J Syst Sci Complex (2018) 31: 820–840

Adjusted Empirical Likelihood Estimation of Distribution Function and Quantile with Nonignorable Missing Data*

DING Xianwen · TANG Niansheng

DOI: 10.1007/s11424-018-6334-6
Received: 3 January 2017 / Revised: 26 June 2017

Abstract This paper considers the estimation problem of distribution functions and quantiles with nonignorable missing response data. Three approaches are developed to estimate distribution functions and quantiles, i.e., the Horvtiz-Thompson-type method, regression imputation method and augmented inverse probability weighted approach. The propensity score is specified by a semiparametric exponential tilting model. To estimate the tilting parameter in the propensity score, the authors propose an adjusted empirical likelihood method to deal with the over-identified system. Under some regular conditions, the authors investigate the asymptotic properties of the proposed three estimators for distribution functions and quantiles, and find that these estimators have the same asymptotic variance. The jackknife method is employed to consistently estimate the asymptotic variances. Simulation studies are conducted to investigate the finite sample performance of the proposed methodologies.

Keywords Adjusted empirical likelihood, distribution estimation, exponential tilting model, nonignorable missing data, quantile.

1 Introduction

Missing data are often encountered in various fields such as social science, survey sampling and biomedicine. When the missingness data mechanism is nonignorable, it is challenging to make statistical inference on missing data because the propensity score model is unknown. A comprehensive overview on nonignorable missing data analysis can refer to Little and Rubin[1] and Kim and Shao[2].

DING Xianwen
Key Laboratory of Statistical Modeling & Data Analysis of Yunnan Province, Yunnan University, Kunming 650500*, China; Department of Mathematics, Jiangsu University of Technology, Changzhou* 213001*, China.*
Email: dingxianwen@jsut.edu.cn.
TANG Niansheng
Key Laboratory of Statistical Modeling & Data Analysis of Yunnan Province, Yunnan University, Kunming 650500*, China.* Email: nstang@ynu.edu.cn.
*This research was supported by the National Natural Science Foundation of China under Grant Nos. 11671349 and 11601195, the Scientific Research Innovation Team of Yunnan Province under Grant No. 2015HC028, and the Natural Science Foundation of Jiangsu Province of China under Grant No. BK20160289.
◇ *This paper was recommended for publication by Editor SHAO Jun.*

There are considerable literatures on nonignorable missing data analysis. For example, Qin, Leung and Shao[3] proposed a semiparametric empirical likelihood (EL) method to estimate the propensity score under the assumption that the propensity score is of a parametric structure; Zhao and Shao[4] presented an approximate conditional likelihood method in a generalized linear model (GLM) with nonignorable missing responses and covariates. It is well known that statistical inference on the parametric propensity score model is sensitive to the model misspecification. To this end, Kim and Yu[5] proposed a robust semiparametric estimation of mean functions with nonignorable missing response based on an exponential tilting model of the propensity score; Zhao, et al.[6] and Tang, et al.[7] developed an EL approach for generalized estimating equations with nonignorable missing data. As Shao and Wang[8] pointed out that the population may suffer from the well known identification issue without any other assumptions on the propensity score of Kim and Yu[5]. Moreover, the follow-up sample has often to be collected for estimating the tilting parameter in the propensity score[5]. However, in many applications, budget or technical limitations will restrict researchers to collect the additional sample. To address the above mentioned issues, Wang, et al.[9] studied the identification issue of Qin, et al.[3]; Shao and Wang[8] investigated the identification issue of the propensity score model based on the instrumental variable method; Tang, et al.[10] discussed the identifiable condition under the assumption that the missingness data mechanism model of response only depends on response variable; Zhao and Shao[11] studied the identifiability and estimation problem in a GLM with nonignorable responses or covariates; Miao, et al.[12] discussed the identifiability of normal and normal mixture models with nonignorable missing responses; Fang and Shao[13] proposed the penalized validation criterion for model selection with nonignorable nonresponse and studied the procedure to find an instrumental variable, which works well under the assumption that the propensity score model is unspecified.

Estimation of the distribution function and quantiles of a random variable plays a fundamental role in making statistical inference on parameters via the distribution function. Many approaches have been proposed to estimate the distribution function of a random variable without missing data (e.g., see [14–17]). When the response is missing at random (MAR), i.e., the propensity score only depends on the observed data, there exist many works on the estimation of the distribution functions and quantiles of a random variable (e.g., see [18–20]). For nonignorable missing responses, based on the exponential tiling model[5], Zhao, et al.[21] proposed a nonparametric/semiparametric estimation method and an augmented inverse probability weighted imputation method to estimate the distribution function and quantiles of a response variable. The main advantage of Zhao, et al's[21] method is its robustness to the misspecification of the propensity score model.

Motivated by Shao and Wang[8] and Zhao, et al.[21], we here develop three approaches to estimate the distribution function and quantiles of a response variable under the assumption that response is subject to nonignorable missingness, i.e., the Horvtiz-Thompson-type method, the regression imputation method and the augmented inverse probability weighted method. By introducing instrumental variables, the propensity score is specified by a semiparametric exponential tilting model. The adjusted empirical likelihood (AEL) method[22] is employed to estimate the tilting parameter in the propensity score. By adding a few artificial observations to the original data set, the AEL method guarantees the existence of the solution to the calibration conditions with respect to the tilting parameter, and retains some good properties of the EL method[23]. Under some regular conditions, we investigate asymptotic properties of the proposed three estimators for the distribution function and quantiles of a response variable with nonignorable missing, which indicates that the proposed three estimators share the same asymptotic variance to be estimated. The jackknife method is employed to consistently estimate asymptotic variances of the proposed three estimators.

The rest of this paper is organized as follows. Section 2 introduces three approaches to estimate the distribution function and quantiles of a response variable with nonignorable missing, and presents an AEL method to estimate the tilting parameter. Asymptotic properties of the proposed three estimators are investigated in Section 3. The jackknife approach is presented to estimate asymptotic variances in Section 4. Simulation studies are conducted to investigate the performance of the proposed methodologies in Section 5. Section 6 provides a brief discussion on the selection of instrumental variables and an extension of the proposed methodologies. Technical details are given in the Appendix.

2 Methodology

2.1 The Distribution Estimation with Nonignorable Missing Response

Consider an incomplete data set $\{(\boldsymbol{x}_i, Y_i, \delta_i), i = 1, 2, \cdots, n\}$, where $\boldsymbol{x}_i$ is a vector of covariates observed completely, Y_i is a response variable that may be subject to missingness, and δ_i is an indicator of missing response variable Y_i, i.e., $\delta_i = 1$ if Y_i is observed and $\delta_i = 0$ if Y_i is missing. It is assumed that δ_i is independent of δ_j for any $i \neq j$, and δ_i depends both on $\boldsymbol{x}_i$ and Y_i such that the propensity score $\pi(\boldsymbol{x}_i, Y_i) = \Pr(\delta_i = 1|\boldsymbol{x}_i, Y_i)$ for $i = 1, 2, \cdots, n$. Since $\pi(\boldsymbol{x}_i, Y_i)$ depends on Y_i, the above defined missingness data mechanism is nonignorable. Following Kim and Yu[5], we consider the following exponential tilting model for the propensity score function

$$\pi(\boldsymbol{x}_i, Y_i; \gamma) = \frac{1}{1 + \exp\{g(\boldsymbol{x}_i) + \gamma Y_i\}}, \tag{1}$$

where $g(\boldsymbol{x}_i)$ is an unknown function of $\boldsymbol{x}_i$, and γ is the tilting parameter to be estimated. Clearly, when $\gamma = 0$, the above defined missingness data mechanism reduces to MAR.

For the model (1), Shao and Wang[8] showed that the population may be unidentifiable when $g(\cdot)$ and γ are unknown, which is an important and challenging issue in analyzing nonignorable missing data. To overcome the difficulty, an instrumental variable method of Shao and Wang[8] is here adopted. Following Shao and Wang[8], it is assumed that $\boldsymbol{x}_i$ has the following decomposition: $\boldsymbol{x}_i = (\boldsymbol{u}_i^{\mathrm{T}}, \boldsymbol{z}_i^{\mathrm{T}})^{\mathrm{T}}$, where $\boldsymbol{z}_i$ is referred to as the instrumental variable in the sense that it is not directly associated with the propensity score but a useful covariate of Y_i. Hence, for identifiability, we define a new propensity score function via the following exponential tilting model

$$\pi_i := \pi(\boldsymbol{u}_i, Y_i; \gamma) = \frac{1}{1 + \exp\{g(\boldsymbol{u}_i) + \gamma Y_i\}}, \tag{2}$$

where $g(\cdot)$ is an unknown function of $\boldsymbol{u}_i$, and γ is the tilting parameter to be estimated consistently via the observed data.

For any fixed value of γ in its parameter space, if the propensity score (2) is correctly specified, we have $\exp\{g(\boldsymbol{u}_i)\} = E(1-\delta_i|\boldsymbol{u}_i)/E\{\delta_i \exp(\gamma Y_i)|\boldsymbol{u}_i\}$. Thus, a kernel regression estimator of $\exp\{g(\boldsymbol{u}_i)\}$ is given by

$$\exp\{\widehat{g}_\gamma(\boldsymbol{u}_i)\} = \frac{\sum_{j=1}^n (1-\delta_j) K_h(\boldsymbol{u}_j - \boldsymbol{u}_i)}{\sum_{j=1}^n \delta_j \exp(\gamma Y_j) K_h(\boldsymbol{u}_j - \boldsymbol{u}_i)}, \tag{3}$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$, and $K(\cdot)$ is a symmetric kernel function and h is the bandwidth. Based on Equation (3), the propensity score can be consistently estimated by

$$\widehat{\pi}_i(\gamma) = \widehat{\pi}(\boldsymbol{u}_i, Y_i; \gamma) = \frac{1}{1+\exp\{\widehat{g}_\gamma(\boldsymbol{u}_i)\}\exp(\gamma Y_i)}.$$

When the propensity score model (2) is correctly specified and γ is known, a consistent estimator of the distribution function of Y is given by

$$\widehat{F}_{HT}^0(y) = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}_i(\gamma)} I(Y_i \le y), \tag{4}$$

which is also referred to as the Horvitz-Thompson-type estimator in the missing data literature (e.g., see [24]), where $I(\cdot)$ is an indicative function. Also, $\widehat{F}_{HT}^0(y)$ is a consistent estimator of $F(y) = \Pr(Y \le y)$ provided that $\gamma = \gamma_0$, which is the true value of γ.

