




Multiply Robust Estimation of Quantile Treatment Effects with Missing Responses

Xiaorui Wang¹ · Guoyou Qin² · Yanlin Tang¹  · Yinfeng Wang³

Received: 7 March 2023 / Revised: 26 April 2023 / Accepted: 13 September 2023

© School of Mathematical Sciences, University of Science and Technology of China and Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Causal inference and missing data have attracted significant research interests in recent years, while the current literature usually focuses on only one of these two issues. In this paper, we develop two multiply robust methods to estimate the quantile treatment effect (QTE), in the context of missing data. Compared to the commonly used average treatment effect, QTE provides a more complete picture of the difference between the treatment and control groups. The first one is based on inverse probability weighting, the resulting QTE estimator is root- n consistent and asymptotic normal, as long as the class of candidate models of propensity scores contains the correct model and so does that for the probability of being observed. The second one is based on augmented inverse probability weighting, which further relaxes the restriction on the probability of being observed. Simulation studies are conducted to investigate the performance of the proposed method, and the motivated CHARLS data are analyzed, exhibiting different treatment effects at various quantile levels.

Keywords Augmented inverse probability weighting · CHARLS · Missing data · Multiply robust · Quantile treatment effect

Mathematics Subject Classification 62G05 · 62G35

✉ Yanlin Tang
yltang@fem.ecnu.edu.cn

✉ Yinfeng Wang
dairy-2006@163.com

¹ KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai, People's Republic of China

² School of Public Health, Fudan University, Shanghai, People's Republic of China

³ Interdisciplinary Research Institute of Data Science, School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai, People's Republic of China

1 Introduction

Causal inference has attracted more and more attention, which aims to identify the extent and nature of cause-and-effect relationships. For example, in the motivated Chinese Health and Retirement Longitudinal Study (CHARLS) in Sect. 5, we aim to study the causal effect of social activities on cognitive functions, among middle-aged and older adults in China. Due to possible heterogeneity in the data, it is necessary to provide a complete picture of the relationship between social activities and cognitive functions, rather than the conditional mean relationship. Furthermore, approximately 24% of the responses are missing, which further challenges the analysis.

In causal analysis, one most popular technique is the potential outcome model [19, 20], where the average treatment effect (ATE) is an important measurement, which measures the average difference between the treatment group and control group; see fruitful studies on ATE, Rosenbaum and Rubin [19, 20], Lunceford and Davidian [17], Tan [23], Imbens and Rubin [14], Hernan and Rubins [12]. However, the ATE is not informative enough when there is heterogeneity in the data and is sensitive to the outliers. The quantile treatment effect (QTE) gives a more complete picture of the causal effect and is robust to the heavy tails of the responses, thus is of growing interests. To estimate the QTE in the absence of randomization, classical methods to deal with the confounding effect include propensity score, outcome regression, and doubly robust methods (e.g., [7, 8, 18, 27]).

As in the motivated CHARLS data, response missing is often encountered in real applications, particularly in observational studies. Among various missing mechanisms, the most useful one is missing at random (MAR), which assumes that the missingness only depends on the observed values. To avoid inconsistent estimation by simply ignoring the missing data, several classes of methods are proposed, say inverse probability weighting (IPW, [22]), augmented IPW (AIPW, [1]), imputation [21]. Recently, double robust and multiply robust methods are developed under the MAR assumption [3, 9–11, 15], to provide robustness against model misspecification. Xie et al. [25] further extended the multiply robust method to handle QTE estimation.

Current literature for quantile regression focuses on either causal inference [8, 27] or missing data [9, 24]. To our best knowledge, there is no literature which studies both issues simultaneously, possibly due to the complexity introduced by both the confounding effects and response missing. However, in practice, the combination of these two issues is frequently encountered, such as the motivated CHARLS data in Sect. 5. To handle both issues, we need to correctly specify the propensity score (PS) model for confounding adjustment and the probability of being observed (PO) or the outcome regression (OR) to deal with missing data, which is challenging in practice.

In this paper, we develop two multiply robust methods to estimate the QTEs with responses being MAR. The first method is that, based on multiply robust estimations of the PS and PO, the IPW approach is utilized to develop a weighted objective function to estimate the QTEs. The second approach, multiply robust AIPW, is to make further resistance to the model misspecification, which relaxes the restriction on the PO. The contributions can be summarized as follows. Firstly, the proposed methods provide a complete picture of the causal effects, which is helpful to identify possible heterogeneity and thus make personalized intervention; this merit is confirmed

in the motivated CHARLS data. Secondly, the proposed methods are able to adjust the confounding effect and deal with missing data, which are commonly encountered in the real applications, providing consistent QTE estimates. Finally, the proposed two estimators are multiply robust. For the IPW method, the resulting estimator is consistent if the class of candidate models of PS contains the correct model and so does that for the PO. For the AIPW method, the restriction on the PO can be relaxed to that either PO or OR contains a correct specified model. Furthermore, the proof of consistency and asymptotic normality properties is not trivial, especially for the proposed AIPW estimator, which is mainly from the cumulative challenges of the augmentation and confounding adjustment.

The rest parts are organized as follows. In Sect. 2, we introduce the potential outcome framework and some key assumptions, and propose the multiply robust IPW and AIPW estimators of QTE with missing responses. We present the asymptotic properties in Sect. 3. Simulation studies are conducted to investigate the finite-sample performance in Sect. 4. Finally, the proposed method is applied to the motivated CHARLS data, showing that the causal effects of social activities on cognitive functions vary across different quantile levels of the responses. All technical proofs are provided in Section S2 of the online Supplementary Materials.

2 The Proposed Method

2.1 Notation and Identification

Let $Y(1)$ and $Y(0)$ be the potential outcomes for the treated and control groups, \mathbf{X} be the covariate vector, and T be the treatment status where $T = 1$ means treated and $T = 0$ means untreated. Under the consistency assumption [25], we observe the outcome $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$. Let $\eta(\mathbf{X}) = Pr(T = 1 | \mathbf{X})$ be the propensity score (PS), satisfying $0 < \eta(\mathbf{X}) < 1$ for all \mathbf{X} .

We are interested in the τ -quantile treatment effect (τ -QTE), defined as

$$\Delta_q(\tau) = q_1(\tau) - q_0(\tau),$$

where $\tau \in (0, 1)$ is a quantile index, and $q_1(\tau)$ and $q_0(\tau)$ are the τ -quantile of random variables $Y(1)$ and $Y(0)$, respectively.

Assuming that the treatment assignment is strongly ignorable, i.e., $(Y(1), Y(0))$ are independent of T conditional on \mathbf{X} . Under this assumption, $q_1(\tau)$ and $q_0(\tau)$ are identified based on the following moment equalities,

$$E\left(\frac{T}{\eta(\mathbf{X})}[\tau - I\{Y \leq q_1(\tau)\}]\right) = 0, \quad E\left(\frac{1 - T}{1 - \eta(\mathbf{X})}[\tau - I\{Y \leq q_0(\tau)\}]\right) = 0. \tag{2.1}$$

Let $\alpha_0(\tau) = q_0(\tau)$ and $\alpha_1(\tau) = q_1(\tau) - q_0(\tau)$, thus $\alpha_1(\tau)$ is our interested τ -QTE. Equation (2.1) is equivalent to

$$E \left(\frac{\tilde{\mathbf{T}}^\top}{\eta(\mathbf{X})} [\tau - I\{Y \leq \tilde{\mathbf{T}}^\top \tilde{\alpha}_0(\tau)\}] - \frac{(1-T, 0)^\top}{1-\eta(\mathbf{X})} [\tau - I\{Y \leq \tilde{\mathbf{T}}^\top \tilde{\alpha}_0(\tau)\}] \right) = 0,$$

where $\tilde{\mathbf{T}} = (1, T)^\top$, $\tilde{\alpha}_0(\tau) = (\alpha_0(\tau), \alpha_1(\tau))^\top$. Then, $\alpha_1(\tau)$ can be identified.

In practice, the responses are possibly missing. Let R be an indicator of observing Y with $R = 1$ if Y is observed and $R = 0$ otherwise. The MAR mechanism is considered here which assumes $\Pr(R = 1 | Y, \mathbf{X}, T) = \Pr(R = 1 | \mathbf{X}, T)$. Denote the probability of being observed (PO) by $\pi(\mathbf{X}, T) = \Pr(R = 1 | \mathbf{X}, T)$.

