A general framework for quantile estimation with incomplete data

Peisong Han,

University of Michigan, Ann Arbor, USA

Linglong Kong,

University of Alberta, Edmonton, Canada

Jiwei Zhao

State University of New York at Buffalo, USA

and Xingcai Zhou

Nanjing Audit University, People's Republic of China

Summary. Quantile estimation has attracted significant research interest in recent years. How-ever, there has been only a limited literature on quantile estimation in the presence of incomplete data. We propose a general framework to address this problem. Our framework combines the two widely adopted approaches for missing data analysis, the imputation approach and the in-verse probability weighting approach, via the empirical likelihood method. The method proposed is capable of dealing with many different missingness settings. We mainly study three of them: estimating the marginal quantile of a response that is subject to missingness while there are fully observed covariates; estimating the conditional quantile of a fully observed response while the covariates are partially available; estimating the conditional quantile of a response that is subject to missingness with fully observed covariates and extra auxiliary variables. The method proposed allows multiple models for both the missingness probability and the data distribution. The resulting estimators are multiply robust in the sense that they are consistent if any one of these models is correctly specified. The asymptotic distributions are established by using empirical process theory.

Keywords: Empirical likelihood; Imputation; Inverse probability weighting; Missing data; Multiple robustness; Quantile regression

1. Introduction

The population mean of a response variable provides an important central measure of the re-sponse, whereas the population median is an important alternative that is robust to potential outliers. The quantiles, a generalized concept of median, are capable of providing not only central features but also the tail properties of the response distribution (Koenker, 2005). Under mild conditions, the quantile function uniquely determines the underlying distribution (Shorack, 2000). Quantiles provide a more complete picture of the response and are especially useful in the presence of outliers or when the distribution of the response is heavy tailed. A commonly

Address for correspondence: Linglong Kong, Department of Mathematical and Statistical Sciences, University of Alberta, CAB 632, Edmonton, Alberta, T6G 2G1, Canada.
E-mail: lkong@ualberta.ca

encountered challenge in quantile-related analysis is the presence of missing values in the collected data. The most straightforward solution, the complete-case analysis ignoring subjects with missing data, usually leads to substantial bias and/or undermines study efficiency, especially when the rate of missingness is high. There is a large literature dealing with missing data; see, for example, Little and Rubin (2002), Tsiatis (2006), Kim and Shao (2013) and references therein. However, most of the existing methods focus on mean estimation.

Imputation (e.g. Rubin (1987, 1996)) is a widely adopted approach for dealing with missing data. In a quantile regression setting, Yoon (2010) proposed an imputation method where the missing responses are substituted by values drawn from the conditional quantile function of the response at given values of regressors. Chen and Yu (2016) considered an imputation method to deal with missing responses based on semiparametric quantile regression. To deal with missing covariates, Wei *et al*. (2012) developed an iterative imputation procedure assuming that the missingness depends only on the observed covariates. Estimating equations were used by Wei and Yang (2014) to produce consistent linear quantile estimation in the presence of missing covariates through an expectation–maximization type of algorithm. Yang *et al*. (2013) proposed a fractional hot deck imputation based on non-parametric kernel regression.

Inverse probability weighting (IPW) (e.g. Horvitz and Thompson (1952) and Robins *et al*. (1994)) is another popular approach for handling missing data. In quantile regression for longitudinal data with non-ignorable dropouts, Lipsitz *et al*. (1997) considered a set of estimating equations weighted by the inverse of the estimated probability of dropout. Yi and He (2009) investigated a similar method under different model assumptions focusing on median regression. For quantile regression with missing covariates, Sherwood *et al*. (2013) took the IPW approach to study healthcare cost data. Under a non-ignorable missingness mechanism, Zhao *et al*. (2013) proposed an augmented IPW method to estimate the distribution function and quantiles of a response variable. Sun *et al*. (2012) developed an IPW-based method for quantile regression for competing risks data when the failure type is prone to missing values. To handle missing response and/or partially missing covariates, Chen *et al*. (2015) proposed to estimate the probability of missingness non-parametrically.

The imputation and the IPW approaches require modelling the data distribution and the probability of missingness respectively. Properties of the corresponding estimators hold true only if the corresponding model is correctly specified. Thus, most existing methods are vulnerable to model misspecification. Non-parametric modelling may help to reduce the risk of model misspecification. However, it is often impractical because of the curse of dimensionality. Therefore, methods that are robust against model misspecification are highly desired. In the literature of mean regression with missing data, augmented IPW (AIPW) is a popular method (Robins *et al*., 1994; Tsiatis, 2006), where both the probability of missingness and the data distribution are modelled. The resulting estimator is consistent if either model is correct and is called doubly robust. Recently, an even more robust method has been proposed: the so-called multiply robust (MR) method (Han and Wang, 2013; Chan and Yam, 2014; Han 2014a, b, 2016a, b; Chen and Haziza, 2017), where multiple models for the probability of missingness and/or the data distribution can be accounted for, and estimation consistency of the point estimator is guaranteed if any one model is correct. However, the current MR methods were developed in the context of mean regression only and do not directly apply to quantile estimation. The major contribution of this paper is to propose a general framework for quantile estimation with missing data. Our general framework combines both the imputation and the IPW approaches to estimate the marginal or conditional quantiles of the response under a variety of practically important missing data settings. All the resulting estimators are MR. Compared with mean regression, dealing with quantiles is much more difficult since the parameters of interest are no longer esti-

mated by using smooth functions of the data, and thus existing methods and results cannot be simply translated. This complexity is in both the implementation and theoretical investigations. Similarly to Wei *et al.* (2012) and Chen *et al.* (2015), our framework is capable of dealing with independent but non-identically distributed error terms.

An MR method is appealing in many studies where multiple working models may exist. An example is in high dimensional data analysis where there are a large number of fully observed auxiliary variables. In such a case model building for the probability of missingness and the data distribution relies on variable selection techniques that require some tuning parameters. Different levels of tuning may result in different working models. Although the tuning parameters may be selected on the basis of some criteria, e.g. the generalized information criterion (Fan and Tang, 2013), such a selection brings additional uncertainty in the working model specification and usually does not lead to one model that rules out the possibility of all others. A more natural approach, which was discussed in Robins *et al.* (2007), is to postulate multiple models, each with different subsets of auxiliary variables and possibly different link functions. Another example is whether to model the distribution of a variable in the original scale or after a transformation, if the former is highly skewed and the latter is approximately normal. In this case two working models, one for the original scale and one for the transformed scale, may be postulated. For both of the aforementioned examples, our proposed method provides an innovative way of combining multiple working models into estimation.

Our development relies on the empirical likelihood method (Owen, 1988, 2001; Qin and Lawless, 1994), which has been successfully applied to address missing data problems and has attracted considerable research interest (e.g. Wang and Rao (2002), Qin and Zhang (2007), Chen *et al.* (2008), Qin *et al.* (2008, 2009), Wang and Chen (2009), Tan (2010), Han and Wang (2013), Chan and Yam (2014) and Han (2014a, b, 2016a)). However, the existing literature mainly focuses on mean estimation with missing data. Another contribution of this paper is to demonstrate the effectiveness of the empirical likelihood method in quantile estimation with missing data. To derive the asymptotic distributions of our proposed estimators, the non-smooth estimating functions are dealt with via empirical process theory (e.g. van der Vaart and Wellner (1996) and Kosorok (2008)).

The rest of the paper is organized as follows. In Section 2, after a brief review of the imputation and the IPW approaches, we present our proposed method. Section 3 establishes the asymptotic properties. Section 4 contains simulation studies to evaluate the finite sample performance of the method proposed. A real data analysis is provided in Section 5. Section 6 gives some discussions. Technical details are provided in Appendix A. The simulation results are given in the on-line supplementary material. R code that was used for our simulation studies can be found at `http://www-personal.umich.edu/~peisong/R-code/quantile-estimation-with-missing-data/missing.covariates.R`. The AIDS Clinical Trials Group (ACTG) 175 data that were used for our data application can be found in the R package `speff2trial`.

2. The method proposed

The method proposed is developed under three commonly encountered scenarios:
   (a) estimating the marginal quantile of a response that is subject to missingness while there are fully observed covariates;
   (b) estimating the conditional quantile of a fully observed response while the covariates are partially available;
   (c) estimating the conditional quantile of a response that is subject to missingness with fully observed covariates and extra auxiliary variables.

To facilitate the presentation, we first review the imputation and the IPW approaches in scenario (a). Although in the existing literature on quantile estimation these two approaches are most often described for quantile regression with missing data, which corresponds to scenarios (b) and (c), the ideas are more straightforwardly demonstrated in scenario (a) for estimating the marginal quantile. This is also one reason why our development starts with scenario (a). The method proposed can be easily generalized to other scenarios that are different from the three discussed in this paper, such as estimating the conditional quantile of a partially observed response while some covariates are fully observed and others are not, with or without additional auxiliary variables.

2.1. A review on imputation and inverse probability weighting approaches

Let Y denote the response of interest that is subject to missingness, $\mathbf{X}$ a vector of covariates and R the indicator of observing Y ($R=1$ if Y is observed and $R=0$ otherwise). The observed data are n independent and identically distributed copies of $(R, RY, \mathbf{X}^{\mathrm{T}})$. We assume that Y is missing at random (Rubin, 1976), which is a commonly adopted missingness mechanism in the literature:

$$P(R=1|Y, \mathbf{X}) = P(R=1|\mathbf{X}). \tag{1}$$

For now we focus on estimating $q_0(\tau) = Q_\tau(Y) = \inf\{y : P(Y \leqslant y) \geqslant \tau\}$: the τth marginal quantile of Y where $0 < \tau < 1$. For ease of notation, we write $q_0(\tau)$ to be q_0 hereafter.

The imputation approach models $f(Y|\mathbf{X})$, the conditional density of Y given $\mathbf{X}$. Under the missingness at random (MAR) mechanism, this model can be fitted on the basis of subjects with Y observed (complete-case analysis). With the estimated $f(Y|\mathbf{X})$, we can then take a set of random draws at the given $\mathbf{X}$ for a subject with missing Y as the imputed responses for this subject: $\{\hat{Y}_\mathbf{X}^l : l = 1, \ldots, L\}$. The imputation estimator of q_0 is the solution to

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ R_i \psi_\tau(Y_i - q) + (1 - R_i) \frac{1}{L} \sum_{l=1}^{L} \psi_\tau(\hat{Y}_{\mathbf{X}_i}^l - q) \right\} \approx 0,$$

where $\psi_\tau(r) = \tau - I(r < 0)$ and we use '$\approx$' as in Chen *et al.* (2015) to indicate that the exact solution may not exist because of the non-smoothness of the function ψ_τ. The imputation estimator is consistent only if $f(Y|\mathbf{X})$ is correctly modelled.

The IPW approach models the probability in equation (1), denoted by $\pi(\mathbf{X})$ hereafter. With $\pi(\mathbf{X})$ estimated by $\hat{\pi}(\mathbf{X})$, the IPW estimator solves

$$\frac{1}{n} \sum_{i=1}^{n} \frac{R_i}{\hat{\pi}(\mathbf{X}_i)} \psi_\tau(Y_i - q) \approx 0,$$

and is consistent only if $\pi(\mathbf{X})$ is correctly modelled. To improve estimation efficiency of the IPW estimator, we may also consider the AIPW estimator solving

$$\frac{1}{n} \sum_{i=1}^{n} \left[\frac{R_i}{\hat{\pi}(\mathbf{X}_i)} \psi_\tau(Y_i - q) - \left\{ \frac{R_i}{\hat{\pi}(\mathbf{X}_i)} - 1 \right\} \frac{1}{L} \sum_{l=1}^{L} \psi_\tau(\hat{Y}_{\mathbf{X}_i}^l - q) \right] \approx 0,$$

where an augmentation term is added to extract more information from the observed covariates. The AIPW estimator was originally proposed in the context of mean estimation (Robins *et al.*, 1994). In addition to potential gains in efficiency, the AIPW estimator is consistent if either $\pi(\mathbf{X})$ or $f(Y|\mathbf{X})$ is correctly modelled: a property known as double robustness (e.g. Bang and Robins (2005)).

2.2. *Estimation of marginal quantiles*

Our goal is to combine the imputation and the IPW approaches to improve the robustness of estimation consistency against possible model misspecifications further. In practice, since the true data-generating process is usually unknown, multiple models for both $\pi(\mathbf{X})$ and $f(Y \mid \mathbf{X})$ may be postulated and none rules out the possibility of others. Specifically, let $\mathcal{P} = \{\pi^j(\mathbf{X}; \boldsymbol{\alpha}^j) : j = 1, \ldots, J\}$ and $\mathcal{F} = \{f^k(Y \mid \mathbf{X}; \boldsymbol{\gamma}^k) : k = 1, \ldots, K\}$ denote collections of models for $\pi(\mathbf{X})$ and $f(Y \mid \mathbf{X})$ respectively, where $\boldsymbol{\alpha}^j$ and $\boldsymbol{\gamma}^k$ are the corresponding parameters. Our proposed method provides a novel way to combine these models simultaneously so that the resulting estimator is consistent if any one model is correctly specified.

Let $m = \Sigma_{i=1}^n R_i$ be the number of subjects with data completely observed. Without loss of generality, these subjects are indexed by $i = 1, \ldots, m$. Our method is composed of the following steps.

Step 1: calculate $\hat{\boldsymbol{\alpha}}^j$, $j = 1, \ldots, J$, by maximizing the binomial likelihood

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

Step 2: calculate $\hat{\boldsymbol{\gamma}}^k$, $k = 1, \ldots, K$, by maximizing $\Pi_{i=1}^n f^k(Y_i \mid \mathbf{X}_i; \boldsymbol{\gamma}^k)^{R_i}$.
Step 3: calculate $\hat{q}_L^k$, $k = 1, \ldots, K$, as an imputation estimator of q_0 by solving

$$\frac{1}{n} \sum_{i=1}^n \left[R_i \psi_\tau(Y_i - q) + (1 - R_i) \frac{1}{L} \sum_{l=1}^L \psi_\tau\{Y_i^l(\hat{\boldsymbol{\gamma}}^k) - q\} \right] \approx 0,$$

where $\{Y^l(\boldsymbol{\gamma}^k) : l = 1, \ldots, L\}$ denotes a set of random draws of size L from $f^k(Y \mid \mathbf{X}; \boldsymbol{\gamma}^k)$.
Step 4: for the complete cases $i = 1, \ldots, m$, calculate weights

$$\hat{w}_i = \frac{1}{m} \frac{1}{1 + \hat{\boldsymbol{\rho}}^{\mathrm{T}} \hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})}, \tag{2}$$

where $\hat{\boldsymbol{\rho}}$ minimizes

$$F_n(\boldsymbol{\rho}) = -\frac{1}{n} \sum_{i=1}^n R_i \log\{1 + \boldsymbol{\rho}^{\mathrm{T}} \hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})\}. \tag{3}$$

Here $\hat{\boldsymbol{\alpha}}^{\mathrm{T}} = \{(\hat{\boldsymbol{\alpha}}^1)^{\mathrm{T}}, \ldots, (\hat{\boldsymbol{\alpha}}^J)^{\mathrm{T}}\}$, $\hat{\boldsymbol{\gamma}}^{\mathrm{T}} = \{(\hat{\boldsymbol{\gamma}}^1)^{\mathrm{T}}, \ldots, (\hat{\boldsymbol{\gamma}}^K)^{\mathrm{T}}\}$, $\hat{\mathbf{q}}_L^{\mathrm{T}} = (\hat{q}_L^1, \ldots, \hat{q}_L^K)$,

$$\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})^{\mathrm{T}} = \left(\pi_i^1(\hat{\boldsymbol{\alpha}}^1) - \hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1), \ldots, \pi_i^J(\hat{\boldsymbol{\alpha}}^J) - \hat{\theta}^J(\hat{\boldsymbol{\alpha}}^J), \frac{1}{L} \sum_{l=1}^L \psi_\tau\{Y_i^l(\hat{\boldsymbol{\gamma}}^1) - \hat{q}_L^1\} \right.$$

$$\left. - \hat{\eta}_L^1(\hat{q}_L^1, \hat{\boldsymbol{\gamma}}^1), \ldots, -\frac{1}{L} \sum_{l=1}^L \psi_\tau\{Y_i^l(\hat{\boldsymbol{\gamma}}^K) - \hat{q}_L^K\} - \hat{\eta}_L^K(\hat{q}_L^K, \hat{\boldsymbol{\gamma}}^K) \right),$$

$$\hat{\theta}^j(\boldsymbol{\alpha}^j) = \frac{1}{n} \sum_{i=1}^n \pi_i^j(\boldsymbol{\alpha}^j)$$

and

$$\hat{\eta}_L^k(q, \boldsymbol{\gamma}^k) = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{L} \sum_{l=1}^L \psi_\tau\{Y_i^l(\boldsymbol{\gamma}^k) - q\} \right].$$

Searching for $\hat{\boldsymbol{\rho}}$ in this step is a convex minimization problem and can be easily implemented by using the algorithm given in Han (2014b).