It is well known that the Horvitz-Thompson-type estimator is sensitive to the misspecification of the propensity score. To address the issue, an imputation method is adopted to construct a consistent estimator of the distribution function under nonignorable missingness data mechanism assumption. When the propensity score model (2) is correctly specified, a regression-imputation-based estimator of the distribution function is defined by

$$\widehat{F}_{RI}^0(y) = \frac{1}{n}\sum_{i=1}^n \left\{\delta_i I(Y_i \le y) + (1-\delta_i)\widehat{F}_0(y|\boldsymbol{u}_i)\right\}, \tag{5}$$

where $\widehat{F}_0(y|\boldsymbol{u}_i)$ is a consistent estimator of $F_0(y|\boldsymbol{u}_i) = \Pr(Y_i \le y|\boldsymbol{u} = \boldsymbol{u}_i, \delta_i = 0)$. Under the propensity score (2), we have

$$F_0(y|\boldsymbol{u}_i) = \frac{E\{(1-\delta_i)I(Y_i \le y)|\boldsymbol{u}_i\}}{E\{(1-\delta_i)|\boldsymbol{u}_i\}} = \frac{E\{\delta_i \exp(\gamma Y_i) I(Y_i \le y)|\boldsymbol{u}_i\}}{E\{\delta_i \exp(\gamma Y_i)|\boldsymbol{u}_i\}}.$$

Thus, a nonparametric kernel estimator of $F_0(y|\boldsymbol{u}_i)$ is given by

$$\widehat{F}_0(y|\boldsymbol{u}_i) := \widehat{F}_0(y|\boldsymbol{u}_i;\gamma) = \frac{\sum_{j=1}^n \delta_j \exp(\gamma Y_j) I(Y_j \le y) K_l(\boldsymbol{u}_j - \boldsymbol{u}_i)}{\sum_{j=1}^n \delta_j \exp(\gamma Y_j) K_l(\boldsymbol{u}_j - \boldsymbol{u}_i)}, \tag{6}$$

where $K_l(\cdot) = l^{-1}K(\cdot/l)$, and $K(\cdot)$ is a symmetric kernel function and l is the bandwidth. Note that the bandwidth l may be different from the bandwidth h defined in Equation (3). Using the standard kernel regression theory, it can be shown that $\Pr\{\lim_{n\to\infty} \widehat{F}_0(y|\boldsymbol{u}_i)\} = F_0(y|\boldsymbol{u}_i)$, which indicates that $\widehat{F}_0(y|\boldsymbol{u}_i)$ is a consistent estimator of $F_0(y|\boldsymbol{u}_i)$. Further, by the law of the iterated expectations, we have $E\{\delta_i I(Y_i \le y)) + (1-\delta_i)F_0(y|\boldsymbol{u}_i)\} = F(y)$, which shows that the estimator $\widehat{F}^0_{RI}(y)$ given in Equation (5) is a consistent estimator of $F(y)$.

The augmented inverse probability weighted (AIPW) estimation under the MAR assumption of missingness data mechanism has been studied well over the past decades (e.g., see [25–28]). A prominent property of the AIPW estimation is that it can reduce the bias of estimator (e.g., see [21]). To this end, under nonignorable missingness data mechanism assumption, we consider the following AIPW estimator of $F(y)$:

$$\widehat{F}^0_{IP}(y) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i}{\widehat{\pi}_i(\gamma)}I(Y_i \le y) + \left(1 - \frac{\delta_i}{\widehat{\pi}_i(\gamma)}\right)\widehat{F}_0(y|\boldsymbol{u}_i)\right\}, \tag{7}$$

where $\widehat{F}_0(y|\boldsymbol{u}_i)$ is given in Equation (6). Following the similar argument as $\widehat{F}^0_{RI}(y)$, it is easily shown that $\widehat{F}^0_{IP}(y)$ is a consistent estimator of $F(y)$.

2.2 The Estimation of Tilting Parameter

In many applications, the tilting parameter γ is usually unknown, and has to be estimated consistently from the sample. To this end, following Shao and Wang[8], we consider the following calibration condition:

$$\varphi_i = \varphi_i(\gamma) = \varphi(\boldsymbol{u}_i, \boldsymbol{z}_i, Y_i; \gamma) = \left(\frac{\delta_i}{\pi\left(\boldsymbol{u}_i, Y_i; \gamma\right)} - 1\right)d(\boldsymbol{z}_i) \quad \text{for} \quad i = 1, 2, \cdots, n, \tag{8}$$

where $d(\boldsymbol{z}_i)$ is an arbitrary user-specified vector with dimension $\nu \ge 2$. If $\pi(\boldsymbol{u}_i, Y_i; \gamma)$ is correctly specified, we have $E\{\varphi(\boldsymbol{u}_i, \boldsymbol{z}_i, Y_i; \gamma)\} = 0$. But, it is quite challenging to select the optimal vector $d(\boldsymbol{z}_i)$. Some methods have been developed to solve the problem in recent years. For example, see Chang and Kott[29]. However, there are considerable questions to be studied on the selection of the optimal vector $d(\boldsymbol{z}_i)$ under Equation (2).

The GMM and EL methods are two powerful tools to deal with over-identified system. Moreover, the GMM estimator has the well-established asymptotic properties and its computation is easily implemented, thus Shao and Wang[8] presented a two-step GMM procedure to estimate the tilting parameter γ when the propensity score model is correctly specified. Also, as discussed in Newey and Smith[30], the asymptotic bias of the EL estimator does not increase with the number of moment restrictions, and the bias corrected EL estimator has higher order effect relative to other bias corrected estimators. In addition, the EL method can easily incorporate the auxiliary information from the sample into the considered moment restrictions, which can help to improve the estimation efficiency. However, when the sample size is small or the dimension of estimating equations is high, the ordinary EL[23] estimator may not be the solutions to estimating equations. To overcome the difficulty, Chen, et al.[22] proposed an AEL method via adding a few artificial observations to the original data set, which guarantees the existence of the solutions under some mild conditions, and retains some good properties of the ordinary EL estimator. Motivated by the above idea, the AEL method is employed to evaluate the consistent estimation of the tilting parameter.

We define $\overline{\varphi}_n = \overline{\varphi}_n(\gamma) = n^{-1}\sum_{i=1}^n \varphi_i(\gamma)$ for any given γ. For some positive constant a_n, we define $\varphi_{n+1} = \varphi_{n+1}(\gamma) = -a_n\overline{\varphi}_n$. Let $p_1, p_2, \cdots, p_n, p_{n+1}$ be the probability weights allocated to $\varphi_1, \varphi_2, \cdots, \varphi_n, \varphi_{n+1}$. The adjusted profile EL function is given by

$$L(\gamma) \quad = \quad \sup\left\{\prod_{i=1}^{n+1} p_i : p_i \geq 0, \sum_{i=1}^{n+1} p_i = 1, \sum_{i=1}^{n+1} p_i\varphi_i = 0\right\}.$$

Following the similar arguments as Qin and Lawless[31] and Chen, et al.[22], the maximum EL estimator $\widehat{\gamma}_{el}$ of γ can be obtained by

$$\widehat{\gamma}_{el} = \arg\max_{\gamma} \ell(\gamma) = \arg\max_{\gamma}\left\{-2\sum_{i=1}^{n+1}\log\{1+\lambda(\gamma)^{\mathrm{T}}\varphi_i\}\right\}, \tag{9}$$

where $\lambda(\gamma)$ satisfies $\sum_{i=1}^{n+1}\varphi_i/\{1+\lambda^{\mathrm{T}}(\gamma)\varphi_i\} = 0$. Because $\varphi_i = \varphi_i(\gamma)$ contains unknown function $g(\boldsymbol{u}_i)$, the resultant AEL ratio statistic $\ell(\gamma)$ cannot be directly used to make statistical inference on γ. To address the issue, we replace $\exp\{g(\boldsymbol{u}_i)\}$ in $\varphi_i(\gamma)$ by its kernel regression estimator $\exp\{\widehat{g}(\boldsymbol{u}_i)\}$ given in Equation (3).

Remark 2.1 As discussed in [22], the AEL method retains the first order property of the ordinary EL estimator provided that $a_n = o_p(n^{2/3})$, and works well for $a_n = \max\{1, \log(n)/2\}$ in many applications. Liu and Chen[32] further studied the optimal selection of a_n in implementing the AEL method. Here, we simply adopt $a_n = \max\{1, \log(n)/2\}$[22] and find that it works well in simulation studies.

Substituting $\widehat{\gamma}_{el}$ into Equations (4), (5) and (7), we obtain three estimators of $F(y)$ (denoted by $\widehat{F}_{HT}(y)$, $\widehat{F}_{RI}(y)$ and $\widehat{F}_{IP}(y)$) at a fixed point y of interest. Note that $\widehat{F}_{\varsigma}(y)$ is a genuine distribution function for $\varsigma = HT, RI$ or IP because it is nondecreasing, right-continuous, and $\widehat{F}_{\varsigma}(\infty) = 1$ and $\widehat{F}_{\varsigma}(-\infty) = 0$. Thus, we only require checking that the indicative function $I(Y \leq y)$ shares the above three properties.

2.3 Dimension Reduction

The above proposed estimation procedure requires the estimation of the propensity score $\pi_i(\gamma)$ and the imputed conditional distribution function $F_0(y|\boldsymbol{u}_i)$. To this end, we propose the kernel estimator of $\pi_i(\gamma)$ and $F_0(y|\boldsymbol{u}_i)$ given in Equations (3) and (6), respectively. However, in some applications, when the dimension of $\boldsymbol{u}_i$ is high, the kernel-based estimator may suffer from the well-known curse of dimensionality. To this end, Shao and Wang[8] suggested a non-parametric method (i.e., the generalized additive model or the sufficient reduction technique), which can be used to obtain a consistent estimator of $\exp\{g(\boldsymbol{u}_i)\}$ when the dimension of $\boldsymbol{u}_i$ is high. For the estimation of $F_0(y|\boldsymbol{u}_i)$ under the high-dimension assumption of $\boldsymbol{u}_i$, motivated by [20, 33], we propose the following dimension reduction technique.

