Statistica Sinica **24** (2014), 1277-1299
doi:http://dx.doi.org/10.5705/ss.2011.302

QUANTILE REGRESSION WITH COVARIATES MISSING AT RANDOM

Ying Wei and Yunwen Yang

Columbia University and Drexel University

Abstract: Regression quantiles can be underpowered or biased when there are missing values in some covariates. We propose a method that produces consistent linear quantile estimation in the presence of missing covariates. The proposed method corrects bias by constructing unbiased estimating equations that simultaneously hold at all the quantile levels. It utilizes all the available data, and produces uniformly consistent estimators. An iterative EM-type algorithm is provided for solving the estimating equations. The finite sample performance of the method is investigated in a simulation study. Finally, the methodology is applied to data from the National Health and Nutrition Examination Survey.

Key words and phrases: Missing data, missing at random, quantile regression, regression quantiles.

1. Introduction

Quantile regression Koenker and Bassett (1978) has been increasingly popular due to its flexibility in modelling the relationship between a response variable and its covariates. For example, quantile regression has been shown to be favoured in obesity studies Terry, Wei, and Essenman (2007), since understanding how the covariates impact the upper quantiles of body mass index is more important than that on the mean level. These studies often rely on conducting surveys among a representative population. Inevitably, such survey data contain missing observations. For example, the respondents may be reluctant to answer one or more survey items Brick and Kalton (1996). Ignoring incomplete observations reduces estimation efficiency and, more importantly, can lead to biased estimation (Little and Rubin (1992)). Appropriate statistical methods are needed to correct bias, which we will illustrate below in a small National Health and Nutrition Examination Survey (NHANES) study. In that study, the researchers are interested in investigating the association between nutritional intake and waist-circumference, a known risk factor for cardiovascular diseases. The effect of total-fat intake on the upper quantiles of waist-circumference are seriously underestimated if some of the relatively heavier subjects choose not to report certain food items.

In this paper, we propose a method to consistently estimate a linear quantile model with some covariates missing at random(MAR). We assume the model

$$Q_{y_i}(\tau; \mathbf{x}_i, \mathbf{z}_i) = \mathbf{x}_i^\top \boldsymbol{\beta}_{0,\tau} + \mathbf{z}_i^\top \boldsymbol{\gamma}_{0,\tau}, \quad i = 1, \ldots, n, \tag{1.1}$$

holds true at all quantile levels $\tau \in (0, 1)$, where $Q_{y_i}(\tau; \mathbf{x}_i, \mathbf{z}_i)$ is the τ-th conditional quantile of the response variable y_i given $\mathbf{x}_i$ and $\mathbf{z}_i$. We assume that the covariate $\mathbf{z}_i$ is q-dimensional including a vector of one's, and $\mathbf{x}_i$ is p-dimensional. We assume that the covariate $\mathbf{z}_i$ and response y_i are completely observed, but some values of $\mathbf{x}_i$ are missing at random, where the probability of missingness may depend on y_i or $\mathbf{z}_i$, but not on $\mathbf{x}_i$ itself. If the missingness of $\mathbf{x}_i$ depends on the covariates only, the regression quantiles based on the observed data only are still unbiased (Lipstiz et al. (1997)) despite some loss of efficiency. If the missingness depends on y_i, then using the observed data alone can lead to substantial bias. The current work is aimed at obtaining consistent estimators of quantile coefficients $(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau)$ when some covariates are missing at random.

Dealing with missing covariates is a long-standing research topic in statistics. For mean regression, classical approaches include complete-case analysis (when missing is completely at random), conditional mean imputation (Afifi and Elashoff (1969a,b)), likelihood based approaches including multiple imputations(Rubin (1978, 1987) and Rubin and Schaferm (1990)). Little (1992) has a nice review of these approaches. In quantile regression the area remains largely undeveloped, mainly because most approaches have relied on parametric likelihood and cannot be applied directly to quantile regression. In addition, due to the non-additivity of the quantile regression objective function, conditional mean imputation does not lead to unbiased estimation. The estimating equation approach of Robins, Rotnitzky, and Zhao (1994) has become a popular method in recent years due to its efficiency and robustness, but it does not avoid distributional assumptions. Among works on handling missing data in quantile regression, Lipstiz et al. (1997) considered an inverse weighting approach to correct the bias due to longitudinal drop-outs. For the same type of data, Yi and He (2009) extended the inverse probability weighted generalized estimating equations proposed by Robins, Rotnitzky, and Zhao (1995) to correct the bias from longitudinal drop outs. Wei, Ma, and Carroll (2012) considered multiple imputation approach when missingness is completely at random. Our approach is based on constructing unbiased estimating equations that hold for all quantile levels. The joint modelling approach is an extension of Wei and Carroll (2009). Although the methods differ in their objective functions and computational algorithms, and inthe derivation of the asymptotic results, to our best knowledge, this is the first attempt at correcting the bias due to MAR covariates in quantile regression.

The rest of the paper is organized as follows. The proposed method is discussed in Section 2, where we also develop an algorithm and establish the asymptotic properties of the resulting estimators. A simulation study is presented in Section 3. The proposed method is applied to NHANES data in Section 4 to show how it can handle missing covariates. Proofs are included in the Appendix.

2. Methods

Let δ_i be the binary indicator for the existence of $\mathbf{x}_i$. We construct the unbiased estimating function φ_τ

$$\begin{aligned}\varphi_\tau(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau, y_i, \mathbf{x}_i, \mathbf{z}_i, \delta_i) = &\ \delta_i \Psi_\tau(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau, y_i, \mathbf{x}_i, \mathbf{z}_i) \\ &+(1-\delta_i) \int_{\mathbf{x}} \Psi_\tau(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau, y_i, \mathbf{x}_i, \mathbf{z}_i) f(\mathbf{x}|\delta_i, y_i, \mathbf{z}_i) d\mathbf{x}, \quad (2.1)\end{aligned}$$

where $(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau)$ are unknown parameters,

$$\Psi_\tau(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau, y, \mathbf{x}, \mathbf{z}) = \{\tau - I(y - \mathbf{x}^\top \boldsymbol{\beta}_\tau - \mathbf{z}^\top \boldsymbol{\gamma}_\tau < 0)\}(\mathbf{x}^\top, \mathbf{z}^\top)^\top;$$

and $f(\mathbf{x}|\delta_i, y_i, \mathbf{z}_i)$ is the conditional density of $\mathbf{x}$ given the observed $(\delta_i, y_i, \mathbf{z}_i)$. Here $\Psi_\tau(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau, y, \mathbf{x}, \mathbf{z})$ is the original estimating function for quantile regression, and (2.1) sets out its conditional expectation given the observed data. Let $(\boldsymbol{\beta}_{0,\tau}, \boldsymbol{\gamma}_{0,\tau})$ be the true conditional quantile coefficients. Then one can show that

$$E_{y_i}\Big\{\varphi_\tau(\boldsymbol{\beta}_{0,\tau}, \boldsymbol{\gamma}_{0,\tau}, y_i, \mathbf{x}_i, \mathbf{z}_i, \delta_i)|\delta_i, \mathbf{x}_i, \mathbf{z}_i\Big\} = \mathbf{0}$$

for any $(\delta_i, \mathbf{x}_i, \mathbf{z}_i)$, so the estimating function $\varphi_\tau()$ is unbiased and leads to the unbiased sample estimating equations

$$\sum_{i=1}^n \varphi_\tau(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau, y_i, \mathbf{x}_i, \mathbf{z}_i, \delta_i) = 0. \quad (2.2)$$

The equations depend on the conditional density $f(\mathbf{x}|\delta_i, y_i, \mathbf{z}_i)$, which needs to be estimated. By Bayes theorem, we decompose this density as

$$f(\mathbf{x}|\delta_i, y_i, \mathbf{z}_i) = \frac{p(\delta_i|\mathbf{x}, y_i, \mathbf{z}_i) f(y_i|\mathbf{x}, \mathbf{z}_i) f(\mathbf{x}|\mathbf{z}_i)}{\int_{\mathbf{x}} p(\delta_i|\mathbf{x}, y_i, \mathbf{z}_i) f(y_i|\mathbf{x}, \mathbf{z}_i) f(\mathbf{x}|\mathbf{z}_i) d\mathbf{x}}.$$

This decomposition implicitly assumes that the probabililty of observation $p(\delta_i|\mathbf{x}, y_i, \mathbf{z}_i)$ is bounded away from zero for any $(y, \mathbf{z})$, a common assumption in missing data methods. Under the assumption that the $\mathbf{x}$ are missing at random, δ_i does not depend on $\mathbf{x}$ conditioning on the observed y_i and $\mathbf{z}_i$, $p(\delta_i|\mathbf{x}, y_i, \mathbf{z}_i) = p(\delta_i|y_i, \mathbf{z}_i)$. Consequently, the representation of $f(\mathbf{x}|\delta_i, y_i, \mathbf{z}_i)$ can be simplified to

$$f(\mathbf{x}|\delta_i, y_i, \mathbf{z}_i) = f(\mathbf{x}|y_i, \mathbf{z}_i) = \frac{f(y_i|\mathbf{x}, \mathbf{z}_i) f(\mathbf{x}|\mathbf{z}_i)}{\int_{\mathbf{x}} f(y_i|\mathbf{x}, \mathbf{z}_i) f(\mathbf{x}|\mathbf{z}_i) d\mathbf{x}}. \quad (2.3)$$

The first conditional density $f(y_i|\mathbf{x},\mathbf{z}_i)$ is unspecified under the quantile regression framework following Wei and Carroll (2009), we assume that the linear model (1.1) holds for all quantile levels. This holds for the location-scale model. We denote the true quantile coefficient process as $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\gamma}_0(\tau)$, while $\boldsymbol{\beta}_{0,\tau}$ and $\boldsymbol{\gamma}_{0,\tau}$ denote the true quantile coefficients at the τth quantile. Under the joint modelling assumption, the conditional density $f(y_i|\mathbf{x},\mathbf{z}_i)$ is then

$$f(y_i|\mathbf{x},\mathbf{z}_i) = \left[\frac{\partial\{\mathbf{x}^\top\boldsymbol{\beta}_0(\tau)+\mathbf{z}_i^\top\boldsymbol{\gamma}_0(\tau)\}}{\partial\tau}\right]^{-1}\Bigg|_{\tau=\tau_{y_i}\{\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0(\tau)\}},$$

where $\tau_{y_i}\{\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma}_0(\tau)\} = \inf\{\tau\in(0,1):\mathbf{x}^\top\boldsymbol{\beta}_0(\tau)+\mathbf{z}_i^\top\boldsymbol{\gamma}_0(\tau) > y_i\}$ denotes the quantile level of y_i.