2.2 Multiply Robust IPW Estimator

To simultaneously handle both confoundedness and missingness, we introduce PS $\eta(\mathbf{X}) = \Pr(T = 1 | \mathbf{X})$ to account for the confoundedness and PO $\pi(\mathbf{X}, T) = \Pr(R = 1 | \mathbf{X}, T)$ to deal with the missingness, estimating the τ -QTE by minimizing

$$\begin{aligned} & Q_{n,\tau}(\mathbf{a}; \pi, \eta) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{R_i}{\pi(\mathbf{X}_i, T_i)} \left\{ \frac{T_i}{\eta(\mathbf{X}_i)} \rho_\tau(Y_i - \tilde{\mathbf{T}}_i^\top \mathbf{a}) + \frac{1-T_i}{1-\eta(\mathbf{X}_i)} \rho_\tau(Y_i - \tilde{\mathbf{T}}_i^\top \mathbf{a}) \right\}, \quad (2.2) \end{aligned}$$

with respect to $\mathbf{a} = (a_0, a_1)^\top$ where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ corresponds to the check loss function. To deal with the missingness, we use $1/\pi(\mathbf{X}_i, T_i)$ to weight each observed subject, which helps to recover the population information based on the biased sample $\{i : R_i = 1\}$. To handle the confounding effect, $\eta(\mathbf{X}_i)$ is used to balance the weighted distribution of the covariates in the two groups.

In practice, $\eta(\mathbf{X})$ and $\pi(\mathbf{X}, T)$ are unknown, and need to be estimated. Generally, we can use parametric methods to estimate $\eta(\mathbf{X})$ and $\pi(\mathbf{X}, T)$, assuming that PS and PO follow a generalized linear model, say the logistic model, and then obtain their maximum likelihood estimates. However, the parametric methods require to specify the models for PS and PO correctly, which is usually infeasible in practice, and we need some robustness against model misspecifications. In this paper, we follow the idea in Han [11] and Han et al. [9], obtaining the estimates through the multiply robust approach.

We first present the estimator for $\pi(\cdot)$. Let $\mathcal{P}_\pi = \{\pi^j(\mathbf{X}, T; \boldsymbol{\theta}^j) : j = 1, \dots, J\}$ denote the set of candidate models for $\pi(\mathbf{X}, T)$, where $\boldsymbol{\theta}^j$ is the corresponding parameter vector. We use $\hat{\boldsymbol{\theta}}^j$ to denote the estimator of $\boldsymbol{\theta}^j$, usually taking to be the maximizer of the binomial likelihood

$$\prod_{i=1}^n \{\pi^j(\mathbf{X}_i, T_i; \boldsymbol{\theta}^j)\}^{R_i} \{1 - \pi^j(\mathbf{X}_i, T_i; \boldsymbol{\theta}^j)\}^{1-R_i}.$$

Let $m = \sum_{i=1}^n R_i$ be the number of subjects observed, and without loss of generality, we assume that $R_1 = \dots = R_m = 1$, $R_{m+1} = \dots = R_n = 0$. Let $\omega_1(\mathbf{X}, T) =$

$1/\pi(\mathbf{X}, T)$, then it is easy to verify that

$$E\left(\omega_1(\mathbf{X}, T)[\pi^j(\mathbf{X}, T; \boldsymbol{\theta}^j) - E\{\pi^j(\mathbf{X}, T; \boldsymbol{\theta}^j)\} \mid R = 1]\right) = 0 \quad (j = 1, \dots, J). \tag{2.3}$$

Therefore, we define the weights $\omega_{1,i}, i = 1, \dots, m$, subject to

$$\begin{aligned} \omega_{1,i} \geq 0 \quad (i = 1, \dots, m), \quad \frac{1}{n} \sum_{i=1}^m \omega_{1,i} &= 1, \\ \frac{1}{n} \sum_{i=1}^m \omega_{1,i} \{ \pi^j(\mathbf{X}_i, T_i; \hat{\boldsymbol{\theta}}^j) - \hat{\zeta}^{1,j}(\hat{\boldsymbol{\theta}}^j) \} &= 0 \quad (j = 1, \dots, J), \end{aligned} \tag{2.4}$$

where $\hat{\zeta}^{1,j}(\boldsymbol{\theta}^j) = n^{-1} \sum_{i=1}^n \pi^j(\mathbf{X}_i, T_i; \boldsymbol{\theta}^j)$. We obtain the empirical likelihood estimate of $\omega_{1,i}$ by maximizing $\prod_{i=1}^m \omega_{1,i}$ subject to the constraints in (2.4). We then give

$$\hat{\pi}(\mathbf{X}_i, T_i) = \frac{1}{\hat{\omega}_{1,i}}, \quad i = 1, \dots, m.$$

Next, we obtain the PS estimates for the treatment group and the control group, respectively. Let $\mathcal{P}_\eta = \{ \eta^l(\mathbf{X}; \boldsymbol{\gamma}^l) : l = 1, \dots, L \}$ denote the set of candidate models for $\eta(\mathbf{X})$, where $\boldsymbol{\gamma}^l$ is the corresponding parameter vector. We use $\hat{\boldsymbol{\gamma}}^l$ to denote the estimator of $\boldsymbol{\gamma}^l$, usually taking to be the maximizer of the binomial likelihood

$$\prod_{i=1}^n \{ \eta^l(\mathbf{X}_i; \boldsymbol{\gamma}^l) \}^{T_i} \{ 1 - \eta^l(\mathbf{X}_i; \boldsymbol{\gamma}^l) \}^{1-T_i}.$$

Step 1, we get the PS estimates for the treatment group. Denote $\mathcal{T}_n = \{ i : T_i = 1 \}$, and let

$$\hat{\eta}_{\mathcal{T}_n}(\mathbf{X}_i) = \frac{1}{\hat{\omega}_{2,i}}, \quad i \in \mathcal{T}_n,$$

where $\{ \hat{\omega}_{2,i}, i \in \mathcal{T}_n \}$ are obtained by maximizing $\prod_{i \in \mathcal{T}_n} \omega_{2,i}$, subject to

$$\begin{aligned} \omega_{2,i} \geq 0 \quad (i \in \mathcal{T}_n), \quad \frac{1}{n} \sum_{i \in \mathcal{T}_n} \omega_{2,i} &= 1, \\ \frac{1}{n} \sum_{i \in \mathcal{T}_n} \omega_{2,i} \{ \eta^l(\mathbf{X}_i; \hat{\boldsymbol{\gamma}}^l) - \zeta^{2,l}(\hat{\boldsymbol{\gamma}}^l) \} &= 0 \quad (l = 1, \dots, L), \end{aligned}$$

with $\zeta^{2,l}(\boldsymbol{\gamma}^l) = n^{-1} \sum_{i=1}^n \eta^l(\mathbf{X}_i; \boldsymbol{\gamma}^l)$.

Step 2, we get the PS estimates for the control group. Let $\mathcal{T}_n^c = \{1, \dots, n\} \setminus \mathcal{T}_n$ and let

$$1 - \widehat{\eta}_{\mathcal{T}_n^c}(\mathbf{X}_i) = \widehat{\Pr}(T_i = 0 \mid \mathbf{X}_i) = \frac{1}{\widehat{\omega}_{3,i}}, \quad i \in \mathcal{T}_n^c,$$

where $\{\widehat{\omega}_{3,i}, i \in \mathcal{T}_n^c\}$ are obtained by maximizing $\prod_{i \in \mathcal{T}_n^c} \omega_{3,i}$, subject to

$$\begin{aligned} \omega_{3,i} &\geq 0 \quad (i \in \mathcal{T}_n^c), \quad \frac{1}{n} \sum_{i \in \mathcal{T}_n^c} \omega_{3,i} = 1, \\ \frac{1}{n} \sum_{i \in \mathcal{T}_n^c} \omega_{3,i} [\{1 - \eta^l(\mathbf{X}_i; \widehat{\boldsymbol{\gamma}}^l)\} - \{1 - \zeta^{2,l}(\widehat{\boldsymbol{\gamma}}^l)\}] &= 0 \quad (l = 1, \dots, L). \end{aligned}$$

Given $\{\widehat{\eta}(\mathbf{X}_i), i = 1, \dots, n\} = \{\widehat{\eta}_{\mathcal{T}_n}(\mathbf{X}_i) : i \in \mathcal{T}_n\} \cup \{\widehat{\eta}_{\mathcal{T}_n^c}(\mathbf{X}_i) : i \in \mathcal{T}_n^c\}$ and $\{\widehat{\pi}(\mathbf{X}_i, T_i), i = 1, \dots, n\}$, we propose to estimate the τ -QTE by minimizing

$$\begin{aligned} Q_{n,\tau}(\mathbf{a}; \widehat{\pi}, \widehat{\eta}) &= \frac{1}{n} \sum_{i=1}^n \frac{R_i}{\widehat{\pi}(\mathbf{X}_i, T_i)} \left\{ \frac{T_i}{\widehat{\eta}(\mathbf{X}_i)} \rho_\tau(Y_i - \widetilde{\mathbf{T}}_i^\top \mathbf{a}) + \frac{1 - T_i}{1 - \widehat{\eta}(\mathbf{X}_i)} \rho_\tau(Y_i - \widetilde{\mathbf{T}}_i^\top \mathbf{a}) \right\}, \quad (2.5) \end{aligned}$$

with respect to \mathbf{a} . Denote the minimizer as $\widehat{\mathbf{a}}(\tau) = (\widehat{a}_0(\tau), \widehat{a}_1(\tau))^\top$, then $\widehat{a}_1(\tau)$ is the proposed estimator for the τ -QTE $\alpha_1(\tau)$, while $\widehat{a}_0(\tau)$ is the estimator for $\alpha_0(\tau)$, the marginal τ -th quantile of the control group. In Sect. 3, we will prove that, if \mathcal{P}_π and \mathcal{P}_η contain the correct PO and PS models, respectively, $\widehat{a}_1(\tau)$ is root- n consistent to the τ -QTE and asymptotic normal.