Step 5: calculate the estimator $\hat{q}_{\mathrm{MR}}^L$ by solving

$$\sum_{i=1}^m \hat{w}_i \psi_\tau(Y_i - q) \approx 0.$$

In step 1, let $\hat{\boldsymbol{\alpha}}^j \to^{\mathrm{p}} \boldsymbol{\alpha}_*^j$ as $n \to \infty$. Then, $\pi^j(\mathbf{X}; \boldsymbol{\alpha}_*^j) = \pi(\mathbf{X})$ only if $\pi^j(\mathbf{X}; \boldsymbol{\alpha}^j)$ is a correctly specified model for $\pi(\mathbf{X})$. In step 2, the calculation is based on the complete cases, which is justified by the MAR mechanism. Let $\hat{\boldsymbol{\gamma}}^k \to^{\mathrm{p}} \boldsymbol{\gamma}_*^k$ as $n \to \infty$. Then, $f^k(Y \mid \mathbf{X}; \boldsymbol{\gamma}_*^k) = f(Y \mid \mathbf{X})$ only if $f^k(Y \mid \mathbf{X}; \boldsymbol{\gamma}^k)$ is a correctly specified model for $f(Y \mid \mathbf{X})$. In step 3, let $\hat{q}_L^k \to^{\mathrm{p}} q_*^k$ as $n \to \infty$. Note that q_*^k does not depend on L. Then, $q_*^k = q_0$ only if $f^k(Y \mid \mathbf{X}; \boldsymbol{\gamma}^k)$ is a correctly specified model for $f(Y \mid \mathbf{X})$.

Step 4 calculates a set of weights $\{w_i : i = 1, \ldots, m\}$ on the complete cases. These weights account for all the postulated models simultaneously, as seen from the expression of $\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})$. The derivation of these weights is based on the following rationale. For any function $b(\mathbf{X})$ where the relevant expectations exist, it is easy to verify that

$$E(w(\mathbf{X})[b(\mathbf{X}) - E\{b(\mathbf{X})\}] \mid R = 1) = 0, \tag{4}$$

where $w(\mathbf{X}) = \pi(\mathbf{X})^{-1}$. We take $b(\mathbf{X})$ to be $\pi^j(\mathbf{X}; \boldsymbol{\alpha}^j)$ and $a^k(\mathbf{X}; \boldsymbol{\gamma}^k) = E^k\{\psi_\tau(Y - q_0) \mid \mathbf{X}; \boldsymbol{\gamma}^k\} = \tau - P^k(Y < q_0 \mid \mathbf{X}; \boldsymbol{\gamma}^k)$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$, where $E^k(\cdot \mid \mathbf{X}; \boldsymbol{\gamma}^k)$ and $P^k(\cdot \mid \mathbf{X}; \boldsymbol{\gamma}^k)$ are the conditional expectation and the conditional probability under the density $f^k(Y \mid \mathbf{X}; \boldsymbol{\gamma}^k)$. Using these particular functions as $b(\mathbf{X})$ and the $\hat{\boldsymbol{\alpha}}^j$, $\hat{\boldsymbol{\gamma}}^k$ and $\hat{q}_L^k$ from steps 1, 2 and 3 respectively, a sample version of result (4) may be constructed as

$$\left.\begin{array}{c} w_i \geqslant 0 \qquad (i = 1, \ldots, m), \\ \displaystyle\sum_{i=1}^m w_i = 1, \\ \displaystyle\sum_{i=1}^m w_i \hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}}) = \mathbf{0}. \end{array}\right\} \tag{5}$$

Here the conditional expectation $E\{w(\mathbf{X})(\cdot) \mid R = 1\}$ is estimated by putting a discrete probability measure $\{w_i : w_i \geqslant 0$ for $i = 1, \ldots, m$ and $\Sigma_{i=1}^m w_i = 1\}$ on the complete cases, the unconditional expectation $E(\cdot)$ is estimated by the unweighted sample average over the whole sample and $E^k(\cdot \mid \mathbf{X}; \boldsymbol{\gamma}^k)$ is estimated by averaging over the L random draws taken from $f^k(Y \mid \mathbf{X}; \boldsymbol{\gamma}^k)$. Since $\{w_i : i = 1, \ldots, m\}$ is a discrete probability measure, it is natural to consider the $\hat{w}_i$ that maximize $\Pi_{i=1}^m w_i$ under the constraints in expression (5). This is the typical formulation of an empirical likelihood problem, and the solution is given by equation (2) (Qin and Lawless, 1994; Owen, 2001) with $\hat{\rho}$ solving

$$\frac{1}{m} \sum_{i=1}^m \frac{\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})}{1 + \boldsymbol{\rho}^{\mathrm{T}} \hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})} = \mathbf{0}.$$

Since $\hat{\rho}$ must also satisfy $1 + \hat{\boldsymbol{\rho}}^{\mathrm{T}} \hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}}) > 0$ for $i = 1, \ldots, m$ to make $\hat{w}_i$ positive, it is easy to see that $\hat{\rho}$ is actually the minimizer of the convex function $F_n(\boldsymbol{\rho})$ in equation (3). Following Han (2014b), it can be shown that the minimizer of $F_n(\boldsymbol{\rho})$ exists and is unique if $\mathbf{0}$ is inside the convex hull of $\{\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}}) : i = 1, \ldots, m\}$, which is true at least when n is large because of the moment equality (4). Thus, step 4 usually has very good numerical performance. Refer to Chen *et al.* (2002) for more discussion on the implementation and the convergence of the algorithm.

Our proposed estimator in step 5 has the same structure as the IPW estimator with weight $\hat{w}_i$ in replacement of $1/\hat{\pi}(\mathbf{X}_i)$. The calculation of $\hat{w}_i$ does not distinguish models for $\pi(\mathbf{X})$ and

$f(Y \mid \mathbf{X})$, but rather treats them equally as functions of $\mathbf{X}$. When one model is correctly specified, $\hat{w}_i$ does account for this information and leads to consistency of $\hat{q}^L_{\mathrm{MR}}$. For the IPW estimator, because of IPW, the numerical performance can be quite unstable when the estimated values of $\pi(\mathbf{X})$ for some complete cases are close to 0. Our proposed estimator considerably mitigates this issue. The maximization of $\Pi^m_{i=1} w_i$ greatly prevents the occurrence of extreme weights, and thus leads to more stable numerical performances. A formal numerical investigation of this property for mean regression can be found in Han (2014b).

Computationwise, step 3 has the same computational complexity as the imputation method and step 5 has the same complexity as the IPW method. Compared with the imputation and the IPW approaches, the only additional computation that is needed in our procedure is step 4, which is a convex minimization and thus is not computationally complex. Step 4 can be easily implemented by using a Newton–Raphson-type algorithm (Han, 2014b).

2.3. Quantile regression with missing covariates

We now consider quantile regression with missing covariates. In this case, our parameter of interest $\boldsymbol{\beta}_0(\tau)$ is defined by $Q_\tau(Y \mid \mathbf{X}) = \mathbf{X}^{\mathrm{T}} \boldsymbol{\beta}_0(\tau)$, where $Q_\tau(Y \mid \mathbf{X}) = \inf\{y : P(Y \leqslant y \mid \mathbf{X}) \geqslant \tau\}$ is the τth conditional quantile of Y given $\mathbf{X}$. For ease of notation, we write $\boldsymbol{\beta}_0(\tau)$ to be $\boldsymbol{\beta}_0$. Now, Y is fully observed but certain components of $\mathbf{X}$ are subject to missingness. Write $\mathbf{X} = (\mathbf{X}_1^{\mathrm{T}}, \mathbf{X}_2^{\mathrm{T}})^{\mathrm{T}}$, where $\mathbf{X}_1$ is always observed and $\mathbf{X}_2$ may be missing. Let R be the indicator of observing $\mathbf{X}_2$. Then the observed data are n independent and identically distributed copies of $(R, Y, \mathbf{X}_1^{\mathrm{T}}, R\mathbf{X}_2^{\mathrm{T}})$. The MAR mechanism becomes $P(R = 1 \mid Y, \mathbf{X}) = P(R = 1 \mid Y, \mathbf{X}_1)$. Let $\pi(Y, \mathbf{X}_1)$ denote this probability and $f(\mathbf{X}_2 \mid Y, \mathbf{X}_1)$ the conditional density of $\mathbf{X}_2$ given Y and $\mathbf{X}_1$. Again, assume that there are two sets of models $\mathcal{P} = \{\pi^j(Y, \mathbf{X}_1; \boldsymbol{\alpha}^j) : j = 1, \ldots, J\}$ and $\mathcal{F} = \{f^k(\mathbf{X}_2 \mid Y, \mathbf{X}_1; \boldsymbol{\gamma}^k) : k = 1, \ldots, K\}$ for $\pi(Y, \mathbf{X}_1)$ and $f(\mathbf{X}_2 \mid Y, \mathbf{X}_1)$ respectively.

For this problem, our method is composed of steps that are similar to those discussed in Section 2.2. Step 1 and step 2 still calculate $\hat{\boldsymbol{\alpha}}^j$ and $\hat{\boldsymbol{\gamma}}^k$ respectively, but now based on $\pi^j(Y, \mathbf{X}_1; \boldsymbol{\alpha}^j)$ and $f^k(\mathbf{X}_2 \mid Y, \mathbf{X}_1; \boldsymbol{\gamma}^k)$ instead. Step 3 now calculates $\hat{\boldsymbol{\beta}}^k_L$ as an imputation estimator of $\boldsymbol{\beta}_0$ by solving

$$\frac{1}{n} \sum_{i=1}^{n} \left[R_i \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}) + (1 - R_i) \frac{1}{L} \sum_{l=1}^{L} \mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^k) \psi_\tau\{Y_i - \mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^k)^{\mathrm{T}} \boldsymbol{\beta}\} \right] \approx \mathbf{0},$$

where $\mathbf{X}^l(\boldsymbol{\gamma}^k) = \{\mathbf{X}_1^{\mathrm{T}}, \mathbf{X}_2^l(\boldsymbol{\gamma}^k)^{\mathrm{T}}\}^{\mathrm{T}}$ and $\mathbf{X}_2^l(\boldsymbol{\gamma}^k)$ denotes the lth random draw from $f^k(\mathbf{X}_2 \mid Y, \mathbf{X}_1; \boldsymbol{\gamma}^k)$, $l = 1, \ldots, L$. Step 4 still calculates weights $\hat{w}_i$ on complete cases $i = 1, \ldots, m$, but with $\hat{\mathbf{g}}^L_{1i}(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})$ replaced by

$$\hat{\mathbf{g}}^L_{2i}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_L, \hat{\boldsymbol{\gamma}})^{\mathrm{T}} = \left(\pi_i^1(\hat{\boldsymbol{\alpha}}^1) - \hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1), \ldots, \pi_i^J(\hat{\boldsymbol{\alpha}}^J) - \hat{\theta}^J(\hat{\boldsymbol{\alpha}}^J), \frac{1}{L} \sum_{l=1}^{L} \mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^1) \psi_\tau\{Y_i - \mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^1)^{\mathrm{T}} \hat{\boldsymbol{\beta}}^1_L\} \right.$$

$$\left. - \hat{\boldsymbol{\eta}}^1_L(\hat{\boldsymbol{\beta}}^1_L, \hat{\boldsymbol{\gamma}}^1), \ldots, \frac{1}{L} \sum_{l=1}^{L} \mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^K) \psi_\tau\{Y_i - \mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^K)^{\mathrm{T}} \hat{\boldsymbol{\beta}}^K_L\} - \hat{\boldsymbol{\eta}}^K_L(\hat{\boldsymbol{\beta}}^K_L, \hat{\boldsymbol{\gamma}}^K) \right),$$

where $\hat{\boldsymbol{\beta}}^{\mathrm{T}}_L = \{(\hat{\boldsymbol{\beta}}^1_L)^{\mathrm{T}}, \ldots, (\hat{\boldsymbol{\beta}}^K_L)^{\mathrm{T}}\}$, $\pi_i^j(\boldsymbol{\alpha}^j) = \pi^j(Y_i, \mathbf{X}_{1i}; \boldsymbol{\alpha}^j)$,

$$\hat{\theta}^j(\boldsymbol{\alpha}^j) = \frac{1}{n} \sum_{i=1}^{n} \pi_i^j(\boldsymbol{\alpha}^j),$$

$$\hat{\boldsymbol{\eta}}^k_L(\boldsymbol{\beta}, \boldsymbol{\gamma}^k) = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{L} \sum_{l=1}^{L} \mathbf{X}_i^l(\boldsymbol{\gamma}^k) \psi_\tau\{Y_i - \mathbf{X}_i^l(\boldsymbol{\gamma}^k)^{\mathrm{T}} \boldsymbol{\beta}\} \right].$$

Step 5 calculates our proposed estimator $\hat{\beta}_{\mathrm{MR}}^L$ by solving

$$\sum_{i=1}^{m} \hat{w}_i \mathbf{X}_i \psi_\tau (Y_i - \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}) \approx \mathbf{0}.$$

For the current problem, $\hat{\mathbf{g}}_{2i}^L(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_L, \hat{\boldsymbol{\gamma}})$ in step 4 is constructed by taking $b(Y, \mathbf{X}_1)$ in the equality $E(\pi(Y, \mathbf{X}_1)^{-1}[b(Y, \mathbf{X}_1) - E\{b(Y, \mathbf{X}_1)\}] \mid R = 1) = 0$ to be $\pi^j(Y, \mathbf{X}_1; \boldsymbol{\alpha}^j)$ and $\mathbf{a}^k(Y, \mathbf{X}_1; \boldsymbol{\gamma}^k) = E^k\{\mathbf{X}\psi_\tau(Y - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_0) \mid Y, \mathbf{X}_1; \boldsymbol{\gamma}^k\}$, with $\hat{\boldsymbol{\alpha}}^j$, $\hat{\boldsymbol{\gamma}}^k$ and $\hat{\boldsymbol{\beta}}_L^k$ from steps 1, 2 and 3 plugged in.

2.4. Quantile regression with missing responses

Finally we consider quantile regression of Y on $\mathbf{X}$ where Y is subject to missingness and $\mathbf{X}$ is fully observed. The parameter of interest $\boldsymbol{\beta}_0(\tau)$ is still defined by $Q_\tau(Y \mid \mathbf{X}) = \mathbf{X}^{\mathrm{T}} \boldsymbol{\beta}_0(\tau)$. When the full data vector is $(Y, \mathbf{X}^{\mathrm{T}})$ and Y is missing at random, a simple complete-case analysis leads to a consistent estimator of $\boldsymbol{\beta}_0$. Therefore, we consider a more complex yet practically more important setting where, in addition to Y and $\mathbf{X}$, some auxiliary variables $\mathbf{S}$ are also available. These auxiliary variables are usually not of main study interest and thus do not enter the quantile regression model $Q_\tau(Y \mid \mathbf{X}; \boldsymbol{\beta})$. However, they can help to explain the missingness mechanism and to build a more plausible model for the conditional distribution of Y. The observed data are now n independent and identically distributed copies of $(R, RY, \mathbf{X}^{\mathrm{T}}, \mathbf{S}^{\mathrm{T}})$. The MAR mechanism becomes $P(R = 1 \mid Y, \mathbf{X}, \mathbf{S}) = P(R = 1 \mid \mathbf{X}, \mathbf{S})$. Let $\pi(\mathbf{X}, \mathbf{S})$ denote this probability and $f(Y \mid \mathbf{X}, \mathbf{S})$ the conditional density of Y given $\mathbf{X}$ and $\mathbf{S}$. Again, assume that there are two sets of models $\mathcal{P} = \{\pi^j(\mathbf{X}, \mathbf{S}; \boldsymbol{\alpha}^j) : j = 1, \ldots, J\}$ and $\mathcal{F} = \{f^k(Y \mid \mathbf{X}, \mathbf{S}; \boldsymbol{\gamma}^k) : k = 1, \ldots, K\}$ for $\pi(\mathbf{X}, \mathbf{S})$ and $f(Y \mid \mathbf{X}, \mathbf{S})$ respectively.