The density $f(y_i|\mathbf{x},\mathbf{z}_i)$ depends on the unknown coefficient functions $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\gamma}_0(\tau)$, which are of infinite dimension. For any $\boldsymbol{\beta}(\tau)\in R^p\times(0,1)$, we approximate it by natural linear splines, expanding from a series of quantile coefficients on a fine grid of quantile levels. Specifically, we choose quantile levels $\tau_k = k/(k_n+1)$, $k=1,\ldots,k_n$, where k_n is the number of quantile levels. We then define $\widetilde{\boldsymbol{\beta}}(\tau)$ as a p-dimensional piecewise linear function on [0,1], that satisfies $\widetilde{\boldsymbol{\beta}}(\tau_k) = \boldsymbol{\beta}(\tau_k)$ and $\widetilde{\boldsymbol{\beta}}'(0) = \widetilde{\boldsymbol{\beta}}'(1) = \mathbf{0}$. If $\boldsymbol{\beta}(\tau)$ is a smooth function, then it can be well approximated by $\widetilde{\boldsymbol{\beta}}(\tau)$ given a sufficient number of quantile levels k_n. We can approximate $\boldsymbol{\gamma}(\tau)$ in the same way. Let $f\{y|\mathbf{x},\mathbf{z},\boldsymbol{\beta}(\tau),\boldsymbol{\gamma}(\tau)\}$ be the conditional density function of y given $(\mathbf{x},\mathbf{z})$ that is induced from the quantile function $\mathbf{x}^\top\boldsymbol{\beta}(\tau)+\mathbf{z}^\top\boldsymbol{\gamma}(\tau)$. We then take

$$\begin{aligned}
f\{y|\mathbf{x},\mathbf{z},\boldsymbol{\beta}(\tau),\boldsymbol{\gamma}(\tau)\} &\approx \widetilde{f}\{y|\mathbf{x},\mathbf{z},\widetilde{\boldsymbol{\beta}}(\tau),\widetilde{\boldsymbol{\gamma}}(\tau)\}\\
&\widehat{=} \left[\frac{\partial\{\mathbf{x}^\top\widetilde{\boldsymbol{\beta}}(\tau)+\mathbf{z}_i^\top\widetilde{\boldsymbol{\gamma}}(\tau)\}}{\partial\tau}\right]^{-1}\Bigg|_{\tau=\tau_{y_i}\{\widetilde{\boldsymbol{\beta}}(\tau),\widetilde{\boldsymbol{\gamma}}(\tau)\}}\\
&= \textstyle\sum_{k=1}^{k_n}\dfrac{\tau_{k+1}-\tau_k}{(\mathbf{x}^\top\boldsymbol{\beta}_{\tau_{k+1}}+\mathbf{z}^\top\boldsymbol{\gamma}_{\tau_{k+1}})-(\mathbf{x}^\top\boldsymbol{\beta}_{\tau_k}+\mathbf{z}^\top\boldsymbol{\gamma}_{\tau_k})}\\
&\quad\times I\{(\mathbf{x}^\top\boldsymbol{\beta}_{\tau_k}+\mathbf{z}^\top\boldsymbol{\gamma}_{\tau_k})\le y<(\mathbf{x}^\top\boldsymbol{\beta}_{\tau_{k+1}}+\mathbf{z}^\top\boldsymbol{\gamma}_{\tau_{k+1}})\},
\end{aligned}\tag{2.4}$$

where $\boldsymbol{\beta}_{\tau_k}$ and $\boldsymbol{\gamma}_{\tau_k}$ are unknown parameters at the τ_kth quantile. WE thus incorporate the unknown coefficient functions $\{\boldsymbol{\beta}(\tau),\boldsymbol{\gamma}(\tau)\}$ into the conditional density $f(y_i|\mathbf{x},\mathbf{z}_i)$ with a finite number of unknown coefficients. Replacing $f(y_i|\mathbf{x},\mathbf{z}_i)$ by $\widetilde{f}\{y|\mathbf{x},\mathbf{z},\widetilde{\boldsymbol{\beta}}(\tau),\widetilde{\boldsymbol{\gamma}}(\tau)\}$, we approximate the estimating function (2.1) by

$$\begin{aligned}
\widetilde{\varphi}_\tau(\boldsymbol{\beta}_\tau,\boldsymbol{\gamma}_\tau,y_i,\mathbf{x}_i,\mathbf{z}_i,\delta_i) &= \delta_i\Psi_\tau(\boldsymbol{\beta}_\tau,\boldsymbol{\gamma}_\tau,y_i,\mathbf{x}_i,\mathbf{z}_i)\\
&\quad+(1-\delta_i)\int_{\mathbf{x}}\Psi_\tau(\boldsymbol{\beta}_\tau,\boldsymbol{\gamma}_\tau,y_i,\mathbf{x},\mathbf{z}_i)\widetilde{f}(\mathbf{x}|y_i,\mathbf{z}_i)d\mathbf{x},
\end{aligned}$$

where

$$\widetilde{f}(\mathbf{x}|y_i, \mathbf{z}_i) = \frac{\widetilde{f}\{y_i|\mathbf{x}, \mathbf{z}_i, \widetilde{\boldsymbol{\beta}}(\tau), \widetilde{\boldsymbol{\gamma}}(\tau)\} f(\mathbf{x}|\mathbf{z}_i)}{\int_{\mathbf{x}} \widetilde{f}\{y_i|\mathbf{x}, \mathbf{z}_i, \widetilde{\boldsymbol{\beta}}(\tau), \widetilde{\boldsymbol{\gamma}}(\tau)\} f(\mathbf{x}|\mathbf{z}_i) d\mathbf{x}},$$

and $\widetilde{f}\{y_i|\mathbf{x}, \mathbf{z}_i, \widetilde{\boldsymbol{\beta}}(\tau), \widetilde{\boldsymbol{\gamma}}(\tau)\}$ is defined at (2.4). We then take k_n sets of working estimating equations,

$$\widetilde{S}_{n,\tau}(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau) = n^{-1} \sum_{i=1}^{n} \widetilde{\varphi}_\tau(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau, y_i, x_i, z_i, \delta_i) = 0, \text{ for } \tau = \frac{1}{k_n+1}, \ldots, \frac{k_n}{k_n+1}, \tag{2.5}$$

that need to be solved on the k_n quantile levels simultaneously. In the next subsection, we outline an iterative algorithm to obtain the $\boldsymbol{\beta}_\tau$ and $\boldsymbol{\gamma}_\tau$ estimates on a grid of quantile levels.

2.1. Main estimation algorithm for solving (2.5)

Let ν be the indicator of iteration steps. The algorithm is outlined as follows.

Step 1 Set the initial values of $(\boldsymbol{\beta}_{\tau_k}^{(0)}, \boldsymbol{\gamma}_{\tau_k}^{(0)})_{k=1}^{k_n}$ based on the quantile regression omitting the cases with missing data.

Step 2 Update the distribution $f^{(\nu)}(\mathbf{x}|y_i, \mathbf{z}_i)$ for missing $\mathbf{x}_i$ based on $(\boldsymbol{\beta}_{\tau_k}^{(\nu-1)}, \boldsymbol{\gamma}_{\tau_k}^{(\nu-1)})_{k=1}^{k_n}$ from the previous iteration, so

$$f^{(\nu)}(\mathbf{x}|y_i, \mathbf{z}_i) = \frac{f(y_i|\mathbf{x}, \mathbf{z}_i, \boldsymbol{\beta}^{(\nu-1)}(\tau), \boldsymbol{\gamma}^{(\nu-1)}(\tau)) f(\mathbf{x}|\mathbf{z}_i)}{\int_x f(y_i|\mathbf{x}, \mathbf{z}_i, \boldsymbol{\beta}^{(\nu-1)}(\tau), \boldsymbol{\gamma}^{(\nu-1)}(\tau)) f(\mathbf{x}|\mathbf{z}_i) dx},$$

where $\boldsymbol{\beta}^{(\nu)}(\tau)$ and $\boldsymbol{\gamma}^{(\nu)}(\tau)$ are natural linear splines expanded from $\boldsymbol{\beta}^{(\nu)}$'s and $\boldsymbol{\gamma}^{(\nu)}$'s and, as defined at (2.4), $f(y_i|\mathbf{x}, \mathbf{z}_i, \boldsymbol{\beta}^{(\nu)}(\tau), \boldsymbol{\gamma}^{(\nu)}(\tau))$ is the conditional density of y_i given $(\mathbf{x}, \mathbf{z}_i)$ in the ν-th iteration.

Step 3 Update $(\boldsymbol{\beta}_{\tau_k}^{(\nu)}, \boldsymbol{\gamma}_{\tau_k}^{(\nu)})_{k=1}^{k_n}$ based on the new estimating functions with $f^{(\nu)}(\mathbf{x}|y_i, \mathbf{z}_i)$. Numerical integrations can be used to perform this step. Let $\tilde{x}_i = (\tilde{x}_{i,1}, \ldots, \tilde{x}_{i,m})$ be a fine grid of possible $\mathbf{x}_i$ values. The estimating equations can be written as

$$\sum_{i=1}^{n} \Big\{ \delta_i \Psi_{\tau_k}(\boldsymbol{\beta}_{\tau_k}, \boldsymbol{\gamma}_{\tau_k}, y_i, \mathbf{x}_i, \mathbf{z}_i) + (1-\delta_i) \sum_{j=1}^{m} \Big[\Psi_{\tau_k}(\beta_{\tau_k}, \gamma_{\tau_k}, y_i, \tilde{\mathbf{x}}_{i,j}, \mathbf{z}_i) \\ f^{(\nu)}(\tilde{\mathbf{x}}_{i,j}|y_i, z_i) \times (\tilde{\mathbf{x}}_{i,j+1} - \tilde{\mathbf{x}}_{i,j}) \Big] \Big\} = 0, \tag{2.6}$$

where $k = 1, \ldots, k_n$. Solving (2.6) can be achieved by weighted quantile regression with response y_i, the covariates $(\tilde{\mathbf{x}}_{i,j}, \mathbf{z}_i)$, and weights δ_i or $(1-\delta_i) f^{(\nu)}(\tilde{\mathbf{x}}_{i,j}|y_i, z_i) \times (\tilde{\mathbf{x}}_{i,j+1} - \tilde{\mathbf{x}}_{i,j})$.

Step 4 Repeat Steps 2 and 3 until the algorithm converges.