2.3 Multiply Robust AIPW Estimator

In our objective function (2.2), we only specify the model for PO to address the missing issues. To gain more resistance against the model misspecification, we develop an augmented inverse probability weighted (AIPW) method to estimate the τ -QTE by solving

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \frac{R_i}{\pi(\mathbf{X}_i, T_i)} \left\{ \frac{(T_i, T_i)^\top}{\eta(\mathbf{X}_i)} \psi_\tau(Y_i - \widetilde{\mathbf{T}}_i^\top \mathbf{a}) + \frac{(1 - T_i, 0)^\top}{1 - \eta(\mathbf{X}_i)} \psi_\tau(Y_i - \widetilde{\mathbf{T}}_i^\top \mathbf{a}) \right\} - \left\{ \frac{R_i}{\pi(\mathbf{X}_i, T_i)} - 1 \right\} \\ &\quad \times \left\{ \frac{(T_i, T_i)^\top}{\eta(\mathbf{X}_i)} \frac{1}{S} \sum_{s=1}^S \psi_\tau(Y_{\mathbf{X}_i, T_i}^s - \widetilde{\mathbf{T}}_i^\top \mathbf{a}) + \frac{(1 - T_i, 0)^\top}{1 - \eta(\mathbf{X}_i)} \frac{1}{S} \sum_{s=1}^S \psi_\tau(Y_{\mathbf{X}_i, T_i}^s - \widetilde{\mathbf{T}}_i^\top \mathbf{a}) \right\} \\ &\approx 0, \end{aligned}$$

where the set $\{Y_{\mathbf{X}_i, T}^s\}_{s=1}^S$ is a random sample of size S from the conditional density $f_{Y|\mathbf{X}, T}(\cdot)$. The first term is the score function of the IPW estimator [see expression (2.2)], the augmentation term is motivated by the imputation estimator, which is added

to obtain more information from the outcome regression (OR) and provide double protection against model misspecifications.

Note that $f_{Y|X,T}(\cdot)$ is unknown, we need estimate $f_{Y|X,T}(\cdot)$ to draw random samples. To gain some resistance against the model misspecifications, we also adopt the multiply robust idea. Let $\mathcal{P}_f = \{f_{Y|X,T}^k(y; \xi^k) : k = 1, \dots, K\}$ denote the set of candidate models for $f_{Y|X,T}(\cdot)$, where ξ^k is the corresponding parameter vector. We use $\widehat{\xi}^k$ to denote the estimator of ξ^k , usually taking to be the maximizer of the likelihood function

$$\prod_{i=1}^n f_{Y|X_i, T_i}^k(Y_i; \xi^k)^{R_i}. \tag{2.6}$$

Let $\omega(\mathbf{X}, T) = 1/\pi(\mathbf{X}, T)$, it is easy to verify that

$$\begin{aligned} & E\left(\omega(\mathbf{X}, T)[\pi^j(\mathbf{X}, T; \theta^j) - E\{\pi^j(\mathbf{X}, T; \theta^j)\}] \mid R = 1\right) = 0, (j = 1, \dots, J), \\ & E\left\{\omega(\mathbf{X}, T) \left(\left[\frac{(T, T)^\top}{\eta(\mathbf{X})} \frac{1}{S} \sum_{s=1}^S \psi_\tau\{Y^s(\xi^k) - \widetilde{\mathbf{T}}^\top \mathbf{a}\} \right. \right. \right. \\ & \quad \left. \left. \left. + \frac{(1-T, 0)^\top}{1-\eta(\mathbf{X})} \frac{1}{S} \sum_{s=1}^S \psi_\tau\{Y^s(\xi^k) - \widetilde{\mathbf{T}}^\top \mathbf{a}\} \right] \right. \right. \\ & \quad \left. \left. - E\left[\frac{(T, T)^\top}{\eta(\mathbf{X})} \frac{1}{S} \sum_{s=1}^S \psi_\tau\{Y^s(\xi^k) - \widetilde{\mathbf{T}}^\top \mathbf{a}\} \right. \right. \right. \\ & \quad \left. \left. \left. + \frac{(1-T, 0)^\top}{1-\eta(\mathbf{X})} \frac{1}{S} \sum_{s=1}^S \psi_\tau\{Y^s(\xi^k) - \widetilde{\mathbf{T}}^\top \mathbf{a}\} \right] \right) \mid R = 1 \right\} \\ & \triangleq E\left(\omega(\mathbf{X}, T) \left[\mathbf{U}^k(\mathbf{X}, T; \xi^k, \eta(\mathbf{X}), \mathbf{a}) - E\{\mathbf{U}^k(\mathbf{X}, T; \xi^k, \eta(\mathbf{X}), \mathbf{a})\} \mid R = 1\right] \right) = 0, \\ & (k = 1, \dots, K), \end{aligned} \tag{2.7}$$

where $\{Y^s(\xi^k)\}_{s=1}^S$ denotes a random sample of size S from $f_{Y|X,T}^k(y; \xi^k)$.

Based on the sample versions of the above two equations, we use the empirical likelihood method to estimate τ -QTE by solving

$$\sum_{i=1}^m \widehat{w}_i \left\{ \frac{(T_i, T_i)^\top}{\widehat{\eta}(\mathbf{X}_i)} \psi_\tau(Y_i - \widetilde{\mathbf{T}}_i^\top \mathbf{a}) + \frac{(1-T_i, 0)^\top}{1-\widehat{\eta}(\mathbf{X}_i)} \psi_\tau(Y_i - \widetilde{\mathbf{T}}_i^\top \mathbf{a}) \right\} \triangleq M_{n,\tau}(\mathbf{a}; \widehat{\omega}, \widehat{\eta}) \approx 0, \tag{2.8}$$

where the weighted estimating equation is solved by converting it into the weighted quantile regression and call the package `quantreg`, and $\widehat{\eta}(\mathbf{X}_i)$ is obtained by the multiply robust method in Sect. 2.2 and \widehat{w}_i 's are calculated by the following steps.

- We first calculate $\widehat{\xi}^k, k = 1, 2, \dots, K$, by maximizing the expression (2.6). And we draw a random sample $\{Y^s(\widehat{\xi}^k)\}_{s=1}^S$ from $f_{Y|X,T}^k(y; \widehat{\xi}^k)$.
- Based on the estimator $\widehat{\eta}(\mathbf{X}_i)$, calculate $\widehat{\mathbf{a}}_S^k(\tau), k = 1, 2, \dots, K$, by solving

$$\frac{1}{n} \sum_{i=1}^n R_i \left\{ \frac{(T_i, T_i)^\top}{\widehat{\eta}(\mathbf{X}_i)} \psi_\tau(Y_i - \widetilde{\mathbf{T}}_i^\top \mathbf{a}) + \frac{(1-T_i, 0)^\top}{1-\widehat{\eta}(\mathbf{X}_i)} \psi_\tau(Y_i - \widetilde{\mathbf{T}}_i^\top \mathbf{a}) \right\} + (1-R_i)$$

$$\begin{aligned} & \times \left[\frac{(T_i, T_i)^\top}{\widehat{\eta}(\mathbf{X}_i)} \frac{1}{S} \sum_{s=1}^S \psi_\tau\{Y_i^s(\widehat{\xi}^k) - \widehat{\mathbf{T}}_i^\top \mathbf{a}\} + \frac{(1 - T_i, 0)^\top}{1 - \widehat{\eta}(\mathbf{X}_i)} \frac{1}{S} \sum_{s=1}^S \psi_\tau\{Y_i^s(\widehat{\xi}^k) - \widehat{\mathbf{T}}_i^\top \mathbf{a}\} \right] \\ & \approx 0, \end{aligned}$$

where the weighted estimating equation is solved by converting it into the weighted quantile regression and call the package `quantreg`.