Similarly, we can follow the previous steps to derive our proposed estimator. Step 1 and step 2 still calculate $\hat{\boldsymbol{\alpha}}^j$ and $\hat{\boldsymbol{\gamma}}^k$ respectively, but now based on $\pi^j(\mathbf{X}, \mathbf{S}; \boldsymbol{\alpha}^j)$ and $f^k(Y \mid \mathbf{X}, \mathbf{S}; \boldsymbol{\gamma}^k)$ instead. Step 3 now calculates $\hat{\boldsymbol{\beta}}_L^k$ as an imputation estimator of $\boldsymbol{\beta}_0$ by solving

$$\frac{1}{n} \sum_{i=1}^{n} \left[R_i \mathbf{X}_i \psi_\tau (Y_i - \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}) + (1 - R_i) \frac{1}{L} \sum_{l=1}^{L} \mathbf{X}_i \psi_\tau \{Y_i^l(\hat{\boldsymbol{\gamma}}^k) - \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}\} \right] \approx \mathbf{0}, \tag{6}$$

where $\{Y^l(\boldsymbol{\gamma}^k) : l = 1, \ldots, L\}$ is a set of random draws of size L from $f^k(Y \mid \mathbf{X}, \mathbf{S}; \boldsymbol{\gamma}^k)$. Step 4 calculates weights $\hat{w}_i$ on complete cases $i = 1, \ldots, m$ with

$$\hat{\mathbf{g}}_{3i}^L(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_L, \hat{\boldsymbol{\gamma}})^{\mathrm{T}} = \left(\pi_i^1(\hat{\boldsymbol{\alpha}}^1) - \hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1), \ldots, \pi_i^J(\hat{\boldsymbol{\alpha}}^J) - \hat{\theta}^J(\hat{\boldsymbol{\alpha}}^J), \frac{1}{L} \sum_{l=1}^{L} \mathbf{X}_i \psi_\tau \{Y_i^l(\hat{\boldsymbol{\gamma}}^1) - \mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_L^1\} \right.$$

$$\left. - \hat{\boldsymbol{\eta}}_L^1(\hat{\boldsymbol{\beta}}_L^1, \hat{\boldsymbol{\gamma}}^1), \ldots, \frac{1}{L} \sum_{l=1}^{L} \mathbf{X}_i \psi_\tau \{Y_i^l(\hat{\boldsymbol{\gamma}}^K) - \mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_L^K\} - \hat{\boldsymbol{\eta}}_L^K(\hat{\boldsymbol{\beta}}_L^K, \hat{\boldsymbol{\gamma}}^K) \right),$$

where $\pi_i^j(\boldsymbol{\alpha}^j) = \pi^j(\mathbf{X}_i, \mathbf{S}_i; \boldsymbol{\alpha}^j)$,

$$\hat{\theta}^j(\boldsymbol{\alpha}^j) = \frac{1}{n} \sum_{i=1}^{n} \pi_i^j(\boldsymbol{\alpha}^j),$$

$$\hat{\boldsymbol{\eta}}_L^k(\boldsymbol{\beta}, \boldsymbol{\gamma}^k) = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{L} \sum_{l=1}^{L} \mathbf{X}_i \psi_\tau \{Y_i^l(\boldsymbol{\gamma}^k) - \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}\} \right].$$

Step 5 calculates our proposed estimator $\hat{\beta}_{\mathrm{MR}}^L$ by solving

$$\sum_{i=1}^{m} \hat{w}_i \mathbf{X}_i \psi_\tau (Y_i - \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}) \approx \mathbf{0}.$$

In step 4, $\hat{\mathbf{g}}_{3i}^L(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_L, \hat{\boldsymbol{\gamma}})$ is constructed by taking $b(\mathbf{X}, \mathbf{S})$ in the equality $E(\pi(\mathbf{X}, \mathbf{S})^{-1}[b(\mathbf{X}, \mathbf{S}) -$

$E\{b(\mathbf{X}, \mathbf{S})\}] \mid R = 1) = 0$ to be $\pi^j(\mathbf{X}, \mathbf{S}; \boldsymbol{\alpha}^j)$ and $\mathbf{a}^k(\mathbf{X}, \mathbf{S}; \boldsymbol{\gamma}^k) = E^k\{\mathbf{X}\psi_\tau(Y - \mathbf{X}^{\mathrm{T}}\beta_0) \mid \mathbf{X}, \mathbf{S}; \boldsymbol{\gamma}^k\} = \mathbf{X}\{\tau - P^k(Y < \mathbf{X}^{\mathrm{T}}\beta_0 \mid \mathbf{X}, \mathbf{S}; \boldsymbol{\gamma}^k)\}$, with $\hat{\boldsymbol{\alpha}}^j$, $\hat{\boldsymbol{\gamma}}^k$ and $\hat{\beta}_L^k$ from steps 1, 2 and 3 plugged in.

2.5. Some remarks

The covariates $\mathbf{X}$ in scenario (a) for estimating marginal quantiles play the same role as the auxiliary variables $\mathbf{S}$ in scenario (c). Although they are not the variables of main interest, they help to explain the missingness mechanism and to build models for the distribution of the response Y. In scenario (b), for simplicity, we did not consider any auxiliary variables when describing the method proposed. The presence of auxiliary variables in this scenario pertains to a direct application of the current method by adding those variables to all the models.

Although we have considered three scenarios only, the framework proposed covers other practically important situations. For example, the application of our method to the case where both the response of interest and part of the covariates are subject to missingness with or without the presence of auxiliary variables is straightforward. A detailed coverage of this scenario is omitted to avoid redundancy.

The method proposed is closely connected to the calibration idea in the survey sampling literature (e.g. Deville and Särndal (1992) and Wu and Sitter (2001)). The constraints in expression (5) imply that

$$\sum_{i=1}^{m} w_i \pi^j(\mathbf{X}_i; \hat{\boldsymbol{\alpha}}^j) = \hat{\theta}^j(\hat{\boldsymbol{\alpha}}^j) \qquad (j = 1, \ldots, J),$$

$$\sum_{i=1}^{m} w_i \left[\frac{1}{L} \sum_{l=1}^{L} \psi_\tau \{Y_i^l(\hat{\boldsymbol{\gamma}}^k) - \hat{q}_L^k\} \right] = \hat{\eta}_L^k(\hat{q}_L^k, \hat{\boldsymbol{\gamma}}^k) \qquad (k = 1, \ldots, K).$$

Therefore, the weights w_i introduce a calibration on the complete cases so that the weighted average based on the complete cases matches the unweighted average based on the whole sample. Functions $\pi^j(\mathbf{X}; \boldsymbol{\alpha}^j)$ and $E^k\{\psi_\tau(Y - q_0) \mid \mathbf{X}; \boldsymbol{\gamma}^k\}$ serve as the calibration variables here. In survey sampling, calibration is used mainly to improve estimation efficiency by incorporating known population information. In our case, the calibration mainly helps to achieve estimation consistency if one model is correctly specified, with the unknown population information consistently estimated by the unweighted sample average over the whole sample.

In scenarios (b) and (c) a possible issue is the compatibility between the model of interest $Q_\tau(Y \mid \mathbf{X}) = \mathbf{X}^{\mathrm{T}}\beta_0(\tau)$ and the working models for $f(\mathbf{X}_2 \mid Y, \mathbf{X}_1)$ or for $f(Y \mid \mathbf{X}, \mathbf{S})$. The models are incompatible if they do not correspond to a genuine distribution. Since $f(Y \mid \mathbf{X})$ is fully determined by the quantile process (Wei *et al.*, 2012), the compatibility is actually between $f(Y \mid \mathbf{X})$ and the working models. This issue is well known in the missing data literature (e.g. Robins *et al.* (1995), Davidian *et al.* (2005) and Tsiatis (2006)). Tsiatis (2006) discussed some methods to ensure model compatibility, but none of them seems to work very effectively here for quantile estimation, unless the joint distribution of $(Y, \mathbf{X}_1, \mathbf{X}_2)$ in scenario (b) or $(Y, \mathbf{X}, \mathbf{S})$ in scenario (c) is simple, such as a normal distribution. In practice, the compatibility issue usually does not lead to a very serious consequence, because the models for $f(\mathbf{X}_2 \mid Y, \mathbf{X}_1)$ or for $f(Y \mid \mathbf{X}, \mathbf{S})$ are working models and do not have to be correctly specified. Our simulation studies in Section 4 show that the method proposed still has a good numerical performance when no working model is correct, consistent with existing findings for MR estimators for mean regression (e.g. Han (2014b) and Chen and Haziza (2017)).

3. Asymptotic results

In this section, we establish the asymptotic properties of the estimators proposed, including

consistency and asymptotic normality. We focus on scenario (a). Derivation for scenarios (b) and (c) involves only straightforward modifications.

3.1. Scenario (a)

We impose the following regularity conditions.

Condition 1. The parameter space $\mathcal{Q}$ for q is compact and q_0 is in the interior of $\mathcal{Q}$.

Condition 2. $P(Y \leqslant y) = \int_{-\infty}^{y} g(s)\,\mathrm{d}s$ where $0 < g(\cdot) < \infty$ in a neighbourhood of q_0.

Condition 3. q_0 is the unique τth quantile.

Condition 4. $E\|\mathbf{X}\|^4 < \infty$.

Condition 5. $\pi^j(\mathbf{X}; \boldsymbol{\alpha}^j)$ has bounded derivatives in $\mathbf{X}$ up to the second order and is continuously differentiable in $\boldsymbol{\alpha}^j$; $\inf_{\mathbf{X}} \inf_{\boldsymbol{\alpha}^j} \pi^j(\mathbf{X}; \boldsymbol{\alpha}^j) > 0$.

Conditions 1 and 2 are often assumed for quantile estimation; condition 3 guarantees the identifiability; conditions 4 and 5 are needed when formulating Donsker classes for certain sets of functions using empirical process theory. For commonly seen models for $\pi(\mathbf{X})$, such as logistic regression, probit regression or other members of generalized linear models, a sufficient condition for condition 5 is that the support for $\mathbf{X}$ is bounded and the parameter space for $\boldsymbol{\alpha}^j$ is compact. Such a condition does not depend on the correctness of model specification and, similarly to the positivity assumption for propensity score in the causal inference literature (e.g. Rosenbaum and Rubin (1983)), does not impose a stringent restriction for implementation of the proposed method in practice.

The following theorem states that the multiple robustness of the proposed estimator $\hat{q}_{\mathrm{MR}}^L$: $\hat{q}_{\mathrm{MR}}^L$ is consistent if any one of the working parametric models is correctly specified. The proof is given in Appendix A.

Theorem 1. Under conditions 1–5, when $\mathcal{P}$ contains a correctly specified model for $\pi(\mathbf{X})$ or $\mathcal{F}$ contains a correctly specified model for $f(Y \mid \mathbf{X})$, $\hat{q}_{\mathrm{MR}}^L \to^{\mathrm{p}} q_0$ as $n \to \infty$.

We now derive the asymptotic distribution of $\hat{q}_{\mathrm{MR}}^L$ when $\mathcal{P}$ contains a correctly specified model for $\pi(\mathbf{X})$, the typical setting for developing semiparametric theory for missing data analysis (e.g. Tsiatis (2006)). We need further to impose the following regularity conditions.

Condition 6. Without loss of generality, we assume that $\pi^1(\mathbf{X}; \boldsymbol{\alpha}^1)$ is the correct model for $\pi(\mathbf{X})$ and

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}^1 - \boldsymbol{\alpha}_*^1) = n^{-1/2} \sum_{i=1}^{n} E(\boldsymbol{\Phi}_1^{\otimes 2})^{-1} \boldsymbol{\Phi}_{1i} + o_p(1),$$

where $\boldsymbol{\Phi}_1$ is the corresponding score function.

Condition 7. The matrix

$$\boldsymbol{G}_1^L = E\left\{ \frac{\mathbf{g}_1^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)^{\otimes 2}}{\pi^1(\mathbf{X}; \boldsymbol{\alpha}_*^1)} \right\}$$

is invertible, where $\mathbf{g}_1^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)$ is given by expression (9) in Appendix A.

Theorem 2. Under conditions 1–7, $\sqrt{n}(\hat{q}_{\mathrm{MR}}^L - q_0)$ has an asymptotic normal distribution with mean 0 and variance $\mathrm{var}(Z_1)$, where

$$Z_1 = g(q_0)^{-1}\{Q_1^L(\alpha_0^1) - E(Q_1^L \Phi_1^T)E(\Phi_1^{\otimes 2})^{-1}\Phi_1\},$$

$$Q_1^L(\alpha^1) = \frac{R}{\pi^1(\alpha^1)}\{\psi_\tau(Y - q_0) - A_1^L(G_1^L)^{-1}g_1^L(\alpha_*, q_*, \gamma_*)\},$$

$$A_1^L = E\left\{\frac{\psi_\tau(Y - q_0)}{\pi^1(\alpha_0^1)}g_1^L(\alpha_*, g_*, \gamma_*)^T\right\}.$$

3.2.   Scenarios (b) and (c)

Because of the similarity between the regularity conditions that are needed for scenarios (b) and (c) and those for scenario (a), we list only the conditions for scenario (b) and then the results for scenarios (b) and (c). The sketched proofs for the results for scenario (b) are given in Appendix A and the proofs for scenario (c) have been omitted.

Condition 8.   The parameter space $\mathcal{B}$ for β is compact and β_0 is in the interior of $\mathcal{B}$.

Condition 9. $P(Y \leqslant y) = \int_{-\infty}^{y} g(s)\,\mathrm{d}s$ where $0 < g(X^T\beta) < \infty$ for β in a neighbourhood of β_0.

Condition 10. β_0 is the unique τth quantile regression coefficient.

Condition 11. $E\|X\|^4 < \infty$ and $E\{g(X^T\beta_0)XX^T\}$ is invertible.

Condition 12. $\pi^j(Y, X_1; \alpha^j)$ has bounded derivatives in (Y, X_1) up to the second order and is continuously differentiable in α^j; $\inf_{(Y,X_1)} \inf_{\alpha^j} \pi^j(Y, X_1; \alpha^j) > 0$.

Condition 13.   Assume that $\pi^1(Y, X_1; \alpha^1)$ is the correct model for $\pi(Y, X_1)$ and

$$\sqrt{n}(\hat{\alpha}^1 - \alpha_*^1) = n^{-1/2}\sum_{i=1}^{n} E(\Phi_2^{\otimes 2})^{-1}\Phi_{2i} + o_p(1),$$

where Φ_2 is the corresponding score function.

Condition 14.   The matrix

$$G_2^L = E\left\{\frac{g_2^L(\alpha_*, \beta_*, \gamma_*)^{\otimes 2}}{\pi^1(Y, X_1; \alpha_*^1)}\right\}$$

is invertible, where $g_2^L(\alpha_*, \beta_*, \gamma_*)$ is given by equation (29) in Appendix A.

For scenario (b) we have the following theoretical results.

Theorem 3.   Under conditions 8–12, when $\mathcal{P}$ contains a correctly specified model for $\pi(Y, X_1)$ or $\mathcal{F}$ contains a correctly specified model for $f(X_2 \,|\, Y, X_1)$, $\hat{\beta}_{MR}^L \to^p \beta_0$ as $n \to \infty$.

Theorem 4.   Under conditions 8–14 $\sqrt{n}(\hat{\beta}_{MR}^L - \beta_0)$ has an asymptotic normal distribution with mean $\mathbf{0}$ and variance $\mathrm{var}(Z_2)$, where

$$Z_2 = E\{g(X^T\beta_0)XX^T\}^{-1}\{Q_2^L(\alpha_0^1) - E(Q_2^L \Phi_2^T)E(\Phi_2^{\otimes 2})^{-1}\Phi_2\},$$

$$Q_2^L(\alpha^1) = \frac{R}{\pi^1(\alpha^1)}\{X\psi_\tau(Y - X^T\beta_0) - A_2^L(G_2^L)^{-1}g_2^L(\alpha_*, \beta_*, \gamma_*)\},$$

$$A_2^L = E\left\{\frac{X\psi_\tau(Y - X^T\beta_0)}{\pi^1(\alpha_0^1)}g_2^L(\alpha_*, \beta_*, \gamma_*)^T\right\}.$$

For scenario (c) we have the following results.

Theorem 5. Under conditions similar to conditions 8–12 when $\mathcal{P}$ contains a correctly specified model for $\pi(\mathbf{X}, \mathbf{S})$ or $\mathcal{F}$ contains a correctly specified model for $f(Y \mid \mathbf{X}, \mathbf{S})$, $\hat{\beta}_{\mathrm{MR}}^L \to^{\mathrm{p}} \beta_0$ as $n \to \infty$.

Theorem 6. Under conditions similar to conditions 8–14 $\sqrt{n}(\hat{\beta}_{\mathrm{MR}}^L - \beta_0)$ has an asymptotic normal distribution with mean $\mathbf{0}$ and variance $\mathrm{var}(\mathbf{Z}_3)$, where

$$\mathbf{Z}_3 = E\{g(\mathbf{X}^{\mathrm{T}}\beta_0)\mathbf{X}\mathbf{X}^{\mathrm{T}}\}^{-1}\{\mathbf{Q}_3^L(\alpha_0^1) - E(\mathbf{Q}_3^L\mathbf{\Phi}_3^{\mathrm{T}})E(\mathbf{\Phi}_3^{\otimes 2})^{-1}\mathbf{\Phi}_3\},$$

$$\mathbf{Q}_3^L(\alpha^1) = \frac{R}{\pi^1(\alpha^1)}\{\mathbf{X}\psi_\tau(Y - \mathbf{X}^{\mathrm{T}}\beta_0) - \mathbf{A}_3^L(\mathbf{G}_3^L)^{-1}\mathbf{g}_3^L(\alpha_*, \beta_*, \gamma_*)\},$$

$$\mathbf{A}_3^L = E\left\{\frac{\mathbf{X}\psi_\tau(Y - \mathbf{X}^{\mathrm{T}}\beta_0)}{\pi^1(\alpha_0^1)}\mathbf{g}_3^L(\alpha_*, \beta_*, \gamma_*)^{\mathrm{T}}\right\},$$

$$\mathbf{g}_3^L(\alpha_*, \beta_*, \gamma_*)^{\mathrm{T}} = \left(\pi^1(\alpha_*^1) - \theta_*^1, \ldots, \pi^J(\alpha_*^J) - \theta_*^J, \frac{1}{L}\sum_{l=1}^L \mathbf{X}\psi_\tau\{Y^l(\gamma_*^1) - \mathbf{X}^{\mathrm{T}}\beta_*^1\} - \eta_*^1,\right.$$

$$\left.\ldots, \frac{1}{L}\sum_{l=1}^L \mathbf{X}\psi_\tau\{Y^l(\gamma_*^K) - \mathbf{X}^{\mathrm{T}}\beta_*^K\} - \eta_*^K\right),$$

and $\mathbf{\Phi}_3$ is the score function corresponding to the estimation of α^1.

In all the asymptotic distributions we have kept L finite, where L is the number of random draws from the data distribution models. The asymptotic distributions when $L \to \infty$ can be easily derived on the basis of the results presented in this section.

Compared with existing methods for quantile estimation with missing data, all of which allow only one model for the probability of missingness and/or one model for the data distribution, the consistency results in this section show that our proposed method by accommodating multiple working models has more protection on estimation consistency and thus provides a highly desirable alternative. No existing method can achieve the same level of robustness as ours. Results on asymptotic distribution show how the asymptotic variance depends on the multiple working models. Similarly to the mean regression case (e.g. Han (2014b)), the dependence is rather complex, which makes a general comparison of efficiency between estimators using different working models very difficult. But the derivations provide some guidance on how the empirical process theory is applied, and the results give formulations of the asymptotic variance, both of which are important for investigations of efficiency under some specific situations such as those which were considered in Han (2018) for mean regression.