In the algorithm, we implicitly assumed that $f(\mathbf{x}|\mathbf{z})$ is known, which needs to be estimated in most applications. In later sections, we present an algorithm adapted from Robins, Rotnitzky, and Zhao (1994) to obtain consistent estimation of $f(\mathbf{x}|\mathbf{z})$ parametrically under the assumption that $\mathbf{x}$ is missing at random. Other estimation methods exist, such as Titterington and Mill (1983), Wang (2008), Dubnicka (2009), Iacus and Torre (2002), and Huang and Salleb-Aouissi (2009). Once $f(\mathbf{x}|\mathbf{z})$ is estimated, we replace $f(\mathbf{x}|\mathbf{z})$ in Step 2 with its estimator $\widehat{f}_n(\mathbf{x}|\mathbf{z})$. The rest of the algorithm remains unchanged. We let $\widehat{S}_{n,\tau}(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau)$ be the counterpart of $\widetilde{S}_{n,\tau}(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau)$, replacing the true density $f(\mathbf{x}|\mathbf{z}_i)$ by its estimator $\widehat{f}_n(\mathbf{x}|\mathbf{z}_i)$, and take

$$\widehat{\mathbf{S}}_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = (\widehat{S}_{n,\tau_1}(\boldsymbol{\beta}_{\tau_1}, \boldsymbol{\gamma}_{\tau_1})^\top, \widehat{S}_{n,\tau_2}(\boldsymbol{\beta}_{\tau_2}, \boldsymbol{\gamma}_{\tau_2})^\top, \ldots, \widehat{S}_{n,\tau_{k_n}}(\boldsymbol{\beta}_{\tau_{k_n}}, \boldsymbol{\gamma}_{\tau_{k_n}})^\top)$$

as the entire k_n sets of working estimation equations at the quantile levels $\Omega = \{\tau_1, \ldots, \tau_{k_n}\}$. Here $\boldsymbol{\beta} = (\boldsymbol{\beta}_{\tau_1}^\top, \ldots, \boldsymbol{\beta}_{\tau_{k_n}}^\top)^\top$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{\tau_1}^\top, \ldots, \boldsymbol{\gamma}_{\tau_{k_n}}^\top)^\top$ are two concatenated coefficient vectors.

2.2. Asymptotics

Let $(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n)$ be the solution of $\widehat{\mathbf{S}}_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = 0$. Let $\widehat{\boldsymbol{\beta}}_n(\tau)$ and $\widehat{\boldsymbol{\gamma}}_n(\tau)$ be the natural linear splines expanded from $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\gamma}}_n$, respectively. In this section, we derive the uniform consistency of $(\widehat{\boldsymbol{\beta}}_n(\tau), \widehat{\boldsymbol{\gamma}}_n(\tau)$. We make the following assumption on the conditional density $f(\mathbf{x}|\mathbf{z})$ and its estimator $\widehat{f}_n(\mathbf{x}|\mathbf{z})$.

Assumption 0.

(i) The conditional density $f(\mathbf{x}|\mathbf{z})$ is bounded away from zero and infinity for all $(\mathbf{x}, \mathbf{z})$.

(ii) There exists a consistent estimator $\widehat{f}_n(\mathbf{x}|\mathbf{z})$ of $f(\mathbf{x}|\mathbf{z})$, such that,

$$\max_i \widehat{f}_n(\mathbf{x}|\mathbf{z}_i) < M\|\mathbf{x}\|^{-(p+2)}. \tag{2.7}$$

(iii) $E\{\max_i \|\mathbf{z}_i\|\} < \infty$

Let $S_\tau(\boldsymbol{\beta}, \boldsymbol{\gamma}) = E_{y_i}\{\varphi_\tau(\boldsymbol{\beta}, \boldsymbol{\gamma}, y_i, \mathbf{x}_i, \mathbf{z}_i, \delta_i)|\mathbf{x}_i, \mathbf{z}_i, \delta_i\}$ be the expectation of the estimating equations at the τth quantile level, evaluated under the *true* density function $f(\mathbf{x}|\delta, y, \mathbf{z})$, and $\widetilde{S}_\tau(\boldsymbol{\beta}, \boldsymbol{\gamma})$ be its counterpart replacing the density component $f(y|\mathbf{x}, \mathbf{z})$ by its approximation $\widetilde{f}\{y|\mathbf{x}, \mathbf{z}, \widetilde{\boldsymbol{\beta}}_0(\tau), \widetilde{\boldsymbol{\gamma}}_0(\tau)\}$ $\widetilde{f}$ as at (2.4).

Assumption 1. The true coefficient $(\boldsymbol{\beta}_{0,\tau}, \boldsymbol{\gamma}_{0,\tau})$ is the unique solution to the equation $S_\tau(\boldsymbol{\beta}, \boldsymbol{\gamma}) = 0$, for all $\tau \in (0, 1)$, and there exists a $(\boldsymbol{\beta}_\tau^*, \boldsymbol{\gamma}_\tau^*)$ that uniquely solves the equation $\widetilde{S}_\tau(\boldsymbol{\beta}, \boldsymbol{\gamma}) = 0$, for all $\tau \in (0, 1)$.

Assumption 2. There is a compact set $\Theta \in \mathbf{R}^{p+q}$, such that

$$|\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n)| \leq \inf_{(\boldsymbol{\theta}, \boldsymbol{\gamma}) \in \boldsymbol{\Theta}} |\widehat{\mathbf{S}}_n(\boldsymbol{\beta}, \boldsymbol{\gamma})| + o_p(1).$$

Assumption 1 is the identifiability condition commonly assumed in the quantile regression literature, while Assumption 2 is used to ensure that the solution to the approximated working estimating equations is confined to a compact set $\mathbf{\Theta}$.

If $\{\boldsymbol{\beta}_0(\tau), \boldsymbol{\gamma}_0(\tau)\}$ are the true quantile coefficient functions, for any $(\mathbf{x}, \mathbf{z})$, $(\mathbf{x}^\top \boldsymbol{\beta}_0(\tau) + \mathbf{z}^\top \boldsymbol{\gamma}_0(\tau))$ defines a conditional quantile function. We set $h(\tau, \mathbf{x}, \mathbf{z}) = 1/(\mathbf{x}^\top \boldsymbol{\beta}_0'(\tau) + \mathbf{z}^\top \boldsymbol{\gamma}_0'(\tau))$, the density of y given $(\mathbf{x}, \mathbf{z})$ at the τ-th quantile. We call this the conditional quantile density function, and its reciprocal is known as the sparsity function (Koenker and Xiao (2004); Welsh (1988)).

Assumption 3. The coefficients $(\boldsymbol{\beta}_0(\tau), \boldsymbol{\gamma}_0(\tau))$ are smooth functions on $(0, 1)$ and, for any $(\mathbf{x}, \mathbf{z})$,

(i) $0 < h(\tau, \mathbf{x}, \mathbf{z}) < \infty$, and $\lim_{\tau \to 0} h(\tau, \mathbf{x}, \mathbf{z}) = \lim_{\tau \to 1} h(\tau, \mathbf{x}, \mathbf{z}) = 0$;

(ii) there exist constants M and $\nu_1, \nu_2 > -1$ such that its first derivative is bounded by

$$\sup_{\mathbf{x}} |h'(\tau, \mathbf{x}, \mathbf{z})| < M\tau^{\nu_1}(1-\tau)^{\nu_2}. \tag{2.8}$$

Assumption 3 has the coefficient functions smooth enough to be well approximated by normalized B-splines. It also implicitly assumes that the conditional density $f(y|\mathbf{x}, \mathbf{z})$ is continuous, bounded away from zero and infinity, and diminishes to zero as quantile level goes to zero and one. These conditions are common, although in different forms. The inequality (2.8) holds for a fairly wide range of distributions, including those in the exponential family.

Theorem 1. *Under Assumptions* 0–3, *for* $k_n \to \infty$, *and* $k_n n^{-1} \to 0$,

$$\sup_{\tau \in [1/(k_n+1),\ k_n/(k_n+1)]} \|\widehat{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}_0(\tau)\| = o_p(1);$$
$$\sup_{\tau \in [1/(k_n+1),\ k_n/(k_n+1)]} \|\widehat{\boldsymbol{\gamma}}_n(\tau) - \boldsymbol{\gamma}_0(\tau)\| = o_p(1).$$

3. Simulations

3.1. Models and settings

To show the performance of the proposed methods, we conducted simulations study based on the location-scale model

$$y_i = 1 + 2x_i + 0.5z_i + (1 + 0.5x_i)e_i. \tag{3.1}$$

Here the error term e_i is the standard normal. Under (3.1), the intercept function is $1 + \Phi^{-1}(\tau)$, where $\Phi^{-1}(\tau)$ is the quantile function of the standard normal. The slope function associated with x is $2 + 0.5\Phi^{-1}(\tau)$, and that of z is constantly 0.5

at any quantile level. We considered the following distributions of the covariates (x, z).

Setting 1: $z \sim$ Bernoulli(0.5) ; and $x \sim N(4.5 + z, 1)$,

Setting 2: $z \sim$ Uniform(0, 1); and $x \sim N(4.5 + z, 1)$,

Setting 3: $z \sim$ Uniform(0, 1); and $x \sim 4.5 + z + (\chi_1^2 - 1)/\sqrt{2}$.

In Setting 3, the covariate x is chi-square, but its conditional mean and variance given z are the same as in Setting 2. We assumed that the covariate x_i could be missing with probability $P(\delta_i = 0|y_i) = (1 + \exp(-12 + 1.6y_i))^{-1}$. That results in approximately 20% missing observations in x.

3.2. Estimation of the conditional distribution $f(x|z)$

To estimate the conditional distribution $f(x|z)$, we assume that x and z are linearly associated as

$$x = a + bz + e, \quad e \sim N(0, \sigma^2),$$

where a, b, and σ are unknown parameters. Under the MAR assumption, the missingness in x is independent of the underlying x values when conditioning on both y and z. However, it could be related to x conditioning on z only and, as a result, standard density estimation using the completely observed (x, z) only could lead to substantial bias. Several methods have been proposed to estimate (x, z) under this scenario, such as the ignorable likelihood approach (Little and Zhang (1992)) and the estimating equations approach in Robins, Rotnitzky, and Zhao (1994). We adapt the latter method, and take the estimating equations for a, b, σ as

$$\frac{\delta}{\pi(x)}\Psi(x, z, a, b, \sigma^2) - \frac{\delta - \pi(x)}{\pi(x)}E\{\Psi(x, z, a, b, \sigma^2)|y\} = 0,$$

where δ is the binary indicator for the existence of x, $\pi(x)$ is the probability of x being observed, and the $\Psi(x, z, a, b, \sigma^2)$ are the estimating functions for (a, b, σ) when all the data are observed. Specifically, we estimate $\pi(x)$ using the logistic regression of δ on y and z, and estimate $E\{\Psi(x, z, a, b, \sigma^2)|y\} \widehat{=} \phi(y, z)$ by regressing $\Psi(x, z, \tilde{a}, \tilde{b}, \tilde{\sigma}^2)$ against y and z, where $\tilde{a}, \tilde{b}, \tilde{\sigma}^2$ are naive estimates using the completely observed data only. To ensure the flexibility of $\phi(y, z)$, we assume a partly linear model $\phi(y, z) = g(y) + \eta z$, where $g(y)$ is nonparametric function and η is the linear coefficient of z. With estimated π and ϕ, we obtain the estimate of (a, b, σ), which we denote as $(\hat{a}_1, \hat{b}_1, \hat{\sigma}_1)$. Then the conditional density $f(x|z)$ can be estimated by

$$\widehat{f}(x|z) = \phi(\hat{a}_1 + \hat{b}_1 z, \hat{\sigma}_1^2),$$

where $\phi()$ is the density of standard normal.

We applied this procedure to all three settings. Since the algorithm assumes the normality of x given z, it produces consistent estimation of $f(x|z)$ in Settings 1 and 2, but may be biased when applied to Setting 3. We use the first two settings illustrate the performance of proposed methods with finite sample sizes, and Setting 3 to investigate the robustness of the proposed estimator against misspecified conditional distributions of $x|z$. One can assess the normality assumption by regressing x_i against y_i and z_i on the complete set, and use a QQ plot to check whether the residuals look normal.

