Biometrics 59, 1–19 DOI: 10.1111/j.1541-0420.2005.00454.x
December 2006

Quantile Regression in the Presence of Monotone Missingness with Sensitivity Analysis

Minzhao Liu[1,*], Michael J. Daniels[2,], and Michael G. Perri[3,***]**

[1]Department of Statistics, University of Florida, FL 32601, USA.

[2]Department of Integrative Biology, Division of Statistics and Scientific Computation

The University of Texas at Austin, 141MC Patterson Hall, Austin, TX 78712, USA.

[3]College of Public Health and Health Professions, University of Florida, Gainesville, FL 32611, USA.

email: liuminzhao@ufl.edu

**email:* mjdaniels@austin.utexas.edu

***email:* mperri@phhp.ufl.edu

Summary: In this paper, we develop methods for longitudinal quantile regression when there is monotone missingness. In particular, we propose pattern mixture models with a constraint that provides an interpretation form for the marginal quantile regression parameters. Our approach allows sensitivity analysis which is an essential component in inference for incomplete data. To facilitate computation of the likelihood, we propose a novel way to obtain analytic forms for required integrals. The model is applied to data from a clinical trial on weight management.

Key words: Monotone missingness; Non-ignorable missingness; Quantile regression; Sensitivity analysis.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Quantile regression is used to study the relationship between a response and covariates when one (or several) quantiles are of interest as opposed to mean regression. The dependence between upper or lower quantiles of the response variable and the covariates often vary differentially relative to that of the mean. How quantiles depend on covariates is of interest in econometrics, educational studies, biomedical studies, and environment studies (Yu and Moyeed, 2001; Buchinsky, 1994, 1998; He et al., 1998; Koenker and Machado, 1999; Wei et al., 2006; Yu et al., 2003). A comprehensive review of applications of quantile regression was presented in Koenker (2005).

Quantile regression is more robust to outliers than mean regression and provides information about how covariates affect quantiles, which offers a more complete description of the conditional distribution of the response. Different effects of covariates can be assumed for different quantiles.

The traditional frequentist approach was proposed by Koenker and Bassett (1978) for a single quantile with estimators derived by minimizing a loss function. The popularity of this approach is due to its computational efficiency, well-developed asymptotic properties, and straightforward extensions to simultaneous quantile regression and random effect models. However, asymptotic inference may not be accurate for small sample sizes and the approach does not naturally extend to missing data.

Bayesian approaches offer exact inference in small samples. Motivated by the loss (check) function, Yu and Moyeed (2001) proposed an asymmetric Laplace distribution for the error term, such that maximizing the posterior distribution is equivalent to minimizing the check function. Also semiparametric methods have been proposed for median regression. Walker and Mallick (1999) used a diffuse finite Pólya Tree prior for the error term. Kottas and Gelfand (2001) modeled the error by two families of median zero distribution using a mixture

Dirichlet process priors, which is very useful for unimodal error distributions. Hanson and Johnson (2002) adopted mixture of Pólya Tree prior in median regression, which is more robust in terms of multimodality and skewness. Other recent approaches include quantile pyramid priors, mixture of Dirichlet process priors of multivariate normal distributions and infinite mixture of Gaussian densities which place quantile constraints on the residuals (Hjort and Petrone, 2007; Hjort and Walker, 2009; Kottas and Krnjajić, 2009; Reich et al., 2010).

The above methods focus on complete data. There are only a few articles about quantile regression with missingness. Wei et al. (2012) proposed a multiple imputation method for quantile regression model when there are some covariates missing at random (MAR). They impute the missing covariates by specifying its conditional density given observed covariates and outcomes, which come from the estimated conditional quantile regression and specification of conditional density of missing covariates given observed ones. However, they focus more on the missing covariates than missing outcomes. Bottai and Zhen (2013) illustrated an imputation method using estimated conditional quantiles of missing outcomes given observed data. Their approach does not make distributional assumptions. They assumed the missing data mechanism (MDM) is MAR. However, because their imputation method is not derived from a joint distribution, the joint distribution under such conditionals may not exist. In addition, their approach does not allow for missing not at random (MNAR).