For mean regression with missing data, the MR estimators are locally efficient (Han and Wang, 2013; Han, 2014b, 2016a), i.e. these estimators achieve the semiparametric efficiency bound when both the probability of missingness and the data distribution are correctly modelled. For quantile regression with non-identically distributed error terms that is considered in this paper, as pointed out in Chen *et al.* (2015), the semiparametric efficiency bound has not been derived in the literature, and thus it is unclear whether our proposed estimators are locally efficient. Deriving the efficiency bound is an interesting yet challenging topic that deserves future investigation.

4. Simulation studies

4.1. Quantile regression with missing responses
We first consider quantile regression with missing responses when the covariates and some

auxiliary variables are fully observed: the scenario in Section 2.4. The simulation model has three covariates: $X_1 \sim$ exponential(1), $X_2 \sim N(0, 1)$ and $X_3 \sim$ Bernoulli(0.5). Given the covariates, the response Y is generated as $Y = -1 + X_1 + X_2 + X_3 + (1 + X_1)\epsilon_Y$ and, given the covariates and response, the auxiliary variable S is generated as $S = -1 + X_1 + X_2 + X_3 - Y + (1 + X_1)\epsilon_S$, where $\epsilon_Y \sim N(0, 1)$, $\epsilon_S \sim N(0, 1)$ and $X_1, X_2, X_3, \epsilon_Y$ and ϵ_S are mutually independent. Here, the error distribution of Y is heteroscedastic. The missingness mechanism for Y is given by logit$\{P(R = 1 \mid Y, \mathbf{X}, S)\} = 0.5 + 0.25X_1 + 0.5X_2 + 0.25X_3 + 0.25S$, which leads to a rate of missingness of approximately 33%. For this simulation model, we have $Q_\tau(Y \mid \mathbf{X}) = (1, \mathbf{X}^T)\boldsymbol{\beta}_0(\tau)$ with $\boldsymbol{\beta}_0(\tau) = (-1 + Q_\tau(\epsilon_Y), 1 + Q_\tau(\epsilon_Y), 1, 1)^T$. Also, it is easy to verify that $Y \mid \mathbf{X}, S$ has a normal distribution with mean $-1 + X_1 + X_2 + X_3 - 0.5S$ and variance $(1 + X_1)^2/2$.

The correct models for $\pi(\mathbf{X}, S)$ and $f(Y \mid \mathbf{X}, S)$ are given by logit$\{\pi^1(\boldsymbol{\alpha}^1)\} = \alpha_1^1 + \alpha_2^1 X_1 + \alpha_3^1 X_2 + \alpha_4^1 X_3 + \alpha_5^1 S$ and

$$f^1(\boldsymbol{\gamma}^1) = \frac{1}{\sqrt{(2\pi)}(\gamma_6^1 + \gamma_7^1 X_1)} \exp\left[-\frac{\{Y - (\gamma_1^1 + \gamma_2^1 X_1 + \gamma_3^1 X_2 + \gamma_4^1 X_3 + \gamma_5^1 S)\}^2}{2(\gamma_6^1 + \gamma_7^1 X_1)^2} \right]$$

respectively. The following two incorrect models are also considered in our simulation studies:

$$\text{logit}\{\pi^2(\boldsymbol{\alpha}^2)\} = \alpha_1^2 + \alpha_2^2 X_1 + \alpha_3^2 X_3,$$

$$f^2(\boldsymbol{\gamma}^2) = \frac{1}{\sqrt{(2\pi)}\gamma_3^2} \exp\left[-\frac{\{Y - (\gamma_1^2 + \gamma_2^2 S)\}^2}{2(\gamma_3^2)^2} \right].$$

Note that $f^2(\boldsymbol{\gamma}^2)$ corresponds to the normal density function with mean $\gamma_1^2 + \gamma_2^2 S$ and constant variance $(\gamma_3^2)^2$. We consider two sample sizes $n = 200$ and $n = 500$ and the results are summarized on the basis of 1000 replications. We have done simulations based on both $L = 10$ and $L = 50$ but we report only results based on $L = 10$ as $L = 50$ led to very similar results. Tables 1, 2 and 3 in the on-line supplementary material contain results for $\tau = 0.25$, 0.5, 0.75 respectively. The IPW estimator $\hat{\boldsymbol{\beta}}_{\text{IPW}}$ solving $n^{-1}\Sigma_{i=1}^n R_i \hat{\pi}(\mathbf{X}_i, S_i)^{-1} \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^T \boldsymbol{\beta}) \approx \mathbf{0}$ and the imputation estimator $\hat{\boldsymbol{\beta}}_{\text{IM}}^L$ solving equation (6) are also calculated to serve as the benchmark for comparison.

When only one model for $\pi(\mathbf{X}, S)$ is used, $\hat{\boldsymbol{\beta}}_{\text{MR}}^L$ (i.e. MR-1000 and MR-0100 in the tables) has a performance that is very similar to the corresponding $\hat{\boldsymbol{\beta}}_{\text{IPW}}$ (IPW-1000 and IPW-0100): both have negligible bias and similar root-mean-square error (RMSE) and median absolute error (MAE) when the correct model $\pi^1(\boldsymbol{\alpha}^1)$ is used, and similar bias, RMSE and MAE when the incorrect model $\pi^2(\boldsymbol{\alpha}^2)$ is used. When only the correct model $f^1(\boldsymbol{\gamma}^1)$ is used, $\hat{\boldsymbol{\beta}}_{\text{MR}}^L$ (MR-0010) and $\hat{\boldsymbol{\beta}}_{\text{IM}}^L$ (IM-0010) both have negligible bias, but the former has slightly larger RMSE and MAE. When only the incorrect model $f^2(\boldsymbol{\gamma}^2)$ is used, $\hat{\boldsymbol{\beta}}_{\text{IM}}^L$ (IM-0001) has large bias due to its inconsistency, but $\hat{\boldsymbol{\beta}}_{\text{MR}}^L$ (MR-0001) has surprisingly small bias as if it was consistent. An explanation of this small bias will be given below.

When two or more models are used to construct $\hat{\boldsymbol{\beta}}_{\text{MR}}^L$, the results suggest that $\hat{\boldsymbol{\beta}}_{\text{MR}}^L$ has negligible bias whenever one model is correctly specified, either for $\pi(\mathbf{X}, S)$ or for $f(Y \mid \mathbf{X}, S)$. This is in full agreement with our theory that $\hat{\boldsymbol{\beta}}_{\text{MR}}^L$ is MR. In addition, when $\hat{\boldsymbol{\beta}}_{\text{MR}}^L$ is consistent, the particular models that were used seem to have little effect on the RMSE and MAE. When only the incorrect models $\pi^2(\boldsymbol{\alpha}^2)$ and $f^2(\boldsymbol{\gamma}^2)$ are used, $\hat{\boldsymbol{\beta}}_{\text{MR}}^L$ (MR-0101) has surprisingly small bias as if it was consistent.

The surprisingly small bias of $\hat{\boldsymbol{\beta}}_{\text{MR}}^L$ in the cases of MR-0001 and MR-0101 may be explained as follows. The equation $E(\pi(\mathbf{X}, S)^{-1}[b(\mathbf{X}, S) - E\{b(\mathbf{X}, S)\}] \mid R = 1) = 0$ holds for an arbitrary function $b(\mathbf{X}, S)$, assuming that relevant expectations exist. When $b(\mathbf{X}, S)$ is taken to be $E^2\{\mathbf{X}\psi_\tau(Y - \mathbf{X}^T\boldsymbol{\beta}) \mid \mathbf{X}, S; \boldsymbol{\gamma}^2\}$, the conditional expectation of the estimating function

$\mathbf{X}\psi_\tau(Y - \mathbf{X}^T\beta)$ under the incorrect working model $f^2(\gamma^2)$, the $\hat{w}_i$ that were used for calculating $\hat{\beta}^L_{\mathrm{MR}}$ are still derived on the basis of a legitimate empirical version of the above equation. Although in this case consistency of $\hat{\beta}^L_{\mathrm{MR}}$ can no longer be theoretically established, the numerical performance may still be reasonably good because, even being incorrect, the working model for $f(Y\,|\,\mathbf{X}, S)$ still captures a large degree of dependence of Y on $(\mathbf{X}, S)$, and the resulting $\hat{w}_i$ make good use of this modelled dependence through calibration. It is the calibration of the estimating functions of β that helps to reduce the bias of $\hat{\beta}^L_{\mathrm{MR}}$. Calibrating moments of covariates alone may not help much. The small bias of MR estimators using incorrectly specified data distribution models has also been observed for mean regression (Han, 2014b, 2016a; Chen and Haziza, 2017).

4.2.   *Quantile regression with missing covariates*

We now consider quantile regression with missing covariates: the scenario in Section 2.3. There are two covariates: $X_1 \sim$ exponential(0.2) and $X_2 \sim N(0, 1)$. Given the covariates, the response Y is generated as $Y = 1 + X_1 + X_2 + (1 + X_1)\epsilon_Y$, where $\epsilon_Y \sim N(0, 1)$ and X_1, X_2 and ϵ_Y are mutually independent. Here the error distribution of Y is again heteroscedastic. The missingness mechanism for X_2 is given by logit$\{P(R = 1\,|\,Y, \mathbf{X})\} = -2 + 0.5X_1 + 0.25Y$, under which the rate of missingness is approximately 38%. For this simulation model, we have $Q_\tau(Y\,|\,\mathbf{X}) = (1, \mathbf{X}^T)\beta_0(\tau)$ with $\beta_0(\tau) = (1 + Q_\tau(\epsilon_Y), 1 + Q_\tau(\epsilon_Y), 1)^T$. It is easy to verify that $X_2\,|\,Y, X_1$ has a normal distribution with mean $(-1 - X_1 + Y)/\{1 + (1 + X_1)^2\}$ and variance $(1 + X_1)^2/\{1 + (1 + X_1)^2\}$.

For $\pi(X_1, Y)$ we consider two working models logit$\{\pi^1(\alpha^1)\} = \alpha^1_1 + \alpha^1_2 X_1 + \alpha^1_3 Y$ and logit$\{\pi^2(\alpha^2)\} = \alpha^2_1 + \alpha^2_2 Y$, where $\pi^2(\alpha^2)$ is incorrectly specified. For $f(X_2\,|\,Y, X_1)$, a correctly specified model would replace all numbers in the mean and variance by unknown parameters, but such a complex model would rarely, if ever, be considered in practice. Besides, estimation of those parameters would be difficult because of the complex dependence on Y and X_1 and that the variance needs to be positive. Instead, in practice one would more probably specify a model where the mean depends on Y and X_1 linearly and the variance is constant. Therefore, in the simulation we consider an incorrect model

$$f^2(\gamma^2) = \frac{1}{\sqrt{(2\pi)\gamma^2_4}} \exp\left[-\frac{\{X_2 - (\gamma^2_1 + \gamma^2_2 X_1 + \gamma^2_3 Y)\}^2}{2(\gamma^2_4)^2}\right],$$

which is an ordinary least square regression of X_2 on Y and X_1. To illustrate the multiple-robustness property of the estimator proposed, we also consider a correct model $f^1(\gamma^1) = f(X_2\,|\,Y, X_1)$ with γ^1 completely known. This correct model is also used for the imputation estimator so it is still a fair comparison between our proposed estimator and the imputation estimator.

Table 4 in the on-line supplementary material contains simulation results for $\tau = 0.25$ and $\tau = 0.75$ summarized on the basis of 1000 replications with $n = 500$ and $L = 10$. It is seen that, using the same incorrect model $\pi^2(\alpha^2)$ or $f^2(\gamma^2)$, the IPW estimator IPW-0100 or the imputation estimator IM-0001 has a considerably worse performance than the MR estimator MR-0100 or MR-0001 respectively. In other cases the MR estimators have small bias, even for MR-0101 where only the two incorrect models $\pi^2(\alpha^2)$ and $f^2(\gamma^2)$ are used, which is consistent with findings from the missing response setting and from existing literature on mean regression with missing data.

5.   Data application

As an application of the method proposed, we analyse the data that were collected on 2139

human immunodeficiency virus infected subjects enrolled in ACTG protocol 175 (Hammer *et al*., 1996). The ACTG 175 study evaluates treatment with either a single nucleoside or two nucleosides in human immunodeficiency virus infected subjects whose CD4 cell counts (a measure of immunologic status) are from 200 to 500 mm^{-3}. Following the analysis of Davidian *et al*. (2005) and Zhang *et al*. (2008), we consider two arms for the treatment: the arm with standard zidovudine, ZDV, monotherapy (ZDV only) and the arm with three newer treatments (ZDV+ didanosine, ddI, ZDV+ zalcitabine, ddC, and ddI only). The two arms have 532 and 1607 subjects respectively. We study the treatment arm effect (trt, $0 \equiv$ ZDV only) on the τth quantile of the CD4 cell count measured at 96 ± 5 weeks post baseline, CD4$_{96}$, adjusting for the baseline CD4 cell count, CD4$_0$, and certain baseline characteristics, including continuous covariates age (*age*, years) and weight (*weight*, kilograms) and binary covariates race (*race*; $0 \equiv$ white), gender (*gender*; $0 \equiv$ female), antiretroviral history (*history*; $0 \equiv$ naive; $1 \equiv$ experienced) and whether the subject is off treatment before 96 weeks (*offtrt*; $0 \equiv$ no). Therefore, we want to fit the linear quantile regression model

$$Q_\tau(\text{CD4}_{96} \,|\, \mathbf{X}) = \beta_1(\tau) + \beta_2(\tau)\text{trt} + \beta_3(\tau)\text{CD4}_0 + \beta_4(\tau)\text{age} + \beta_5(\tau)\text{weight} + \beta_6(\tau)\text{race}$$
$$+ \beta_7(\tau)\text{gender} + \beta_8(\tau)\text{history} + \beta_9(\tau)\text{offtrt}.$$

The data can be found in the R package `speff2trial`. The average age of the subjects is 35 years old with a standard deviation 8.7 years old. There are 1522 white subjects and 617 non-whites, and 1171 males and 368 females. Among the patients, 1253 have antiretroviral history, and 776 are off treatment before 96 weeks.

The variable CD4$_{96}$ is missing for approximately 37% of the subjects because of dropout during the study. However, at baseline and during the follow-up, full measurements on additional variables correlated with CD4$_{96}$ are obtained. These include the CD4 cell count at 20 ± 5 weeks, CD4$_{20}$, and the CD8 cell count, another measure of immunologic status, at both the baseline, CD8$_0$, and 20 ± 5 weeks, CD8$_{20}$. We use these as the auxiliary variables. As argued by Davidian *et al*. (2005), it is reasonable to assume that CD4$_{96}$ is missing at random.

Fig. 1 contains histograms for both CD4$_{96}$ and its logarithm. Although the distribution of CD4$_{96}$ is apparently right skewed, the log-transformation does not result in any better symmetry, and thus both make the normality assumption inappropriate. For analysis, one could assume either a left-truncated normal distribution for CD4$_{96}$ or a right-truncated normal distribution for log(CD4$_{96}$), and there is no apparent reason for favouring one of them. With these two candidate models, the imputation method requires choosing one from them, but our proposed method can simultaneously accommodate both. In our analysis, we assume that CD4$_{96}$ has a normal distribution left truncated at 0 and log(CD4$_{96}$) has a normal distribution right truncated at the logarithm of the maximum of observed CD4$_{96}$, where both means depend on the main effects of the eight covariates $\mathbf{X}$ and the three auxiliary variables $\mathbf{S}$. The truncated regression models are fitted by using the R package `truncreg`. The $\pi(\mathbf{X}, \mathbf{S})$ is modelled by a logistic regression containing all main effects of $\mathbf{X}$ and $\mathbf{S}$.

Final data analysis results for $\tau = 0.25, 0.50, 0.75$ are summarized in Table 1. To make comparisons, results for the IPW estimator, the imputation estimators with truncated normal distributions for CD4$_{96}$ (imputation estimator 1) and log(CD4$_{96}$) (imputation estimator 2), and the complete-case analysis are also included. For the two imputation estimators and our proposed estimator, the number of random draws is taken to be $L = 20$. The standard errors of all the estimators are calculated by using the bootstrap method with 200 resamplings.