3.3. Results and comparisons

We simulated 100 Monte-Carlo samples from each of the three settings with $n = 200, 500$, and 2,000. For each Monte carlo sample, we estimated the quantile coefficients using four approaches.

(1) The proposed iterative estimation algorithm was used to obtain the conditional quantile coefficient estimators at 40 evenly spaced quantile levels, when the density $f(x|z)$ was estimated with the algorithm in Section 3.2.

(2) The proposed iterative estimation algorithm was used with the *true* density $f(x|z)$.

(3) Unadjusted quantile regression was done using the completely observed data only.

(4) We estimated the coefficients by solving the estimation equations

$$\sum_{i=1}^{n} \frac{\delta_i \psi_\tau(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau, y_i, x_i, z_i)}{\pi(y_i, z_i)} = 0,$$

where $\pi(y_i, z_i)$ is the probability of the existence of x_i conditional on (y_i, z_i), estimated by a logistic regression of δ_i over y_i and z_i.

The comparison between the first two approaches help in understanding the impact of estimating $f(x|z)$. We compare the performances of the four estimators based on the X coefficient, similar results are found for the Z coefficients.

In Table 1, we present the mean square errors, mean biases, and empirical standard errors of the estimated X coefficients under Setting 1 from the four approaches at quantile levels 0.1, 0.5, and 0.9. Tables 2 and 3 present the same entries under Settings 2 and 3, respectively. In all the tables, RQ stands for the unadjusted quantile regression; IPW stands for the inverse probability weighting approach; MAR are estimates from the proposed method with estimated $f(x|z)$; and MARt are the estimates from the proposed method but using the true $f(x|z)$.

In Settings 1 and 2, where $f(x|z)$ is consistently estimated, the proposed estimation and the IPW method corrected bias to a large extent. The standard errors of the former are consistently smaller than those of latter, which suggests that the proposed method is more efficient than IPW, as it uses all the available observations in the estimating equations. The MAR and IPW estimations performed worse at lower quantiles. By design, the probability of missing X is fairly substantial at the lower quantiles of Y. The probability of missing x is as high as 90% when $Y = 6.1$(the 0.1th quantile). In Tables 1 and 2, one sees smaller biases as the sample sizes increase from 200 to 2,000.

The mean square errors of IPW estimates and the proposed estimates are smaller than those from the unadjusted quantile regression in most cases. However, when the quantile level is 0.1 and sample sizes are 200 and 500, the mean square errors from the IPW and the proposed method are larger than those from unadjusted quantile regressions. These methods need to estimate the conditional densities/probabilities, which adds extra variability. When sample sizes are small and the local probability of missingness is large, the extra variances can outweigh the benefits of the bias correction. The proposed estimates with true $f(x|z)$ (MAR_t) have smaller mean square errors than the unadjusted regression quantiles, which further indicates that the inflated mean square errors are due to the variability in estimating the conditional density $f(x|z)$. Therefore, in applications, we suggest evaluating the standard errors of the estimates using the bootstrap, to decide on an estimator.

The misspecified $f(x|z)$ in the Setting 3 undermines the performance of the proposed estimates, especially at lower quantiles. However, the mean square errors of the proposed MAR estimates are still smaller that those of the unadjusted quantile estimates and IPW estimates at median and upper quantiles.

4. Application: National Health and Nutrition Examination Survey

We applied our method to part of the nutrition data from National Health and Nutrition Examination Survey (NHANES) 2005-2006. The data were from 289 male African American adults between 25 and 45 years old. Daily nutrition intake was recorded, in addition to various body size measures. Since the waist circumference is known to be highly associated with the risk of cardiovascular diseases, it is of interest to understand how daily nutrition intake is related to waist circumference. For illustrative purposes, we consider the following linear quantile model to evaluate the association between the total-fat intake and the waist circumference(WC) while controlling for age:

$$Q_\tau(\mathrm{WC}) = \mu_\tau + \beta_{1,\tau}(\text{TOTAL-FAT} - \overline{\text{TOTAL-FAT}}) + \beta_{2,\tau}\mathrm{AGE}. \tag{4.1}$$

Here TOTAL-FAT is the actual daily total fat intake, ranging from 54.0 to 450.4 gm, $\overline{\text{TOTAL-FAT}}$ is the average total fat intake, and AGE is a binary indicator for the younger age group. We also considered a model including an interaction

Table 1. **(Setting 1)** The mean square erros, mean biases and emprical standard errors of the estimated X coefficients at quantile levels 0.1, 0.5 and 0.9 using different methods at different sample sizes.

	$n = 200$			$n = 500$			$n = 2{,}000$		
	$\tau=0.1$	$\tau=0.5$	$\tau=0.9$	$\tau=0.1$	$\tau=0.5$	$\tau=0.9$	$\tau=0.1$	$\tau=0.5$	$\tau=0.9$
	Setting 1: $z \sim Uniform(0,1)$								
	Mean Squre Errors								
RQ	0.466	0.228	0.222	0.375	0.177	0.125	0.342	0.134	0.063
IPW	0.779	0.211	0.218	0.637	0.304	0.171	0.406	0.221	0.046
MAR	0.666	0.185	0.206	0.493	0.090	0.076	0.413	0.042	0.030
MARt	0.446	0.116	0.177	0.255	0.051	0.071	0.083	0.017	0.018
	Mean Biases								
RQ	-0.598	-0.382	-0.224	-0.585	-0.381	-0.220	-0.576	-0.353	-0.206
IPW	-0.561	-0.194	-0.114	-0.542	-0.139	-0.075	-0.285	-0.034	-0.022
MAR	-0.499	-0.218	-0.159	-0.452	-0.182	-0.082	-0.442	-0.124	-0.074
MARt	-0.131	0.055	-0.085	-0.211	-0.067	-0.029	-0.042	0.033	-0.004
	Empirical Standard Errors								
RQ	0.328	0.288	0.417	0.191	0.179	0.281	0.098	0.095	0.143
IPW	0.684	0.416	0.455	0.585	0.533	0.409	0.572	0.468	0.216
MAR	0.645	0.370	0.427	0.548	0.241	0.265	0.468	0.163	0.156
MARt	0.659	0.337	0.414	0.459	0.218	0.265	0.290	0.128	0.136

between age group and total fat intake (not presented here); interaction was not significant at the quantiles of interest, so (4.1) was then employed.

We evaluated Model (4.1) at the 0.1th, 0.5th, and 0.9th quantiles, and found the total fat intake not associated with the lower quantile of waist circumference, but that it significantly impacts its median and upper quantile. Every unit increase of total fat intake results in an increase of the 0.9th quantile of waist-circumference by 0.10, and an increase of median by 0.06. The estimated TOTAL-FAT coefficients and their 95% confidence intervals are in Table 4. The age group is also significant at all the quantiles, the subjects in the younger age group tending to smaller waist circumference.

AS some overweight respondents may be reluctant to report their daily intakes, we mimicked this by assuming that those with waist circumferences larger than 90 inches (the median) do not report their total-fat intake values with probability 0.4. We then re-estimated regression quantiles using the remaining 80% "completely observed data". Although the intercept and age group coefficients wew not affected, TOTAL-FAT effect was seriously under-estimated: the TOTAL-FAT coefficient at the 0.9th quantile decreased from 0.10 to 0.01 and was no longer significant. A similar pattern was found at median. The estimated coefficients using the 80% "completely observed data", as well as their 95% confidence intervals, are listed in Table 4.

Table 2. **(Setting 2)** The mean square erros, mean biases and emprical standard errors of the estimated X coefficients at quantile levels 0.1, 0.5 and 0.9 using different methods at different sample sizes.

	$n = 200$			$n = 500$			$n = 2{,}000$		
	τ=0.1	τ=0.5	τ=0.9	τ=0.1	τ=0.5	τ=0.9	τ=0.1	τ=0.5	τ=0.9
	Setting 2: $z \sim Bernoulli(0.5)$								
	Mean Squre Errors								
RQ	0.406	0.216	0.257	0.370	0.190	0.114	0.337	0.132	0.071
IPW	0.797	0.182	0.221	0.593	0.158	0.098	0.420	0.145	0.040
MAR	0.703	0.176	0.200	0.408	0.090	0.073	0.298	0.031	0.026
MARt	0.401	0.105	0.198	0.282	0.054	0.066	0.082	0.018	0.017
	Mean Biases								
RQ	-0.538	-0.360	-0.214	-0.585	-0.402	-0.210	-0.571	-0.348	-0.221
IPW	-0.598	-0.164	-0.091	-0.498	-0.173	-0.076	-0.373	-0.028	-0.034
MAR	-0.476	-0.196	-0.108	-0.391	-0.178	-0.100	-0.326	-0.076	-0.057
MARt	-0.113	0.012	-0.059	-0.233	-0.094	-0.069	-0.059	0.030	-0.009
	Empirical Standard Errors								
RQ	0.338	0.294	0.461	0.179	0.170	0.263	0.099	0.102	0.149
IPW	0.661	0.427	0.463	0.678	0.615	0.313	0.529	0.380	0.198
MAR	0.690	0.372	0.434	0.573	0.255	0.251	0.437	0.160	0.153
MARt	0.626	0.325	0.441	0.481	0.212	0.248	0.284	0.134	0.132

We also applied the proposed procedure to estimate coefficients, pretending that the 20% of total-fat intake values were missing. As in the simulations, we choose 40 evenly spaced quantile levels. Figure 1 displays the qq-plots of the logarithm of TOTAL-FAT in the two age groups, and the residuals from regressing log transformed totalfat against age and waist circumference. These QQ plots suggest that the log transformed TOTAL-FAT follows an approximate normal distribution. We hence applied the estimation procedure in Subsection 3.2 to the log of TOTAL-FAT, and subsequently estimated the conditional distributions of TOTAL-FAT given age groups. The resulting estimated TOTAL-FAT coefficients are in Table 4, together with their 95% bootstrap confidence intervals with 50 bootstrap samples. The proposed estimation largely corrected the bias in TOTAL-FAT that was introduced from the removed observations. The estimated coefficient associated with the total fat intake was 0.08 for the median and 0.12 for the 0.9th quantile, close to the original estimates.

We plot in Figure 2 the three sets of regression lines at the 0.9th quantile. The solid lines are the original regression quantiles using all the data; the dotted lines are the regression quantiles using the 80% "completely observed" data only; the dashed lines are regression quantiles using the proposed method. Here the plus points are those data points whose total fat intakes were treated as missing,

Table 3. **(Setting 3)** The mean square erros, mean biases and emprical standard errors of the estimated X coefficients at quantile levels 0.1, 0.5 and 0.9 using different methods at different sample sizes.