Let $\mathcal{S}$ be a continuous function from $\mathcal{R}^{d_{\boldsymbol{u}}}$ to $\mathcal{R}$ with $d_{\boldsymbol{u}}$ being the dimension of $\boldsymbol{u}$, such that $\mathcal{S} = \mathcal{S}(\boldsymbol{u})$ is univariate and $F_0(y|\boldsymbol{u}_i) = E\{I(Y_i \leq y)|\boldsymbol{u} = \boldsymbol{u}_i, \delta_i = 0\} = E\{I(Y_i \leq y)|\mathcal{S}(\boldsymbol{u}_i), \delta_i = 0\} := F_0(y|\mathcal{S}_i)$ with $\mathcal{S}_i = \mathcal{S}(\boldsymbol{u}_i)$. Thus, the kernel-assisted estimator of $F_0(y|\mathcal{S}_i)$ is given by

$$\widehat{F}_0(y|\mathcal{S}_i;\gamma) = \frac{\sum_{j=1}^n \delta_j \exp(\gamma Y_j) I(Y_j \leq y) K_l(\mathcal{S}_j - \mathcal{S}_i)}{\sum_{j=1}^n \delta_j \exp(\gamma Y_j) K_l(\mathcal{S}_j - \mathcal{S}_i)},$$

which is structurally identical to $\widehat{F}_0(y|\boldsymbol{u}_i)$ except that $\boldsymbol{u}_i$ is replaced by $\mathcal{S}_i$. Given the estimator $\widehat{\gamma}_{el}$, we obtain the following two dimension reduction estimators of $F(y)$:

$$\widehat{F}_{RI}(y;\mathcal{S}_i)=\frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i I(Y_i\le y)+(1-\delta_i)\widehat{F}_0(y|\mathcal{S}_i;\widehat{\gamma}_{el})\right\},$$

$$\widehat{F}_{IP}(y;\mathcal{S}_i)=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i}{\widehat{\pi}_i(\widehat{\gamma}_{el})}I(Y_i\le y)+\left(1-\frac{\delta_i}{\widehat{\pi}_i(\widehat{\gamma}_{el})}\right)\widehat{F}_0(y|\mathcal{S}_i;\widehat{\gamma}_{el})\right\}.$$

In applications, we assume that the working index $\mathcal{S}=\mathcal{S}(\boldsymbol{u};\boldsymbol{\beta})$ contains an unknown parameter vector $\boldsymbol{\beta}$. Given an estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, the dimension reduction estimators $\widehat{F}_{RI}(y;\mathcal{S}_i)$ and $\widehat{F}_{IP}(y;\mathcal{S}_i)$ are constructed by $\widehat{F}_{RI}(y;\widehat{\mathcal{S}}_i)$ and $\widehat{F}_{IP}(y;\widehat{\mathcal{S}}_i)$ with $\widehat{\mathcal{S}}_i=\mathcal{S}(\boldsymbol{u}_i;\widehat{\boldsymbol{\beta}})$, respectively. Following the same arguments as Hu, et al.[20], we can show that $\widehat{F}_{RI}(y;\widehat{\mathcal{S}}_i)$ and $\widehat{F}_{IP}(y;\widehat{\mathcal{S}}_i)$ are asymptotically equivalent to $\widehat{F}_{RI}(y)$ and $\widehat{F}_{IP}(y)$ provided that $\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}=O_p(n^{-1/2})$, respectively.

3 Asymptotic Properties

In this section, we establish asymptotic properties of the proposed estimators of $F(y)$. We first present the large sample property of the tilting parameter estimator $\widehat{\gamma}_{el}$.

Theorem 3.1 *Suppose that condition A given in the Appendix holds. Let γ_0 be the true value of γ, $A=E\{\varphi_i(\gamma_0)\varphi_i^{\rm T}(\gamma_0)\}$ and $B=E\{\partial\varphi_i(\gamma)/\partial\gamma|_{\gamma=\gamma_0}\}=E\{\partial\varphi_i(\gamma_0)/\partial\gamma\}$. Then, we have $\widehat{\gamma}_{el}\xrightarrow{\mathcal{P}}\gamma_0$ and $\sqrt{n}(\widehat{\gamma}_{el}-\gamma_0)\xrightarrow{\mathcal{L}}N(0,(B^{\rm T}A^{-1}B)^{-1})$, where $\xrightarrow{\mathcal{P}}$ and $\xrightarrow{\mathcal{L}}$ stand for convergence in probability and convergence in distribution, respectively.*

Theorem 3.1 establishes the consistency and the asymptotic normality of the MELE $\widehat{\gamma}_{el}$. It follows from the proof of Theorem 3.1 that the asymptotic expansion of $\widehat{\gamma}_{el}$ has the form $\sqrt{n}(\widehat{\gamma}_{el}-\gamma_0)=n^{-1/2}\sum_{i=1}^{n}\psi_i(\gamma_0)+o_p(1)$, where $\psi_i(\gamma_0)=(B^{\rm T}A^{-1}B)^{-1}B^{\rm T}A^{-1}\varphi(\boldsymbol{u}_i,\boldsymbol{z}_i,Y_i;\gamma_0)$ is the influence function.

When the propensity score model (2) is correctly specified, we have

Theorem 3.2 *Suppose that the conditions given in the Appendix hold and the propensity score model* (2) *is correctly specified. For any given y, we obtain*

$$\sqrt{n}\{\widehat{F}_{\varsigma}(y)-F(y)\}\xrightarrow{\mathcal{L}}N(0,\sigma^2(y))\quad\text{for}\quad\varsigma=HT,RI,IP,$$

where $\sigma^2(y)=\mathrm{var}(\eta_i)$, $\eta_i=\{\frac{\delta_i}{\pi(\boldsymbol{u}_i,Y_i;\gamma_0)}\{I(Y_i\le y)-F_0(y|\boldsymbol{u}_i)\}+F_0(y|\boldsymbol{u}_i)-F(y)+\psi_i(\gamma_0)H\}$ and $H=E\{(1-\delta)\{Y-m_Y(\boldsymbol{u})\}\{I(Y\le y)-F_0(y|\boldsymbol{u})\}\}$ with $m_Y(\boldsymbol{u})=E(Y|\boldsymbol{u},\delta=0)$.

Remark 3.3 Theorem 3.2 shows that the proposed three estimators of $F(y)$ are $\sqrt{n}-$consistent, and asymptotically distributed as the normal distribution with the same mean and variance, which indicates that three estimators are asymptotically equivalent that is similar to the one given in [33, 34]. But they may have different higher order properties.

For $0<\tau<1$, let ξ_τ be the τth quantile, and $\widehat{\xi}_{\tau,\varsigma}$ be the τth sample quantile for $\varsigma=HT,IP$ and RI. Thus, $\xi_\tau=F^{-1}(\tau)=\inf\{y:F^{-1}(y)\ge\tau\}$ and $\widehat{\xi}_{\tau,\varsigma}=\widehat{F}_{\varsigma}^{-1}(\tau)=\inf\{y:\widehat{F}_{\varsigma}^{-1}(y)\ge\tau\}$ for $\varsigma=HT,IP$ and RI. The following theorem presents the asymptotic property of the proposed estimator $\widehat{\xi}_{\tau,\varsigma}$ with continuous distribution.

Theorem 3.4 *Suppose that the conditions given in the Appendix hold, and the probability density function* $F^0(y)$ *of* $F(y)$ *is bounded away from* 0. *If the estimator of* γ *is given in Equation* (9), *we have*

$$\sqrt{n}(\widehat{\xi}_{\tau,\varsigma} - \xi_\tau) \xrightarrow{\mathcal{L}} N(0, \sigma^2(\xi_\tau)/(F^0(\xi_\tau))^2) \quad for \quad \varsigma = HT, IP, RI,$$

where $\sigma^2(\cdot)$ *is the variance function given in Theorem* 3.2.

4 Estimation of Asymptotic Variance

Theorems 3.2 and 3.4 show that the asymptotic variances of the proposed estimators have the complicated expressions. It is quite difficult to estimate directly them. Many approaches have been developed to overcome the difficulty. For example, Shao and Sitter[35] presented a Bootstrap sampling method for the variance estimation for incomplete data; Wang, et al.[34] gave a jackknife variance estimator when some responses are missing at random; Wang and Chen[36] used a Bootstrap procedure to estimate asymptotic variance in which the bootstrap data set is imputed by the same way as the original data set.

The jackknife procedure is here adopted to approximate the asymptotic variances of $\widehat{F}_{HT}(y)$, $\widehat{F}_{RI}(y)$ and $\widehat{F}_{IP}(y)$. Let $\widehat{F}_\varsigma^{(-i)}(y)$ be the reduced form of $\widehat{F}_\varsigma(y)$ based on the reduced sample data $\{(Y_j, \boldsymbol{x}_j, \delta_j)\}_{j\neq i}$ for $i = 1, 2, \cdots, n$ and $\varsigma = HT, RI$ or IP. Let J_{ni}^ς be the jackknife pseudo-values, i.e., $J_{ni}^\varsigma = n\widehat{F}_\varsigma(y) - (n-1)\widehat{F}_\varsigma^{(-i)}(y)$ for $i = 1, 2, \cdots, n$. Thus, the jackknife variance estimator of $\sigma_\varsigma^2(y)$ is defined as $\widehat{\sigma}_\varsigma^2(y) = \frac{1}{n}\sum_{i=1}^n (J_{ni}^\varsigma - \overline{J}_n^\varsigma)^2$, where $\overline{J}_n^\varsigma = n^{-1}\sum_{i=1}^n J_{ni}^\varsigma$. The $100(1-\alpha)\%$ normal-approximation-based conference interval for $F(y)$ at a fixed point y of interest is given by $\widehat{F}_\varsigma(y) \pm u_{1-\alpha/2}\sqrt{\widehat{\sigma}_\varsigma^2(y)/n}$, where $u_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. For the variance estimation of $\widehat{\xi}_{\tau,\varsigma}$, we require estimating $F^0(y)$ consistently from the observed sample, which can be conducted by using the Horvtiz-Thompson-type estimation given in [6]. In particular, the kernel-assisted estimator of $F^0(y)$ is given by $\widehat{F}_{HTt}^0(y) = n^{-1}\sum_{i=1}^n \delta_i K_h(y - Y_i)/\widehat{\pi}_i(\widehat{\gamma}_{el})$. Consequently, the consistent variance estimator for $\widehat{\xi}_{\tau,\varsigma}$ is given by $n^{-1}\widehat{\sigma}^2(\widehat{\xi}_{\tau,\varsigma})\{\widehat{F}_{HTt}^0(\widehat{\xi}_{\tau,\varsigma})\}^{-2}$ for $\varsigma = HT, RI$ and IP, respectively.