- Finally, \widehat{w}_i can be obtained by maximizing $\prod_{i=1}^m \omega_i$ subject to

$$\begin{aligned} \omega_i & \geq 0 \quad (i = 1, \dots, m), \quad \sum_{i=1}^m \omega_i = 1, \\ \sum_{i=1}^m \omega_i \{ \pi^j(\mathbf{X}_i, T_i; \widehat{\theta}^j) - \widehat{\zeta}^{1,j}(\widehat{\theta}^j) \} & = 0 \quad (j = 1, \dots, J), \\ \sum_{i=1}^m \omega_i \{ \mathbf{U}^k(\mathbf{X}_i, T_i; \widehat{\xi}^k, \widehat{\eta}(\mathbf{X}_i), \widehat{\mathbf{a}}_S^k(\tau)) - \widehat{\zeta}^k(\widehat{\xi}^k, \widehat{\eta}(\mathbf{X}_i), \widehat{\mathbf{a}}_S^k(\tau)) \} & = 0 \\ & (k = 1, \dots, K), \end{aligned} \tag{2.9}$$

where $\widehat{\zeta}^{1,j}(\theta^j)$ is defined in (2.4) and

$$\widehat{\zeta}^k(\xi^k, \eta(\mathbf{X}), \mathbf{a}_S^k(\tau)) = n^{-1} \sum_{i=1}^n \mathbf{U}^k(\mathbf{X}_i, T_i; \xi^k, \eta(\mathbf{X}), \mathbf{a}_S^k(\tau)).$$

Denote the solution of Eq. (2.8) as $\check{\mathbf{a}}(\tau) = (\check{a}_0(\tau), \check{a}_1(\tau))^\top$, then $\check{a}_1(\tau)$ is the proposed multiply robust AIPW estimator of the τ -QTE $\alpha_1(\tau)$. In Sect. 3, we will prove that, as long as \mathcal{P}_η contains the correct model for PO, and so does either \mathcal{P}_π for PS or \mathcal{P}_f for OR, the τ -QTE $\check{a}_1(\tau)$ is root- n consistent and asymptotic normal.

3 Asymptotic Properties

In this section, we present the asymptotic properties of the proposed two QTE estimators. Without loss of generality, let $\pi^1(\mathbf{X}, T; \theta^1)$, $\eta^1(\mathbf{X}; \boldsymbol{\gamma}^1)$ and $f_{Y|\mathbf{X},T}^1(y; \xi^1)$ be the correctly specified models for $\pi(\mathbf{X}, T)$, $\eta(\mathbf{X})$ and $f_{Y|\mathbf{X},T}(y)$, respectively, with θ_0^1 , $\boldsymbol{\gamma}_0^1$ and ξ_0^1 being the true parameter vectors. Let Φ_1 and Φ_2 be the score functions for θ^1 and $\boldsymbol{\gamma}^1$, respectively, $\Phi_{1,i}$ and $\Phi_{2,i}$ be the score functions evaluated at i th subject. The following conditions are needed for the asymptotic properties of the proposed estimators.

- A1. The treatment assignment is strongly ignorable: $(Y(1), Y(0))$ are independent of T conditional on \mathbf{X} . We also assume that $\sum_{i=1}^n T_i/n \rightarrow c_1$ for some $0 < c_1 < 1$.
- A2. The response is missing at random, $\Pr(R_i = 1 \mid \mathbf{X}_i, T_i, Y_i) = \Pr(R_i = 1 \mid \mathbf{X}_i, T_i)$. We also assume that $\sum_{i=1}^n R_i/n \rightarrow c_2$ for some $0 < c_2 < 1$.

- A3. The $\alpha_0(\tau)$ and $\alpha_0(\tau) + \alpha_1(\tau)$ are the unique τ quantile of potential outcomes $Y(0)$ and $Y(1)$, respectively. The density functions of $Y(0)$ and $Y(1)$, $f_{Y(0)}(\cdot)$, $f_{Y(1)}(\cdot)$ at $\alpha_0(\tau)$ and $\alpha_0(\tau) + \alpha_1(\tau)$ are bounded away from zero. Assume that $E\|\mathbf{X}\|^4 < \infty$.
- A4. For $j = 1, \dots, J, l = 1, \dots, L, \pi^j(\mathbf{X}, T; \boldsymbol{\theta}^j)$ and $\eta^l(\mathbf{X}; \boldsymbol{\gamma}^l)$ have bounded derivatives in \mathbf{X} up to the second order, and are continuously differentiable in $\boldsymbol{\theta}^j$ and $\boldsymbol{\gamma}^l$, respectively. Assume that $\inf_{\mathbf{X}, T} \inf_{\boldsymbol{\theta}^j} \pi^j(\mathbf{X}, T; \boldsymbol{\theta}^j) > 0, \inf_{\mathbf{X}} \inf_{\boldsymbol{\gamma}^l} \eta^l(\mathbf{X}; \boldsymbol{\gamma}^l) > 0$.
- A5. The matrices \mathbf{D} and $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3, \mathbf{G}_4$, defined in Section S1 of the online Supplementary Materials, are invertible.

Regularity assumptions A1–A5 for the major results are mild. Assumptions A1 and A2 are commonly used in literatures of causal inference and missing data, respectively. Assumptions A3 and A4 are similar to the conditions adopted in Han et al. [9], in order to ensure the consistency of the proposed estimator. Assumption A5 is a technical assumption needed for the asymptotic normality.

Before presenting the asymptotic properties of the proposed estimators, we first give a lemma which shows that the IPW estimation is consistent when the true $\eta(\mathbf{X})$ and $\pi(\mathbf{X}, T)$ are given.

Lemma 3.1 *Let $\tilde{\mathbf{a}}(\tau) = (\tilde{\alpha}_0(\tau), \tilde{\alpha}_1(\tau))^\top$ be the minimizer of (2.2) with true $\eta(\mathbf{X})$ and $\pi(\mathbf{X}, T)$. Under assumptions A1 and A2, we have that $\tilde{\mathbf{a}}(\tau)$ is a consistent estimator of $\tilde{\alpha}_0(\tau)$.*

Theorem 3.2 *Assume that \mathcal{P}_π contains a correctly specified model for $\pi(\mathbf{X}, T)$ and so does \mathcal{P}_η for $\eta(\mathbf{X})$. Under assumptions A1–A4, we have that $\widehat{\mathbf{a}}(\tau) \xrightarrow{P} \tilde{\alpha}_0(\tau)$ as $n \rightarrow \infty$.*

Theorem 3.3 *Assume that the conditions in Theorem 3.2 hold. Under the additional assumption A5, we have that*

$$n^{1/2}\{\widehat{\mathbf{a}}(\tau) - \tilde{\alpha}_0(\tau)\} = \mathbf{D}^{-1}n^{-1/2} \sum_{i=1}^n \left[V_{1,i}(\boldsymbol{\theta}_0^1, \boldsymbol{\gamma}_0^1) + E \left\{ \frac{\partial V_1(\boldsymbol{\theta}_0^1, \boldsymbol{\gamma}_0^1)}{\partial \boldsymbol{\theta}^1}, \frac{\partial V_1(\boldsymbol{\theta}_0^1, \boldsymbol{\gamma}_0^1)}{\partial \boldsymbol{\gamma}^1} \right\} \begin{pmatrix} \{E(\boldsymbol{\Phi}_1^{\otimes 2})\}^{-1} \boldsymbol{\Phi}_{1,i} \\ \{E(\boldsymbol{\Phi}_2^{\otimes 2})\}^{-1} \boldsymbol{\Phi}_{2,i} \end{pmatrix} \right] + o_p(1),$$

where \mathbf{D} and $V_{1,i}(\boldsymbol{\theta}_0^1, \boldsymbol{\gamma}_0^1)$ are defined in Section S1 of the online Supplementary Materials, $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^\top$ for a vector/matrix \mathbf{A} . Furthermore, we can derive that

$$n^{1/2}\{\widehat{\mathbf{a}}(\tau) - \tilde{\alpha}_0(\tau)\} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_1),$$

where

$$\boldsymbol{\Sigma}_1 = \mathbf{D}^{-1} \text{Cov} \left[V_1(\boldsymbol{\theta}_0^1, \boldsymbol{\gamma}_0^1) + E \left\{ \frac{\partial V_1(\boldsymbol{\theta}_0^1, \boldsymbol{\gamma}_0^1)}{\partial \boldsymbol{\theta}^1}, \frac{\partial V_1(\boldsymbol{\theta}_0^1, \boldsymbol{\gamma}_0^1)}{\partial \boldsymbol{\gamma}^1} \right\} \begin{pmatrix} \{E(\boldsymbol{\Phi}_1^{\otimes 2})\}^{-1} \boldsymbol{\Phi}_1 \\ \{E(\boldsymbol{\Phi}_2^{\otimes 2})\}^{-1} \boldsymbol{\Phi}_2 \end{pmatrix} \right] (\mathbf{D}^{-1})^\top.$$