Yuan and Yin (2010) introduced a fully parametric Bayesian quantile regression approach for longitudinal data with non-ignorable missing data. They used shared latent subject-specific random effects to explain the within-subject correlation and to associate the response process with missing data process, and applied multivariate normal priors on the random terms to match the traditional quantile regression check function with penalties. However, the quantile regression coefficients are conditional on the random effects, which is not of interest if we are interested in interpreting regression coefficients unconditionally. In addition, they

are conditional on random effects, which tie together the responses and missingness process, so they have a slightly different interpretation than regular random effects in longitudinal methods. Moreover, due to their full parametric specification for the full data, their model does not allow for sensitivity analysis, which is a key component in inference for incomplete data (NAS 2010).

Pattern mixture models were originally proposed to model missing data in Rubin (1977). Later mixture models were extended to handle MNAR in longitudinal data. For discrete dropout times, Little (1993, 1994) proposed a general method by introducing a finite mixture of multivariate distribution for longitudinal data. When there are many possible dropout time, Roy (2003) proposed to group them by latent classes.

Roy and Daniels (2008) extended Roy (2003) to generalized linear models and proposed a pattern mixture model for data with non-ignorable dropout, borrowing ideas from Heagerty (1999). But their approach only estimates the marginal covariate effects on the mean. We will use related ideas for quantile regression models which allow for non-ignorable missingness and sensitivity analysis.

The structure of this article is as follows. First, we introduce a quantile regression method to address monotone non-ignorable missingness in section 2, including sensitivity analysis and computational details. We use simulation studies to evaluate the performance of the model in section 3. We apply our approach to data from a recent clinical trial in section 4. Finally, discussion and conclusions are given in section 5.

2. Model

In this section, we first introduce some notation, then describe our proposed quantile regression model in section 2.1. We provide details on MAR and MNAR and computation in sections 2.2 and 2.3 respectively.

Under monotone dropout, without loss of generality, denote $S_i \in \{1, 2, \ldots, J\}$ to be the

number of observed $Y_{ij}'s$ for subject i, and $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iJ})^T$ to be the full data response vector for subject i, where J is the maximum follow up time. We assume Y_{i1} is always observed. We are interested in the τ-th marginal quantile regression coefficients $\boldsymbol{\gamma}_j = (\gamma_{j1}, \gamma_{j2}, \ldots, \gamma_{jp})^T$,

$$\Pr(Y_{ij} \leqslant \boldsymbol{x}_i^T \boldsymbol{\gamma}_j) = \tau, \text{ for } j = 1, \ldots, J, \tag{1}$$

where $\boldsymbol{x}_i$ is a $p \times 1$ vector of covariates for subject i.

Let

$$\mathrm{p}_k(Y) = \mathrm{p}(Y|S = k), \quad \mathrm{p}_{\geqslant k}(Y) = \mathrm{p}(Y|S \geqslant k)$$

be the densities of response $\boldsymbol{Y}$ given follow-up time $S = k$ and $S \geqslant k$. And $\Pr_k$ be the corresponding probability given $S = k$.

2.1 *Mixture Model Specification*

We adopt a pattern mixture model to jointly model the response and missingness (Little, 1994; Daniels and Hogan, 2008). Mixture models factor the joint distribution of response and missingness as

$$\mathrm{p}(\boldsymbol{y}, \boldsymbol{S}, |\boldsymbol{x}, \boldsymbol{\omega}) = \mathrm{p}(\boldsymbol{y}|\boldsymbol{S}, \boldsymbol{x}, \boldsymbol{\omega})\mathrm{p}(\boldsymbol{S}|\boldsymbol{x}, \boldsymbol{\omega}).$$

Thus the full-data response follows the distribution given by

$$\mathrm{p}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\omega}) = \sum_{S \in \mathcal{S}} \mathrm{p}(\boldsymbol{y}|\boldsymbol{S}, \boldsymbol{x}, \boldsymbol{\theta})\mathrm{p}(\boldsymbol{S}|\boldsymbol{x}, \boldsymbol{\phi}),$$

where $\mathcal{S}$ is the sample space for dropout time S (i.e., the pattern) and the parameter vector $\boldsymbol{\omega}$ is partitioned as $(\boldsymbol{\theta}, \boldsymbol{\phi})$.