5 Simulation Study

Several simulation studies were conducted to investigate the finite-sample performance of the proposed methods. In the first simulation study, the data set $\{y_i :, i = 1, 2, \cdots, n\}$ was generated from the following model: $y_i = u_i + z_i + \varepsilon_i$ for $i = 1, 2, \cdots, n$, where z_i was generated from the standard normal distribution $N(0, 1)$; given z_i, u_i was generated form the normal distribution $N(z_i, 1)$; ε_i was independently simulated from $N(0, 1)$. The missing indicator δ_i for y_i was generated from the Bernoulli distribution with probability π_i, which was specified by

M1: $\pi_i = 1/\{1 + \exp(\beta_0 + \beta_1 u_i)\}$, where $(\beta_0, \beta_1) = (-0.85, 0.15)$;

M2: $\pi_i = 1/\{1 + \exp(\beta_0 + \beta_1 u_i + \gamma y_i)\}$, where $(\beta_0, \beta_1, \gamma) = (-1.05, -0.55, -0.1)$;

M3: $\pi_i = 1/\{1 + \exp(\beta_0 + \beta_1 \sin(u_i) + \gamma y_i)\}$, where $(\beta_0, \beta_1, \gamma) = (-0.85, -0.55, 0.15)$;

M4: $\pi_i = 1/\{1 + \exp(\beta_0 + \beta_1 u_i + \gamma \sin(y_i))\}$, where $(\beta_0, \beta_1, \gamma) = (-0.95, -0.45, -0.15)$;

M5: $\pi_i = 1/\{1 + \exp(\beta_0 + \beta_1 u_i + \gamma y_i u_i)\}$, where $(\beta_0, \beta_1, \gamma) = (-0.45, 0.15, -0.15)$;

M6: $\pi_i = 1 - \exp\{-\exp(\beta_0 + \beta_1 u_i + \gamma y_i)\}$, where $(\beta_0, \beta_1, \gamma) = (0.25, 0.35, 0.15)$.

The propensity score M1 was MAR, while M2–M6 represented nonignorable missing data mechanism. Moreover, M1–M3 satisfied the propensity score model given in Equation (2), while M4–M6 did not satisfy the presented propensity score model and were designed to investigate the robustness of the proposed estimators under misspecified propensity score assumption. The average missing proportion was about 30%.

Based on the above generated data set, the previous developed estimation procedure was used to evaluate the estimation of $F(y)$ at two points of y (e.g., $y = 0.5$ and 2.5). To compute the nonparametric kernel-assisted estimates given in Equations (3) and (6), we took the Gaussian kernel function to be $K(u) = \exp(-u^2/2)/(2\pi)^{1/2}$ and selected the bandwidth to be $l = h = \widehat{\sigma}_u n^{-1/3}$, where $\widehat{\sigma}_u$ is the standard deviation of observations $\{u_i, i = 1, 2, \cdots, n\}$. To estimate the tilting parameter γ via the proposed AEL method, we considered the following calibration condition: $\varphi_i(\gamma) = \{\delta_i/\widehat{\pi}_i(\gamma) - 1\}(1, z_i, z_i^2)^{\mathrm{T}}$. Results for 1,000 replications with $n = 60$ and 120 were presented in Table 1, where "Bias" was the absolute difference between the true value and the mean of the 1,000 estimates for $F(y)$, "RMSE" was the root mean square error between the true value and the estimates based on 1,000 replications, "SD" was the standard deviation of the estimates based on 1,000 replications, and "CP" was the empirical coverage probability with the nominal level 95%. Since the values of SD and RMSE were close to each other, to save space, we only reported the values of RMSE. In addition, we calculated the 95% pointwise confidence intervals of $F(y)$ for M2 with $n = 150$ based on the Wald statistic in which the asymptotic variance was estimated by the proposed jackknife method. Because $\widehat{F}_{RI}(y)$ and $\widehat{F}_{IP}(y)$ had almost the same pointwise intervals, we only presented the 95% pointwise confidence intervals of $\widehat{F}_{IP}(y)$ in Figure 1. Also, we computed three estimations $\widehat{\xi}_{\tau,\zeta}$ of quantiles ξ_τ for $\zeta = HT, RI$ and IP at $\tau = 0.5$, 0.75 and 0.9 with $n = 200$. Results were given in Table 2.

Examination of Table 1 indicated that (i) the proposed three estimators had the similar standard deviation under all the considered cases, which was consistent with that given in Theorem 3.2; (ii) the inverse probability weighted estimator $\widehat{F}_{HT}(y)$ produced relatively larger bias than that obtained by the imputation procedures (i.e., $\widehat{F}_{RI}(y)$ and $\widehat{F}_{IP}(y)$) even if the propensity score was correctly specified, which implied that the proposed imputation methods can reduce the bias of estimator; (iii) the proposed two imputation estimators performed well under all the considered cases; (iv) as n increased, the SD and RMSE values of three estimators decreased as expected; (v) the empirical CPs based on the proposed three estimators were close to the nominal level 95% even if sample size was small or the propensity score was misspecified.

Table 1 Performance of estimators for $F(y)$ in the first simulation study

n	y	case	$\widehat{F}_{HT}$			$\widehat{F}_{RI}$			$\widehat{F}_{IP}$		
			Bias	RMSE	CP	Bias	RMSE	CP	Bias	RMSE	CP
60	0.5	M1	0.017	0.072	0.950	0.005	0.073	0.958	0.007	0.071	0.952
		M2	0.039	0.081	0.930	0.008	0.071	0.956	0.005	0.069	0.962
		M3	0.022	0.073	0.956	0.013	0.072	0.958	0.008	0.070	0.962
		M4	0.032	0.077	0.938	0.009	0.071	0.952	0.007	0.069	0.956
		M5	0.015	0.070	0.950	0.005	0.073	0.956	0.007	0.071	0.952
		M6	0.041	0.082	0.924	0.020	0.074	0.952	0.016	0.071	0.950
	2.5	M1	0.017	0.519	0.960	0.002	0.050	0.918	0.000	0.050	0.930
		M2	0.034	0.062	0.948	0.000	0.046	0.942	0.001	0.046	0.948
		M3	0.018	0.053	0.964	0.002	0.050	0.926	0.003	0.049	0.922
		M4	0.024	0.057	0.958	0.003	0.046	0.936	0.003	0.046	0.934
		M5	0.021	0.055	0.958	0.007	0.051	0.934	0.007	0.050	0.940
		M6	0.028	0.057	0.960	0.004	0.047	0.944	0.003	0.047	0.944
120	0.5	M1	0.015	0.054	0.940	0.008	0.054	0.936	0.009	0.052	0.938
		M2	0.030	0.061	0.924	0.006	0.052	0.944	0.005	0.051	0.942
		M3	0.017	0.056	0.938	0.012	0.055	0.940	0.008	0.053	0.938
		M4	0.029	0.060	0.928	0.013	0.054	0.934	0.010	0.052	0.938
		M5	0.012	0.054	0.950	0.007	0.056	0.942	0.009	0.054	0.934
		M6	0.031	0.060	0.924	0.017	0.054	0.936	0.015	0.052	0.938
	2.5	M1	0.017	0.041	0.938	0.005	0.039	0.948	0.007	0.039	0.938
		M2	0.032	0.049	0.922	0.007	0.034	0.962	0.007	0.034	0.956
		M3	0.018	0.043	0.938	0.006	0.038	0.930	0.006	0.038	0.936
		M4	0.025	0.045	0.934	0.006	0.036	0.944	0.005	0.035	0.936
		M5	0.022	0.043	0.936	0.014	0.041	0.946	0.015	0.041	0.940
		M6	0.026	0.044	0.938	0.011	0.036	0.950	0.011	0.036	0.948

Inspection of Figure 1 implied that the confidence intervals based on $\widehat{F}_{IP}(y)$ were slightly narrower than those based on $\widehat{F}_{HT}(y)$ when response of interest was relatively large, which was not surprised in a finite sample because the estimation of the distribution function required more information on sample size when the response point became relatively large and the estimator $\widehat{F}_{HT}(y)$ did not fully extract the information contained in the observed sample. Also, the estimated curve of $F(y)$ based on $\widehat{F}_{IP}(y)$ was closer to the true curve of $F(y)$ than that based on $\widehat{F}_{HT}(y)$. Examination of Table 2 implied that the proposed quantile estimators yielded the similar results as those for the estimators of distribution functions, which confirmed the feasibility and validity of the proposed methodologies.

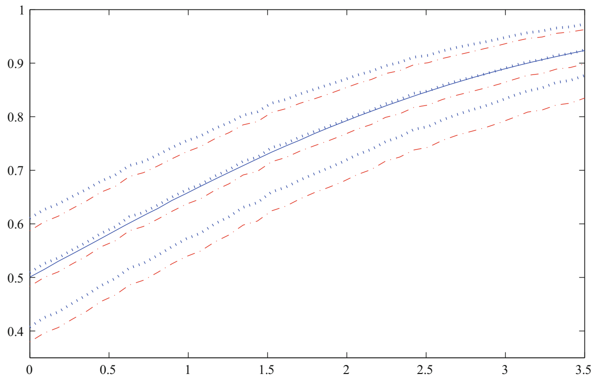

Figure 1 Normal-approximation-based 95% pointwise confidence interval of $F(y)$ with $n = 150$

Notes: In Figure 1, solid curve: The true value of $F(y)$; dotted curve: The estimated curve of $F(y)$ via $\widehat{F}_{IP}(y)$ against its normal-approximation-based confidence interval; dot-dashed curve: The estimated curve of $F(y)$ via $\widehat{F}_{HT}(y)$ against its normal-approximation-based confidence interval in the first simulation study.