Finally, for the τ -QTE, let $\mathbf{e}_1 = (1, 0)^\top$, we have

$$n^{1/2}\{\widehat{a}_1(\tau) - \alpha_1(\tau)\} \xrightarrow{d} N(0, \mathbf{e}_1^\top \boldsymbol{\Sigma}_1 \mathbf{e}_1).$$

Theorem 3.2 proves that the proposed IPW estimator is consistent, as long as \mathcal{P}_π contains a correctly specified model for $\pi(\mathbf{X}, T)$ and so does \mathcal{P}_η for $\eta(\mathbf{X})$, which implies the robustness of $\widehat{\mathbf{a}}(\tau)$. Theorem 3.3 gives the representation of the estimated QTE, where the first part in $V_{1,i}(\boldsymbol{\theta}_0^1, \boldsymbol{\gamma}_0^1)$ is the representation as if the true $\pi(\mathbf{X}, T)$ and $\eta(\mathbf{X})$ were known, and the rest parts account for the multiple robust estimation of $\pi(\mathbf{X}, T)$ and $\eta(\mathbf{X})$. The expression of the covariance matrix in Theorem 3.3 is complicated, thus resampling bootstrap would be a good choice for uncertainty quantification of the QTE.

Theorem 3.4 *Assume that \mathcal{P}_η contains the correct model for PO, and so does either \mathcal{P}_π for PS or \mathcal{P}_f for OR. Under assumptions A1–A4, we have $\check{\mathbf{a}}(\tau) \xrightarrow{P} \tilde{\boldsymbol{\alpha}}_0(\tau)$ as $n \rightarrow \infty$.*

Theorem 3.5 *Assume that the conditions in Theorem 3.4 hold. Under the additional assumption A5, we have*

$$n^{1/2}\{\check{a}_1(\tau) - \alpha_1(\tau)\} \xrightarrow{d} N(0, \mathbf{e}_1^\top \boldsymbol{\Sigma}_2 \mathbf{e}_1),$$

where $\mathbf{e}_1 = (1, 0)^\top$ and

$$\boldsymbol{\Sigma}_2 = \mathbf{D}^{-1} \text{Cov} \left[V_2(\boldsymbol{\theta}_0^1, \boldsymbol{\gamma}_0^1) + \mathbb{E} \left\{ \frac{\partial V_2(\boldsymbol{\theta}_0^1, \boldsymbol{\gamma}_0^1)}{\partial \boldsymbol{\theta}^1}, \frac{\partial V_2(\boldsymbol{\theta}_0^1, \boldsymbol{\gamma}_0^1)}{\partial \boldsymbol{\gamma}^1} \right\} \begin{pmatrix} \{\mathbb{E}(\boldsymbol{\Phi}_1^{\otimes 2})\}^{-1} \boldsymbol{\Phi}_1 \\ \{\mathbb{E}(\boldsymbol{\Phi}_2^{\otimes 2})\}^{-1} \boldsymbol{\Phi}_2 \end{pmatrix} \right] (\mathbf{D}^{-1})^\top,$$

where \mathbf{D} and $V_2(\boldsymbol{\theta}_0^1, \boldsymbol{\gamma}_0^1)$ are defined in Section S1 of the online Supplementary Materials.

Theorem 3.4 demonstrates that the proposed AIPW estimator is multiply robust, i.e., as long as \mathcal{P}_η contains the correct model for PO, and so does either \mathcal{P}_π for PS or \mathcal{P}_f for OR, the resulting estimator $\check{\mathbf{a}}(\tau)$ is consistent. Theorem 3.5 gives the asymptotic normality of the proposed AIPW estimator. In fact, just as in Theorem 3.3, we first get the representation of $\check{\mathbf{a}}(\tau)$, and we only present the asymptotic distribution of $\check{a}_1(\tau)$ for brevity. The proof of Theorem 3.5 is more challenging than Theorem 3.3 due to the additional outcome regression model. Compared with the literature of missing data, Theorem 3.5 is challenging due to confounding adjustment.

4 Simulation Study

In this section, simulation studies are conducted to evaluate the finite performance of the proposed τ -QTE estimators. On the one hand, simulation studies are used to illustrate the effectiveness of the proposed estimators in dealing with the confounding effect and missingness. On the other hand, the proposed multiply robust IPW and AIPW estimators have a certain level of resistance to the model misspecifications.

The data are generated as follows. First, we generate the covariates as $X_{i,1} \sim \text{Unif}(-0.25, 0.25)$, $X_{i,2} \sim \text{Binom}(0.5)$, $X_{i,3} \sim \text{Unif}(-0.5, 0.5)$, $X_{i,4} \sim \text{Binom}(0.5) \times$

0.6. Second, the treatment indicator T_i is generated from a logistic regression model as

$$\text{logit}\{\eta(\mathbf{X}_i; \boldsymbol{\gamma})\} = \gamma_0 + \gamma_1 X_{i,1} + \gamma_2 X_{i,2},$$

where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)^\top = (0.5, -1.5, -1)^\top$. Third, the response is generated from the outcome regression model

$$Y_i = \alpha_0 + \alpha_1 T_i + \alpha_2 X_{i,1} + \alpha_3 X_{i,2} + \alpha_4 X_{i,3} + \alpha_5 X_{i,4} + \varepsilon_i, i = 1, \dots, n,$$

where $(\alpha_0, \alpha_1, \dots, \alpha_5)^\top = (3.5, 2, 1, 1, 0.5, 2)^\top$, ε_i 's are random errors. Finally, the missing mechanism follows

$$\text{logit}\{\pi(\mathbf{X}_i, T_i; \boldsymbol{\theta})\} = \theta_0 + \theta_1 X_{i,3} + \theta_2 X_{i,4} + \theta_3 T_i,$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3)^\top$. For the random errors, we consider two cases: (1) $\varepsilon_i \sim N(0, 1)$, (2) $\varepsilon_i \sim t(3)$, both leading to τ -QTE $\alpha_1 = 2$. To investigate the impact of different missing probabilities, we choose $\boldsymbol{\theta} = (2.5, 0.5, -3, 1)^\top$ and $(1.5, 0.5, -3, 1)^\top$, which on average result in 15% and 30% missing proportions, respectively.

According to the data generating process, the correct working models for PS and PO are given by $\text{logit}\{\eta^1(\mathbf{X}_i; \boldsymbol{\gamma}^1)\} = \gamma_0^1 + \gamma_1^1 X_{i,1} + \gamma_2^1 X_{i,2}$ and $\text{logit}\{\pi^1(\mathbf{X}_i, T_i; \boldsymbol{\theta}^1)\} = \theta_0^1 + \theta_1^1 X_{i,3} + \theta_2^1 X_{i,4} + \theta_3^1 T_i$, respectively; for the AIPW estimator, the correct working model for $f_{Y|\mathbf{X},T}(\cdot)$ is given by the probability density function of $N(\xi_0^1 + \xi_1^1 T_i + \xi_2^1 X_{i,1} + \xi_3^1 X_{i,2} + \xi_4^1 X_{i,3} + \xi_5^1 X_{i,4}, 1)$ or non-central $t(3)$ distribution with non-centrality parameter $\xi_0^1 + \xi_1^1 T_i + \xi_2^1 X_{i,1} + \xi_3^1 X_{i,2} + \xi_4^1 X_{i,3} + \xi_5^1 X_{i,4}$. The additional incorrect models for PS and PO are $\text{logit}\{\eta^2(\mathbf{X}_i; \boldsymbol{\gamma}^2)\} = \gamma_0^2 + \gamma_1^2 \exp(X_{i,1})$ and $\text{logit}\{\pi^2(\mathbf{X}_i, T_i; \boldsymbol{\theta}^2)\} = \theta_0^2 + \theta_1^2 \exp(X_{i,3})$, respectively, and $N(\xi_0^2 + \xi_1^2 T_i, 1)$ for OR with normal errors and non-central $t(3)$ distribution with non-centrality parameter $\xi_0^2 + \xi_1^2 T_i$ for OR with $t(3)$ errors, respectively. The parameters in OR are derived by the quasi-maximum likelihood estimate, that is, the ordinary least squares regression, based on complete-case analysis; we also tried to estimate those parameters by the maximum likelihood method, but it heavily depends on initial value and converges very slowly. We consider two imputation sample sizes $S = 10$ and 50 , and only report the results based on $S = 10$, as those from $S = 50$ are similar.