Furthermore, the conditional distribution of response within patterns can be decomposed as

$$\mathrm{p}(\boldsymbol{y}_{obs}, \boldsymbol{y}_{mis}|\boldsymbol{S}, \boldsymbol{\theta}) = \mathrm{p}(\boldsymbol{y}_{mis}|\boldsymbol{y}_{obs}, \boldsymbol{S}, \boldsymbol{\theta}_E)\mathrm{p}(\boldsymbol{y}_{obs}|\boldsymbol{S}, \boldsymbol{\theta}_{y,O}), \tag{2}$$

where $\boldsymbol{\theta}_E$ indexes the parameters in the extrapolation distribution, the first term on the

right hand side and $\boldsymbol{\theta}_{y,O}$ indexes parameters in the distribution of observed responses, the second term on the right hand side.

We assume models within pattern to be multivariate normal distributions and specify a sequential model parametrization for each multivariate normal distribution; we specify the model in this way as in (3) to have multivariate normal distributions on the full data within patterns and that MAR exists (Wang and Daniels, 2011) (More details are presented in section 2.2). These distributions include a constraint such that the marginal quantile regression models from equation 1 hold. In particular, we specify the distributions conditional on $S = k$ as:

$$\begin{aligned}
&\mathrm{p}_k(y_{i1}) = N(\Delta_{i1} + \boldsymbol{x}_{i1}^T \boldsymbol{\beta}_1^{(k)}, \sigma_1^{(k)}), k = 1, \ldots, J, \\
&\mathrm{p}_k(y_{ij}|\boldsymbol{y}_{ij^-}) = \begin{cases} \mathrm{N}\big(\Delta_{ij} + \boldsymbol{y}_{ij^-}^T \boldsymbol{\beta}_{y,j-1}^{(\geqslant j)}, \sigma_j^{(\geqslant j)}\big), & k \geqslant j; \\ \mathrm{N}\big(\chi(\boldsymbol{x}_{ij}, \boldsymbol{y}_{ij^-}), \sigma_j^{(k)}\big), & k < j; \end{cases}, \text{ for } 2 \leqslant j \leqslant J, \\
&S_{ij} = k \sim \text{Multinomial}(1, \boldsymbol{\phi}),
\end{aligned} \tag{3}$$

where, $\boldsymbol{y}_{ij^-} = (y_{i1}, \ldots, y_{i(j-1)})^T$ is the response history for subject i up to time point $(j-1)$, $\boldsymbol{x}_{ij}$ is a $p \times 1$ covariate vector, $\boldsymbol{\beta}_{y,j-1}^{(k)} = \big(\beta_{y_1,j-1}^{(k)}, \ldots, \beta_{y_{j-1},j-1}^{(k)}\big)^T$ are autoregressive coefficients, $\sigma_j^{(k)}$ is the conditional standard deviation of response component j, and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_J)$ is the multinomial probability vector for the number of observed responses. In addition, $\chi(\boldsymbol{x}_{ij}, \boldsymbol{y}_{ij^-})$ is a function of $(\boldsymbol{x}_{ij}, \boldsymbol{y}_{ij^-})$, which stands for the mean of the unobserved data distribution and allows sensitivity analysis by putting different assumptions on χ. Particularly, in this section we specify

$$\chi(\boldsymbol{x}_{ij}, \boldsymbol{y}_{ij^-}) = \Delta_{ij} + h_0^{(k)} + \boldsymbol{y}_{ij^-}^T \boldsymbol{\beta}_{y,j-1}^{(k)}, \tag{4}$$

where $h_0^{(k)}$ is a sensitivity parameter which indicates the intercept shift. More details about sensitivity parameters are shown in section 2.2.