	$n=200$			$n=500$			$n=2{,}000$		
	$\tau=0.1$	$\tau=0.5$	$\tau=0.9$	$\tau=0.1$	$\tau=0.5$	$\tau=0.9$	$\tau=0.1$	$\tau=0.5$	$\tau=0.9$
	Setting 3: Misspecified model								
	Mean Squre Errors								
RQ	0.348	0.188	0.442	0.260	0.119	0.145	0.253	0.067	0.039
IPW	0.638	0.230	0.407	0.487	0.091	0.136	0.542	0.301	0.070
MAR	0.764	0.175	0.356	0.663	0.080	0.130	1.033	0.045	0.038
MARt	0.764	0.209	0.389	0.416	0.073	0.133	0.093	0.018	0.032
	Mean Biases								
RQ	-0.443	-0.175	-0.099	-0.436	-0.225	-0.130	-0.489	-0.227	-0.089
IPW	-0.311	0.039	0.029	-0.211	-0.031	-0.027	-0.178	0.061	0.039
MAR	-0.669	-0.049	-0.004	-0.684	-0.073	-0.004	-0.944	-0.094	0.034
MARt	-0.169	0.048	-0.020	-0.224	-0.019	-0.020	-0.091	-0.006	0.014
	Empirical Standard Errors								
RQ	0.392	0.398	0.661	0.265	0.262	0.360	0.155	0.125	0.178
IPW	0.739	0.480	0.640	0.669	0.301	0.370	0.718	0.548	0.263
MAR	0.565	0.418	0.600	0.445	0.274	0.362	0.379	0.190	0.193
MARt	0.862	0.458	0.627	0.608	0.271	0.366	0.292	0.134	0.179

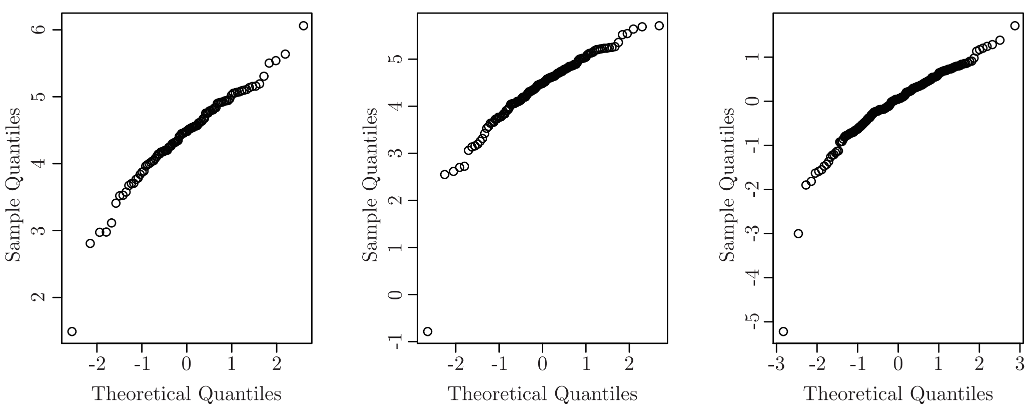

Figure 1. The qqplots to assess the normality of the log transformed totalfat. (a) is the QQ plot for the observed log transformed totalfat in the younger age group, (b) is that in the old age group. (c) is the QQ plot of the residuals from regressing the log transformed totalfat against age group and waist circumference.

and the circled ones are the remaining 80% of data points. From Figure 2, one sees a stronger association between fat intake and waist circumference among the heavier subjects. The slope is seriously underestimated after excluding some of

Table 4. Estimated TOTAL-FAT effect and their 95% confidence intervals in the NHANES study at quantile levels 0.5 and 0.9. RQ_{com} refers to the regression quantile estimates using the entire data; RQ_{obs} refers to the regression quantile estimates using 80% completely observed data; and RQ_{mis} refers to the estimates obtained from the proposed method assuming 20% total fat intakes are missing.

τ	RQ_{com}	RQ_{obs}	RQ_{mis}
0.5	0.06*	0.04	0.08*
	(0.02,0.09)	(-0.02,0.09)	(0.04,0.12)
0.9	0.10*	0.01	0.12*
	(0.01,0.17)	(-0.04,0.14)	(0.05,0.22)

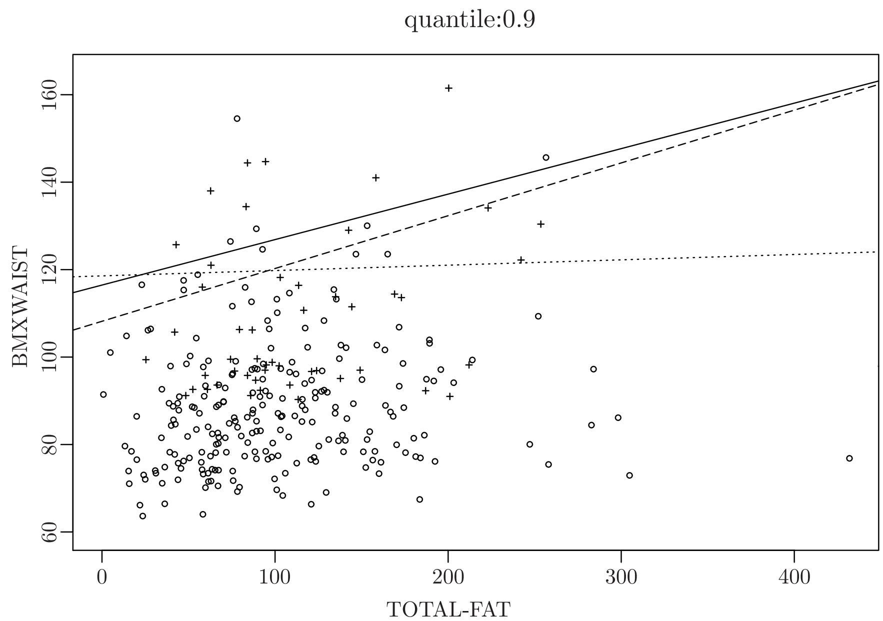

Figure 2. Comparison of the estimated regressions lines at 0.9th quantile. The plus signs indicate those data points whose TOTAL-FAT values are artificially treated as missing, and the circles are the rest of the 80% data points. The solid line is the original regression quantile using all of the data; the dotted line is the regression quantile only using the 80% “completely observed” data; the dashed line is the regression quantile using the proposed method.

them from the estimation, and the estimates using the proposed method successfully correct the bias.

5. Discussion

We have proposed to construct unbiased estimation equations for parameter estimation in a quantile regression model when some covariates contain missing values. Under suitable conditions, the estimator was shown to be uniformly consistent. In addition, since the method fully utilizes the entire data set, it improves the estimation efficiency.

Dealing with the missing covariates in the quantile regression context is challenging because the conditional density y given the covariates is unspecified. Thus, the classical parametric likelihood based approaches cannot be applied directly. Here, we adopted the similar joint modelling approach in Wei and Carroll (2009) to circumvent the difficulty, though the objectives of the two papers are different: we aim to correct the bias in regression quantiles from missing data, while Wei and Carroll (2009) handle mis-measured covariates. The proposed method differs from previous work in that estimating equations are constructed differently, the estimation algorithms need to be adapted, and the asymptotic results need to be studied separately.

The validity of the method relies on a correct specification of the conditional density $f(\mathbf{x}|\mathbf{z})$. In the simulation study and a data application, we modeled it parametrically for robustness against the possible misspecification, and obtained reasonably good results. Although nonparametric estimation methods are available to further improve flexibility, they are usually complex and known to have slow convergence rates. In addition, we assumed the conditional quantile functions to be linear at all quantile levels. This assumption holds readily for location-scale models. If needed, one can easily relax the linear quantile function to an arbitrary nonlinear or even nonparametric function. The algorithm remains largely unchanged after setting the linear function to the new regression function in the quantile regression estimating function. Thus, let $g_\tau(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}_\tau)$ be the conditional quantile function, with $\boldsymbol{\beta}$ containing linear or non-linear coefficients. If g_τ contains a nonparametric component, it can be approximated by appropriate spline functions, and then $\boldsymbol{\beta}$ includes the spline coefficients. We need to generalize the Ψ_τ function in (2.1) to

$$\Psi_\tau(\boldsymbol{\beta}_\tau, y, \mathbf{x}, \mathbf{z}) = \{\tau - I(y - g_\tau(y, \mathbf{x}, \mathbf{z}, \boldsymbol{\beta}_\tau) < 0)\}\partial g_\tau(y, \mathbf{x}, \mathbf{z}, \boldsymbol{\beta}_\tau)/\partial \boldsymbol{\beta}_\tau.$$

To improve the robustness of the estimation, one could appeal to Robins's (1994) estimating function

$$\Omega(Y, X, Z, \delta, \boldsymbol{\beta}) = \frac{\delta}{\pi(Y, Z)}\Psi(Y, X, Z, \boldsymbol{\beta}) + \frac{\delta - \pi(Y, Z)}{\pi(Y, Z)}E\{\Psi(Y, X, Z, \boldsymbol{\beta}) \mid Y, Z\},$$

where $\pi(Y, Z)$ is the probability of X being observed, estimated using logistic regression, and $E\{\Psi(Y, X, Z, \boldsymbol{\beta}) \mid Y, Z\}$ evaluated using a similar approach. The resulting estimators inherit double-robustness from Robin's estimating equation: if either $\pi(Y, Z)$ or $E\{\Psi(Y, X, Z, \beta)|Y, Z\}$ are estimated "consistently", the resulting estimators are consistent. However, such estimators may be less efficient from variability introduced by estimating $\pi(Y,Z)$ and its correlation with $E\{\Psi(Y,X,Z,\beta)|Y,Z\}$. Future research is needed on asymptotic behaviours.

Acknowledgement

The authors thank Professor Xuming He for some helpful suggestions. The research of Ying Wei was partly supported by the National Science Foundation (DMS-0906568 and DMS-1209023).

Appendix: Proof of Theorem 1

We take $(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau)$ as the unknown coefficient at the τ-th quantile, and write $\boldsymbol{\beta} = (\boldsymbol{\beta}_{\tau_1}^\top, \ldots, \boldsymbol{\beta}_{\tau_{k_n}}^\top)^\top$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{\tau_1}^\top, \ldots, \boldsymbol{\gamma}_{\tau_{k_n}}^\top)^\top$ as the concatenated coefficient vectors. Consequently, $\widetilde{\boldsymbol{\beta}}(\tau)$ and $\widetilde{\boldsymbol{\gamma}}(\tau)$ are natural linear splines expanded from $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. We take $(\boldsymbol{\beta}_{0,\tau}, \boldsymbol{\gamma}_{0,\tau})$ as the true coefficient at the τ-th quantile, $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ as the concatenated true coefficient vector at k_n quantile levels, and $\{\boldsymbol{\beta}_0(\tau), \boldsymbol{\gamma}_0(\tau)\}$ as the entire true coefficient processes.