To the best of our knowledge, this is the first work discussing causal inference on quantiles with missing responses, so we only compare the proposed IPW and AIPW estimators, denoted as IPW.1111 and AIPW.111111, with four different classes of estimates, and the descriptions of all estimators are summarized in Table 1. (I) TM-type estimators, the minimizer of the objective function (2.5), where $\widehat{\eta}(\mathbf{X}_i)$ and $\widehat{\pi}(\mathbf{X}_i, T_i)$ are obtained by the empirical likelihood method similarly to those in IPW.1111, but including only one model for $P(T_i = 1 | \mathbf{X}_i)$ and $P(R_i = 1 | \mathbf{X}_i, T_i)$. (II) TMO-type estimators, the solution of the estimation Eq. (2.8), where the weights are obtained by the empirical likelihood method similarly to those in AIPW.111111, and we include five representative TMO-type estimators. The TMO-type estimators provide double

protection for missing issues and only require that either PO or OR model is correct. (III) T-type estimators, the minimizer of an objective function similar to (2.5), where $\widehat{\eta}(\mathbf{X}_i)$ is obtained in the same way as the TM-type estimator, while $\widehat{\pi}(\mathbf{X}_i, T_i) = 1$ if $R_i = 1$. The T-type estimators consider the causal effect but ignores the feature of missingness. (IV) Naive estimator, denoted as Naive.0000, the minimizer of an objective function similar to (2.5), where $\widehat{\eta}(\mathbf{X}_i) = T_i$ and $\widehat{\pi}(\mathbf{X}_i, T_i) = 1$ if $R_i = 1$. The naive estimator ignores the features of causal effect and missingness.

A sample size of $n = 400$ and two quantile levels $\tau = 0.3, 0.5$ are considered, and all simulation results are summarized based on 1000 replicates. Table 2 summarizes the finite-sample performances of the QTE estimators, including the bias (BIAS), the root mean squared error (RMSE), as well as the empirical coverage probabilities (ECP) and the empirical mean lengths (EML) with nominal level 95%. The confidence intervals are constructed by the bootstrap method, which is implemented by the `summary.rq` function of the R package `quantreg`, and the number of resampling replicates is 500. From Table 2, we have the following findings. First, when both PS and PO candidate models include a correctly specified model, the obtained estimates TM.1010 have apparently small BIAS and RMSE than the estimators under model misspecification. The proposed IPW.1111 has almost the same performance as TM.1010, and performs even better in some scenarios. Second, when the PS model is correctly specified, the estimators are reasonably good even if the PO model is misspecified, though the BIAS and RMSE are relatively larger than IPW.1111 and TM.1010, see the performances of TM.1001, T.1100, and T.1000. However, if the PS model is misspecified, no matter whether the PO model is correctly specified, the BIAS and RMSE are much larger than IPW.1111. Third, the proposed AIPW estimator has lower RMSE than the IPW estimator. As long as \mathcal{P}_η contains the correct model for PO, and so does either \mathcal{P}_π for PS or \mathcal{P}_f for OR, the resulting estimators are close to TMO.101010 and AIPW.111111, which reflects that the AIPW estimator provides more protection for model specification. Similar to the TM-type estimator, if the PS model is misspecified, even if the models for PO and OR are correct, the BIAS and RMSE are much larger than other estimators. Fourth, the ECPs of IPW.1111, AIPW.111111, TM.1010 and TMO.111111 estimator (except TMO.110101), are close to the nominal level, while others lead to low ECPs. Table 3 summarizes the results for $t(3)$ error, and the comparison results are similar.

Finally, we would like to give a comment on the comparisons between AIPW.111111 and IPW.1111. From the view of theory, the AIPW approach is more efficient than the IPW approach, as we use more information in the AIPW. However, even if the true outcome model is included, there are several sources that lead to additional variation in the estimator: (i) the estimation of the parameters in the true outcome model; (ii) the inclusion of the misspecified outcome model in the empirical likelihood approach; (iii) sampling from the outcome models. Therefore, the AIPW approach does not benefit from augmentation when the missing rate is low (say 15%), but performs slightly better when the missing rate achieves 30%.

Table 1 Descriptions of several QTE estimators

Method	Models for PS $\eta^1(\mathbf{X}; \gamma^1)$	$\eta^2(\mathbf{X}; \gamma^2)$	Models for PO $\pi^1(\mathbf{X}_i, T_i; \theta^1)$	$\pi^2(\mathbf{X}_i, T_i; \theta^2)$	Models for OR $f^1(\mathbf{X}_i, T_i; \xi^1)$	$f^2(\mathbf{X}_i, T_i; \xi^2)$
IPW.1111	✓	✓	✓	✓		
TM.1010	✓		✓			
TM.1001	✓			✓		
TM.0110		✓	✓			
TM.0101		✓		✓		
AIPW.111111	✓	✓	✓	✓	✓	✓
TMO.111010	✓	✓	✓		✓	
TMO.111001	✓	✓	✓			✓
TMO.110110	✓	✓		✓	✓	
TMO.110101	✓	✓		✓		✓
TMO.011010		✓	✓		✓	
T.1100	✓	✓				
T.1000	✓					
T.0100	✓					
Naive.0000		✓				

^a The notation “✓” indicates that the model is used in the corresponding method, where model with superscript 1 is the correct one

Table 2 BIAS($\times 100$), RMSE($\times 100$), ECP(%) and EML for normal error, with $n = 400$, based on 1000 replicates

τ	Methods	15% missing				30% missing			
		BIAS	RMSE	ECP	EML	BIAS	RMSE	ECP	EML
0.3	IPW.1111	-1.03	18.31	0.95	0.78	-0.35	21.55	0.95	0.88
	TM.1010	-1.05	18.35	0.95	0.78	-0.48	21.51	0.95	0.88
	TM.1001	4.46	19.16	0.96	0.79	8.12	22.91	0.94	0.86
	TM.0110	-26.39	32.00	0.72	0.75	-25.59	32.55	0.77	0.84
	TM.0101	-21.01	27.90	0.80	0.75	-17.48	27.19	0.85	0.82
	AIPW.111111	-1.30	18.21	0.96	0.79	-0.02	21.32	0.95	0.88
	TMO.111010	-1.23	18.16	0.95	0.78	-0.13	21.23	0.95	0.88
	TMO.111001	-1.11	18.37	0.95	0.78	-0.36	21.58	0.95	0.88
	TMO.110110	-0.69	18.08	0.96	0.77	-0.25	20.96	0.94	0.84
	TMO.110101	4.61	19.33	0.96	0.79	8.69	23.52	0.94	0.86
	TMO.011010	-26.39	31.95	0.71	0.75	-25.38	32.36	0.78	0.84
	T.1100	4.50	19.11	0.96	0.79	8.27	23.22	0.93	0.86
	T.1000	4.43	19.12	0.96	0.78	8.28	23.21	0.94	0.86
	T.0100	-20.92	27.78	0.80	0.75	-17.40	27.04	0.85	0.82
	Naive.0000	-23.81	29.90	0.75	0.74	-20.27	28.87	0.82	0.81
0.5	IPW.1111	-0.23	17.63	0.96	0.76	0.18	20.23	0.96	0.87
	TM.1010	-0.28	17.67	0.96	0.76	0.19	20.18	0.96	0.87
	TM.1001	5.56	18.38	0.95	0.75	9.78	21.87	0.94	0.83
	TM.0110	-26.00	31.79	0.71	0.73	-25.61	32.49	0.77	0.83
	TM.0101	-20.07	26.87	0.82	0.72	-16.24	25.25	0.87	0.79
	AIPW.111111	-0.37	17.62	0.96	0.76	0.37	19.97	0.96	0.87
	TMO.111010	-0.27	17.70	0.96	0.76	0.38	19.77	0.96	0.87
	TMO.111001	-0.27	17.65	0.96	0.76	0.11	20.52	0.96	0.87
	TMO.110110	0.24	17.50	0.96	0.75	-0.18	19.98	0.96	0.84
	TMO.110101	5.65	18.40	0.95	0.75	9.81	21.94	0.94	0.83
	TMO.011010	-26.10	31.86	0.71	0.73	-25.16	31.93	0.77	0.83
	T.1100	5.56	18.30	0.95	0.75	9.89	22.12	0.93	0.83
	T.1000	5.57	18.31	0.95	0.75	9.82	22.03	0.94	0.83
	T.0100	-19.98	26.78	0.82	0.72	-16.12	25.19	0.87	0.79
	Naive.0000	-23.78	29.72	0.76	0.72	-20.01	27.89	0.82	0.79