For identifiability of the distribution of the observed data, we use the following restrictions

(without loss of generality),

$$\sum_{k=1}^{J} \beta_{l1}^{(k)} = 0, l = 1, \ldots, p.$$

In (3) and (4), the conditional distributions $\mathrm{p}_k(y_{ij}|\boldsymbol{y}_{ij^-})$ (and all there respective parameters) are identified for $k \geqslant j$, but not for $k < j$. Δ_{ij} are subject/time specific intercepts determined by the parameters in (3), (4) and (1); more details are given below. The parameters $h_0^{(k)}$s are sensitivity parameters (not identified by the observed data) which we will discuss in section 2.2.

In (3) and (4), Δ_{ij} are functions of $\tau, \boldsymbol{x}_{ij}, \boldsymbol{\beta}, \boldsymbol{h}, \boldsymbol{\sigma}, \boldsymbol{\gamma}_j, \boldsymbol{\phi}$ and are determined by the marginal quantile regressions,

$$\tau = \Pr(Y_{ij} \leqslant \boldsymbol{x}_{ij}^T \boldsymbol{\gamma}_j) = \sum_{k=1}^{J} \phi_k \Pr_k(Y_{ij} \leqslant \boldsymbol{x}_{ij}^T \boldsymbol{\gamma}_j) \text{ for } j = 1, \tag{5}$$

and

$$\begin{aligned} \tau = \Pr(Y_{ij} \leqslant \boldsymbol{x}_{ij}^T \boldsymbol{\gamma}_j) &= \sum_{k=1}^{J} \phi_k \Pr_k(Y_{ij} \leqslant \boldsymbol{x}_{ij}^T \boldsymbol{\gamma}_j) \\ &= \sum_{k=1}^{J} \phi_k \int \cdots \int \Pr_k(Y_{ij} \leqslant \boldsymbol{x}_{ij}^T \boldsymbol{\gamma}_j | \boldsymbol{y}_{ij^-}) \mathrm{p}_k(y_{i(j-1)} | \boldsymbol{y}_{i(j-1)^-}) \\ &\quad \cdots \mathrm{p}_k(y_{i2}|y_{i1}) \mathrm{p}_k(y_{i1}) dy_{i(j-1)} \cdots dy_{i1}. \text{ for } j = 2, \ldots, J. \end{aligned} \tag{6}$$

Details on computing the Δ_{ij} will be given in section 2.3.

The idea in the above specification is to model the marginal quantile regressions directly and then to embed them in the likelihood through restrictions in the mixture model. The mixture model in (3) allows the marginal quantile regression coefficients to differ by quantiles; otherwise, the quantile lines would be parallel to each other. Moreover, the mixture model also allows sensitivity analysis for the missing data. This is an essential component of the analysis of missing data (as discussed in the introduction) and is not permitted in most previous approaches.

2.2 *Missing Data Mechanism and Sensitivity Analysis*

Mixture models specified as in Section 2.1 are not identified by the observed data. Specific forms of missingness induce constraints to identify the distributions for incomplete patterns, in particular, the extrapolation distribution in (2). In this section, we explore ways to embed the missingness mechanism and sensitivity parameters in mixture models for our setting.

In the mixture model in (3), MAR holds (Molenberghs et al., 1998; Wang and Daniels, 2011) if and only if, for each $j \geqslant 2$ and $k < j$:

$$\mathrm{p}_k(y_j|y_1,\ldots,y_{j-1}) = \mathrm{p}_{\geqslant j}(y_j|y_1,\ldots,y_{j-1}).$$

When $2 \leqslant j \leqslant J$ and $k < j$, Y_j is not observed, thus $h_0^{(k)}$ and $\sigma_j^{(k)}$, $\boldsymbol{\beta}_{y,j-1}^{(k)} = \left(\beta_{y_1,j}^{(k)},\ldots,\beta_{y_{j-1},j-1}^{(k)}\right)^T$ can not be identified from the observed data. Denote

$$\log\sigma_j^{(k)} = \log\sigma_j^{(\geqslant j)} + \delta_j^{(k)},$$
$$\boldsymbol{\beta}_{y,j-1}^{(k)} = \boldsymbol{\beta}_{y,j-1}^{(\geqslant j)} + \boldsymbol{\eta}_{j-1}^{(k)},$$

where $\boldsymbol{\eta}_{j-1}^{(k)} = \left(\eta_{y_1,j-1}^{(k)},\ldots,\eta_{y_{j-1},j-1}^{(k)}\right)$ for $k < j$. Then $\boldsymbol{\xi}_s = (h_0^{(k)}, \boldsymbol{\eta}_{j-1}^{(k)}, \delta_j^{(k)})$ is a set of sensitivity parameters (Daniels and Hogan, 2008), where $k < j, 2 \leqslant j \leqslant J$.