From Table 1, for all three values of τ and all methods under comparison, patients receiving the three newer treatments have a significantly higher CD4 cell count at 96 ± 5 weeks adjusting
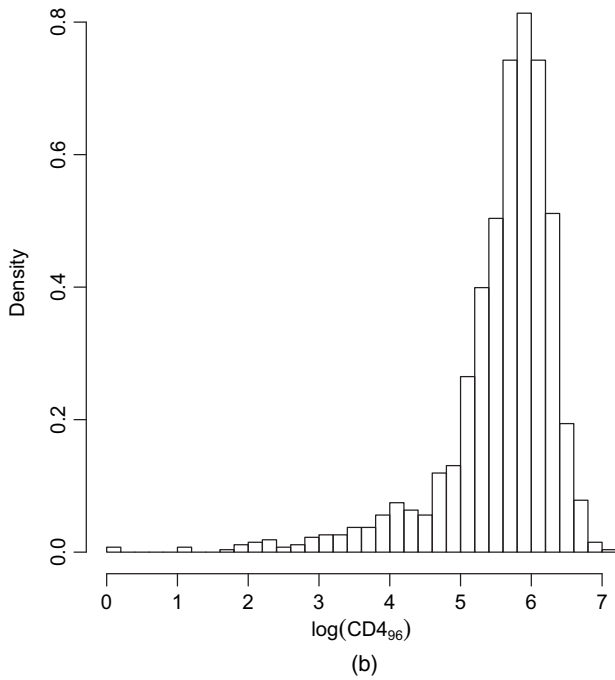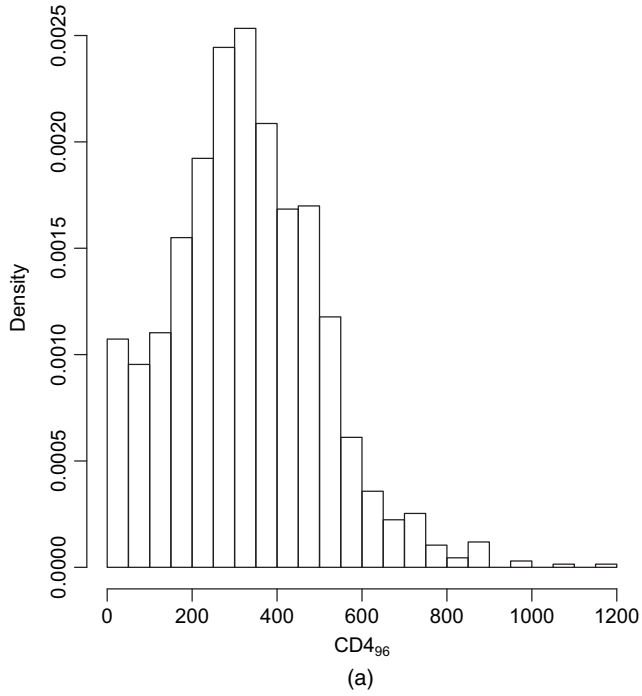
(a)

(b)

Fig. 1. Histograms of (a) CD4$_{96}$ and (b) log(CD4$_{96}$) based on complete cases for the ACTG 175 data

Table 1. Analysis results for the ACTG 175 data ($n = 2139$)†

Covariate	Results for MR estimator			Results for IPW estimator			Results for imputation estimator 1			Results for imputation estimator 2			Results for complete-case analysis		
	Est	SE	p	Est	SE	p	Est	SE	p	Est	SE	p	Est	SE	p
$\tau = 0.25$															
intercept	−6.80	44.79	0.88	−9.21	46.70	0.84	−16.53	29.95	0.58	0.73	29.99	0.98	−41.28	40.11	0.30
trt	60.99	10.96	<0.01	60.50	11.63	<0.01	57.80	7.92	<0.01	50.96	7.76	<0.01	65.85	9.98	<0.01
$CD4_0$	0.70	0.04	<0.01	0.70	0.04	<0.01	0.69	0.03	<0.01	0.63	0.03	<0.01	0.73	0.03	<0.01
age	−0.27	0.63	0.66	−0.27	0.70	0.70	−0.01	0.48	0.98	−0.14	0.41	0.73	−0.12	0.64	0.86
weight	0.29	0.43	0.49	0.31	0.49	0.54	0.23	0.29	0.42	0.23	0.31	0.46	0.41	0.41	0.32
race	−15.61	10.85	0.15	−14.82	11.48	0.20	−12.25	7.83	0.12	−8.04	8.10	0.32	−16.06	10.01	0.11
gender	−14.96	12.23	0.22	−14.58	12.71	0.25	−10.03	8.54	0.24	−7.26	8.92	0.42	−9.99	11.04	0.37
history	−42.90	11.77	<0.01	−42.83	12.06	<0.01	−37.99	7.41	<0.01	−29.09	7.82	<0.01	−39.41	9.83	<0.01
offtrt	−96.26	13.06	<0.01	−96.11	13.73	<0.01	−90.77	8.83	<0.01	−110.11	9.22	<0.01	−96.56	14.27	<0.01
$\tau = 0.50$															
intercept	112.44	32.02	<0.01	105.56	31.35	<0.01	74.21	31.99	0.02	53.43	34.72	0.12	80.23	38.66	0.04
trt	60.56	12.06	<0.01	59.70	12.26	<0.01	60.77	8.57	<0.01	59.35	9.23	<0.01	67.24	11.37	<0.01
$CD4_0$	0.73	0.04	<0.01	0.73	0.04	<0.01	0.75	0.03	<0.01	0.77	0.04	<0.01	0.76	0.04	<0.01
age	−1.18	0.57	0.04	−1.20	0.59	0.04	−0.66	0.54	0.22	−0.64	0.55	0.24	−1.02	0.61	0.10
weight	0.28	0.32	0.37	0.23	0.34	0.50	0.21	0.32	0.51	0.47	0.36	0.20	0.21	0.40	0.59
race	−25.76	10.76	0.02	−22.95	10.52	0.03	−16.65	8.10	0.04	−16.90	9.06	0.06	−23.34	10.20	0.02
gender	−28.43	14.10	0.04	−16.33	14.26	0.25	−12.90	11.81	0.27	−17.15	12.20	0.16	−11.57	13.57	0.39
history	−52.53	10.03	<0.01	−53.55	11.12	<0.01	−45.15	7.89	<0.01	−47.67	8.72	<0.01	−50.12	10.07	<0.01
offtrt	−89.63	10.37	<0.01	−89.04	11.43	<0.01	−89.81	9.24	<0.01	−113.69	9.91	<0.01	−87.40	12.44	<0.01
$\tau = 0.75$															
Intercept	128.02	42.87	<0.01	123.26	42.39	<0.01	103.53	29.74	<0.01	47.67	35.11	0.17	84.30	39.59	0.03
trt	51.16	14.41	<0.01	50.72	14.67	<0.01	60.06	9.48	<0.01	68.57	9.92	<0.01	55.58	11.40	<0.01
$CD4_0$	0.76	0.06	<0.01	0.77	0.06	<0.01	0.81	0.04	<0.01	0.94	0.05	<0.01	0.82	0.05	<0.01
age	0.35	0.56	0.53	0.17	0.58	0.77	0.05	0.44	0.91	−0.31	0.51	0.54	0.16	0.56	0.78
weight	0.29	0.35	0.41	0.35	0.38	0.36	0.47	0.29	0.11	0.90	0.30	<0.01	0.60	0.38	0.12
race	−35.29	12.73	0.01	−34.47	12.55	0.01	−27.02	9.54	<0.01	−28.76	10.88	0.01	−32.87	10.50	<0.01
gender	−8.64	13.39	0.52	−4.43	13.90	0.75	−11.72	11.79	0.32	−19.99	13.56	0.14	−6.48	13.69	0.64
history	−40.41	10.24	<0.01	−40.75	10.89	<0.01	−43.15	7.75	<0.01	−44.63	8.76	<0.01	−44.61	9.43	<0.01
offtrt	−88.40	14.74	<0.01	−85.73	15.50	<0.01	−89.00	11.30	<0.01	−105.71	13.87	<0.01	−79.79	14.59	<0.01

†Standard errors are calculated based on the bootstrap method with 200 resamplings. $L = 20$ for our proposed MR estimator and for the imputation estimators. Est, estimated value; SE, standard error; p, p-value.

for the baseline CD4 cell count and other covariates. In other words, the three newer treatments significantly slow the progression of human immunodeficiency virus disease compared with the treatment of ZDV alone. Our proposed method and the IPW method produce very similar estimates for the treatment arm effect, and the estimated effect is smaller for patients whose CD4 cell count at 96 ± 5 weeks is at the third quartile. Imputation estimator 1 fails to catch the difference in treatment arm effect for different quantiles, and imputation estimator 2 gives an increasing estimated effect as τ varies from 0.25 to 0.75: a trend that is opposite to what the MR and the IPW methods reveal. In addition, the two imputation estimates for some regression coefficients are quite different occasionally, showing the sensitivity of the imputation method to the selection of working models. The complete-case analysis seems to overestimate the treatment arm effect.

Given the disparity between the two imputation estimates and between them and the MR and the IPW estimates, and also given the good performance of the MR estimator under complete model misspecification as shown in our simulation studies, results based on our MR method should be more trustworthy. The closeness between the MR and the IPW estimates indicates the suitability of the logistic regression that we used to model $\pi(\mathbf{X}, \mathbf{S})$. However, when $\tau = 0.5$, the MR method detects significance in the effect of gender that the IPW method fails to detect.

6. Discussion

We have proposed a general framework for quantile estimation with missing data and have investigated the estimation method in three scenarios covering both marginal quantile estimation and quantile regression with missing response or covariates. The estimators proposed are robust against possible model misspecifications. The method proposed can be easily generalized to many other missing data scenarios that were not discussed here. As shown and explained in Section 4, the estimators proposed still have relatively good performance when no model is correctly specified due to the nature of calibration. This is very appealing in practice, as the true data-generating process is usually unknown, which makes the conclusions drawn based on our method more trustworthy compared with those based on methods that are sensitive to model misspecifications.

It is well known that the IPW approach is sensitive to near-zero values of the estimated missingness probability (e.g. Kang and Schafer (2007), Robins *et al.* (2007) and Cao *et al.* (2009)). Our proposed estimators mitigate this issue. First, unlike the IPW approach where the inverse of the estimated probability of missingness is used as the weight, our method uses the probability of missingness as the calibration variable. Second, through maximizing $\Pi_{i=1}^{m} w_i$ where the w_i are positive and sum to 1, the occurrence of extreme weights is prevented. This is because $\Pi_{i=1}^{m} w_i$ increases if the values of w_i become more evenly distributed rather than concentrating on a few subjects. Therefore, it is unlikely that some subjects receive extremely large weights that dominate others. Han (2014b) provided a detailed numerical investigation of this property for mean regression with missing responses.

Our method is more general than the recent methods in Wei *et al.* (2012) and Chen *et al.* (2015). The method of Wei *et al.* (2012) deals with missing covariate problems where the missingness depends only on the other fully observed covariates but not on the response, whereas our method allows the dependence on both. The method of Chen *et al.* (2015) requires the fully observed components of the data vector to have the same distribution across all subjects. This requirement is not needed by our method. Because of these limitations, neither the method of Wei *et al.* (2012) nor that of Chen *et al.* (2015) applies to the simulation settings that were considered in Section 4.

For the method proposed, the dimension of $\hat{\rho}$ minimizing equation (3) is the same as the number of working models used. The calculation of $\hat{\rho}$ is affected when this number increases, especially if some models are highly correlated. As a result, the numerical performance of $\hat{\beta}_{\mathrm{MR}}^L$ may be affected. Therefore, although theoretically the method proposed can accommodate an arbitrary number of models as long as this number does not increase with the sample size, for the purpose of numerical performance, the models should still be carefully chosen so that only the most reasonable are used. Detailed investigation on how the numerical performance changes as the number of models increases will be a topic of future research.

In this paper, the working models were all parametric. As one referee pointed out, the robustness could be further improved by using working models that are non-parametric or built on the basis of machine learning techniques. In this case, the final estimator should still be consistent when one working model is correctly specified, but the rate of convergence for the working models will affect the asymptotic distribution (e.g. Robins *et al.* (1995)). We shall make a future investigation on this topic.

Acknowledgements

Appendix A

A.1. Proof of theorem 1

Under condition 3, $E\{\psi_\tau(Y-q)\}=0$ has q_0 as the unique solution, which guarantees the identifiability of q_0. On the basis of theorem 5.9 in van der Vaart (1998), we need to check only the uniform convergence

$$\sup_{|q-q_0|<\epsilon}\left|\sum_{i=1}^m \hat{w}_i\psi_\tau(Y_i-q)-E\{\psi_\tau(Y-q)\}\right|=o_p(1). \tag{7}$$

We first look at the case where $\mathcal{P}$ contains a correctly specified model for $\pi(\mathbf{X})$. Without loss of generality, let $\pi^1(\mathbf{X};\boldsymbol{\alpha}^1)$ be the correctly specified model, and let $\boldsymbol{\alpha}_0^1$ denote the true value of $\boldsymbol{\alpha}^1$ so that $\pi^1(\mathbf{X};\boldsymbol{\alpha}_0^1)=\pi(\mathbf{X})$. It is easy to check that

$$\frac{1}{m}\sum_{i=1}^m\frac{\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}},\hat{\mathbf{q}}_L,\hat{\boldsymbol{\gamma}})}{1+\boldsymbol{\rho}^{\mathrm{T}}\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}},\hat{\mathbf{q}}_L,\hat{\boldsymbol{\gamma}})}=\frac{\hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1)}{m}\sum_{i=1}^m\frac{\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}},\hat{\mathbf{q}}_L,\hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}{1+(\hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1)\rho_1-1,\hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1)\rho_2,\ldots,\hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1)\rho_{J+K})^{\mathrm{T}}\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}},\hat{\mathbf{q}}_L,\hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}.$$

Therefore, if we define $\hat{\lambda}_1=\hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1)\hat{\rho}_1-1$ and $\hat{\lambda}_t=\hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1)\hat{\rho}_t, t=2,\ldots,J+K$, then $\hat{\boldsymbol{\lambda}}^{\mathrm{T}}=(\hat{\lambda}_1,\ldots,\hat{\lambda}_{J+K})$ solves

$$\frac{1}{m}\sum_{i=1}^m\frac{\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}},\hat{\mathbf{q}}_L,\hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}{1+\boldsymbol{\lambda}^{\mathrm{T}}\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}},\hat{\mathbf{q}}_L,\hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}=\mathbf{0}, \tag{8}$$

and

$$\hat{w}_i = \frac{1}{m} \frac{\hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1)/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}{1 + \hat{\boldsymbol{\lambda}}^{\mathrm{T}} \hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}.$$

Now let $\boldsymbol{\alpha}_*^j$, $\boldsymbol{\gamma}_*^k$, q_*^k, θ_*^j and η_*^k denote the probability limits of $\hat{\boldsymbol{\alpha}}^j$, $\hat{\boldsymbol{\gamma}}^k$, $\hat{q}_L^k$, $\hat{\theta}^j(\hat{\boldsymbol{\alpha}}^j)$ and $\hat{\eta}_L^k(\hat{q}_L^k, \hat{\boldsymbol{\gamma}}^k)$ respectively, as $n \to \infty$. Note that q_*^k and η_*^k do not depend on L. It is clear that $\boldsymbol{\alpha}_*^1 = \boldsymbol{\alpha}_0^1$, $\theta_*^j = E\{\pi^j(\boldsymbol{\alpha}_*^j)\}$ and $\eta_*^k = E[\psi_\tau\{Y^l(\boldsymbol{\gamma}_*^k) - q_*^k\}]$. Write $\boldsymbol{\alpha}_*^{\mathrm{T}} = \{(\boldsymbol{\alpha}_*^1)^{\mathrm{T}}, \dots, (\boldsymbol{\alpha}_*^J)^{\mathrm{T}}\}$, $\boldsymbol{\gamma}_*^{\mathrm{T}} = \{(\boldsymbol{\gamma}_*^1)^{\mathrm{T}}, \dots, (\boldsymbol{\gamma}_*^K)^{\mathrm{T}}\}$, $\mathbf{q}_*^{\mathrm{T}} = (q_*^1, \dots, q_*^K)$ and

$$\mathbf{g}_1^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)^{\mathrm{T}} = \left(\pi^1(\boldsymbol{\alpha}_*^1) - \theta_*^1, \dots, \pi^J(\boldsymbol{\alpha}_*^J) - \theta_*^J, \frac{1}{L}\sum_{l=1}^L \psi_\tau\{Y^l(\boldsymbol{\gamma}_*^1) - q_*^1\} - \eta_*^1, \dots, \right.$$
$$\left. \frac{1}{L}\sum_{l=1}^L \psi_\tau\{Y^l(\boldsymbol{\gamma}_*^K) - q_*^K\} - \eta_*^K \right). \tag{9}$$

Since

$$E\left\{ \frac{R}{\pi(\mathbf{X})} \mathbf{g}_1^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*) \right\} = \mathbf{0}$$

and $\boldsymbol{\alpha}_*^1 = \boldsymbol{\alpha}_0^1$, $\mathbf{0}$ is the solution to

$$E\left\{ \frac{R\mathbf{g}_1^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)/\pi^1(\boldsymbol{\alpha}_*^1)}{1 + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{g}_1^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)/\pi^1(\boldsymbol{\alpha}_*^1)} \right\} = \mathbf{0}$$

as an equation of $\boldsymbol{\lambda}$. Thus, from the theory of empirical likelihood (e.g. Owen (2001)), $\hat{\boldsymbol{\lambda}} = o_p(1)$.

