T\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})/\pi_i^1(\hat{\alpha}^1)}\hat{\xi}_i(\beta) \\
&= \frac{\nu_*^1}{m}\sum_{i=1}^n \frac{R_i/\pi_i^1(\hat{\alpha}^1)}{1 + \hat{\lambda}^T\hat{g}_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma})/\pi_i^1(\hat{\alpha}^1)}\hat{\xi}_i(\beta) \\
&= \frac{1}{n}\sum_{i=1}^n \frac{R_i}{\pi_i^1(\alpha_*^1)}\hat{\xi}_i(\beta) + o_p(1).
\end{aligned}$$

Refer to Zhao [4], since

$$\begin{aligned}
\hat{\xi}_i(\beta) &= [Z_i - \hat{\mu}^T(T_i)X_i]\varepsilon_i + [\mu(T_i) - \hat{\mu}(T_i)]^TX_i\varepsilon_i \\
&\quad + [Z_i - \hat{\mu}^T(T_i)X_i]X_i^T[\theta(T_i) - \hat{g}(T_i) + \hat{\mu}(T_i)\beta] \\
&\quad + [\mu(T_i) - \hat{\mu}(T_i)]^TX_iX_i^T[\theta(T_i) - \hat{g}(T_i) + \hat{\mu}(T_i)\beta], \\
\xi_i(\beta) &= [Z_i - \mu^T(T_i)X_i][Z_i - \mu^T(T_i)X_i]^T\beta + [Z_i - \mu^T(T_i)X_i]\varepsilon_i,
\end{aligned}$$

and $E[X_i\varepsilon_i] = 0$, $E[(Z_i - \mu^T(T_i)X_i)X_i)^T] = 0$, we have

$$\begin{aligned}
\hat{\xi}_i(\beta) - \xi_i(\beta) &= [\mu(T_i) - \hat{\mu}(T_i)]^TX_i\varepsilon_i \\
&\quad + [\mu(T_i) - \hat{\mu}(T_i)]^TX_iX_i^T[\theta(T_i) - \hat{g}(T_i) + \hat{\mu}(T_i)\beta] \\
&\quad + [Z_i - \mu^T(T_i)X_i][X_i^T\theta(T_i) - X_i^T\hat{g}(T_i) \\
&\quad + X_i^T\hat{\mu}(T_i)\beta - Z_i^T\beta + X_i^T\mu(T_i)\beta].
\end{aligned}$$

Combine conditions C1,C4,C5 and Lemma 4.1, we have

$$\|\frac{1}{n}\sum_{i=1}^n \frac{R_i}{\pi_i^1(\alpha_*^1)}\hat{\xi}_i(\beta) - \frac{1}{n}\sum_{i=1}^n \frac{R_i}{\pi_i^1(\alpha_*^1)}\xi_i(\beta)\| \xrightarrow{P} 0.$$

That is, $\frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi_i^1(\alpha_*^1)}\hat{\xi}_i(\beta)\stackrel{P}{\longrightarrow}\frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi_i^1(\alpha_*^1)}\xi_i(\beta)$. Then we have

$$\begin{aligned}\sum_{i=1}^{m}\hat{\omega}_i\hat{\xi}_i(\beta)&=\frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi_i^1(\alpha_*^1)}\hat{\xi}_i(\beta)+o_p(1)\\&=\frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi_i^1(\alpha_*^1)}\xi_i(\beta)+o_p(1)\\&\stackrel{P}{\longrightarrow}E[\frac{R}{\pi(V,S)}\xi(\beta)]=0.\end{aligned}$$

Therefore, when $n\to\infty$, β is the solution of the formula (2.13), which shows that $\hat{\beta}_{MR}$ is a consistent estimator of β.

Next, suppose that $\mathscr{F}$ contains a model that correctly specifies $E(Y|V,T,S)$. Without loss of generality, let $a^1(\gamma^1)$ be the true model and γ_0^1 be the true value of γ^1, that is $a^1(\gamma_0^1)=E(Y|V,T,S)$, and $\gamma_*^1=\gamma_0^1$. A previous constraint is actually

$$\sum_{i=1}^{m}\hat{\omega}_i\hat{U}_i^1(\hat{\beta}^1,\hat{\gamma}^1)=\frac{1}{n}\sum_{i=1}^{n}\hat{U}_i^1(\hat{\beta}^1,\hat{\gamma}^1),$$

and $\hat{\beta}^1\stackrel{P}{\longrightarrow}\beta_*^1=\beta$, so we get $\frac{1}{n}\sum_{i=1}^{n}\hat{U}_i^1(\hat{\beta}^1,\hat{\gamma}^1)\stackrel{P}{\longrightarrow}0$.