When $\boldsymbol{\xi}_s = \boldsymbol{\xi}_{s0} = \mathbf{0}$, MAR holds. If $\boldsymbol{\xi}_s$ is fixed at $\boldsymbol{\xi}_s \neq \boldsymbol{\xi}_{s0}$, the missingness mechanism is MNAR. We can vary $\boldsymbol{\xi}_s$ around $\mathbf{0}$ to examine the impact of different MNAR mechanisms.

In general, each pattern $S = k$ has its own set of sensitivity parameters $\boldsymbol{\xi}_s^{(k)}$. However, to keep the number of sensitivity parameters at a manageable level (Daniels and Hogan, 2008) and without loss of generality in what follows, we assume $\boldsymbol{\xi}_s$ does not depend on pattern.

2.3 *Computation*

In section 2.3.1, we provide details on calculating Δ_{ij} in (3) for $j = 1,\ldots,J$. Then we show how to obtain maximum likelihood estimates in section 2.3.2.

2.3.1 *Calculation of* Δ . From equation (5) and (6), Δ_{ij} depends on subject-specific covariates $\boldsymbol{x}_{ij}$, thus Δ_{ij} needs to be calculated for each subject. We now illustrate how to calculate Δ_{ij} given all the other parameters $\boldsymbol{\xi} = (\boldsymbol{\xi}_m, \boldsymbol{\xi}_s)$.

- Δ_{i1} : Expand equation (5):

$$\tau = \sum_{k=1}^{J} \phi_k \Phi \left(\frac{\boldsymbol{x}_{i1}^T \boldsymbol{\gamma}_1 - \Delta_{i1} - \boldsymbol{x}_{i1}^T \boldsymbol{\beta}_1^{(k)}}{\sigma_1^{(k)}} \right),$$

  where Φ is the standard normal CDF. Because the above equation is continuous and monotone in Δ_{i1}, it can be solved by a standard numerical root-finding method (e.g. bisection method) with minimal difficulty.

- $\Delta_{ij}, 2 \leqslant j \leqslant J$:

  First we introduce a lemma:

LEMMA 2.1: *An integral of a normal CDF with mean b and standard deviation a over another normal distribution with mean μ and standard deviation σ can be simplified to a closed form in terms of normal CDF:*

$$\int \Phi \left(\frac{x-b}{a} \right) d\Phi(x; \mu, \sigma) = \begin{cases} 1 - \Phi \left(\frac{b-\mu}{\sigma} \Big/ \sqrt{\frac{a^2}{\sigma^2} + 1} \right) & a > 0, \\ \Phi \left(\frac{b-\mu}{\sigma} \Big/ \sqrt{\frac{a^2}{\sigma^2} + 1} \right) & a < 0, \end{cases}$$

where $\Phi(x; \mu, \sigma)$ stands for a CDF of normal distribution with mean μ and standard deviation σ.

Given the result in Lemma 2.1, to solve equation (6), we propose a recursive approach. For

the first multiple integral in equation (6), apply lemma 2.1 once to obtain:

$$\begin{aligned}
\Pr_1(Y_{ij} \leqslant \boldsymbol{x}_{ij}^T\boldsymbol{\gamma}_j) &= \int \dots \int \Pr(Y_{ij} \leqslant \boldsymbol{x}_{ij}^T\boldsymbol{\gamma}_j | S=1, \boldsymbol{x}_{ij}, \boldsymbol{Y}_{ij^-}) \\
&\quad dF(Y_{i(j-1)}|S=1, \boldsymbol{x}_{ij}, \boldsymbol{Y}_{i(j-1)^-}) \cdots dF(Y_{i1}|S=1, \boldsymbol{x}_{ij}), \\
&= \int \dots \int \Phi\left(\frac{\boldsymbol{x}_{ij}^T\boldsymbol{\gamma}_j - \Delta_{ij} - h_0^{(1)} - \boldsymbol{y}_{ij^-}^T\boldsymbol{\beta}_{y,j-1}^{(1)}}{\sigma_j^{(1)}}\right) \\
&\quad dF(Y_{i(j-1)}|S=1, \boldsymbol{x}_{ij}, \boldsymbol{Y}_{i(j-1)^-}) \cdots dF(Y_{i1}|S=1, \boldsymbol{x}_{ij}), \\
&= \int \dots \int \Phi\left(\frac{Y_{i(j-2)} - b^*}{a^*}\right) dF(Y_{i(j-2)}|S=1, \boldsymbol{x}_{ij}, \boldsymbol{Y}_{i(j-2)^-}) \\
&\quad \cdots dF(Y_{i1}|S=1, \boldsymbol{x}_{ij}).
\end{aligned}$$

Then, by recursively applying lemma 2.1 $(j-1)$ times, each multiple integral in equation (6) can be simplified to single normal CDF. Thus we can easily solve for Δ_{ij} using standard numerical root-finding method as for $j = 1$.

2.3.2 *Maximum Likelihood Estimation.* The observed data likelihood for an individual i with follow-up time $S_i = k$ is

$$\begin{aligned}
L_i(\boldsymbol{\xi}|\boldsymbol{y}_i, S_i = k) &= \phi_k \mathrm{p}_k(y_{ik}|y_{i1}, \dots, y_{i(k-1)}) \mathrm{p}_k(y_{i(k-1)}|y_{i1}, \dots, y_{i(k-2)}) \cdots \mathrm{p}_k(y_{i1}) \qquad (7) \\
&= \phi_k \mathrm{p}_{\geqslant k}(y_{ik}|y_{i1}, \dots, y_{i(k-1)}) \mathrm{p}_{\geqslant k-1}(y_{i(k-1)}|y_{i1}, \dots, y_{i(k-2)}) \cdots \mathrm{p}_k(y_{i1}),
\end{aligned}$$

where $\boldsymbol{y}_i = (y_{i1}, \dots, y_{ik})$.

We use derivative-free optimization algorithms by quadratic approximation to compute the maximum likelihood estimates (Bates et al., 2012). Denote $J(\boldsymbol{\xi}) = -\log L = -\log \sum_{i=1}^n L_i$. Then maximizing the likelihood is equivalent to minimizing the target function $J(\boldsymbol{\xi})$. Under an MAR assumption, we fix $\boldsymbol{\xi}_s = \mathbf{0}$, while under MNAR assumption, $\boldsymbol{\xi}_s$ can be chosen as desired.

During each step of the algorithm, Δ_{ij} has to be calculated for each subject and at each time, as well as partial derivatives for each parameter.

As an example of the speed of the algorithm, for 100 bivariate outcomes and 5 covariates, it

takes about 1.9 seconds to get convergence using R version 2.15.3 (2013-03-01) (R Core Team, 2013) and platform: x86_64-apple-darwin9.8.0/x86_64 (64-bit). Main parts of the algorithm are coded in Fortran such as calculation of numerical derivatives and log-likelihood to quicken computations. We have incorporated those functions implementing the algorithm into the new R (R Core Team, 2013) package "qrmissing".

We use the bootstrap (Efron, 1979; Efron and Tibshirani, 1993; Davison and Hinkley, 1997) to construct confidence interval and make inferences. We resample subjects and use bootstrap percentile intervals to form confidence intervals.

2.3.3 *Goodness of Fit Check.* A simple goodness-of-fit check can be done by examining normal QQ plots of the fitted residuals from the model. This visual test can help to diagnose if the parametric assumptions are suitable for model.