Note that

$$\sup_{|q-q_0|<\epsilon} \left| \sum_{i=1}^m \hat{w}_i \psi_\tau(Y_i - q) - E\{\psi_\tau(Y - q)\} \right| \leqslant \sup_{|q-q_0|<\epsilon} \left| \sum_{i=1}^m \hat{w}_i \psi_\tau(Y_i - q) - \frac{1}{n}\sum_{i=1}^n \frac{R_i}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)} \psi_\tau(Y_i - q) \right| \tag{10}$$

$$+ \sup_{|q-q_0|<\epsilon} \left| \frac{1}{n}\sum_{i=1}^n \frac{R_i}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)} \psi_\tau(Y_i - q) - \frac{1}{n}\sum_{i=1}^n \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_0^1)} \psi_\tau(Y_i - q) \right| \tag{11}$$

$$+ \sup_{|q-q_0|<\epsilon} \left| \frac{1}{n}\sum_{i=1}^n \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_0^1)} \psi_\tau(Y_i - q) - E\{\psi_\tau(Y - q)\} \right|. \tag{12}$$

Since $\hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1) - m/n = o_p(1)$, we have

$$\hat{w}_i = \frac{1}{n} \frac{1}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)} + o_p(1);$$

therefore inequality (10) is $o_p(1)$, and

$$(11) \leqslant \sup_{|q-q_0|<\epsilon} \left| \frac{1}{n}\sum_{i=1}^n \frac{R_i \psi_\tau(Y_i - q)}{\pi_i^1(\boldsymbol{\alpha}_0^1)^2} \left. \frac{\partial \pi_i^1(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^{\mathrm{T}}} \right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0^1} \right| |\hat{\boldsymbol{\alpha}}^1 - \boldsymbol{\alpha}_0^1| = o_p(1).$$

For expression (12), we have

$$\sup_{|q-q_0|<\epsilon} \left| \frac{1}{n}\sum_{i=1}^n \frac{R_i \psi_\tau(Y_i - q)}{\pi_i^1(\boldsymbol{\alpha}_0^1)} - E\{\psi_\tau(Y - q)\} - \frac{1}{n}\sum_{i=1}^n \frac{R_i \psi_\tau(Y_i - q_0)}{\pi_i^1(\boldsymbol{\alpha}_0^1)} + E\{\psi_\tau(Y - q_0)\} \right| = o_p(n^{-1/2}),$$

which is straightforward if we can prove two results: first,

$$\left\{ \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_0^1)} \psi_\tau(Y_i - q) : |q - q_0| < \epsilon \right\}$$

forms a Donsker class and, second,

$$\frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_0^1)}\psi_\tau(Y_i - q_0)$$

is L_2 continuous at q_0.

For the first result, we define $\mathcal{E}_1 = \{I(Y < q), |q - q_0| < \epsilon\}$ and $\mathcal{E}_2 = \{\pi^j(\boldsymbol{\alpha}_0^j), \forall j, \mathbf{X} \in \mathcal{X}\}$, where $\mathcal{X} = \{\mathbf{X} : E\|\mathbf{X}\|^4 \leqslant C\}$ and C is a large positive number. Note that $\mathcal{E}_1$ is a Vapnik–Chervonenkis class of functions (by theorem 2.6.4 of van der Vaart and Wellner (1996)) with bounded uniform entropy integral. So $\mathcal{E}_1$ belongs to the Donsker class. For $\mathcal{E}_2$, note that the set $\{\mathbf{X} : \mathbf{X} \in \mathcal{X}\}$ can be covered by N_1 balls with L_2-radius ϵ_1 such that $N_1 \lesssim 1/\epsilon_1$, where '$\lesssim$' means that the left-hand side of it is bounded by a positive constant multiplying the right-hand side of it. Since $\pi^j(\mathbf{X}; \boldsymbol{\alpha}^j)$ has bounded derivatives up to order 2, for any $\mathbf{X}$, there is an $\tilde{\mathbf{X}}$ which belongs to the same L_2-ball as $\mathbf{X}$ and $\|\pi^j(\mathbf{X}; \boldsymbol{\alpha}^j) - \pi^j(\tilde{\mathbf{X}}; \boldsymbol{\alpha}^j)\|_{L_2} \lesssim \|\mathbf{X} - \tilde{\mathbf{X}}\|_{L_2}$. Therefore, $N_{[]}(\epsilon, \mathcal{E}_2, L_2) \lesssim 1/\epsilon$, and $\mathcal{E}_2$ has bounded uniform entropy integral and forms a Donsker class. Since R_i is bounded, from the preservation of Donsker classes (corollary 9.32 of Kosorok (2008)),

$$\left\{ \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_0^1)}\psi_\tau(Y_i - q) : |q - q_0| < \epsilon \right\}$$

forms a Donsker class.

For the second result, note that

$$\sup_{|q - q_0| < \epsilon} E\left\{ \frac{R_i \psi_\tau(Y_i - q)}{\pi_i^1(\boldsymbol{\alpha}_0^1)} - \frac{R_i \psi_\tau(Y_i - q_0)}{\pi_i^1(\boldsymbol{\alpha}_0^1)} \right\}^2 \leqslant 2 \sup_{|q - q_0| < \epsilon} E\{\pi_i^1(\boldsymbol{\alpha}_0^1)\}^{-2} g(q_0)^2 |q - q_0|^2 \to 0.$$

Therefore, we have shown that

$$\frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_0^1)}\psi_\tau(Y_i - q_0)$$

is L_2 continuous at q_0. Combining with the fact that

$$\frac{1}{n}\sum_{i=1}^n \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_0^1)}\psi_\tau(Y_i - q_0) = O_p(n^{-1/2})$$

and $E\{\psi_\tau(Y - q_0)\} = 0$, we have that expression (12) is $O_p(n^{-1/2})$ and we obtain the uniform convergence.

Now we consider the case where $\mathcal{F}$ contains a correctly specified model for $f(Y | \mathbf{X})$. Without loss of generality, let $f^1(Y | \mathbf{X}; \boldsymbol{\gamma}^1)$ be the correctly specified model, and let $\boldsymbol{\gamma}_0^1$ denote the true value of $\boldsymbol{\gamma}^1$ so that $f^1(Y | \mathbf{X}; \boldsymbol{\gamma}_0^1) = f(Y | \mathbf{X})$. We then have $\boldsymbol{\gamma}_*^1 = \boldsymbol{\gamma}_0^1$. In addition, we have $\hat{\boldsymbol{\gamma}}^1 \to^p \boldsymbol{\gamma}_0^1$ and $\hat{q}_L^1 \to^p q_0$. Similarly to the situation where $\mathcal{P}$ contains a correctly specified model for $\pi(\mathbf{X})$, we have $\hat{\mathbf{g}}_1^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}}) \to^p \mathbf{g}_1^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)$ and we denote $\boldsymbol{\rho}_*$ as the probability limit of $\hat{\boldsymbol{\rho}}$.

Note that one of the constraints in expression (5) is actually

$$\sum_{i=1}^m \hat{w}_i\left[\frac{1}{L}\sum_{l=1}^L \psi_\tau\{Y_i^l(\hat{\boldsymbol{\gamma}}^1) - \hat{q}_L^1\}\right] = \frac{1}{n}\sum_{i=1}^n \left[\frac{1}{L}\sum_{l=1}^L \psi_\tau\{Y_i^l(\hat{\boldsymbol{\gamma}}^1) - \hat{q}_L^1\}\right],$$

which leads the left-hand side in equation (7) to

$$\sup_{|q - q_0| < \epsilon}\left|\sum_{i=1}^m \hat{w}_i\psi_\tau(Y_i - q) - E\{\psi_\tau(Y - q)\}\right|$$

$$\leqslant \sup_{|q - q_0| < \epsilon}\left|\sum_{i=1}^m \hat{w}_i\left[\psi_\tau(Y_i - q) - \frac{1}{L}\sum_{l=1}^L \psi_\tau\{Y_i^l(\hat{\boldsymbol{\gamma}}^1) - \hat{q}_L^1\}\right]\right| + \sup_{|q - q_0| < \epsilon}|\hat{\eta}_L^1 - E\{\psi_\tau(Y - q)\}|$$

$$\leqslant \sup_{|q - q_0| < \epsilon}\left|\sum_{i=1}^m \hat{w}_i\left[\psi_\tau(Y_i - q) - \frac{1}{L}\sum_{l=1}^L \psi_\tau\{Y_i^l(\hat{\boldsymbol{\gamma}}^1) - \hat{q}_L^1\}\right]\right.$$

$$\left. - \frac{1}{m}\frac{1}{1 + \boldsymbol{\rho}_*^{\mathrm{T}}\mathbf{g}_1^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)}\sum_{i=1}^n R_i\left[\psi_\tau(Y_i - q) - \frac{1}{L}\sum_{l=1}^L \psi_\tau\{Y_i^l(\hat{\boldsymbol{\gamma}}^1) - \hat{q}_L^1\}\right]\right| \tag{13}$$

$$+ \left| \frac{1}{m} \frac{1}{1 + \boldsymbol{\rho}_*^{\mathrm{T}} \mathbf{g}_1^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)} \sum_{i=1}^{n} R_i \frac{1}{L} \sum_{l=1}^{L} [\psi_\tau\{Y_i^l(\hat{\boldsymbol{\gamma}}^1) - \hat{q}_L^1\} - \psi_\tau\{Y_i^l(\boldsymbol{\gamma}_0^1) - q_0\}] \right| \tag{14}$$

$$+ \sup_{|q-q_0|<\epsilon} \left| \frac{1}{m} \frac{1}{1 + \boldsymbol{\rho}_*^{\mathrm{T}} \mathbf{g}_1^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)} \sum_{i=1}^{n} R_i \left[\psi_\tau(Y_i - q) - \frac{1}{L} \sum_{l=1}^{L} \psi_\tau\{Y_i^l(\boldsymbol{\gamma}_0^1) - q_0\} \right] \right.$$
$$\left. - \frac{n}{m} \frac{1}{1 + \boldsymbol{\rho}_*^{\mathrm{T}} \mathbf{g}_1^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)} E[R\{\psi_\tau(Y - q) - \psi_\tau(Y - q_0)\}] \right| \tag{15}$$

$$+ \sup_{|q-q_0|<\epsilon} \left| \frac{n}{m} \frac{1}{1 + \boldsymbol{\rho}_*^{\mathrm{T}} \mathbf{g}_1^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)} E[R\{\psi_\tau(Y - q) - \psi_\tau(Y - q_0)\}] \right| \tag{16}$$

$$+ \left| \frac{1}{n} \sum_{i=1}^{n} \frac{1}{L} \sum_{l=1}^{L} \psi_\tau\{Y_i^l(\hat{\boldsymbol{\gamma}}^1) - \hat{q}_L^1\} - \frac{1}{n} \sum_{i=1}^{n} \frac{1}{L} \sum_{l=1}^{L} \psi_\tau\{Y_i^l(\boldsymbol{\gamma}_0^1) - q_0\} \right| \tag{17}$$

$$+ \left| \frac{1}{n} \sum_{i=1}^{n} \frac{1}{L} \sum_{l=1}^{L} \psi_\tau\{Y_i^l(\boldsymbol{\gamma}_0^1) - q_0\} - E\{\psi_\tau(Y - q_0)\} \right| \tag{18}$$

$$+ \sup_{|q-q_0|<\epsilon} |E\{\psi_\tau(Y - q_0)\} - E\{\psi_\tau(Y - q)\}|. \tag{19}$$

We can show that expression (13) is $o_p(1)$ by using the idea of proving result (11) and the derivatives and subderivatives (He and Shao, 1996); expression (14) is $o_p(1)$ and expression (17) is $o_p(1)$ by using the idea of proving result (11) and subderivatives; expression (16) is $o_p(1)$ and expression (19) is $o_p(1)$ by condition 3; expression (18) is $o_p(1)$ by the weak law of large numbers. For expression (15), similarly to the techniques that were used for proving result (12), it follows that expression (15) is $O_p(n^{-1/2})$ by noting that

$$\left\{ R_i\left[\psi_\tau(Y_i - q) - \frac{1}{L}\sum_{l=1}^{L}\psi_\tau\{Y_i^l(\boldsymbol{\gamma}_0^1) - q_0\}\right]; |q - q_0| < \epsilon \right\}$$

forms a Donsker class and that

$$R_i \left[\psi_\tau(Y_i - q_0) - \frac{1}{L}\sum_{l=1}^{L}\psi_\tau\{Y_i^l(\boldsymbol{\gamma}_0^1) - q_0\} \right]$$

is L_2 continuous at q_0. This completes the proof.

A.2. Proof of theorem 2
First, from expression (8) we have

$$\mathbf{0} = \frac{1}{n} \sum_{i=1}^{n} R_i \frac{\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}{1 + \boldsymbol{\lambda}^{\mathrm{T}} \hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)} - \frac{1}{n} \sum_{i=1}^{n} R_i \frac{\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)} \tag{20}$$

$$+ \frac{1}{n} \sum_{i=1}^{n} R_i \frac{\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)} - \frac{1}{n} \sum_{i=1}^{n} R_i \frac{\hat{\mathbf{g}}_{1i}^L(\boldsymbol{\alpha}_*, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})}{\pi_i^1(\boldsymbol{\alpha}_*^1)} \tag{21}$$

$$+ \frac{1}{n} \sum_{i=1}^{n} R_i \frac{\hat{\mathbf{g}}_{1i}^L(\boldsymbol{\alpha}_*, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})}{\pi_i^1(\boldsymbol{\alpha}_*^1)} - \frac{1}{n} \sum_{i=1}^{n} R_i \frac{\hat{\mathbf{g}}_{1i}^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \hat{\boldsymbol{\gamma}})}{\pi_i^1(\boldsymbol{\alpha}_*^1)} \tag{22}$$

$$+ \frac{1}{n} \sum_{i=1}^{n} R_i \frac{\hat{\mathbf{g}}_{1i}^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \hat{\boldsymbol{\gamma}})}{\pi_i^1(\boldsymbol{\alpha}_*^1)} - \frac{1}{n} \sum_{i=1}^{n} R_i \frac{\hat{\mathbf{g}}_{1i}^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)}{\pi_i^1(\boldsymbol{\alpha}_*^1)} \tag{23}$$

$$+ \frac{1}{n} \sum_{i=1}^{n} R_i \frac{\hat{\mathbf{g}}_{1i}^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)}{\pi_i^1(\boldsymbol{\alpha}_*^1)}.$$

Note that a regular Taylor series expansion can be applied to expressions (20) and (21). For expression (20), we have

$$(20) = -\frac{1}{n}\sum_{i=1}^{n} R_i \frac{\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})^{\otimes 2}}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)^2}\hat{\boldsymbol{\lambda}} + o_p(n^{-1/2}).$$

For expression (21), denote $T_i(\boldsymbol{\alpha}, \mathbf{q}, \boldsymbol{\gamma}) = \hat{\mathbf{g}}_{1i}^L(\boldsymbol{\alpha}, \mathbf{q}, \boldsymbol{\gamma})/\pi_i^1(\boldsymbol{\alpha}^1)$. Note that

$$\frac{\partial T_i(\boldsymbol{\alpha}_*, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})}{\partial \boldsymbol{\alpha}^1} = \frac{\pi_i^1(\boldsymbol{\alpha}_*^1)\partial \hat{\mathbf{g}}_{1i}^L(\boldsymbol{\alpha}_*, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})/\partial \boldsymbol{\alpha}^1 - \hat{\mathbf{g}}_{1i}^L(\boldsymbol{\alpha}_*, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})(\partial \pi_i^1(\boldsymbol{\alpha}_*^1)/\partial \boldsymbol{\alpha}^1)^{\mathrm{T}}}{\pi_i^1(\boldsymbol{\alpha}_*^1)^2},$$

where $\partial \hat{\mathbf{g}}_{1i}^L(\mathbf{0}, \boldsymbol{\alpha}_*, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})/\partial \boldsymbol{\alpha}^1$ has non-zero values only in the first row, and

$$\frac{\partial T_i(\boldsymbol{\alpha}_*, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})}{\partial \boldsymbol{\alpha}^j} = \frac{1}{\pi_i^1(\boldsymbol{\alpha}_*^1)}\frac{\partial \hat{\mathbf{g}}_{1i}^L(\boldsymbol{\alpha}_*, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})}{\partial \boldsymbol{\alpha}^j},$$

where $\partial \hat{\mathbf{g}}_{1i}^L(\boldsymbol{\alpha}_*, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})/\partial \boldsymbol{\alpha}^j$ has non-zero values only in the jth row, $j = 2, \ldots, J$, we have

$$(21) = -\frac{1}{n}\sum_{i=1}^{n} R_i \frac{\hat{\mathbf{g}}_{1i}^L(\boldsymbol{\alpha}_*, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})(\partial \pi_i^1(\boldsymbol{\alpha}_*^1)/\partial \boldsymbol{\alpha}^1)^{\mathrm{T}}}{\pi_i^1(\boldsymbol{\alpha}_*^1)^2}(\hat{\boldsymbol{\alpha}}^1 - \boldsymbol{\alpha}_*^1) + o_p(n^{-1/2}).$$

For expression (22), similarly to result (12), we can show that $\{R_i T_i(\boldsymbol{\alpha}_*, \mathbf{q}, \hat{\boldsymbol{\gamma}}) : \|\mathbf{q} - \mathbf{q}_*\| < \epsilon\}$ forms a Donsker class and $R_i T_i(\boldsymbol{\alpha}_*, \mathbf{q}_*, \hat{\boldsymbol{\gamma}})$ is L_2 continuous at $\mathbf{q}_*$. Therefore, we have

$$(22) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial E\{R_i T_i(\boldsymbol{\alpha}_*, \mathbf{q}_*, \hat{\boldsymbol{\gamma}})\}}{\partial \mathbf{q}}(\hat{\mathbf{q}} - \mathbf{q}_*) + o_p(n^{-1/2}).$$

For expression (23), we can show that $\{R_i T_i(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}) : \|\boldsymbol{\gamma} - \boldsymbol{\gamma}_*\| < \epsilon\}$ forms a Donsker class and $R_i T_i(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)$ is L_2 continuous at $\boldsymbol{\gamma}_*$. Therefore, we have

$$(23) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial E\{R_i T_i(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)\}}{\partial \boldsymbol{\gamma}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_*) + o_p(n^{-1/2}).$$