The nominal level of the confidence interval is 95%

^aThe description of different methods are given in Table 1

5 Analysis of CHARLS Data

Cognitive function is a critical dimension of the life quality in later life, the decline of which disrupts daily life function and happiness. It is thus important to study the determinants of cognitive function in order to know how to delay and/or slow down its eventual decline. A growing body of research suggested that social activities or

Table 3 BIAS($\times 100$), RMSE($\times 100$), ECP(%) and EML for $t(3)$ error, with $n = 400$, based on 1000 replicates

τ	Methods	15% missing				30% missing			
		BIAS	RMSE	ECP	EML	BIAS	RMSE	ECP	EML
0.3	IPW.1111	0.61	22.23	0.96	0.91	0.48	25.34	0.94	1.01
	TM.1010	0.60	22.27	0.95	0.90	0.44	25.41	0.94	1.01
	TM.1001	5.98	22.81	0.95	0.90	8.69	26.31	0.92	0.99
	TM.0110	-24.28	32.50	0.80	0.87	-24.17	34.47	0.82	0.97
	TM.0101	-19.32	29.04	0.84	0.86	-16.53	29.55	0.88	0.94
	AIPW.111111	0.63	22.24	0.95	0.90	0.59	25.02	0.94	1.02
	TMO.111010	0.62	22.33	0.95	0.90	0.68	24.89	0.94	1.01
	TMO.111001	0.60	22.29	0.95	0.90	0.36	25.52	0.94	1.02
	TMO.110110	1.71	21.97	0.96	0.89	1.58	24.40	0.95	0.98
	TMO.110101	6.28	22.94	0.95	0.90	9.25	26.89	0.92	0.99
	TMO.011010	-24.41	32.56	0.80	0.87	-24.17	34.31	0.84	0.97
	T.1100	5.90	22.88	0.95	0.90	8.58	26.27	0.92	0.98
	T.1000	5.90	22.85	0.95	0.90	8.59	26.29	0.93	0.98
	T.0100	-19.15	28.96	0.85	0.87	-16.58	29.37	0.87	0.94
	Naive.0000	-22.19	31.00	0.82	0.86	-19.32	30.84	0.85	0.94
0.5	IPW.1111	0.18	20.08	0.96	0.85	0.04	23.74	0.96	0.98
	TM.1010	0.14	20.10	0.96	0.85	0.03	23.75	0.95	0.98
	TM.1001	6.02	21.01	0.95	0.84	9.85	25.10	0.92	0.93
	TM.0110	-25.32	31.93	0.79	0.82	-25.31	33.92	0.82	0.94
	TM.0101	-19.55	27.55	0.85	0.81	-15.31	27.23	0.89	0.89
	AIPW.111111	0.19	20.12	0.97	0.85	0.38	23.58	0.96	0.98
	TMO.111010	0.04	20.09	0.97	0.85	0.29	23.56	0.95	0.98
	TMO.111001	0.17	20.12	0.96	0.85	0.09	23.84	0.96	0.98
	TMO.110110	0.46	19.84	0.96	0.84	0.34	22.90	0.95	0.94
	TMO.110101	6.40	21.20	0.95	0.84	10.60	25.73	0.92	0.93
	TMO.011010	-25.45	32.00	0.78	0.82	-25.22	33.85	0.83	0.94
	T.1100	6.00	21.03	0.95	0.84	9.76	25.00	0.93	0.93
	T.1000	6.03	21.07	0.95	0.84	9.75	25.08	0.92	0.92
	T.0100	-19.57	27.56	0.85	0.81	-15.42	27.27	0.89	0.89
	Naive.0000	-23.15	30.23	0.81	0.80	-19.17	29.42	0.86	0.88

The nominal level of the confidence interval is 95%
^aThe description of different methods are given in Table 1

social engagement among the elderly have a positive effect on cognitive function [13, 16]. In this section, we apply the proposed method to the motivated data set from the Chinese Health and Retirement Longitudinal Study (CHARLS), collected in 2018, and we aim to study the causal effect of social activities on cognitive functions among middle-aged and older adults in China.

Following the analysis of Hu et al. [13], we use episodic memory, a necessary component of reasoning in many cognitive dimensions, as the outcome variable. Episodic memory was evaluated through the means of scores in the immediate and delayed word recall, with scores ranged from 0 to 10. After discarding participants who are either younger than 45 years old, or social activity information (the covariates) missing, 7818 participants are included in our study, with 1889 whose outcome variables are not observed completely, yielding around 24% missingness. The treatment variable is whether to participate in certain common specified activities in China, such as playing chess, card games or Mahjong, interacting with friends, and other social activities. Other covariates (confounders) in the analysis are composed of sociodemographic characteristics and health status related to cognitive function, including age, gender, resident areas (urban or rural), education levels (illiterate, primary education, secondary or above), marital status (currently married or not), smoke (yes or no), hypertension (yes or no), diabetes (yes or no), cardiopathy (yes or no), apoplexy (yes or no), where age is standardized and education levels introduce two dummy variables, education_a (primary education as 1, otherwise 0) and education_b (secondary and above as 1, otherwise 0). The QTE is defined as the quantile difference of cognitive distributions between the treatment (participate in at least one social activity) and control (not participate in any social activities) groups.

To apply our method, we need specify candidate models. For PS $\eta(\mathbf{X})$, we consider two candidate models: $\text{logit}\{\eta^1(\mathbf{X}_i; \boldsymbol{\gamma}^1)\} = \gamma_0^1 + \boldsymbol{\gamma}_1^\top \mathbf{X}_i$ and $\text{logit}\{\eta^2(\mathbf{X}_i^*; \boldsymbol{\gamma}^2)\} = \gamma_0^2 + \boldsymbol{\gamma}_2^\top \mathbf{X}_i^*$, where \mathbf{X}_i only contains all covariates we mentioned above, while \mathbf{X}_i^* includes \mathbf{X}_i and the interactions between the age and other covariates. For PO, we consider $\text{logit}\{\pi^1(\mathbf{X}_i, T_i; \boldsymbol{\theta}^1)\} = \theta_0^1 + \boldsymbol{\theta}_1^\top \mathbf{X}_i + \theta_{11} T_i$ and $\text{logit}\{\pi^2(\mathbf{X}_i^*, T_i; \boldsymbol{\theta}^2)\} = \theta_0^2 + \boldsymbol{\theta}_2^\top \mathbf{X}_i^* + \theta_{21} T_i$. For AIPW estimator, the two models specified for $f_{Y|\mathbf{X},T}(\cdot)$ are given by the probability density functions of $N(\xi_0^1 + \xi_{11} T_i + \boldsymbol{\xi}_1^\top \mathbf{X}_i, \sigma_1^2)$ and $N(\xi_0^2 + \xi_{21} T_i + \boldsymbol{\xi}_2^\top \mathbf{X}_i^*, \sigma_2^2)$. We compare the proposed IPW (IPW.1111) and AIPW (AIPW.111111) estimators with four different classes of estimators as in Sect. 4.

Table 4 summarizes the point estimators, the bootstrap standard error (BSE), and 95% confidence intervals at quantile levels $\tau = 0.3, 0.5$ and 0.7 . The confidence intervals are constructed by 500 bootstrap samples. We have the following findings. First, the positive estimates from all the methods indicate that the social activities have positive effects on cognition, and the effect is significant since none of the 95% confidence intervals includes zero, which is consistent with previous literature [6, 13]. Second, the causal effects of social activities on cognitive functions vary across different quantiles. The QTE at lower quantile is larger than those at the median or higher quantile, which indicates that social activities have greater effects on cognitive improvement of individuals with low cognitive function. This new finding provides important personalized intervention evidence for improving cognitive function. Third, the proposed estimator has a certain discrepancy with T.#### and Naive.0000, showing the necessity of dealing with the confoundedness and missingness. Finally, the proposed IPW.1111 is very close to TM.0101, which indicates the suitability of $\text{logit}\{\eta^2(\mathbf{X}_i^*; \boldsymbol{\gamma}^2)\}$ for PS and $\text{logit}\{\pi^2(\mathbf{X}_i^*, T_i; \boldsymbol{\theta}^2)\}$ for PO; the proposed AIPW.111111 deviates from IPW.1111 in some extent, and according to our theory, the inclusion of the augmentation term plays a role, producing more trustworthy results.