Table 2 Performance of estimators for quantiles in the first simulation study

τ	case	$\widehat{\xi}_{\tau,HT}$			$\widehat{\xi}_{\tau,RI}$			$\widehat{\xi}_{\tau,IP}$		
		Bias	RMSE	CP	Bias	RMSE	CP	Bias	RMSE	CP
0.5	M1	0.009	0.245	0.930	0.006	0.242	0.940	0.000	0.241	0.936
	M2	0.135	0.287	0.898	0.019	0.243	0.942	0.003	0.241	0.940
	M3	0.035	0.255	0.926	0.028	0.250	0.936	0.004	0.247	0.938
	M4	0.069	0.258	0.930	0.002	0.242	0.946	0.006	0.242	0.946
	M5	0.055	0.263	0.926	0.068	0.265	0.920	0.051	0.260	0.926
	M6	0.061	0.260	0.928	0.012	0.261	0.931	0.016	0.251	0.929
0.75	M1	0.024	0.258	0.952	0.017	0.255	0.958	0.002	0.251	0.962
	M2	0.161	0.314	0.904	0.005	0.246	0.952	0.002	0.245	0.944
	M3	0.049	0.258	0.954	0.013	0.249	0.962	0.003	0.247	0.960
	M4	0.066	0.260	0.936	0.018	0.238	0.954	0.018	0.240	0.948
	M5	0.028	0.267	0.952	0.005	0.263	0.954	0.011	0.261	0.958
	M6	0.205	0.325	0.870	0.101	0.262	0.922	0.084	0.255	0.934
0.9	M1	0.063	0.360	0.932	0.024	0.340	0.926	0.005	0.340	0.928
	M2	0.202	0.389	0.898	0.009	0.309	0.928	0.007	0.307	0.930
	M3	0.077	0.357	0.932	0.016	0.323	0.946	0.007	0.330	0.940
	M4	0.122	0.376	0.924	0.053	0.314	0.910	0.048	0.314	0.912
	M5	0.082	0.344	0.932	0.027	0.338	0.924	0.039	0.335	0.922
	M6	0.269	0.432	0.874	0.055	0.310	0.936	0.051	0.309	0.934

To investigate the sensitivity of the estimators to missingness data mechanism assumption, we conducted the second simulation study. In this simulation study, 1,000 data sets were generated as done in the first simulation study with the missingness data mechanism specified by model M2 for $\gamma = 0.0, 0.1, 0.3, 0.5$ and 0.7. Note that the tilting parameter γ was a measure of the amount of the departure from the MAR assumption[5]. Let $\widehat{F}_{HT}^{\rm MAR}(y)$, $\widehat{F}_{RI}^{\rm MAR}(y)$ (Cheng and Chu[18]) and $\widehat{F}_{IP}^{\rm MAR}(y)$ be the corresponding versions of $\widehat{F}_{HT}(y)$, $\widehat{F}_{RI}(y)$ and $\widehat{F}_{IP}(y)$ under the MAR assumption, respectively, where the working propensity score was estimated by $\widetilde{\pi}(u_i) = \sum_{j=1}^n \delta_j K_h(u_j - u_i)/\sum_{j=1}^n K_h(u_j - u_i)$. Let $\widehat{\xi}_{\tau,HT}^{\rm MAR}$, $\widehat{\xi}_{\tau,RI}^{\rm MAR}$ and $\widehat{\xi}_{\tau,IP}^{\rm MAR}$ be the corresponding τth quantile estimators based on $\widehat{F}_{HT}^{\rm MAR}(y)$, $\widehat{F}_{RI}^{\rm MAR}(y)$ and $\widehat{F}_{IP}^{\rm MAR}(y)$, respectively. Results for $F(y)$ with $n = 60$ and $y = 2.5$, and for ξ_τ with $n = 200$ and $\tau = 0.75$ were given in Table 3. The estimated propensity scores for different values of γ were given in Figure 2.

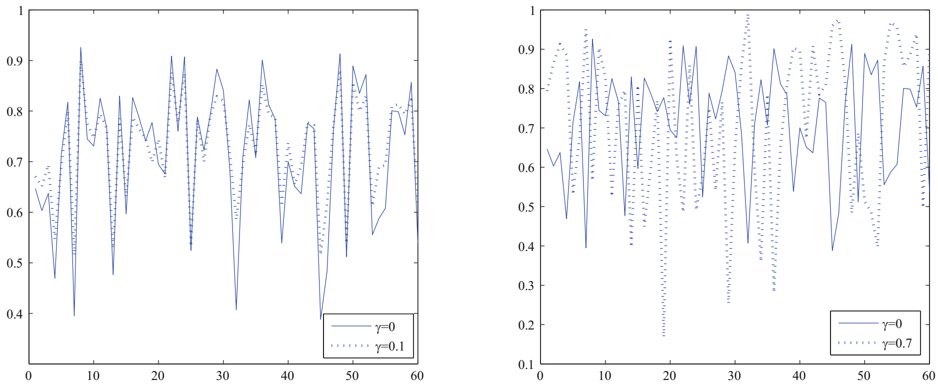

Figure 2 Plots of the estimated propensity scores with (i) $\gamma = 0.0$ and 0.1 (left panel) and (ii) $\gamma = 0.0$ and 0.7 (right panel) in the second simulation study

From Table 3, we observed that the proposed estimators worked well under the correct specification of missingness data mechanism in terms of Bias, RMSE and CPs. However, the proposed estimators behaved poor when missingness data mechanism was misspecified. That is, the proposed estimation procedure implemented under the MAR assumption yielded misleading results when the true missingness data mechanism was nonignorable.

To investigate the performance of the proposed methods in multivariate case, we conducted the third simulation study. In this simulation study, the data set $\{y_i : i = 1, 2, \cdots, n\}$ was generated from the model: $y_i = \exp(u_{i1} + u_{i2} + z_i) + \varepsilon_i$ for $i = 1, 2, \cdots, 200$, where u_{i1} and u_{i2} were independently sampled from the uniform distribution $U(0, 1)$, z_i and ε_i were independently simulated from the standard normal distribution $N(0, 1)$. We assumed that u_{i1}, u_{i2} and z_i were completely observed, but y_i's may subject to missingness. The missing indicator δ_i for y_i was generated from the Bernoulli distribution with probability π_i, which was specified by

M1: $\pi_i = 1/\{1 + \exp(-0.5 - 0.45u_{i1} - 0.35u_{i2})\}$;

M2: $\pi_i = 1/\{1 + \exp(-0.1 - 0.35u_{i1} - 0.35u_{i2} + \gamma y_i)\}$ with $\gamma = -0.1$;

M3: $\pi_i = 1/\{1 + \exp(-0.5 - 0.35u_{i1} - 0.35u_{i2} - 0.1 \sin y_i)\}$;

Table 3 Performance of estimators for $F(y)$ and ξ_τ in the second simulation study

γ	$\widehat{F}_{HT}$			$\widehat{F}_{HT}^{\mathrm{MAR}}$			$\widehat{F}_{RI}$			$\widehat{F}_{RI}^{\mathrm{MAR}}$			$\widehat{F}_{IP}$			$\widehat{F}_{IP}^{\mathrm{MAR}}$		
	Bias	RMSE	CP	Bias	RMSE	CP	Bias	RMSE	CP	Bias	RMSE	CP	Bias	RMSE	CP	Bias	RMSE	CP
0	0.012	0.051	0.945	0.013	0.051	0.945	0.007	0.049	0.940	0.008	0.048	0.941	0.006	0.048	0.935	0.007	0.048	0.944
0.1	0.011	0.050	0.947	0.008	0.048	0.945	0.007	0.049	0.940	0.004	0.047	0.940	0.006	0.048	0.937	0.004	0.047	0.941
0.3	0.007	0.051	0.938	0.004	0.048	0.929	0.007	0.052	0.936	0.005	0.048	0.920	0.005	0.050	0.935	0.004	0.048	0.926
0.5	0.000	0.053	0.936	0.021	0.052	0.895	0.006	0.056	0.943	0.021	0.052	0.890	0.001	0.053	0.936	0.018	0.051	0.905
0.7	0.013	0.060	0.904	0.044	0.066	0.779	0.000	0.063	0.920	0.042	0.064	0.803	0.007	0.061	0.910	0.037	0.063	0.815

γ	$\widehat{\xi}_{\tau,HT}$			$\widehat{\xi}_{\tau,HT}^{\mathrm{MAR}}$			$\widehat{\xi}_{\tau,RI}$			$\widehat{\xi}_{\tau,RI}^{MAR}$			$\widehat{\xi}_{\tau,IP}$			$\widehat{\xi}_{\tau,IP}^{MAR}$		
	Bias	RMSE	CP	Bias	RMSE	CP	Bias	RMSE	CP	Bias	RMSE	CP	Bias	RMSE	CP	Bias	RMSE	CP
0	0.124	0.288	0.929	0.123	0.291	0.922	0.006	0.246	0.943	0.004	0.246	0.943	0.002	0.246	0.941	0.002	0.247	0.939
0.1	0.102	0.274	0.933	0.054	0.261	0.936	0.008	0.245	0.944	0.027	0.248	0.936	0.003	0.245	0.939	0.031	0.250	0.935
0.3	0.079	0.272	0.936	0.074	0.267	0.927	0.014	0.255	0.945	0.116	0.276	0.909	0.006	0.254	0.938	0.112	0.275	0.910
0.5	0.067	0.281	0.935	0.208	0.332	0.840	0.019	0.266	0.943	0.243	0.352	0.809	0.012	0.261	0.940	0.224	0.340	0.820
0.7	0.059	0.277	0.933	0.348	0.430	0.677	0.017	0.267	0.942	0.389	0.460	0.631	0.026	0.266	0.944	0.361	0.440	0.673

M4: $\pi_i = 0.5I(y_i \leq 3.5) + I(y_i > 3.5)$;

M5: $\pi_i = \Phi(-0.05 + 0.85u_{i1} - 0.35u_{i2} - 0.1\sin y_i)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

The missingness data mechanism models M1 and M2 satisfied the propensity score model given in Equation (2), while the models M3–M5 did not satisfy the propensity score model given in Equation (2) and were designed to investigate the robustness of the proposed estimators to the misspecified missingness data mechanisms. The average missing proportion was about 30%.