It is straightforward to see that both $E[R_i T_i(\boldsymbol{\alpha}_*, \mathbf{q}_*, \hat{\boldsymbol{\gamma}})]$ and $E[R_i T_i(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)]$ are 0. Therefore, both expression (22) and expression (23) are 0s. Hence, by defining

$$\mathbf{M}_1^L = E\left\{\frac{\mathbf{g}_1^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*)(\partial \pi^1(\boldsymbol{\alpha}_*^1)/\partial \boldsymbol{\alpha}^1)^{\mathrm{T}}}{\pi^1(\boldsymbol{\alpha}_*^1)}\right\}$$

and from condition 7 we have

$$\sqrt{n}\hat{\boldsymbol{\lambda}} = (\mathbf{G}_1^L)^{-1}\left\{n^{-1/2}\sum_{i=1}^{n}\frac{R_i - \pi_i^1(\boldsymbol{\alpha}_*^1)}{\pi_i^1(\boldsymbol{\alpha}_*^1)}\mathbf{g}_{1i}^L(\boldsymbol{\alpha}_*, \mathbf{q}_*, \boldsymbol{\gamma}_*) - n^{-1/2}\sum_{i=1}^{n}\mathbf{M}_1^L E(\boldsymbol{\Phi}_1^{\otimes 2})^{-1}\boldsymbol{\Phi}_{1i}\right\} + o_p(1).$$

Next, note that

$$0 = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}{1 + \hat{\boldsymbol{\lambda}}^{\mathrm{T}}\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}\psi_\tau(Y_i - \hat{q}_{\mathrm{MR}}^L)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{R_i/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}{1 + \hat{\boldsymbol{\lambda}}^{\mathrm{T}}\hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}\psi_\tau(Y_i - \hat{q}_{\mathrm{MR}}^L) - \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}\psi_\tau(Y_i - \hat{q}_{\mathrm{MR}}^L) \qquad (24)$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}\psi_\tau(Y_i - \hat{q}_{\mathrm{MR}}^L) - \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_*^1)}\psi_\tau(Y_i - \hat{q}_{\mathrm{MR}}^L) \qquad (25)$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_*^1)}\psi_\tau(Y_i - \hat{q}_{\mathrm{MR}}^L) - \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_*^1)}\psi_\tau(Y_i - q_0) \qquad (26)$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_*^1)}\psi_\tau(Y_i - q_0).$$

It can be shown that

$$(24) = -\left\{ \frac{1}{n} \sum_{i=1}^{n} R_i \frac{\psi_\tau(Y_i - \hat{q}_{\mathrm{MR}}^L)}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)^2} \hat{\mathbf{g}}_{1i}^L(\hat{\boldsymbol{\alpha}}, \hat{\mathbf{q}}_L, \hat{\boldsymbol{\gamma}})^{\mathrm{T}} \right\} \hat{\boldsymbol{\lambda}} + o_p(n^{-1/2}),$$

$$(25) = -\left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{R_i \psi_\tau(Y_i - \hat{q}_{\mathrm{MR}}^L)}{\pi_i^1(\boldsymbol{\alpha}_*^1)^2} \left(\frac{\partial \pi_i^1(\boldsymbol{\alpha}_*^1)}{\partial \boldsymbol{\alpha}^1} \right)^{\mathrm{T}} \right\} (\hat{\boldsymbol{\alpha}}^1 - \boldsymbol{\alpha}_*^1) + o_p(n^{-1/2})$$

and

$$(26) = -g(q_0)(\hat{q}_{\mathrm{MR}}^L - q_0) + o_p(n^{-1/2}).$$

Write

$$\mathbf{B}_1^L = E\left\{ \frac{\psi_\tau(Y - q_0)}{\pi^1(\boldsymbol{\alpha}_0^1)} \left(\frac{\partial \pi^1(\boldsymbol{\alpha}_0^1)}{\partial \boldsymbol{\alpha}^1} \right)^{\mathrm{T}} \right\};$$

we then have

$$g(q_0)\sqrt{n}(\hat{q}_{\mathrm{MR}}^L - q_0) = -\mathbf{A}_1^L\sqrt{n}\hat{\boldsymbol{\lambda}} - \mathbf{B}_1^L\sqrt{n}(\hat{\boldsymbol{\alpha}}^1 - \boldsymbol{\alpha}_*^1) + n^{-1/2} \sum_{i=1}^{n} \frac{R_i\psi_\tau(Y_i - q_0)}{\pi_i^1(\boldsymbol{\alpha}_0^1)} + o_p(1)$$

$$= n^{-1/2} \sum_{i=1}^{n} [\mathbf{Q}_{1i}^L(\boldsymbol{\alpha}_0^1) - \{\mathbf{B}_1^L - \mathbf{A}_1^L(\mathbf{G}_1^L)^{-1}\mathbf{M}_1^L\} E(\boldsymbol{\Phi}_1^{\otimes 2})^{-1}\boldsymbol{\Phi}_{1i}] + o_p(1).$$

From the generalized information equality (Newey, 1990), we can show that

$$\mathbf{B}_1^L - \mathbf{A}_1^L(\mathbf{G}_1^L)^{-1}\mathbf{M}_1^L = -E\left\{ \frac{\partial \mathbf{Q}_1^L(\boldsymbol{\alpha}_0^1)}{\partial \boldsymbol{\alpha}^1} \right\} = E(\mathbf{Q}_1\boldsymbol{\Phi}_1^{\mathrm{T}}),$$

and finally that

$$g(q_0)\sqrt{n}(\hat{q}_{\mathrm{MR}}^L - q_0) = n^{-1/2} \sum_{i=1}^{n} \{\mathbf{Q}_{1i}^L(\boldsymbol{\alpha}_0^1) - E(\mathbf{Q}_1^L\boldsymbol{\Phi}_1^{\mathrm{T}})E(\boldsymbol{\Phi}_1^{\otimes 2})^{-1}\boldsymbol{\Phi}_{1i}\} + o_p(1).$$

This completes the proof.

A.3.   Sketched proof of theorem 3

On the basis of theorem 5.9 of van de Vaart (1998), it is enough to check the uniform convergence

$$\sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \epsilon} \left| \sum_{i=1}^{m} \hat{w}_i \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) - E\{\mathbf{X}\psi_\tau(Y - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta})\} \right| = o_p(1). \tag{27}$$

For the case where $\mathcal{P}$ contains a correctly specified model for $\pi(Y, \mathbf{X}_1)$, let $\pi^1(Y, \mathbf{X}_1; \boldsymbol{\alpha}^1)$ be the correctly specified model, and let $\boldsymbol{\alpha}_0^1$ denote the true value of $\boldsymbol{\alpha}^1$ so that $\pi^1(Y, \mathbf{X}_1; \boldsymbol{\alpha}_0^1) = \pi(Y, \mathbf{X}_1)$. Let $\boldsymbol{\alpha}_*^j, \boldsymbol{\gamma}_*^k, \boldsymbol{\beta}_*^k, \theta_*^j$ and η_*^k denote the probability limits of $\hat{\boldsymbol{\alpha}}^j, \hat{\boldsymbol{\gamma}}^k, \hat{\boldsymbol{\beta}}_L^k, \hat{\theta}^j(\hat{\boldsymbol{\alpha}}^j)$ and $\hat{\eta}_L^k(\hat{\boldsymbol{\beta}}_L^k, \hat{\boldsymbol{\gamma}}^k)$ respectively, as $n \to \infty$. Note that $\boldsymbol{\beta}_*^k$ and η_*^k do not depend on L. It is clear that $\boldsymbol{\alpha}_*^1 = \boldsymbol{\alpha}_0^1$, $\theta_*^j = E\{\pi^j(\boldsymbol{\alpha}_*^j)\}$ and $\eta_*^k = E[\mathbf{X}^l(\boldsymbol{\gamma}_*^k)\psi_\tau\{Y - \mathbf{X}^l(\boldsymbol{\gamma}_*^k)^{\mathrm{T}}\boldsymbol{\beta}_*^k\}]$. Define $\hat{\lambda}_1 = \hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1)\hat{\rho}_1 - 1$ and $\hat{\lambda}_t = \hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1)\hat{\rho}_t, t = 2, \ldots, J + K$. Then $\hat{\boldsymbol{\lambda}}^{\mathrm{T}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_{J+K})$ solves

$$\frac{1}{m} \sum_{i=1}^{m} \frac{\hat{\mathbf{g}}_{2i}^L(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_L, \hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}{1 + \boldsymbol{\lambda}^{\mathrm{T}}\hat{\mathbf{g}}_{2i}^L(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_L, \hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)} = \mathbf{0} \tag{28}$$

and

$$\hat{w}_i = \frac{1}{m} \frac{\hat{\theta}^1(\hat{\boldsymbol{\alpha}}^1)/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}{1 + \hat{\boldsymbol{\lambda}}^{\mathrm{T}}\hat{\mathbf{g}}_{2i}^L(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_L, \hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}.$$

Write $\boldsymbol{\alpha}_*^{\mathrm{T}} = \{(\boldsymbol{\alpha}_*^1)^{\mathrm{T}}, \ldots, (\boldsymbol{\alpha}_*^J)^{\mathrm{T}}\}$, $\boldsymbol{\gamma}_*^{\mathrm{T}} = \{(\boldsymbol{\gamma}_*^1)^{\mathrm{T}}, \ldots, (\boldsymbol{\gamma}_*^K)^{\mathrm{T}}\}$, $\boldsymbol{\beta}_*^{\mathrm{T}} = ((\boldsymbol{\beta}_*^1)^{\mathrm{T}}, \ldots, (\boldsymbol{\beta}_*^K)^{\mathrm{T}})$ and

$$
\mathbf{g}_2^L(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*, \boldsymbol{\gamma}_*)^{\mathrm{T}} = \left(\pi^1(\boldsymbol{\alpha}_*^1) - \theta_*^1, \ldots, \pi^J(\boldsymbol{\alpha}_*^J) - \theta_*^J, \frac{1}{L}\sum_{l=1}^{L} \mathbf{X}^l(\boldsymbol{\gamma}_*^1)\psi_\tau\{Y - \mathbf{X}^l(\boldsymbol{\gamma}_*^1)^{\mathrm{T}}\boldsymbol{\beta}_*^1\} - \eta_*^1, \ldots, \right.
$$
$$
\left. \frac{1}{L}\sum_{l=1}^{L} \mathbf{X}^l(\boldsymbol{\gamma}_*^K)\psi_\tau\{Y - \mathbf{X}^l(\boldsymbol{\gamma}_*^K)^{\mathrm{T}}\boldsymbol{\beta}_*^K\} - \eta_*^K \right). \tag{29}
$$

From the theory of empirical likelihood (e.g. Owen (2001)) we have $\hat{\boldsymbol{\lambda}} = o_p(1)$. Observe that

$$
\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\epsilon} \left| \sum_{i=1}^{m} \hat{w}_i \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) - E\{\mathbf{X}\psi_\tau(Y - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta})\} \right|
$$
$$
\leqslant \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\epsilon} \left| \sum_{i=1}^{m} \hat{w}_i \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) - \frac{1}{n}\sum_{i=1}^{n} \frac{R_i}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}\mathbf{X}_i\psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) \right| \tag{30}
$$
$$
+ \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\epsilon} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{R_i}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}\mathbf{X}_i\psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) - \frac{1}{n}\sum_{i=1}^{n} \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_0^1)}\mathbf{X}_i\psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) \right| \tag{31}
$$
$$
+ \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\epsilon} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_0^1)}\mathbf{X}_i\psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) - E\{\mathbf{X}\psi_\tau(Y - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})\} \right|. \tag{32}
$$

Similarly to the arguments for expressions (10) and (11), we have that expression (30) is $o_p(1)$ and that expression (31) is $o_p(1)$. Similarly to the techniques that were used for proving result (12), it follows that expression (32) is $O_p(n^{-1/2})$ by the facts that

$$
\left\{ \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_0^1)}\mathbf{X}_i\psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) : \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| < \epsilon \right\}
$$

forms a Donsker class and that

$$
\frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_0^1)}\mathbf{X}_i\psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})
$$

is L_2 continuous at $\boldsymbol{\beta}_0$. Therefore, result (27) holds.

For the case where $\mathcal{F}$ contains a correctly specified model for $f(\mathbf{X}_2 \mid Y, \mathbf{X}_1)$, let $f^1(\mathbf{X}_2 \mid Y, \mathbf{X}_1; \boldsymbol{\gamma}^1)$ be the correctly specified model, and let $\boldsymbol{\gamma}_0^1$ denote the true value of $\boldsymbol{\gamma}^1$ so that $f^1(\mathbf{X}_2 \mid Y, \mathbf{X}_1; \boldsymbol{\gamma}_0^1) = f(\mathbf{X}_2 \mid Y, \mathbf{X}_1)$. We obtain $\hat{\boldsymbol{\gamma}}^1 \to^p \boldsymbol{\gamma}_0^1$, $\boldsymbol{\gamma}_*^1 = \boldsymbol{\gamma}_0^1$, $\hat{\boldsymbol{\beta}}_L^1 \to^p \boldsymbol{\beta}_0$ and $\hat{\mathbf{g}}_2^L(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_L, \hat{\boldsymbol{\gamma}}) \to^p \mathbf{g}_2^L(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*, \boldsymbol{\gamma}_*)$. Denote $\boldsymbol{\rho}_*$ as the probability limit of $\hat{\boldsymbol{\rho}}$ and $\varpi_* = \boldsymbol{\rho}_*^{\mathrm{T}} \mathbf{g}_2^L(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*, \boldsymbol{\gamma}_*)$. We have

$$
\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\epsilon} \left| \sum_{i=1}^{m} \hat{w}_i \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) - E\{\mathbf{X}\psi_\tau(Y - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta})\} \right|
$$
$$
\leqslant \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\epsilon} \left| \sum_{i=1}^{m} \hat{w}_i \left[\mathbf{X}_i\psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) - \frac{1}{L}\sum_{l=1}^{L} \mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^1)\psi_\tau\{Y_i - \mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^1)^{\mathrm{T}}\hat{\boldsymbol{\beta}}_L^1\} \right] \right.
$$
$$
\left. - \frac{1}{m}\frac{1}{1+\varpi_*}\sum_{i=1}^{n} R_i \left[\mathbf{X}_i\psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) - \frac{1}{L}\sum_{l=1}^{L} \mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^1)\psi_\tau\{Y_i - \mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^1)^{\mathrm{T}}\hat{\boldsymbol{\beta}}_L^1\} \right] \right| \tag{33}
$$
$$
+ \left| \frac{1}{m}\frac{1}{1+\varpi_*}\sum_{i=1}^{n} \frac{R_i}{L}\sum_{l=1}^{L}[\mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^1)\psi_\tau\{Y_i - \mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^1)^{\mathrm{T}}\hat{\boldsymbol{\beta}}_L^1\} - \mathbf{X}_i^l(\boldsymbol{\gamma}_0^1)\psi_\tau\{Y_i - \mathbf{X}_i^l(\boldsymbol{\gamma}_0^1)^{\mathrm{T}}\boldsymbol{\beta}_0\}] \right| \tag{34}
$$
$$
+ \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\epsilon} \left| \frac{1}{m}\frac{1}{1+\varpi_*}\sum_{i=1}^{n} R_i \left[\mathbf{X}_i\psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) - \frac{1}{L}\sum_{l=1}^{L} \mathbf{X}_i^l(\boldsymbol{\gamma}_0^1)\psi_\tau\{Y_i - \mathbf{X}_i^l(\boldsymbol{\gamma}_0^1)^{\mathrm{T}}\boldsymbol{\beta}_0\} \right] \right.
$$
$$
\left. - \frac{n}{m}\frac{1}{1+\varpi_*}E[R\mathbf{X}\{\psi_\tau(Y - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}) - \psi_\tau(Y - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_0)\}] \right| \tag{35}
$$
$$
+ \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\epsilon} \left| \frac{n}{m}\frac{1}{1+\varpi_*}E[R\mathbf{X}\{\psi_\tau(Y - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}) - \psi_\tau(Y - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_0)\}] \right| \tag{36}
$$

$$+\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{L}\sum_{l=1}^{L}[\mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^1)\psi_\tau\{Y_i-\mathbf{X}_i^l(\hat{\boldsymbol{\gamma}}^1)^{\mathrm{T}}\hat{\boldsymbol{\beta}}_L^1\}-\mathbf{X}_i^l(\boldsymbol{\gamma}_0)\psi_\tau\{Y_i-\mathbf{X}_i^l(\boldsymbol{\gamma}_0^1)^{\mathrm{T}}\boldsymbol{\beta}_0\}]\right| \tag{37}$$

$$+\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{L}\sum_{l=1}^{L}\mathbf{X}_i^l(\boldsymbol{\gamma}_0)\psi_\tau\{Y_i-\mathbf{X}_i^l(\boldsymbol{\gamma}_0^1)^{\mathrm{T}}\boldsymbol{\beta}_0\}-E\{\mathbf{X}\psi_\tau(Y-\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_0)\}\right| \tag{38}$$

$$+\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\epsilon}\left|E\{\mathbf{X}\psi_\tau(Y-\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_0)\}-E\{\mathbf{X}\psi_\tau(Y-\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta})\}\right|. \tag{39}$$

Using similar arguments to those for expressions (13), (14) and (17), we can show that expression (33) is $o_p(1)$, expression (34) is $o_p(1)$ and expression (37) is $o_p(1)$. In addition, we have that expression (36) is $o_p(1)$ and expression (39) is $o_p(1)$ by condition 10, expression (38) is $o_p(1)$ by the weak law of large numbers and expression (35) is $O_p(n^{-1/2})$ by observing that

$$\left\{R_i\left[\mathbf{X}_i\psi_\tau(Y_i-\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})-\frac{1}{L}\sum_{l=1}^{L}\mathbf{X}_i^l(\boldsymbol{\gamma}_0^1)\psi_\tau\{Y_i-\mathbf{X}_i^l(\boldsymbol{\gamma}_0^1)^{\mathrm{T}}\boldsymbol{\beta}_0\}\right]:\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\epsilon}\right\}$$

forms a Donsker class and that

$$R_i\left[\mathbf{X}_i\psi_\tau(Y_i-\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})-\frac{1}{L}\sum_{l=1}^{L}\mathbf{X}_i^l(\boldsymbol{\gamma}_0^1)\psi_\tau\{Y_i-\mathbf{X}_i^l(\boldsymbol{\gamma}_0^1)^{\mathrm{T}}\boldsymbol{\beta}_0\}\right]$$

is L_2 continuous at $\boldsymbol{\beta}_0$. Thus, result (27) holds. This completes the proof.