After obtaining the MLE, we use the approach described in section 2.3.1 to get the fitted Δ_{ij} for each subject. Then the fitted residuals can be obtained by plugging in the fitted estimates and $\hat{\Delta}_{ij}$ to obtain,

$$\hat{\epsilon}_{ij} = \begin{cases} (y_{ij} - \hat{\Delta}_{ij} - \boldsymbol{x_{ij}^T}\hat{\boldsymbol{\beta}}_{\mathbf{1}}^{(\boldsymbol{k})})/\hat{\sigma}_1^{(k)}, & j = 1 \\ (y_{ij} - \hat{\Delta}_{ij} - \boldsymbol{y_{ij-}^T}\hat{\boldsymbol{\beta}}_{\boldsymbol{y,j-1}}^{(\geqslant \boldsymbol{j})})/\hat{\sigma}_j^{(\geqslant j)}, & j > 1 \end{cases}.$$

3. Simulation Study

In this section, we compare the performance of our proposed model in section 2.1 with the *rq* function (noted as RQ) in *quantreg* R package (Koenker, 2012) and Bottai's algorithm (Bottai and Zhen, 2013) (noted as BZ). The *rq* function minimizes the loss (check) function $\sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})$ in terms of $\boldsymbol{\beta}$, where the loss function $\rho_\tau(u) = u(\tau - I(u < 0))$ and does not make any distributional assumptions. Bottai and Zhen (2013) impute missing outcomes using the estimated conditional quantiles of missing outcomes given observed data. Their

approach does not make distributional assumptions similar to *rq* and assumes MAR missing data, but does not allow MNAR missingness.

We considered three scenarios corresponding to both MAR and MNAR assumptions for a bivariate response. In the first scenario, Y_2 were missing at random and we used the MAR assumption in our algorithm. In the next two scenarios, Y_2 were missing not at random. However, in the second scenario, we misspecified the MDM for our algorithm and still assumed MAR, while in the third scenario, we used the correct MNAR MDM. For each scenario, we considered three error distributions: normal, student t distribution with 3 degrees of freedom and Laplace distribution. For each error model, we simulated 100 data sets. For each set there are 200 bivariate observations $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})$ for $i = 1, \ldots, 200$. Y_{i1} were always observed, while some of Y_{i2} were missing. A single covariate x was sampled from Uniform(0,2). The three models for the full data response $\boldsymbol{Y}_i$ were:

$$Y_{i1}|R = 1 \sim 2 + x_i + \epsilon_{i1},$$

$$Y_{i1}|R = 0 \sim -2 - x_i + \epsilon_{i1},$$

$$Y_{i2}|R = 1, Y_{i1} \sim 1 - x_i - 1/2Y_{i1} + \epsilon_{i2},$$

where $\epsilon_{i1}, \epsilon_{i2} \overset{\text{i.i.d}}{\sim} \text{N}(0,1)$, t_3 or LP(rate = 1) distribution within each scenario.

For all cases, $\Pr(R = 1) = 0.5$. When $R = 0$, Y_{i2} is not observed, so $\text{p}(Y_{i2}|R = 0, Y_{i1})$ is not identifiable from observed data.

In the first scenario, Y_2 is missing at random, thus $\text{p}(Y_{i2}|R = 0, Y_{i1}) = \text{p}(Y_{i2}|R = 1, Y_{i1})$. In the last two scenarios, Y_2 are missing not at random. We assume $Y_{i2}|R = 0, Y_{i1} \sim 3 - x_i - 1/2Y_{i1} + \epsilon_{i2}$. Therefore, there is a shift of 2 in the intercept between $\text{p}(Y_2|R = 1, Y_1)$ and $\text{p}(Y_2|R = 0, Y_1)$.

Under an MAR assumption, the sensitivity parameter $\boldsymbol{\xi}_s$ is fixed at $\mathbf{0}$ as discussed in section 2.2. For *rq* function from *quantreg* R package, because only $Y_{i2}|R = 1$ is observed, the quantile regression for Y_{i2} can only be fit from the information of $Y_{i2}|R = 1$ vs x.

In scenario 2 under MNAR, we mis-specified the MDM using the wrong sensitivity parameter $\boldsymbol{\xi}_s$ at $\mathbf{0}$. In scenario 3, we assumed there was an intercept shift between distribution of $Y_{i2}|Y_{i1}, R = 1$ and $Y_{i2}|Y_{i1},\ R = 0$, thus fixed $\boldsymbol{\xi}_s$ at its true value.