As defined earlier, $S_\tau(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau)$ is the limiting estimating function of (2.1), and $\widetilde{S}_\tau(\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau)$ is its approximation when replacing the true coefficient function $\{\boldsymbol{\beta}_0(\tau), \boldsymbol{\gamma}_0(\tau)\}$ by its spline approximation $\{\widetilde{\boldsymbol{\beta}}_0(\tau), \widetilde{\boldsymbol{\gamma}}_0(\tau)\}$. We write

$$\mathbf{S}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \{S_{\tau_1}(\boldsymbol{\beta}_{\tau_1}, \boldsymbol{\gamma}_{\tau_1})^\top, \ldots, S_{\tau_{k_n}}(\boldsymbol{\beta}_{\tau_{k_n}}, \boldsymbol{\gamma}_{\tau_{k_n}})^\top\}^\top$$

as the entire set of limiting functions at k_n quantile levels, and define $\widetilde{\mathbf{S}}_\tau(\boldsymbol{\beta}, \boldsymbol{\gamma})$ in the same manner. It is easy to see that

$$\|\widetilde{\mathbf{S}}(\boldsymbol{\beta}, \boldsymbol{\gamma}) - \mathbf{S}(\boldsymbol{\beta}, \boldsymbol{\gamma})\| \leq \sum_{k=1}^{k_n} E_{(y,\mathbf{z})}\Big\{ \int_{\mathbf{x}} \|\Psi_{\tau_k}(\boldsymbol{\beta}_{\tau_k}, \boldsymbol{\gamma}_{\tau_k}, y, \mathbf{x}, \mathbf{z})\| [f(x|y, \mathbf{z}, \boldsymbol{\beta}_0(\tau), \boldsymbol{\gamma}_0(\tau)) \\ - f(x|y, \mathbf{z}, \widetilde{\boldsymbol{\beta}}_0(\tau), \widetilde{\boldsymbol{\gamma}}_0(\tau))] d\mathbf{x} \Big\}.$$

Following the arguments of Lemma 1 of Wei and Carroll (2009), one can show that

$$k_n^{-1} \|\widetilde{\mathbf{S}}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0) - \mathbf{S}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)\| = o(1) \tag{A.1}$$

under Assumptions 0 and 3.

With Assumption 1, $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ is the unique solution of $\mathbf{S}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = 0$. Therefore, the convergence of (A.1) is equivalent to $k_n^{-1}\|\widetilde{\mathbf{S}}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)\| = o(1)$. Since $(\boldsymbol{\beta}_{\tau_k}^*, \boldsymbol{\gamma}_{\tau,k}^*)$ is the unique solution of $\widetilde{\mathbf{S}}_{\tau_k}(\boldsymbol{\theta}) = 0$, and Assumption 1, it follows that $k_n^{-1}\|\widetilde{\mathbf{S}}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0) - \widetilde{\mathbf{S}}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*)\| \to 0$. Due to the continuity of $\widetilde{\mathbf{S}}()$ and the uniqueness of $(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*)$, we have

$$k_n^{-1}\|\boldsymbol{\beta}^* - \boldsymbol{\beta}_0\| \to 0 \text{ and } k_n^{-1}\|\boldsymbol{\gamma}^* - \boldsymbol{\gamma}_0\| \to 0, \tag{A.2}$$

as n goes to infinity.

Now $\widetilde{\mathbf{S}}_n(\boldsymbol{\beta},\boldsymbol{\gamma})$ is the sample version of the estimating function $\widetilde{\mathbf{S}}(\boldsymbol{\beta},\boldsymbol{\gamma})$, and $\widehat{\mathbf{S}}_n(\boldsymbol{\beta},\boldsymbol{\gamma})$ is its approximation replacing the true density $f(\mathbf{x}|\mathbf{z})$ by its estimator $\widehat{f}_n(\mathbf{x}|\mathbf{z})$. We show that the difference between the two sample estimating functions above uniformly converges to zero as n goes to infinity, under the Assumptions 0 and 2.

Let

$$A_i(\boldsymbol{\beta},\boldsymbol{\gamma}) = \int_{\mathbf{x}} f(y_i|\mathbf{x},\mathbf{z}_i,\widetilde{\boldsymbol{\beta}}(\tau),\widetilde{\boldsymbol{\gamma}}(\tau))f(\mathbf{x}|\mathbf{z}_i)d\mathbf{x},$$
$$\widehat{A}_i(\boldsymbol{\beta},\boldsymbol{\gamma}) = \int_{\mathbf{x}} f(y_i|\mathbf{x},\mathbf{z}_i,\widetilde{\boldsymbol{\beta}}(\tau),\widetilde{\boldsymbol{\gamma}}(\tau))\widehat{f}_n(\mathbf{x}|\mathbf{z}_i)d\mathbf{x}.$$

By Assumption 0, $\widehat{f}_n(\mathbf{x}|\mathbf{z}_i)$ converges to $f(\mathbf{x}|\mathbf{z}_i)$ for any $(\mathbf{x},\mathbf{z}_i)$, and $\max_i \widehat{f}_n(\mathbf{x}|\mathbf{z}_i) < M\|\mathbf{x}\|^{-(p+2)}$ for some constant M, so

$$\begin{aligned}&\max_i \int_{\mathbf{x}} f(y_i|\mathbf{x},\mathbf{z}_i,\widetilde{\boldsymbol{\beta}}(\tau),\widetilde{\boldsymbol{\gamma}}(\tau))\widehat{f}_n(\mathbf{x}|\mathbf{z}_i)d\mathbf{x}\\ &\quad< \int_{\mathbf{x}} f(y_i|\mathbf{x},\mathbf{z}_i,\widetilde{\boldsymbol{\beta}}(\tau),\widetilde{\boldsymbol{\gamma}}(\tau))M\|\mathbf{x}\|^{-(p+2)}d\mathbf{x} < \infty.\end{aligned}$$

Then, it follows from the Dominated Convergence Theorem that

$$\begin{aligned}\max_i |\widehat{A}(\boldsymbol{\beta},\boldsymbol{\gamma}) - A(\boldsymbol{\beta},\boldsymbol{\gamma})| &= \max_i \int_{\mathbf{x}} f(y_i|\mathbf{x},\mathbf{z}_i,\widetilde{\boldsymbol{\beta}}(\tau),\widetilde{\boldsymbol{\gamma}}(\tau))\{\widehat{f}_n(\mathbf{x}|\mathbf{z}_i) - f(\mathbf{x}|\mathbf{z}_i)\}d\mathbf{x}\\ &= o_p(1),\end{aligned}$$

for any $(\boldsymbol{\beta},\boldsymbol{\gamma}) \in \Theta\times\Omega$. Following the same argument, we have

$$\sup_{(\boldsymbol{\beta},\boldsymbol{\gamma})\in\Theta\times\Omega} \int_{\mathbf{x}} \|\Psi_\tau(y_i,\mathbf{x},\mathbf{z}_i,\boldsymbol{\beta}_\tau,\boldsymbol{\gamma}_\tau)\||\widehat{f}_n(\mathbf{x}|\mathbf{z}_i) - f(\mathbf{x}|\mathbf{z}_i)|d\mathbf{x} = o_p(1).$$

Consequently, together with the fact that $A_i(\boldsymbol{\beta},\boldsymbol{\gamma}) > 0$ and $\widehat{A}_i(\boldsymbol{\beta},\boldsymbol{\gamma}) > 0$ for any $(\boldsymbol{\beta},\boldsymbol{\gamma})$, we have

$$\begin{aligned}&\sup_{(\boldsymbol{\beta},\boldsymbol{\gamma})\in\Theta\times\Omega} k_n^{-1}\|\widehat{\mathbf{S}}_n(\boldsymbol{\beta},\boldsymbol{\gamma}) - \widetilde{\mathbf{S}}_n(\boldsymbol{\beta},\boldsymbol{\gamma})\|\\ &\le \sup_{(\boldsymbol{\beta},\boldsymbol{\gamma})\in\Theta\times\Omega} n^{-1}k_n^{-1}\sum_{i=1}^n\sum_{k=1}^{k_n}\tfrac{1}{A_i(\boldsymbol{\beta},\boldsymbol{\gamma})}\int_{\mathbf{x}} \|\Psi_\tau(y_i,\mathbf{x},\mathbf{z}_i,\boldsymbol{\beta}_\tau,\boldsymbol{\gamma}_\tau)\||\widehat{f}_n(\mathbf{x}|\mathbf{z}_i) - f(\mathbf{x}|\mathbf{z}_i)|d\mathbf{x}\\ &\quad+ \sup_{(\boldsymbol{\beta},\boldsymbol{\gamma})\in\Theta\times\Omega} k_n^{-1}\sum_{i=1}^n\sum_{k=1}^{k_n}\tfrac{|A_i(\boldsymbol{\beta},\boldsymbol{\gamma}) - \widehat{A}_i(\boldsymbol{\beta},\boldsymbol{\gamma})|}{A_i(\boldsymbol{\beta},\boldsymbol{\gamma})\widehat{A}_i(\boldsymbol{\beta},\boldsymbol{\gamma})}\int_{\mathbf{x}} \|\Psi_\tau(y_i,\mathbf{x},\mathbf{z}_i,\boldsymbol{\beta}_\tau,\boldsymbol{\gamma}_\tau)\|f(\mathbf{x}|\mathbf{z}_i)d\mathbf{x}\\ &= o_p(1)\end{aligned} \tag{A.3}$$

Next we show that

$$\sup_{(\boldsymbol{\beta},\boldsymbol{\gamma})\in\Theta\times\Omega} k_n^{-1}\|\widetilde{\mathbf{S}}_n(\boldsymbol{\beta},\boldsymbol{\gamma}) - \widetilde{\mathbf{S}}(\boldsymbol{\beta},\boldsymbol{\gamma})\| = o_p(1) \tag{A.4}$$

To simplify the notation, let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$, $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$, $\boldsymbol{\theta}_\tau = (\boldsymbol{\beta}_\tau, \boldsymbol{\gamma}_\tau)$, $\boldsymbol{\theta}_{0,\tau} = (\boldsymbol{\beta}_{0,\tau}, \boldsymbol{\gamma}_{0,\tau})$ and $\widetilde{\boldsymbol{\theta}}(\tau) = \{\widetilde{\boldsymbol{\beta}}(\tau), \widetilde{\boldsymbol{\gamma}}(\tau)\}$.