The preceding developed method was used to compute the estimation of $F(y)$ at two fixed points, i.e., $y = 0.5$ and 13.5. In evaluating the nonparametric estimations of $\pi_i(\gamma)$ and $F_0(y|\mathbf{u}_i)$ via Equations (3) and (6), we took the kernel function to be $K(u_1, u_2) = K(u_1)K(u_2)$, and set the bandwidth to be $\ell_1 = h_1 = \widehat{\sigma}_{u1}n^{-1/3}$ and $\ell_2 = h_2 = \widehat{\sigma}_{u2}n^{-1/3}$, respectively, where $K(u) = \exp(-u^2/2)/(2\pi)^{1/2}$ and $\widehat{\sigma}_{uj}$ was the standard deviation of observations $\{u_{ij} : i = 1, 2, \cdots, n\}$ for $j = 1$ and 2. The tilting parameter γ was estimated via the same calibration condition as given in the second simulation study. Results for 1,000 replications with $n = 200$ were reported in Table 4. Also, we computed the corresponding estimations of $F(y)$ at $y = 0.5$ via the proposed estimators $\widehat{F}_{HT}^{\mathrm{MAR}}(y)$, $\widehat{F}_{RI}^{\mathrm{MAR}}(y)$, $\widehat{F}_{IP}^{\mathrm{MAR}}(y)$, $\widehat{F}_{HT}(y)$, $\widehat{F}_{RI}(y)$, $\widehat{F}_{IP}(y)$ based on 1,000 data sets, which were generated via the model M2 with $\gamma = 0.0, -0.05, -0.15$ and -0.25. The values of Bias, RMSE and CP for the 95% confidence level were given in Table 5.

From Table 4, we observed that when the point of interest was large (e.g., $y = 13.5$), the inverse probability weighted estimator $\widehat{F}_{HT}(y)$ yielded larger Bias and RMSE values and lower CP than those obtained by $\widehat{F}_{RI}(y)$ and $\widehat{F}_{IP}(y)$, which indicated that the proposed imputation procedure outperformed other two methods. Table 5 showed the same conclusions as observed in Table 3, hence we omitted the details.

Table 4 Performance of estimators for $F(y)$ in the third simulation study

y	case	$\widehat{F}_{HT}$			$\widehat{F}_{RI}$			$\widehat{F}_{IP}$		
		Bias	RMSE	CP	Bias	RMSE	CP	Bias	RMSE	CP
0.5	M1	0.007	0.027	0.924	0.002	0.028	0.934	0.002	0.028	0.940
	M2	0.011	0.030	0.910	0.002	0.030	0.936	0.005	0.030	0.936
	M3	0.009	0.028	0.922	0.004	0.028	0.936	0.004	0.028	0.938
	M4	0.006	0.031	0.930	0.019	0.040	0.938	0.008	0.035	0.946
	M5	0.008	0.029	0.928	0.003	0.030	0.942	0.002	0.031	0.948
13.5	M1	0.035	0.042	0.776	0.000	0.022	0.926	0.000	0.022	0.914
	M2	0.044	0.049	0.614	0.000	0.020	0.928	0.001	0.020	0.928
	M3	0.037	0.043	0.756	0.002	0.023	0.926	0.000	0.022	0.926
	M4	0.062	0.067	0.328	0.000	0.018	0.940	0.000	0.018	0.940
	M5	0.066	0.071	0.464	0.003	0.024	0.920	0.003	0.024	0.908

Table 5 Performance of estimators for $F(y)$ under different missingness data mechanism in the third simulation study

γ	$\widehat{F}_{HT}^{\mathrm{MAR}}$			$\widehat{F}_{RI}^{\mathrm{MAR}}$			$\widehat{F}_{IP}^{\mathrm{MAR}}$		
	Bias	RMSE	CP	Bias	RMSE	CP	Bias	RMSE	CP
0	0.003	0.030	0.925	0.002	0.030	0.925	0.003	0.031	0.925
−0.05	0.012	0.030	0.915	0.012	0.030	0.915	0.011	0.030	0.930
−0.15	0.022	0.034	0.870	0.022	0.034	0.880	0.022	0.033	0.870
−0.25	0.030	0.037	0.760	0.029	0.037	0.780	0.029	0.037	0.765
γ	$\widehat{F}_{HT}$			$\widehat{F}_{RI}$			$\widehat{F}_{IP}$		
	Bias	RMSE	CP	Bias	RMSE	CP	Bias	RMSE	CP
0	0.002	0.029	0.930	0.002	0.029	0.930	0.002	0.030	0.930
−0.05	0.004	0.029	0.935	0.002	0.030	0.940	0.004	0.030	0.930
−0.15	0.006	0.029	0.930	0.001	0.030	0.935	0.005	0.029	0.935
−0.25	0.010	0.028	0.940	0.002	0.028	0.945	0.008	0.028	0.950

6 Discussion

This paper proposes an inverse probability weighted estimator $\widehat{F}_{HT}(y)$ of $F(y)$ when responses are subject to nonignorable missingness. In applications, if there is the probability density function $f(y)$ of Y, we can extend the preceding proposed inverse probability weighted method by considering the smoothness of $F(y)$. In the case, a modified version of the Horvitz-Thompson-type estimator can be constructed by

$$\widehat{F}_{HT}^{m}(y) = \frac{1}{n^\star}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}(\boldsymbol{u}_i, Y_i; \widehat{\gamma}_{el})}H\left(\frac{y-Y_i}{h_n}\right),$$

where $H(x) = \int_{-\infty}^{x} K(t)dt$, $K(\cdot)$ is the kernel function, $n^\star = \sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}(\boldsymbol{u}_i,Y_i;\widehat{\gamma}_{el})}$ and h_n is the bandwidth. Here, we modify $n^\star = \sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}(\boldsymbol{u}_i,Y_i;\widehat{\gamma}_{el})}$ such that the kernel density estimator $\widehat{f}_n(y) = \frac{1}{n^\star}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}(\boldsymbol{u}_i,Y_i;\widehat{\gamma}_{el})}K_{h_n}(y-Y_i)$ satisfies $\int \widehat{f}_n(y)dy = 1$, where $K_{h_n}(\cdot) = h_n^{-1}K(\cdot/h_n)$. If we choose $d(\boldsymbol{z}_i) = (1, \widetilde{d}(\boldsymbol{z}_i))^{\mathrm{T}}$ in Equation (8) with $\widetilde{d}(\boldsymbol{z}_i)$ being some known function of $\boldsymbol{z}_i$, thus we have $n^\star/n \to 1$ in probability as n goes to infinity, which indicates that $\widehat{F}_{HT}^{m}(y)$ is structurally similar as $\widehat{F}_{HT}(y)$ except that $I(Y_i \le y)$ is replaced by $H(\frac{y-Y_i}{h_n})$. Similarly, we can define the imputation estimators for $F(y)$ by considering the smoothness of $F(y)$. The selection of the bandwidth h_n plays a critical role in estimating $F(y)$, which is challenging when responses are subject to nonignorable missingness. We will investigate the issue in another paper.

To use the proposed methods to estimate the distribution function and quantiles of response under nonignorable missingness data mechanism assumption of responses, we require choosing the instrument variable vector $\boldsymbol{z}$. To this end, we here present a brief discussion on how to choose instrumental variables. Following the argument of [8], an instrument variable z must satisfy the following two conditions[8]: (i) z should be related to response variable y; (ii) z can be excluded from the propensity score model given in Equation (2). If z does not satisfy (i), thus the calibration condition reduces to a single function intrinsically, which is insufficient to estimate unknown function $g(\cdot)$ and γ simultaneously, regardless of whether (ii) is satisfied. On the other hand, if a variable z satisfies the condition (i) but it does not satisfy the condition (ii), we have $E\{\varphi_i(\gamma)\} \neq 0$. Thus, the resulting propensity score estimation based on Equation (8), denoted by $\widetilde{\pi}(\boldsymbol{u}_i, y_i)$, is not a consistent estimator of the true propensity score, which indicates

$$D = \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{\widetilde{\pi}(\boldsymbol{u}_i, Y_i)} \boldsymbol{x}_i - \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \right\| \tag{10}$$

does not converge to zero in probability. Thus, we can choose a variable z such that it leads to the minimum of D in Equation (10) and satisfies the condition (i). But, it is quite challenging to choose the best subset of the instrument variables in some applications. A new criterion may be required to address the issue. We leave it for our future research.

References

[1] Little R J and Rubin D B, *Statistical Analysis with Missing Data*, 2nd Edition, Wiley, New York, 2002.

[2] Kim J K and Shao J, *Statistical Methods for Handling Incomplete Data*, CRC Press, 2013.

[3] Qin J, Leung D, Shao J, Estimation with survey data under nonignorable nonresponse or informative sampling, *Journal of the American Statistical Association*, 2002, **97**(457): 193–200.

[4] Zhao J and Shao J, Approximate conditional likelihood for generalized linear models with general missing data mechanism, *Journal of Systems Science and Complexity*, 2017, **30**(1): 139–153.

[5] Kim J K and Yu C L, A semiparametric estimation of mean functionals with nonignorable missing data, *Journal of the American Statistical Association*, 2011, **106**(493): 157–165.

[6] Zhao H, Zhao P Y, and Tang N S, Empirical likelihood inference for mean functionals with nonignorably missing response data, *Computational Statistics and Data Analysis*, 2013, **66**(66): 101–116.

[7] Tang N, Zhao P, and Zhu H, Empirical likelihood for estimating equations with nonignorably missing data, *Statistica Sinica*, 2014, **24**(2): 723.

[8] Shao J and Wang L, Semiparametric inverse propensity weighting for nonignorable missing data, *Biometrika*, 2016, **103**(1): 175–187.

[9] Wang S, Shao J, and Kim J K, An instrumental variable approach for identification and estimation with nonignorable nonresponse, *Statistica Sinica*, 2014, **24**(3): 1097–1116.

[10] Tang G, Little R J, and Raghunathan T E, Analysis of multivariate missing data with nonignorable nonresponse, *Biometrika*, 2003, **90**(4): 747–764.

[11] Zhao J and Shao J, Semiparametric Pseudo-likelihoods in generalized linear models with nonignorable missing data, *Journal of the American Statistical Association*, 2015, **110**(512): 1577–1590.

[12] Miao W, Ding P, and Geng Z, Identifiability of normal and normal mixture models with nonignorable missing data, *Journal of the American Statistical Association*, 2016, **111**(516): 1673–1683.

[13] Fang F and Shao J, Model selection with nonignorable nonresponse, *Biometrika*, 2016, **103**(4): 861–874.