Table 4 Analysis results of for the effects of social activities on cognition ($n = 7818$)

τ	Methods	Est	BSE	Lbd	Ubd
0.3	IPW.1111	0.650	0.133	0.389	0.911
	TM.1010	0.649	0.126	0.402	0.897
	TM.1001	0.649	0.132	0.391	0.908
	TM.0110	0.639	0.133	0.379	0.900
	TM.0101	0.654	0.139	0.382	0.926
	AIPW.111111	0.498	0.094	0.313	0.682
	TMO.111010	0.501	0.098	0.309	0.693
	TMO.111001	0.507	0.099	0.314	0.701
	TMO.110110	0.498	0.091	0.320	0.676
	TMO.110101	0.498	0.098	0.307	0.690
	TMO.011010	0.501	0.098	0.309	0.693
	T.1100	0.592	0.115	0.366	0.819
	T.1000	0.629	0.110	0.413	0.844
	T.0100	0.592	0.117	0.363	0.822
	Naive.0000	1.567	0.093	1.385	1.750
0.5	IPW.1111	0.575	0.093	0.392	0.757
	TM.1010	0.610	0.093	0.427	0.793
	TM.1001	0.603	0.097	0.412	0.793
	TM.0110	0.586	0.091	0.408	0.765
	TM.0101	0.584	0.096	0.395	0.773
	AIPW.111111	0.438	0.083	0.275	0.602
	TMO.111010	0.462	0.080	0.306	0.619
	TMO.111001	0.438	0.080	0.282	0.595
	TMO.110110	0.441	0.082	0.280	0.601
	TMO.110101	0.432	0.081	0.273	0.590
	TMO.011010	0.462	0.084	0.297	0.628
	T.1100	0.437	0.097	0.248	0.627
	T.1000	0.496	0.100	0.300	0.691
	T.0100	0.437	0.099	0.244	0.631
	Naive.0000	1.303	0.087	1.131	1.474
0.7	IPW.1111	0.424	0.087	0.253	0.595
	TM.1010	0.488	0.080	0.330	0.645
	TM.1001	0.476	0.079	0.322	0.630
	TM.0110	0.440	0.087	0.268	0.611

^aEst: point estimators of QTE; BSE: bootstrap standard error; Lbd: lower bound; Ubd: upper bound. Refer the descriptions of the estimators to Table 1

Table 4 continued

τ	Methods	Est	BSE	Lbd	Ubd
	TM.0101	0.431	0.082	0.269	0.592
	AIPW.111111	0.313	0.071	0.175	0.452
	TMO.111010	0.320	0.060	0.202	0.438
	TMO.111001	0.309	0.062	0.186	0.431
	TMO.110110	0.316	0.067	0.184	0.448
	TMO.110101	0.305	0.069	0.170	0.441
	TMO.011010	0.320	0.066	0.190	0.450
	T.1100	0.336	0.075	0.189	0.484
	T.1000	0.367	0.083	0.205	0.529
	T.0100	0.336	0.078	0.184	0.489
	Naive.0000	0.935	0.073	0.792	1.078

^aEst: point estimators of QTE; BSE: bootstrap standard error; Lbd: lower bound; Ubd: upper bound. Refer the descriptions of the estimators to Table 1

6 Discussion

In this paper, we proposed multiply robust IPW and AIPW estimators for the QTE with missing responses. The proposed IPW estimator is robust against model misspecification for PS and PO, and the proposed AIPW estimator further strengthens this robustness. Further research topics can be proposed. First, the methods proposed can be generalized to estimate QTE at extreme quantile levels [26] with missing responses. Second, under memory constraint, we can extend our method to distributed computing [4, 5]. Third, we can further consider the statistical inference of the QTE [2] with missing responses, testing whether the QTE is zero, or heterogeneous to covariates, or varies across quantiles.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40304-023-00380-4>.

Acknowledgements The authors would like to thank two anonymous reviewers, an associate editor and the editor for constructive comments and helpful suggestions. This publication is based upon work partially supported by National Natural Science Foundation of China grants 11871376 and 11871164, Natural Science Foundation of Shanghai 21ZR1420700 and Open Research Fund of KLATASDS-MOE.

Declarations

Conflict of interest The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

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

2. Cai, Z.W., Fang, Y., Lin, M., Tang, S.F.: Inferences for partially conditional quantile treatment effect model. Technical Report 202005, Department of Economics, University of Kansas (2020)
3. Chen, C.X., Shen, B.Y., Liu, A.Y., Wu, R.L., Wang, M.: A multiple robust propensity score method for longitudinal analysis with intermittent missing data. *Biometrics* **77**(2), 519–532 (2021)
4. Chen, L.J., Li, D.Y., Zhou, C.: Distributed inference for the extreme value index. *Biometrika* **109**(1), 257–264 (2021)
5. Chen, X., Liu, W.D., Zhang, Y.C.: Quantile regression under memory constraint. *Ann. Stat.* **47**(6), 3244–3273 (2019)
6. Christelis, D., Dobrescu, L.I.: The causal effect of social activities on cognition: evidence from 20 European countries. *Soc. Sci. Med.* **247**, 1–9 (2020)
7. Donald, S., Hsu, Y.C.: Estimation and inference for distribution functions and quantile functions in treatment effect models. *J. Econom.* **178**, 383–397 (2014)
8. Firpo, S.: Efficient semiparametric estimation of quantile treatment effects. *Econometrica* **75**(1), 259–276 (2007)
9. Han, P.S., Kong, L.L., Zhao, J.W., Zhou, X.C.: A general framework for quantile estimation with incomplete data. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **81**(2), 305–333 (2019)
10. Han, P.S., Wang, L.: Estimation with missing data: beyond double robustness. *Biometrika* **100**, 417–430 (2013)
11. Han, P.S.: Multiply robust estimation in regression analysis with missing data. *J. Am. Stat. Assoc.* **109**(507), 1159–1173 (2014)
12. Hernán, M.A., Robins, J.M.: Causal Inference: What If. Chapman & Hall/CRC, Boca Raton (2020)
13. Hu, Y.Q., Lei, X.Y., Smith, J.P., Zhao, Y.H.: Effects of social activities on cognitive functions: evidence from CHARLS. In: *Aging in Asia: Findings From New and Emerging Data Initiatives*, pp. 279–308, Chap. 12. National Academies Press (US), Washington (2012)
14. Imbens, G.W., Rubin, D.B.: Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press, Cambridge (2015)
15. Lin, H.Z., Zhou, F.Y., Wang, Q.X., Zhou, L., Qin, J.: Robust and efficient estimation for the treatment effect in causal inference and missing data problems. *J. Econom.* **205**(2), 363–380 (2018)
16. Liu, H.X., Fan, X.J., Luo, H.Y., Zhou, Z.L., Shen, C., Hu, N.B., Zhai, X.M.: Comparison of depressive symptoms and its influencing factors among the elderly in urban and rural areas: evidence from the china health and retirement longitudinal study (CHARLS). *Int. J. Environ. Res. Public Health* **18**(8), 3886 (2021)
17. Lunceford, J.K., Davidian, M.: Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat. Med.* **23**, 2937–2960 (2004)
18. Melly, B.: Estimation of counterfactual distributions using quantile regression. University of St.Gallen, Monograph (discussion paper) (2006)
19. Rosenbaum, P.R., Rubin, D.B.: Reducing bias in observational studies using subclassification on the propensity score. *J. Am. Stat. Assoc.* **79**(387), 516–524 (1984)
20. Rosenbaum, P.R., Rubin, D.B.: The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**(1), 41–55 (1983)
21. Rubin, D.B.: Multiple Imputation for Nonresponse in Surveys. Wiley, New York (2004)
22. Seaman, S.R., White, I.R.: Review of inverse probability weighting for dealing with missing data. *Stat. Methods Med. Res.* **22**, 278–295 (2013)
23. Tan, Z.Q.: Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika* **97**(3), 661–682 (2010)
24. Wei, Y., Ma, Y.Y., Carroll, R.J.: Multiple imputation in quantile regression. *Biometrika* **99**, 423–438 (2012)
25. Xie, Y.Y., Cotton, C., Zhu, Y.Y.: Multiply robust estimation of causal quantile treatment effects. *Stat. Med.* **39**, 4238–4251 (2020)
26. Zhang, Y.C.: Extremal quantile treatment effects. *Ann. Stat.* **46**(6B), 3707–3740 (2018)
27. Zhang, Z.W., Chen, Z., Troendle, J.F., Zhang, J.: Causal inference on quantiles with an obstetric application. *Biometrics* **68**, 697–706 (2012)

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.