Let $g(\alpha_*,\beta_*,\gamma_*)^T=[\pi^1(\alpha_*^1)-\nu_*^1,\cdots,\pi^J(\alpha_*^J)-\nu_*^J,\{U^1(\beta_*^1,\gamma_*^1)-\eta_*^1\}^T,\cdots,\{U^K(\beta_*^K,\gamma_*^K)-\eta_*^K\}^T]$, due to $\frac{1}{n}\sum_{i=1}^{n}\hat{U}_i^1(\hat{\beta}^1,\hat{\gamma}^1)\stackrel{P}{\longrightarrow}\frac{1}{n}\sum_{i=1}^{n}U_i^1(\beta,\gamma_0^1),\|\frac{1}{n}\sum_{i=1}^{n}[\hat{\xi}_i(\beta)-\xi_i(\beta)]\|\stackrel{P}{\longrightarrow}0$, and $E[U^1(\beta,\gamma_0^1)]=0$, then we have

$$\begin{aligned}&\sum_{i=1}^{m}\hat{\omega}_i\hat{\xi}_i(\beta)\\=&\sum_{i=1}^{m}\hat{\omega}_i\{\hat{\xi}_i(\beta)-\hat{U}_i^1(\hat{\beta}^1,\hat{\gamma}^1)\}+\frac{1}{n}\sum_{i=1}^{n}\hat{U}_i^1(\hat{\beta}^1,\hat{\gamma}^1)\\=&\frac{1}{m}\sum_{i=1}^{n}\frac{R_i\hat{\xi}_i(\beta)-\hat{U}_i^1(\hat{\beta}^1,\hat{\gamma}^1)}{1+\hat{\rho}^T\hat{g}_i(\hat{\alpha},\hat{\beta},\hat{\gamma})}+E[U^1(\beta,\gamma_0^1)]+o_p(1)\\=&\frac{1}{P(R=1)}E[\frac{R\xi(\beta)-U^1(\beta,\gamma_0^1)}{1+\rho_*^Tg(\alpha_*,\beta_*,\gamma_*)}]+o_p(1)\\=&\frac{1}{P(R=1)}E\{E[\frac{R\xi(\beta)-U^1(\beta,\gamma_0^1)}{1+\rho_*^Tg(\alpha_*,\beta_*,\gamma_*)}|Y,V,T,S]\}+o_p(1)\stackrel{P}{\longrightarrow}0.\end{aligned}$$

This shows that $\hat{\beta}_{MR}$ is a consistent estimator of β.

So the proof of Theorem 2.1 is completed.

5. Conclusions

In this article, we have proposed the multiple robust estimators for parameters in varying-coefficient partially linear model with missing response at random, and the multiple robustness of our proposals has been shown theoretically under some regular conditions. Our simulation studies fully demonstrate the superiority of our multiple robust estimation method through Table 1 and Table 2. Finally, we point out some problems for the future researches. First, we only discuss the multiple robust estimation process of parameters, and the fitting of nonparametric function curves can be expanded in the future studies. Next, based on the model in this article, if the missing mechanism is nonignorable missing, how to obtain the robust estimation of parameters is also worth studying.

Acknowledgments

The research is supported by NSF projects (ZR2021MA077 and ZR2019MA016) of Shandong Province of China.

Conflict of interest

The authors declare that they have no conflicts of interest to this work.

References

1. J. Fan, T. Huang, Profile Likelihood Inferences on Semiparametric Varying-Coefficient Partially Linear Models, *Bernoulli,* **11** (2005), 1031–1057. https://doi.org/10.3150/bj/1137421639
2. J. You, Y. Zhou, Empirical likelihood for semiparametric varying-coefficient partially linear regression models, *Stat. Probabil. Lett.,* **76** (2006), 412–422. https://doi.org/10.1016/j.spl.2005.08.029
3. Z. Huang, R. Zhang, Empirical likelihood for nonparametric parts in semiparametric varying-coefficient partially linear models, *Stat. Probabil. Lett.,* **79** (2009), 1798–1808. https://doi.org/10.1016/j.spl.2009.05.008
4. P. Zhao, *Infrerence for semiparametric varying coefficient partially linear models*, doctoral dissertation, Beijing: Beijing University of Technology, 2010.