[14] Nadaraya E A, Some new estimates for distribution functions, *Theory of Probability and Its Applications*, 1964, **9**(3): 497–500.

[15] Nadaraya E A, On estimating regression, *Theory of Probability and Its Applications*, 1964, **9**(1): 141–142.

[16] Watson G S, Smooth regression analysis, *Sankhya: The Indian Journal of Statistics Series A*, 1964, **26**(4): 359–372.

[17] Chen J and Wu C, Estimation of distribution function and quantiles using model-calibrated pseudo empirical likelihood method, *Statistica Sinica*, 2002, **12**(4): 1223–1239.

[18] Cheng P E and Chu C K, Kernel estimation of distribution functions and quantiles with missing data, *Statistica Sinica*, 1996, **6**(1): 63–78.

[19] Liu X, Liu P, and Zhou Y, Distribution estimation with auxiliary information for missing data, *Journal of Statistical Planning and Inference*, 2011, **141**(2): 711–724.

[20] Hu Z, Follmann D A, and Qin J, Dimension reduced kernel estimation for distribution function with incomplete data, *Journal of Statistical Planning and Inference*, 2011, **141**(9): 3084–3093.

[21] Zhao P Y, Tang M L, and Tang N S, Robust estimation of distribution functions and quantiles with non-ignorable missing data, *Canadian Journal of Statistics*, 2013, **41**(4): 575–595.

[22] Chen J, Variyath A M, and Abraham B, Adjusted empirical likelihood and its properties, *Journal of Computational and Graphical Statistics*, 2008, **17**(2): 426–443.

[23] Owen A B, *Empirical Likelihood*, CRC Press, Florida, 2001.

[24] Horvitz D G and Thompson D J, A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association*, 1952, **47**(260): 663–685.

[25] Kang J D Y and Schafer J L, Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data, *Statistical Science*, 2006, **22**(4): 574–580.

[26] Robins J M, Rotnitzky A, and Zhao L P, Estimation of regression coefficients when some regressors are not always observed, *Journal of the American Statistical Association*, 1994, **89**(427): 846–866.

[27] Qin J, Shao J, and Zhang B, Efficient and doubly robust imputation for covariate-dependent missing responses, *Journal of the American Statistical Association*, 2008, **103**(482): 797–810.

[28] Ogburn E L, Rotnitzky A, and Robins J M, Doubly robust estimation of the local average treatment effect curve, *Journal of the Royal Statistical Society: Series B* (*Statistical Methodology*), 2015, **77**(2): 373–396.

[29] Chang T, Kott P S, Using calibration weighting to adjust for nonresponse under a plausible model, *Biometrika*, 2008, **95**(3): 555–571.

[30] Newey W K and Smith R J, Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica*, 2004, **72**(1): 219–255.

[31] Qin J and Lawless J, Empirical likelihood and general estimating equations, *Annals of Statistics*, 1994, **22**(1): 300–325.

[32] Liu Y and Chen J, Adjusted empirical likelihood with high-order precision, *Annals of Statistics*, 2010, **38**(3): 1341–1362.

[33] Chen X, Wan A T K, and Zhou Y, Efficient quantile regression analysis with missing observations, *Journal of the American Statistical Association*, 2015, **110**(510): 723–741.

[34] Wang Q, Linton O, and Härdle W, Semiparametric regression analysis with missing response at random, *Journal of the American Statistical Association*, 2004, **99**(466): 334–345.

[35] Shao J and Sitter R R, Bootstrap for imputed survey data, *Journal of the American Statistical Association*, 1996, **91**(435): 1278–1288.

[36] Wang D and Chen S X, Empirical likelihood for estimating equations with missing values, *Annals of Statistics*, 2009, **37**(1): 490–517.

Appendix

To show the proof of the presented theorems, we fist need some regularity conditions.

Assumption A The estimating equation $\varphi_i(\gamma)$ satisfies (i) $E\{\varphi_i(\gamma)\varphi_i^{\mathrm{T}}(\gamma)\}$ is positive definite; (ii) the second derivatives of $\varphi_i(\gamma)$, say $\partial^2\varphi_i(\gamma)/\partial\gamma^2$, is continuous in a neighborhood $\mathcal{A}$ of the true value γ_0, $|\partial\varphi_i(\gamma)/\partial\gamma|$ is bounded by some integrable function $G(\boldsymbol{x}, Y)$ in $\mathcal{A}$; (iii) $E\{||\varphi_i(\gamma)||^{\kappa}\}$ is bounded for some $\kappa > 2$ and $\gamma \in \Theta$; (iv) matrix $\Gamma(\gamma_0) = E[\partial\varphi_i(\gamma_0)/\partial\gamma]$ is of full rank.

Assumption B (i) The marginal probability density function $f(\boldsymbol{u})$ is bounded away from ∞ in the support of $\boldsymbol{u}$ and the second derivative of $f(\boldsymbol{u})$ is continuous and bounded; (ii) the probability function $\pi(\boldsymbol{u}, Y; \gamma)$ satisfies $\min_{1\le i\le n}\pi(\boldsymbol{u}_i, Y_i; \gamma) \ge c_0 > 0$ a.s. for some positive constant c_0; (iii) the kernel function $K(\cdot)$ is a probability density function such that (a) it is bounded and has compact support, (b) it is symmetric with $\int \omega^2 K(\omega)d\omega < \infty$, (c) $K(\cdot) \ge d_1$ for some $d_1 > 0$ in some closed interval centered at zero, (d) $nh \to \infty$ and $nh^4 \to 0$ as $n \to \infty$, and (e) $nl \to \infty$ and $nl^4 \to 0$ as $n \to \infty$.

Remark A.1 Assumption A is the regularity condition for the EL and AEL estimator (e.g., see [22, 23, 31]); while Assumption B is commonly adopted in the missing data literatures (e.g., see [1, 8, 21]).

Proof of Theorem 3.1 Because the result is similar to Qin and Lawless[31], Newey and Smith[30] and Variyath, et al.[22], we only present the sketch of the proof. First, the consistency of $\widehat{\gamma}_{el}$ can be obtained by using the same argument as given in Theorem 3.1 of Newey and Smith[30], we here omit the details. Next we show the asymptotic normality of $\widehat{\gamma}_{el}$. Because $\widehat{\gamma}_{el}$ is a consistent estimator of γ and the estimating equation $\varphi_i(\gamma)$ is a continuous function, $\widehat{\gamma}_{el}$ is a stationary point of $\ell(\gamma)$. Hence, $\widehat{\gamma}_{el}$ is the solution to the following equations:

$$Q_{1,n+1}(\gamma, \boldsymbol{\lambda}) = \frac{1}{n+1}\sum_{i=1}^{n+1}\frac{\varphi_i(\gamma)}{1+\boldsymbol{\lambda}^{\mathrm{T}}\varphi_i(\gamma)} = \mathbf{0},$$

$$Q_{2,n+1}(\gamma, \boldsymbol{\lambda}) = \frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{1+\boldsymbol{\lambda}^{\mathrm{T}}\varphi_i(\gamma)}(\partial\varphi_i(\gamma)/\partial\gamma)^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{0}.$$

Expanding $Q_{1,n+1}(\widehat{\gamma}_{el},\widehat{\boldsymbol{\lambda}})$ and $Q_{2,n+1}(\widehat{\gamma}_{el},\widehat{\boldsymbol{\lambda}})$ at $(\gamma,\boldsymbol{\lambda})=(\gamma_0,\mathbf{0})$ and ignoring the higher order terms leads to

$$Q_{1,n+1}(\widehat{\gamma}_{el},\widehat{\boldsymbol{\lambda}})=Q_{1,n+1}(\gamma_0,\mathbf{0})+\frac{\partial Q_{1,n+1}(\gamma_0,\mathbf{0})}{\partial\gamma}(\widehat{\gamma}_{el}-\gamma_0)+\frac{\partial Q_{1,n+1}(\gamma_0,\mathbf{0})}{\partial\boldsymbol{\lambda}}\widehat{\boldsymbol{\lambda}}=\mathbf{0},$$

$$Q_{2,n+1}(\widehat{\gamma}_{el},\widehat{\boldsymbol{\lambda}})=Q_{2,n+1}(\gamma_0,\mathbf{0})+\frac{\partial Q_{1,n+1}(\gamma_0,\mathbf{0})}{\partial\gamma}(\widehat{\gamma}_{el}-\gamma_0)+\frac{\partial Q_{2,n+1}(\gamma_0,\mathbf{0})}{\partial\boldsymbol{\lambda}}\widehat{\boldsymbol{\lambda}}=\mathbf{0}.$$

Note that at $(\gamma_0,\mathbf{0})$, we have

$$\frac{\partial Q_{1,n+1}(\gamma_0,\mathbf{0})}{\partial\gamma}=\frac{\partial Q_{2,n+1}(\gamma_0,\mathbf{0})}{\partial\boldsymbol{\lambda}}=\frac{1}{n+1}\sum_{i=1}^{n+1}\frac{\partial\varphi_i(\gamma_0)}{\partial\gamma}\xrightarrow{\mathcal{P}}B,$$

$$\frac{\partial Q_{1,n+1}(\gamma_0,\mathbf{0})}{\partial\boldsymbol{\lambda}}=-\frac{1}{n+1}\sum_{i=1}^{n+1}\varphi_i(\gamma_0)\varphi_i^{\mathrm{T}}(\gamma_0)\xrightarrow{\mathcal{P}}-A,\ \frac{\partial Q_{2,n+1}(\gamma_0,\mathbf{0})}{\partial\gamma}=0.$$

Combining the above equations yields

$$\sqrt{n}(\widehat{\gamma}_{el}-\gamma_0)=(B^{\mathrm{T}}A^{-1}B)^{-1}B^{\mathrm{T}}A^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_i(\gamma_0)+o_p(1).$$

Note that the choice of a_n together with the moment conditions makes the above equation hold. Thus, the proof of Theorem 3.1 is completed. ▮