If $\Psi(\boldsymbol{\theta}, y_i, \mathbf{x}_i, \mathbf{z}_i) = \{\Psi_{\tau_1}(\boldsymbol{\theta}_{\tau_1}, y_i, \mathbf{x}_i, \mathbf{z}_i)^\top, \ldots, \Psi_{\tau_{k_n}}(\boldsymbol{\theta}_{\tau_{k_n}}, y_i, \mathbf{x}_i, \mathbf{z}_i)^\top\}^\top$ then, by definition,

$$\begin{aligned}\widetilde{\mathbf{S}}_n(\boldsymbol{\theta}) &= n^{-1}\sum_{i=1}^n \delta_i \Psi(\boldsymbol{\theta}, y_i, \mathbf{x}_i, \mathbf{z}_i) + \frac{1}{n}\sum_{i=1}^n (1-\delta_i)\int_{\mathbf{x}} \Psi(\boldsymbol{\theta}, y_i, \mathbf{x}, \mathbf{z}_i) f\{\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\boldsymbol{\theta}}(\tau)\}d\mathbf{x}\\ &\widehat{=}\ \widetilde{S}_n^{(1)}(\boldsymbol{\theta}) + \widetilde{S}_n^{(2)}(\boldsymbol{\theta}).\end{aligned}$$

We can write the limiting function as $\widetilde{S}(\boldsymbol{\theta}) = S^{(1)}(\boldsymbol{\theta}) + \widetilde{S}^{(2)}(\boldsymbol{\theta})$. Following Lemma 8.4 in Wei and He (2006), $\sup_{\boldsymbol{\theta}} k_n^{-1}\|S_n^{(1)}(\boldsymbol{\theta}) - S^{(1)}(\boldsymbol{\theta})\| = o_p(1)$. Hence we only need to show that $\sup_{\boldsymbol{\theta}} k_n^{-1}\|\widetilde{S}_n^{(2)}(\boldsymbol{\theta}) - \widetilde{S}^{(2)}(\boldsymbol{\theta})\| = o_p(1)$, which is equivalent to

$$\mathrm{pr}\left(\sup_{\boldsymbol{\theta}\in\Theta\times\Omega} k_n^{-1}\|\widetilde{S}_n^{(2)}(\boldsymbol{\theta}) - \widetilde{S}^{(2)}(\boldsymbol{\theta})\| > \epsilon\right) \to 0 \tag{A.5}$$

for any $\epsilon > 0$. Without loss of generality, we assume the parameter space $\Theta\times\Omega = \bigcup_k\{\boldsymbol{\theta} : |\boldsymbol{\theta}_{\tau_k} - \boldsymbol{\theta}_{0,\tau_k}| < 1\}$, and partition it into L_n disjoint small cubes Γ_l with diameters less than $q_n = C_1 k_n/n$, for some constant C_1. Let ξ_l be the center of the l^{th} cube Γ_l. The probability of the left side of (A.5) is bounded by the sum, $P_1 + P_2$, of

$$\begin{aligned}P_1 &= \mathrm{pr}\Big(\max_{1\le l\le L_n}\sup_{\boldsymbol{\theta}\in\Gamma_l} k_n^{-1}\|\widetilde{S}_n^{(2)}(\boldsymbol{\theta}) - \widetilde{S}_n^{(2)}(\xi_l) - \widetilde{S}^{(2)}(\boldsymbol{\theta}) + \widetilde{S}^{(2)}(\xi_l)\| > \frac{\varepsilon}{2}\Big);\\ P_2 &= \mathrm{pr}\Big(\max_{1\le l\le L_n} k_n^{-1}\|\widetilde{S}_n^{(2)}(\xi_l) - \widetilde{S}^{(2)}(\xi_l)\| > \frac{\varepsilon}{2}\Big).\end{aligned}$$

Note that

$$\begin{aligned}&\|\widetilde{S}_n^{(2)}(\boldsymbol{\theta}) - \widetilde{S}_n^{(2)}(\xi_l)\|\\ &\le n^{-1}\sum_{i=1}^n \left\|\int_{\mathbf{x}} \{\Psi(\boldsymbol{\theta}, y_i, \mathbf{x}, \mathbf{z}_i) - \Psi(\xi_l, y_i, \mathbf{x}, \mathbf{z}_i)\} f\{\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\xi}_l(\tau)\}d\mathbf{x}\right\|\\ &\quad + n^{-1}\sum_{i=1}^n \left\|\int_{\mathbf{x}} \Psi(\boldsymbol{\theta}, y_i, \mathbf{x}, \mathbf{z}_i)\left[f\{\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\boldsymbol{\theta}}(\tau)\} - f(\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\xi}_l(\tau))\right]d\mathbf{x}\right\|\\ &= SS_1 + SS_2.\end{aligned}$$

Here

$$\begin{aligned}&\max_{1\le l\le L_n}\sup_{\boldsymbol{\theta}\in\Gamma_l} SS_1\\ &= \max_{1\le l\le L_n}\sup_{\boldsymbol{\theta}\in\Gamma_l} n^{-1}\sum_{i=1}^n \left\|\int_{\mathbf{x}} \{\Psi(\boldsymbol{\theta}, y_i, \mathbf{x}, \mathbf{z}_i) - \Psi(\xi_l, y_i, \mathbf{x}, \mathbf{z}_i)\} f\{\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\xi}_l(\tau)\}d\mathbf{x}\right\|\end{aligned}$$

$$
\begin{aligned}
&\leq \max_{1\leq l\leq L_n} \sup_{\boldsymbol{\theta}\in\Gamma_l} 2n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{k_n}\int_{\mathbf{x}} \|\Psi_{\tau_k}(\boldsymbol{\theta}_{\tau_k}, y_i, \mathbf{x}, \mathbf{z}_i) - \Psi_{\tau_k}(\xi_{l,\tau_k}, y_i, \mathbf{x}, \mathbf{z}_i)\| \\
&\quad f\{\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\xi}_l(\tau)\}d\mathbf{x} \\
&\leq \max_{1\leq l\leq L_n} \sup_{\boldsymbol{\theta}\in\Gamma_l} 2n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{k_n}\int_{\mathbf{x}} I\{|(\mathbf{x}^\top, \mathbf{z}_i^\top)\xi_{l,\tau_k} - y_i| \leq |(\mathbf{x}^\top, \mathbf{z}_i^\top)(\xi_{l,\tau_k} - \boldsymbol{\theta}_{\tau_k})|\} \\
&\quad \|(\mathbf{x}, \mathbf{z}_i)\| f\{\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\xi}_l(\tau)\}d\mathbf{x} \\
&\leq 2n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{k_n}\int_{\mathbf{x}} I\{|(\mathbf{x}^\top, \mathbf{z}_i^\top)\xi_{l,\tau_k} - y_i| \leq \|(\mathbf{x}^\top, \mathbf{z}_i^\top)\|q_n\}\|(\mathbf{x}, \mathbf{z}_i)\| f\{\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\xi}_l(\tau)\}d\mathbf{x}.
\end{aligned}
$$

Under Assumption 0, and the fact that $f(y_i|\mathbf{x}, \mathbf{z}_i, \widetilde{\xi}_l(\tau))$ is bounded,

$$
\max_i \int_{\mathbf{x}} \|(\mathbf{x}^\top, \mathbf{z}_i^\top)\| f\{\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\xi}_l(\tau)\}d\mathbf{x} < \infty
$$

for any l. Since q_n goes to zero as n goes to infinity,

$$
\int_{\mathbf{x}} I\{|(\mathbf{x}^\top, \mathbf{z}_i^\top)\xi_{l,\tau_k} - y_i| \leq \|(\mathbf{x}^\top, \mathbf{z}_i^\top)\|q_n\} f\{\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\xi}_l(\tau)\}d\mathbf{x} = o_p(1)
$$

follows the Dominate Convergence Theorem. Consequently we have $\max_{1\leq l\leq L_n} \sup_{\boldsymbol{\theta}\in\Gamma_l} k_n^{-1} SS_1 = o_p(1)$. On the other hand,

$$
\begin{aligned}
&\max_{1\leq l\leq L_n} \sup_{\boldsymbol{\theta}\in\Gamma_l} SS_2 \\
&= \max_{1\leq l\leq L_n} \sup_{\boldsymbol{\theta}\in\Gamma_l} n^{-1}\sum_{i=1}^{n}\left\|\int_{\mathbf{x}} \Psi(\boldsymbol{\theta}, y_i, \mathbf{x}, \mathbf{z}_i)\left[f\{\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\boldsymbol{\theta}}(\tau)\} - f(\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\xi}_l(\tau))\right]d\mathbf{x}\right\| \\
&\leq \sup_{\boldsymbol{\theta}\in\Gamma_l} 2n^{-1}\sum_{i=1}^{n} k_n \int_{\mathbf{x}} \|(\mathbf{x}^\top, \mathbf{z}_i^\top)\| \left[f\{\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\boldsymbol{\theta}}(\tau)\} - f(\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\xi}_l(\tau))\right]d\mathbf{x}.
\end{aligned}
$$

Since

$$
\sup_{\boldsymbol{\theta}\in\Gamma_l} \max_i |f\{\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\boldsymbol{\theta}}(\tau)\} - f(\mathbf{x}|y_i, \mathbf{z}_i, \widetilde{\xi}_l(\tau))| = o_p(1),
$$

and $\max_i \int \|(\mathbf{x}^\top, \mathbf{z}_i^\top)\|\widehat{f}_n(\mathbf{x}|\mathbf{z}_i)d\mathbf{x} < \infty$ under Assumption 0, we have

$$
\max_{1\leq l\leq L_n} \sup_{\boldsymbol{\theta}\in\Gamma_l} k_n^{-1} SS_2 = o_p(1).
$$

Combining the convergence of SS_1 and SS_2, we have $\sup_{\boldsymbol{\theta}\in\Gamma_l} k_n^{-1}\|S_n^{(2)}(\boldsymbol{\theta}) - S_n^{(2)}(\xi_l)\| = o_p(1)$. Following a similar argument, we can also show that $\sup_{\boldsymbol{\theta}\in\Gamma_l} \|S(\boldsymbol{\theta}) - S(\xi_l)\| = o(1)$. It then follows that $P_1 = o(1)$.

Under the assumption that $E(\max_i \|\mathbf{z}_i\|) < \infty$, we have $\text{prob}(\max_i \|\mathbf{z}_i\| > Ln^{1/2}k_n^{-1/2}) \to 0$ for some constant L. Consequently, $P_2 = o(1)$ is equivalent to

$$\text{pr}\Big(\max_{1\le l\le L_n} k_n^{-1}\|\widetilde{S}_n^{(2)}(\xi_l) - \widetilde{S}^{(2)}(\xi_l)\| > \frac{\varepsilon}{2}, \text{ and } \max_i z_i \le n^{1/2}k_n^{-1/2}\Big) = o(1).$$

Let

$$\mathcal{X}_i(l,k,m) = 1 - \delta_i \int_{\mathbf{x}} (\tau_k - I\{y_i - (\mathbf{x}^\top, \mathbf{z}_i^\top)\xi_{l,k}\})x_m f(\mathbf{x}|y_i, \mathbf{z}_i; \widetilde{\xi}_l(\tau))d\,\mathbf{x}\},$$

for $m = 1, \dots, p$, and

$$\mathcal{Z}_i(l,k,m) = 1 - \delta_i \int_{\mathbf{x}} (\tau_k - I\{y_i - (\mathbf{x}^\top, \mathbf{z}_i^\top)\xi_l\})z_{i,m} f(\mathbf{x}|y_i, \mathbf{z}_i; \widetilde{\xi}_l(\tau))$$
$$d\,\mathbf{x} I\{z_{i,m} \le n^{1/2}k_n^{-1/2}\}, \quad m = 1, \dots, q.$$