5. J. Feng, R. Zhang, Y. Lu, Inference on varying-coefficient partially linear regression model, *Acta Math. Appl. Sin-E,* **31** (2015), 139–156. https://doi.org/10.1007/s10255-015-0457-5

6. R. Little, D. Rubin, *Statistical analysis with missing data*, 1 Eds., New York: Wiley Press, 1986.

7. D. Horvitz, D. Thompson, A generalization of sampling without replacement from a finite universe, *J. Am. Stat. Assoc.,* **47** (1952), 663–685. https://doi.org/10.1080/01621459.1952.10483446

8. F. Yates, The analysis of replicated experiments when the field results are incomplete, *Exp. Agr.,* **1** (1933), 129–142.

9. P. Cheng, Nonparametric estimation of mean functionals with data missing at random, *J. Am. Stat. Assoc.,* **89** (1994), 81–87. https://doi.org/10.1080/01621459.1994.10476448

10. Q. Wang, J. Rao, Empirical likelihood for linear regression models under imputation for missing responses, *Can. J. Stat.,* **29** (2001), 596–608. https://doi.org/10.2307/3316009

11. Q. Wang, J. Rao, Empirical likelihood-based inference under imputation for missing response data, *Ann. Stat.,* **30** (2002), 896–924. https://doi.org/10.1214/aos/1028674845

12. Q. Wang, O. Linton, W. Hardle, Semi-parametric regression analysis with missing response at random, *J. Am. Stat. Assoc.,* **99** (2004), 334–345. https://doi.org/10.1198/016214504000000449

13. J. Robins, A. Rotnitzky, L. Zhao, Estimation of Regression Coefficients When Some Regressors Are Not Always Observed, *J. Am. Stat. Assoc.,* **89** (1994), 846–866. https://doi.org/10.1080/01621459.1994.10476818

14. J. Robins, A. Rotnitzky, Semiparametric efficiency in multivariate regression models with missing data, *J. Am. Stat. Assoc.,* **90** (1995), 122–129. https://doi.org/10.1080/01621459.1995.10476494

15. D. Scharfstein, A. Rotnitzky, R. Robins, Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models, *J. Am. Stat. Assoc.,* **94** (1999), 1096–1120. https://doi.org/10.1080/01621459.1999.10473862

16. J. Kang, J. Schafer, Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data (with Discussion), *Statistics,* **22** (2007), 523–539. https://doi.org/10.1214/07-STS227

17. J. Qin, J. Shao, B. Zhang, Efficient and doubly robust imputation for covariate-dependent missing responses, *J. Am. Stat. Assoc.,* **103** (2008), 797–810. https://doi.org/10.1198/016214508000000238

18. W. Cao, A. Tsiatis, M. Davidian, Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data, *Biometrika,* **96** (2009), 723–734. https://doi.org/10.1093/biomet/asp033

19. P. Han, A note on improving the efficiency of inverse probability weighted estimator using the augmentation term, *Stat. Probabil. Lett.,* **82** (2012), 2221–2228. https://doi.org/10.1016/j.spl.2012.08.005

20. A. Rotnitzky, Q. Lei, M. Sued, J. Robins, Improved double-robust estimation in missing data and causal inference models, *Biometrika,* **99** (2012), 439–456. https://doi.org/10.1093/biomet/ass013

21. P. Han, L. Wang, Estimation with missing data: Beyond double robustness, *Biometrika,* **100** (2013), 417–430. https://doi.org/10.1093/biomet/ass087

22. P. Han, Multiply Robust Estimation in Regression Analysis With Missing Data, *J. Am. Stat. Assoc.,* **109** (2014), 1159–1173. https://doi.org/10.1080/01621459.2014.880058

23. Y. Sun, L. Wang, P. Han, Multiply robust estimation in nonparametric regression with missing data, *J. Nonparametr. Stat.,* **32** (2020), 73–92. https://doi.org/10.1080/10485252.2019.1700254

24. P. Han, Calibration and multiple robustness when data are missing not at random, *Stat. Sinica,* **109** (2018), 1725–1740. https://doi.org/10.5705/ss.202015.0408

25. W. Li, S. Yang, P. Han, Robust estimation for moment condition models with data missing not at

random, *J. Stat. Plan. Infer.,* **207** (2020), 246–254. https://doi.org/10.1016/j.jspi.2020.01.001

26. H. White, Maximum likelihood estimation of misspecified models, *Econometrica,* **50** (1982), 1–25. https://doi.org/10.2307/1912526

AIMS Press