A.4. Sketched proof of theorem 4
By expression (28), we have

$$\mathbf{0}=\frac{1}{n}\sum_{i=1}^{n}R_i\frac{\hat{\mathbf{g}}_{2i}^L(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_L,\hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}{1+\boldsymbol{\lambda}^{\mathrm{T}}\hat{\mathbf{g}}_{2i}^L(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_L,\hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}-\frac{1}{n}\sum_{i=1}^{n}R_i\frac{\hat{\mathbf{g}}_{2i}^L(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_L,\hat{\boldsymbol{\gamma}})}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)} \tag{40}$$

$$+\frac{1}{n}\sum_{i=1}^{n}R_i\frac{\hat{\mathbf{g}}_{2i}^L(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_L,\hat{\boldsymbol{\gamma}})}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}-\frac{1}{n}\sum_{i=1}^{n}R_i\frac{\hat{\mathbf{g}}_{2i}^L(\boldsymbol{\alpha}_*,\hat{\boldsymbol{\beta}}_L,\hat{\boldsymbol{\gamma}})}{\pi_i^1(\boldsymbol{\alpha}_*^1)} \tag{41}$$

$$+\frac{1}{n}\sum_{i=1}^{n}R_i\frac{\hat{\mathbf{g}}_{2i}^L(\boldsymbol{\alpha}_*,\hat{\boldsymbol{\beta}}_L,\hat{\boldsymbol{\gamma}})}{\pi_i^1(\boldsymbol{\alpha}_*^1)}-\frac{1}{n}\sum_{i=1}^{n}R_i\frac{\hat{\mathbf{g}}_{2i}^L(\boldsymbol{\alpha}_*,\boldsymbol{\beta}_*,\hat{\boldsymbol{\gamma}})}{\pi_i^1(\boldsymbol{\alpha}_*^1)} \tag{42}$$

$$+\frac{1}{n}\sum_{i=1}^{n}R_i\frac{\hat{\mathbf{g}}_{2i}^L(\boldsymbol{\alpha}_*,\boldsymbol{\beta}_*,\hat{\boldsymbol{\gamma}})}{\pi_i^1(\boldsymbol{\alpha}_*^1)}-\frac{1}{n}\sum_{i=1}^{n}R_i\frac{\hat{\mathbf{g}}_{2i}^L(\boldsymbol{\alpha}_*,\boldsymbol{\beta}_*,\boldsymbol{\gamma}_*)}{\pi_i^1(\boldsymbol{\alpha}_*^1)} \tag{43}$$

$$+\frac{1}{n}\sum_{i=1}^{n}R_i\frac{\hat{\mathbf{g}}_{2i}^L(\boldsymbol{\alpha}_*,\boldsymbol{\beta}_*,\boldsymbol{\gamma}_*)}{\pi_i^1(\boldsymbol{\alpha}_*^1)}.$$

We can show that

$$(40)=-\frac{1}{n}\sum_{i=1}^{n}R_i\frac{\hat{\mathbf{g}}_{2i}^L(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_L,\hat{\boldsymbol{\gamma}})^{\otimes 2}}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)^2}\hat{\boldsymbol{\lambda}}+o_p(n^{-1/2});$$

$$(41)=-\frac{1}{n}\sum_{i=1}^{n}R_i\frac{\hat{\mathbf{g}}_{2i}^L(\boldsymbol{\alpha}_*,\hat{\boldsymbol{\beta}}_L,\hat{\boldsymbol{\gamma}})(\partial\pi_i^1(\boldsymbol{\alpha}_*^1)/\partial\boldsymbol{\alpha}^1)^{\mathrm{T}}}{\pi_i^1(\boldsymbol{\alpha}_*^1)^2}(\hat{\boldsymbol{\alpha}}^1-\boldsymbol{\alpha}_*^1)+o_p(n^{-1/2});$$

expression (42) is $o_p(n^{-1/2})$ and expression (43) is $o_p(n^{-1/2})$. Write

$$\mathbf{M}_2^L=E\left\{\mathbf{g}_2^L(\boldsymbol{\alpha}_*,\boldsymbol{\beta}_*,\boldsymbol{\gamma}_*)\left(\frac{\partial\pi^1(\boldsymbol{\alpha}_*^1)}{\partial\boldsymbol{\alpha}^1}\right)^{\mathrm{T}}\bigg/\pi^1(\boldsymbol{\alpha}_*^1)\right\}.$$

Together with condition 13, we have

$$\sqrt{n}\hat{\boldsymbol{\lambda}} = \frac{(\mathbf{G}_2^L)^{-1}}{\sqrt{n}} \left\{ \sum_{i=1}^n \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_*^1)} \mathbf{g}_{2i}^L(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*, \boldsymbol{\gamma}_*) - \sum_{i=1}^n \mathbf{M}_2^L E(\boldsymbol{\Phi}_2^{\otimes 2})^{-1} \boldsymbol{\Phi}_{2i} \right\} + o_p(1).$$

Meanwhile, note that

$$0 = \frac{1}{n} \sum_{i=1}^n \frac{R_i/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)}{1 + \hat{\boldsymbol{\lambda}}^{\mathrm{T}} \hat{\mathbf{g}}_{2i}^L(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_L, \hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)} \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{\mathrm{MR}}^L)$$

$$= \frac{1}{n} \sum_{i=1}^n \frac{\{R_i/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)\} \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{\mathrm{MR}}^L)}{1 + \hat{\boldsymbol{\lambda}}^{\mathrm{T}} \hat{\mathbf{g}}_{2i}^L(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_L, \hat{\boldsymbol{\gamma}})/\pi_i^1(\hat{\boldsymbol{\alpha}}^1)} - \frac{1}{n} \sum_{i=1}^n \frac{R_i \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{\mathrm{MR}}^L)}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)} \tag{44}$$

$$+ \frac{1}{n} \sum_{i=1}^n \frac{R_i}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)} \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{\mathrm{MR}}^L) - \frac{1}{n} \sum_{i=1}^n \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_*^1)} \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{\mathrm{MR}}^L) \tag{45}$$

$$+ \frac{1}{n} \sum_{i=1}^n \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_*^1)} \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{\mathrm{MR}}^L) - \frac{1}{n} \sum_{i=1}^n \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_*^1)} \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}_0) \tag{46}$$

$$+ \frac{1}{n} \sum_{i=1}^n \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_*^1)} \mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}_0).$$

We can show that

$$(44) = -\left\{ \frac{1}{n} \sum_{i=1}^n R_i \frac{\mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{\mathrm{MR}}^L)}{\pi_i^1(\hat{\boldsymbol{\alpha}}^1)^2} \hat{\mathbf{g}}_{2i}^L(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_L, \hat{\boldsymbol{\gamma}})^{\mathrm{T}} \right\} \hat{\boldsymbol{\lambda}} + o_p(n^{-1/2}),$$

$$(45) = -\left\{ \frac{1}{n} \sum_{i=1}^n R_i \frac{\mathbf{X}_i \psi_\tau(Y_i - \mathbf{X}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{\mathrm{MR}}^L)}{\pi_i^1(\boldsymbol{\alpha}_*^1)^2} \left(\frac{\partial \pi_i^1(\boldsymbol{\alpha}_*^1)}{\partial \boldsymbol{\alpha}^1} \right)^{\mathrm{T}} \right\} (\hat{\boldsymbol{\alpha}}^1 - \boldsymbol{\alpha}_*^1) + o_p(n^{-1/2}),$$

$$(46) = -\left\{ \frac{1}{n} \sum_{i=1}^n \frac{R_i}{\pi^1(\boldsymbol{\alpha}_*^1)} g(\mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}_0) \mathbf{X}_i \mathbf{X}_i^{\mathrm{T}} \right\} (\hat{\boldsymbol{\beta}}_{\mathrm{MR}}^L - \boldsymbol{\beta}_0) + o_p(n^{-1/2}).$$

Write

$$\mathbf{B}_2^L = E\left\{ \frac{\mathbf{X} \psi_\tau(Y - \mathbf{X}^{\mathrm{T}} \boldsymbol{\beta}_0)}{\pi^1(\boldsymbol{\alpha}_0^1)} \left(\frac{\partial \pi^1(\boldsymbol{\alpha}_0^1)}{\partial \boldsymbol{\alpha}^1} \right)^{\mathrm{T}} \right\}$$

and

$$Q_{2i}^L(\boldsymbol{\alpha}_0^1) = \frac{R_i}{\pi_i^1(\boldsymbol{\alpha}_0^1)} \{\mathbf{X}_i \psi_i(Y_i - \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}_0) - \mathbf{A}_2^L(\mathbf{G}_2^L)^{-1} \mathbf{g}_{2i}^L(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*, \boldsymbol{\gamma}_*)\}.$$

From the generalized information equality we have

$$\mathbf{B}_2^L - \mathbf{A}_2^L(\mathbf{G}_2^L)^{-1} \mathbf{M}_2^L = -E\left\{ \frac{\partial \mathbf{Q}_2^L(\boldsymbol{\alpha}_0^1)}{\partial \boldsymbol{\alpha}^1} \right\} = E(\mathbf{Q}_2^L \boldsymbol{\Phi}_2^{\mathrm{T}}),$$

and then

$$E\{g(\mathbf{X}^{\mathrm{T}} \boldsymbol{\beta}_0) \mathbf{X} \mathbf{X}^{\mathrm{T}}\} \sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{MR}}^L - \boldsymbol{\beta}_0) = n^{-1/2} \sum_{i=1}^n \{\mathbf{Q}_{2i}^L(\boldsymbol{\alpha}_0^1) - E(\mathbf{Q}_2^L \boldsymbol{\Phi}_2^{\mathrm{T}}) E(\boldsymbol{\Phi}_2^{\otimes 2})^{-1} \boldsymbol{\Phi}_{2i}\} + o_p(1).$$

This leads to the result.

References

Bang, H. and Robins, J. M. (2005) Doubly robust estimation in missing data and causal inference models. *Biometrics*, **61**, 962–972.

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

Chan, K. C. G. and Yam, S. C. P. (2014) Oracle, multiple robust and multipurpose calibration in a missing response problem. *Statist. Sci.*, **29**, 380–396.

Chen, S. and Haziza, D. (2017) Multiply robust imputation procedures for the treatment of item nonresponse in surveys. *Biometrika*, **104**, 439–453.

Chen, S. X., Leung, D. H. Y. and Qin, J. (2008) Improving semiparametric estimation by using surrogate data. *J. R. Statist. Soc.* B, **70**, 803–823.

Chen, J., Sitter, R. R. and Wu, C. (2002) Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika*, **89**, 230–237.

Chen, X., Wan, A. T. K. and Zhou, Y. (2015) Efficient quantile regression analysis with missing observations. *J. Am. Statist. Ass.*, **110**, 723–741.

Chen, S. and Yu, C. L. (2016) Parameter estimation through semiparametric quantile regression imputation. *Electron. J. Statist.*, **10**, 3621–3647.

Davidian, M., Tsiatis, A. A. and Leon, S. (2005) Semiparametric estimation of treatment effect in a pretest-posttest study with missing data. *Statist. Sci.*, **20**, 261–301.

Deville, J. and Särndal, C. (1992) Calibration estimators in survey sampling. *J. Am. Statist. Ass.*, **87**, 376–382.

Fan, Y. and Tang, C. Y. (2013) Tuning parameter selection in high dimensional penalized likelihood. *J. R. Statist. Soc.* B, **75**, 531–552.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundaker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., Hirsch, M. S. and Merigan, T. C. (1996) A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New Engl. J. Med.*, **335**, 1081–1089.

Han, P. (2014a) A further study of the multiply robust estimator in missing data analysis. *J. Statist. Planng Inf.*, **148**, 101–110.

Han, P. (2014b) Multiply robust estimation in regression analysis with missing data. *J. Am. Statist. Ass.*, **109**, 1159–1173.

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

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

Han, P. (2018) A further study of propensity score calibration in missing data analysis. *Statist. Sin.*, **28**, 1307–1332.

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

He, X. and Shao, Q. (1996) A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *Ann. Statist.*, **24**, 2608–2630.

Horvitz, D. G. and Thompson, D. J. (1952) A generalization of sampling without replacement from a finite universe. *J. Am. Statist. Ass.*, **47**, 663–685.

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

Kim, J. K. and Shao, J. (2013) *Statistical Methods for Handling Incomplete Data*. Boca Raton: Chapman and Hall–CRC.

Koenker, R. (2005) *Quantile Regression*. New York: Cambridge University Press.

Kosorok, M. R. (2008) *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.

Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G. and Zhao, L. P. (1997) Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus. *Appl. Statist.*, **46**, 463–476.

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

Newey, W. K. (1990) Semiparametric efficiency bounds. *J. Appl. Econometr.*, **5**, 99–135.

Owen, A. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

Owen, A. (2001) *Empirical Likelihood*. New York: Chapman and Hall–CRC.

Qin, J. and Lawless, J. (1994) Empirical likelihood and general estimating equations. *Ann. Statist.*, **22**, 300–325.

Qin, J., Shao, J. and Zhang, B. (2008) Efficient and doubly robust imputation for covariate-dependent missing responses. *J. Am. Statist. Ass.*, **103**, 797–810.

Qin, J. and Zhang, B. (2007) Empirical-likelihood-based inference in missing response problems and its application in observational studies. *J. R. Statist. Soc.* B, **69**, 101–122.

Qin, J., Zhang, B. and Leung, D. H. Y. (2009) Empirical likelihood in missing data problems. *J. Am. Statist. Ass.*, **104**, 1492–1503.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994) Estimation of regression coefficients when some regressors are not always observed. *J. Am. Statist. Ass.*, **89**, 846–866.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995) Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Am. Statist. Ass.*, **90**, 106–121.

Robins, J. M., Sued, M., Gomez-Lei, Q. and Rotnitzky, A. (2007) Performance of double-robust estimators when "inverse probability" weights are highly variable. *Statist. Sci.*, **22**, 544–559.

Rosenbaum, P. and Rubin, D. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41–55.

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

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

Rubin, D. B. (1996) Multiple imputation after 18+ years. *J. Am. Statist. Ass.*, **91**, 473–489.

Sherwood, B., Wang, L. and Zhou, X.-H. (2013) Weighted quantile regression for analyzing health care cost data with missing covariates. *Statist. Med.*, **32**, 4967–4979.

Shorack, G. R. (2000) *Probability for Statisticians*. New York: Springer.

Sun, Y., Wang, H. J. and Gilbert, P. B. (2012) Quantile regression for competing risks data with missing cause of failure. *Statist. Sin.*, **22**, 703–728.

Tan, Z. (2010) Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika*, **97**, 661–682.

Tsiatis, A. A. (2006) *Semiparametric Theory and Missing Data*. New York: Springer.

van der Vaart, A. W. (1998) *Asymptotic Statistics*. Cambridge: Cambridge University Press.

van der Vaart, A. W. and Wellner, J. (1996) *Weak Convergence and Empirical Processes: with Applications to Statistics*. New York: Springer.

Wang, D. and Chen, S. X. (2009) Empirical likelihood for estimating equations with missing values. *Ann. Statist.*, **37**, 490–517.

Wang, Q. and Rao, J. N. K. (2002) Empirical likelihood-based inference under imputation for missing response data. *Ann. Statist.*, **30**, 896–924.

Wei, Y., Ma, Y. and Carroll, R. J. (2012) Multiple imputation in quantile regression. *Biometrika*, **99**, 423–438.

Wei, Y. and Yang, Y. (2014) Quantile regression with covariates missing at random. *Statist. Sin.*, **24**, 1277–1299.

Wu, C. and Sitter, R. R. (2001) A model-calibration approach to using complete auxiliary information from survey data. *J. Am. Statist. Ass.*, **96**, 185–193.

Yang, S., Kim, J.-K. and Shin, D. W. (2013) Imputation methods for quantile estimation under missing at random. *Statist. Interfc.*, **6**, 369–377.

Yi, G. Y. and He, W. (2009) Median regression models for longitudinal data with dropouts. *Biometrics*, **65**, 618–625.

Yoon, J. (2010) Quantile regression analysis with missing response with applications to inequality measures and data combination. (Available from `http://dx.doi.org/10.2139/ssrn.2952579`.)

Zhang, M., Tsiatis, A. A. and Davidian, M. (2008) Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics*, **64**, 707–715.

Zhao, P.-Y., Tang, M.-L. and Tang, N.-S. (2013) Robust estimation of distribution functions and quantiles with non-ignorable missing data. *Can. J. Statist.*, **41**, 575–595.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

   'Supplementary material to "A general framework for quantile estimation with incomplete data"'.