A sufficient condition for $P_2 = o(1)$ is that

$$\text{pr}\Big\{\max_{1\le l\le L_n;1\le k\le k_n;1\le m\le p} n^{-1}\Big|\sum_{i=1}^n \mathcal{X}_i(l,k,m) - E\{\mathcal{X}_i(l,k,m)\}\Big| > \frac{\varepsilon}{2}\Big\} = o(1),$$

and $\text{pr}\{\max_{1\le l\le L_n;1\le k\le k_n;1\le m\le q} n^{-1}|\sum_{i=1}^n \mathcal{Z}_i(l,k,m) - E\{\mathcal{Z}_i(l,k,m)\}| > \varepsilon/2\} = o(1)$. Under Assumption 0, and the fact that Γ is a bounded support, $|\mathcal{X}_i(l,k,m)| < Cn^{1/2}k_n^{-1/2}$ and $|\mathcal{Z}_i(l,k,m)| < Cn^{1/2}k_n^{-1/2}$ for all i. Applying Bernstein's inequality to this probability term, we have

$$\begin{aligned}
&\text{pr}\Big(\max_{1\le l\le L_n;1\le k\le k_n;1\le m\le p} n^{-1}\Big|\sum_{i=1}^n \mathcal{X}_i(l,k,m) - E\mathcal{X}_i(l,k,m)\Big| > \varepsilon\Big)\\
&\quad \le k_n \cdot p \cdot L_n \cdot \text{pr}\Big(n^{-1}\Big|\sum_{i=1}^n \mathcal{X}_i(l,k,m) - E\mathcal{X}_i(l,k,m)\Big| > \varepsilon\Big)\\
&\quad \le k_n \cdot \cdot L_n \cdot \exp\Big\{-\frac{n^2\varepsilon^2}{2nC^2nk_n^{-1} + (2/3)Cn^{1/2}k_n^{-1/2}n\varepsilon}\Big\} = o(1).
\end{aligned}$$

The same convergence holds for $\mathcal{Z}_i(l,k,m)$'s. The uniform convergence (A.5) now follows, which in turn implies (A.4). Combining (A.3) and (A.4), we have

$$\sup_{\boldsymbol{\theta}\in\Theta\times\Omega} k_n^{-1}\|\widehat{\mathbf{S}}_n(\boldsymbol{\theta}) - \widetilde{\mathbf{S}}(\boldsymbol{\theta})\| = o_p(1). \tag{A.6}$$

We turn to the uniform convergence of $\widehat{\boldsymbol{\theta}}_n$.

For any $\delta > 0$, we define a compact set $\mathrm{B}_\tau = \{\boldsymbol{\theta}_\tau \in \mathbb{R}^{p+1} : \|\boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{0,\tau}\| < \delta\}$, where $\boldsymbol{\theta}_{0,\tau}$ is the true coefficients at the quantile level τ, and let B_τ^c be its complementary set. We define the distance

$$d_n(\delta) = k_n^{-1}\Big\{\min_{\boldsymbol{\theta}\in\{\Theta\cap B_\tau^c\}\otimes\Omega} \|\widetilde{\mathbf{S}}(\boldsymbol{\theta})\| - \|\widetilde{\mathbf{S}}(\boldsymbol{\theta}_0)\|\Big\} \tag{A.7}$$

between the norm of the limiting estimating equations $\widetilde{\mathbf{S}}$ evaluated at the true coefficients $\boldsymbol{\theta}_0$ and the minimized norm when $\boldsymbol{\theta}$ stays outside of $\mathrm{B}_\tau \otimes \Omega$.

Due to the convergence (A.1), there exist K_δ, such that when $k_n > K_\delta$, we have $k_n^{-1}\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_0\| < \delta/2$, so $\boldsymbol{\theta}^* \in B_\tau \times \Omega$ for $k_n > K_\delta$. On the other hand, the uniqueness of $\boldsymbol{\theta}^*$ ensures that, for any $k_n > K_\delta$, we have

$$d_n^*(\delta) = k_n^{-1}\left\{\min_{\boldsymbol{\theta}\in\{\Theta\cap B_\tau^c\}\otimes\Omega} \|\widetilde{\mathbf{S}}(\boldsymbol{\theta})\| - \|\widetilde{\mathbf{S}}(\boldsymbol{\theta}^*)\|\right\} > 0. \tag{A.8}$$

Following the continuity of $\widetilde{\mathbf{S}}()$, and convergence of $\boldsymbol{\theta}^*$, we also have

$$k_n^{-1}\|\widetilde{\mathbf{S}}(\boldsymbol{\theta}^*) - \widetilde{\mathbf{S}}(\boldsymbol{\theta}_0)\| < \frac{d_n^*(\delta)}{2}, \tag{A.9}$$

for sufficiently larger k_n. Combining these (A.9) and (A.8), we have

$$d_n(\delta) = k_n^{-1}[\min_{\boldsymbol{\theta}\in\Theta\cap B_\tau^c} \|\widetilde{\mathbf{S}}(\boldsymbol{\theta})\| - \|\widetilde{\mathbf{S}}(\boldsymbol{\theta}_0)\|] > \frac{d_n^*(\delta)}{2} > 0, \tag{A.10}$$

for sufficiently large k_n.

The random event

$$E_n = \left\{k_n^{-1} \max_{\boldsymbol{\theta}\in\Theta\times\Omega}[\|\widehat{\mathbf{S}}_n(\boldsymbol{\theta}) - \widetilde{\mathbf{S}}(\boldsymbol{\theta})\|] < \frac{d_n(\delta)}{3}\right\},$$

together with Assumption 2, imply that

$$k_n^{-1}\|\widetilde{\mathbf{S}}(\widehat{\boldsymbol{\theta}}_n)\| \le k_n^{-1}\|\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)\| + \frac{d_n(\delta)}{3}, \tag{A.11}$$

$$k_n^{-1}\|\widehat{\mathbf{S}}_n(\boldsymbol{\theta}_0)\| \le k_n^{-1}\|\widetilde{\mathbf{S}}(\boldsymbol{\theta}_0)\| + \frac{d_n(\delta)}{3}. \tag{A.12}$$

Since $\widehat{\boldsymbol{\theta}}_n$ is the minimizer of $\|\widehat{\mathbf{S}}_n(\boldsymbol{\theta})\|$, we have $\|\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\theta}}_n)\| < \|\widehat{\mathbf{S}}_n(\boldsymbol{\theta}_0)\|$, which, together with (A.11), shows that $k_n^{-1}\|\widetilde{\mathbf{S}}(\widehat{\boldsymbol{\theta}}_n)\| \le k_n^{-1}\|\widehat{S}_n(\boldsymbol{\theta}_0)\| + d_n(\delta)/3 \le k_n^{-1}\|\widetilde{\mathbf{S}}(\boldsymbol{\theta}_0)\|$ $+2d_n(\delta)/3$.

Following (A.6), $\lim_{n\to\infty} \mathrm{pr}(E_n) = 1$, which implies

$$\lim_{n\to\infty} \mathrm{pr}\left\{k_n^{-1}\|\widetilde{\mathbf{S}}(\widehat{\boldsymbol{\theta}}_n)\| \le k_n^{-1}\|\widetilde{\mathbf{S}}(\boldsymbol{\theta}_0)\| + \frac{2d_n(\delta)}{3}\right\} \ge \lim_{n\to\infty} \mathrm{pr}(E_n) = 1.$$

By the definition of B_τ and the fact that $d_n(\delta) > d_n^*(\delta)/2 > 0$, this in turn implies that $\lim_{n\to\infty} \mathrm{pr}(\widehat{\boldsymbol{\theta}}_n \in B_\tau) = 1$, or

$$\sup_{\tau\in[1/(k_n+1),k_n/(k_n+1)]} \|\widehat{\boldsymbol{\theta}}_n(\tau) - \boldsymbol{\theta}_0(\tau)\| = o_p(1).$$

The consistency of $\widehat{\boldsymbol{\theta}}_n(\tau)$ is proved.

References

Afifi, A. A. and Elashoff, R. M. (1969a). Missing observations in multivariate statistics III. Large sample analysis of simple linear regression. *J. Amer. Statist. Assoc.* **64**, 337-358.

Afifi, A. A. and Elashoff, R. M. (1969b). Missing observations in multivariate statistics IV. A note on simple linear regression. *J. Amer. Statist. Assoc.* **64**, 359-365.

Brick, J. M. and Kalton, G. (1996). Handling missing data in survey research. *Statist. Meth. Medical Res.* **4**, 215-238.

Dubnicka, S. R. (2009). Kernel density estimation with missing data and auxiliary variables. *Austral. N. Z. J. Statist.* **51**, 247-270.

Huang, B. and Salleb-Aouissi, A. (2009). Maximum entropy density estimation with incomplete presence-only data. *JMLR Workshop and Conference Proceedings* **5,** 240-247.

Iacus, S. M. and Torre, D. L. (2002). Nonparametric estimation of distribution and density functions in presence of missing data: an Ifs approach. Working Paper.

Koenker, R. and Bassett, G. J. (1978). Regression quantiles. *Econometrica* **46,** 33-50.

Koenker, R. and Xiao, Z. J. (2004). Unit root quantile autoregression inference. *J. Amer. Statist. Assoc.* **99,** 775-787.

Lipstiz, 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. (1992). Regression with missing X's: a review. *J. Amer. Statist. Assoc.* **87**, 1227-1237.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* Wiley, New York.

Little, R. J. A and Zhang, N (2011). Subsample ignorable likelihood for regression analysis with missing data. *Appl. Statist.*, **60**, 591-605.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coeffcients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **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. Amer. Statist. Assoc.* **90,** 106-121.

Rubin, D. B. (1978). Multiple imputation in sample surveys - A phenomeno-logical Bayesian approach to nonresponse. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 22-34.

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

Rubin, D. B. and Schaferm, J. L. (1990). Efficiently creating multiple imputations for incomplete multivariate normal data. *Proceedings of the Statistical Computing Section of the American Statistical Association*, 83-88.

Terry, M. B., Wei, Y. and Essenman, D. (2007). Maternal, birth, and early life influences on adult body size in women (with comments). *Amer. J. Epidemiology* **166,** 5-13.

Titterington, D. M. and Mill, G. M. (1983). Kernel-based density estimates from incomplete data. *J. Roy. Statist. Soc. Ser. B* **45,** 258-266.

Wang, Q. (2008). Probability density estimation with data missing at random when covariables are present. *J. Statist. Plann. Inference* **138,** 568-587.

Wei, Y. and Carroll, R. J. (2009). Quantile regression with measurement error. *J. Amer. Statist. Assoc.* **104,** 1129-1143.

Wei, Y. and He, X. (2006). Conditional growth charts The *Ann. Statist.* **34**, 2069-2591.

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

Welsh, A. H. (1988). Asymptotically efficient estimation of the sparsity function at a point. *Statist. Probab. Lett.* **6**, 427-432.

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

Department of Biostatistics, Columbia University, New York, NY 10032, USA.

E-mail: ying.wei@columbia.edu

Department of Epidemiology and Biostatistics, Drexel University, 3141 Chestnut St, Philadelphia, PA 19104, USA.

E-mail: Yunwen.Yang@drexel.edu

(Received December 2011; accepted August 2013)