Proof of Theorem 3.2 We first present the asymptotic properties of estimator $\widehat{F}_{HT}(y)$. Let

$$I_{1n}=n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\delta_i\{I(Y_i\le y)-F_0(y|\boldsymbol{u}_i)\}}{\pi(\boldsymbol{u}_i,Y_i;\gamma_0)}+\{F_0(y|\boldsymbol{u}_i)-F(y)\}\right\},$$

$$I_{2n}=n^{-1/2}\sum_{i=1}^{n}\left\{\left\{\frac{1}{\widehat{\pi}_i(\gamma_0)}-\frac{1}{\pi(\boldsymbol{u}_i,Y_i;\gamma_0)}\right\}\delta_i\{I(Y_i\le y)-F_0(y|\boldsymbol{u}_i)\}\right\},$$

$$I_{3n}=n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\delta_i}{\widehat{\pi}_i(\gamma_0)}-1\right\}F_0(y|\boldsymbol{u}_i).$$

Then, we have the following decomposition

$$n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\delta_iI(Y_i\le y)}{\widehat{\pi}_i(\gamma_0)}-F(y)\right\}=I_{1n}+I_{2n}+I_{3n}.$$

Following the similar argument as given in [8] and [21], we have $I_{2n}=o_p(1)$ and $I_{3n}=o_p(1)$. Expanding $\widehat{F}_{HT}(y)$ at γ_0 yields

$$\begin{aligned}&\sqrt{n}(\widehat{F}_{HT}(y)-F(y))\\&=n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\delta_iI(Y_i\le y)}{\widehat{\pi}_i(\gamma_0)}-F(y)\right\}+n^{1/2}(\widehat{\gamma}_{el}-\gamma_0)\frac{1}{n}\sum_{i=1}^{n}\delta_iI(Y_i\le y)\frac{\partial\widehat{\pi}_i^{-1}(\gamma)}{\partial\gamma}\bigg|_{\gamma=\widetilde{\gamma}}\\&=n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\delta_i\{I(Y_i\le y)-F_0(y|\boldsymbol{u}_i)\}}{\pi(\boldsymbol{u}_i,Y_i;\gamma_0)}+\{F_0(y|\boldsymbol{u}_i)-F(y)\}\right\}+n^{1/2}(\widehat{\gamma}_{el}-\gamma_0)H+o_p(1)\\&=n^{-1/2}\sum_{i=1}^{n}\eta_i+o_p(1),\end{aligned}$$

where $\eta_i = \{\frac{\delta_i}{\pi(\boldsymbol{u}_i,Y_i;\gamma_0)}\{I(Y_i \le y) - F_0(y|\boldsymbol{u}_i)\} + F_0(y|\boldsymbol{u}_i) - F(y) + \psi_i(\gamma_0)H\}$ and $H = E\{(1-\delta)\{Y - m_Y(\boldsymbol{u})\}\{I(Y \le y) - F_0(y|\boldsymbol{u})\}\}$ with $m_Y(\boldsymbol{u}) = E(Y|\boldsymbol{u}, \delta = 0)$. Thus, the proof of the asymptotic property of $\widehat{F}_{HT}(y)$ is completed.

Next, we show the asymptotic property of $\widehat{F}_{RI}(y)$. By the definition of $\widehat{F}^0_{RI}(y)$, we have

$$\begin{aligned}\sqrt{n}(\widehat{F}^0_{RI}(y) - F(y)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\{\delta_i I(Y_i \le y) + (1-\delta_i)\widehat{F}_0(y|\boldsymbol{u}_i;\gamma_0) - F(y)\}\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_i\{I(Y_i \le y) - F_0(y|\boldsymbol{u}_i)\} + \frac{1}{\sqrt{n}}\sum_{i=1}^n\{F_0(y|\boldsymbol{u}_i) - F(y)\}\\ &\quad +\frac{1}{\sqrt{n}}\sum_{i=1}^n(1-\delta_i)\{\widehat{F}_0(y|\boldsymbol{u}_i;\gamma_0) - F_0(y|\boldsymbol{u}_i)\}\\ &:= J_{1n} + J_{2n} + J_{3n}.\end{aligned}$$

Using the similar arguments as given in the proof of Theorem 1 of Zhao, et al.[21], we have $J_{3n} = n^{-1/2}\sum_{i=1}^n\{1-\pi(\boldsymbol{u}_i,Y_i;\gamma)\}\delta_i\{I(Y_i \le y) - F_0(y|\boldsymbol{u}_i)\}/\pi(\boldsymbol{u}_i,Y_i;\gamma) + o_p(1)$. Then, we have $\sqrt{n}(\widehat{F}^0_{RI}(y) - F(y)) = I_{1n} + o_p(1)$, where I_{1n} is defined in the proof of the asymptotic property of $\widehat{F}_{HT}(y)$. By the definition of $\widehat{F}_{RI}(y)$, we have

$$\begin{aligned}\sqrt{n}(\widehat{F}_{RI}(y) - F(y)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\{\delta_i I(Y_i \le y) + (1-\delta_i)\widehat{F}_0(y|\boldsymbol{u}_i;\widehat{\gamma}_{el}) - F(y)\}\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_i\{I(Y_i \le y) - F_0(y|\boldsymbol{u}_i)\} + \frac{1}{\sqrt{n}}\sum_{i=1}^n\{F_0(y|\boldsymbol{u}_i) - F(y)\}\\ &\quad +\frac{1}{\sqrt{n}}\sum_{i=1}^n(1-\delta_i)\{\widehat{F}_0(y|\boldsymbol{u}_i;\gamma_0) - F_0(y|\boldsymbol{u}_i)\}\\ &\quad +\frac{1}{\sqrt{n}}\sum_{i=1}^n(1-\delta_i)\{\widehat{F}_0(y|\boldsymbol{u}_i;\widehat{\gamma}_{el}) - \widehat{F}_0(y|\boldsymbol{u}_i;\gamma_0)\}\\ &:= J_{1n} + J_{2n} + J_{3n} + J_{4n}.\end{aligned}$$

Taking the Taylor expansion of J_{4n} yields

$$J_{4n} = \sqrt{n}(\widehat{\gamma}_{el} - \gamma_0)\frac{1}{n}\sum_{i=1}^n(1-\delta_i)\frac{\partial\widehat{F}_0(y|\boldsymbol{u}_i)}{\partial\gamma} = \sqrt{n}(\widehat{\gamma}_{el} - \gamma_0)W,$$

where $W = \frac{1}{n}\sum_{i=1}^n(1-\delta_i)\partial\widehat{F}_0(y|\boldsymbol{u}_i)/\partial\gamma$. Following the argument of the proof of Theorem 2 in [21], we have $W = H + o_p(1)$. Combining the above equations leads to $\sqrt{n}(\widehat{F}_{RI}(y) - F(y)) = n^{-1/2}\sum_{i=1}^n\left\{\frac{\delta_i\{I(Y_i\le y)-F_0(y|\boldsymbol{u}_i)\}}{\pi(\boldsymbol{u}_i,Y_i;\gamma_0)} + \{F_0(y|\boldsymbol{u}_i) - F(y)\}\right\} + n^{1/2}(\widehat{\gamma}_{el}-\gamma_0)H + o_p(1)$, which indicates that the asymptotic property of $\widehat{F}_{RI}(y)$ holds.

Finally, we show the asymptotic property of $\widehat{F}_{IP}(y)$. Note that

$$\begin{aligned}\sqrt{n}(\widehat{F}^0_{IP}(y) - F(y)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\frac{\delta_i}{\widehat{\pi}_i(\gamma)}I(Y_i \le y) + \left(1 - \frac{\delta_i}{\widehat{\pi}_i(\gamma)}\right)\widehat{F}_0(y|\boldsymbol{u}_i) - F(y)\right\}\\ &= I_{1n} + I_{2n} + \widetilde{I}_{3n},\end{aligned}$$

where I_{1n} and I_{2n} are defined in the proof of of the asymptotic property of $\widehat{F}_{HT}(y)$, $\widetilde{I}_{3n} = n^{-1/2}\sum_{i=1}^{n}\{\frac{\delta_i}{\widehat{\pi}_i(\gamma_0)} - 1\}\{F_0(y|\boldsymbol{u}_i) - \widehat{F}_0(y|\boldsymbol{u}_i)\}$. Following the similar argument as given in the proof of Theorem 4 in [21], we have $\widetilde{I}_{3n} = o_p(1)$. Combining the above results yields $\sqrt{n}(\widehat{F}^0_{IP}(y) - F(y)) = I_{1n} + o_p(1)$. Expanding $\widehat{F}_{IP}(y)$ at γ_0 results in $\sqrt{n}(\widehat{F}_{IP}(y) - F(y)) = \sqrt{n}(\widehat{F}^0_{IP}(y) - F(y)) + \sqrt{n}(\widehat{\gamma}_{el} - \gamma_0)\partial\widehat{F}^0_{IP}(y)/\partial\gamma + o_p(1)$, where

$$\partial\widehat{F}^0_{IP}(y)/\partial\gamma = \frac{1}{n}\sum_{i=1}^{n}\delta_i\frac{\partial\widehat{\pi}_i^{-1}(\gamma_0)}{\partial\gamma}\{I(Y_i \le y) - \widehat{F}_0(y|\boldsymbol{u}_i)\} + \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\delta_i}{\widehat{\pi}_i(\gamma_0)}\right)\partial\widehat{F}_0(y|\boldsymbol{u}_i)/\partial\gamma.$$

Using the similar arguments as given in the proof of Theorem 5 in [21], we obtain $\partial\widehat{F}^0_{IP}(y)/\partial\gamma|_{\gamma=\gamma_0} = H + o_p(1)$. Thus, we obtain

$$\begin{aligned}\sqrt{n}(\widehat{F}_{IP}(y) - F(y)) &= \sqrt{n}(\widehat{F}^0_{IP}(y) - F(y)) + \sqrt{n}(\widehat{\gamma}_{el} - \gamma_0)H + o_p(1)\\ &= n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\delta_i\{I(Y_i \le y) - F_0(y|\boldsymbol{u}_i)\}}{\pi(\boldsymbol{u}_i, Y_i; \gamma_0)} + \{F_0(y|\boldsymbol{u}_i) - F(y)\}\right\}\\ &\quad + n^{1/2}(\widehat{\gamma}_{el} - \gamma_0)H + o_p(1)\\ &= n^{-1/2}\sum_{i=1}^{n}\eta_i + o_p(1).\end{aligned}$$

Thus, we finish the proof of Theorem 3.2. ∎

Proof of Theorem 3.4 The proof of Theorem 3.4 can be obtained by the same argument as given in [21], we here omit the details. ∎