For each dataset, we fit quantile regression for quantiles $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$. Parameter estimates were evaluated by mean squared error (MSE),

$$\text{MSE}(\gamma_{ij}) = \frac{1}{100}\sum_{k=1}^{100}\left(\hat{\gamma}_{ij}^{(k)} - \gamma_{ij}\right)^2, i = 0, 1$$

where γ_j is the true value for quantile regression coefficient, $\hat{\gamma}_j^{(k)}$ is the maximum likelihood estimates in k-th simulated dataset ($(\gamma_{01}, \gamma_{11})$ for Y_{i1}, $(\gamma_{02}, \gamma_{12})$ for Y_{i2}).

Monte Carlo standard error (MCSE) is used to evaluate the significance of difference between methods. It is calculated by

$$\text{MCSE} = \hat{\text{sd}}(\text{Bias}^2)/\sqrt{N},$$

where $\hat{\text{sd}}$ is the sample standard deviation and Bias $= \hat{\gamma}_{ij} - \gamma_{ij}$ and N is the number of simulations.

Table 1, 2 and 3 present the MSE for coefficients estimates of quantile 0.1, 0.3, 0.5, 0.7, 0.9 under each scenario. Simulation results show estimates from our algorithm and Bottai's approach are closer to the true value for all quantiles from 0.1 to 0.9. As expected, under normal errors, the proposed methods dominates both *rq* and BZ in most cases for MAR, incorrect MAR, and MNAR.

For the heavier tail distributions, t_3 and Laplace distribution, our approach shows better performance in middle quantiles and worse performance for extreme quantiles for observed data Y_1. However, our algorithm provides larger gains over *rq* function for each marginal quantile for the second component Y_2, which is missing for some units, since *rq* implicitly assumes MAR missingness. The difference in MSE becomes larger for the upper quantiles because $Y_2|R = 0$ tends to be larger than $Y_2|R = 1$; therefore, the *rq* method using only the observed Y_2 yields larger bias for upper quantiles. Bottai's approach does much better than

rq function for missing data because it imputes missing responses under MAR. It also has smaller MSE than ours on extreme quantiles when distribution has heavy tail. However, our approach has advantages in the middle quantiles (30% - 70%). We also see more gains over BZ in the quantile regression slope estimates for Y_2.

To assess the goodness of fit, we examined the QQ plot of fitted residuals in model (3) to check the normality assumption on the error term for a random sample of the simulated datasets. When our error assumption is correct (normal), the QQ plot reflects the fitted residuals follow a normal distribution. However, when we misspecified the error distribution, the proposed diagnostic method clearly suggests heavier tail error than normal, and this also demonstrates why our approach has some disadvantages for regression on extreme quantiles when errors are not normal.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

4. Application to the TOURS trial

We apply our quantile regression approach to data from TOURS, a weight management clinical trial (Perri et al., 2008). This trial was designed to test whether a lifestyle modification program could effectively help people to manage their weights in the long term. After finishing the six-month weight loss program, participants were randomly assigned to three treatments groups: face-to-face counseling, telephone counseling and control group. Their weights were recorded at baseline (Y_0), 6 months (Y_1) and 18 months (Y_2). Here, we are interested in how the distribution of weights at six months and eighteen months change with covariates. The regressors of interest include AGE, RACE (black and white) and weight at baseline (Y_0). Weights at the six months (Y_1) were always observed and 13 out of 224 observations

(6%) were missing at 18 months (Y_2). The "Age" covariate was scaled to 0 to 5 with every increment representing 5 years.

We fitted regression models for bivariate responses $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})$ for quantiles (10%, 30%, 50%, 70%, 90%). We ran 1000 bootstrap samples to obtain 95% confidence intervals.

Estimates under MAR and MNAR are presented in Table 4. For weights of participants at six months, weights of whites are generally 4kg lower than those of blacks for all quantiles, and the coefficients of race are negative and significant. Meanwhile, weights of participants are not affected by age since the coefficients are not significant. Difference in quantiles are reflected by the intercept. Coefficients of baseline weight show a strong relationship with weights after 6 months.

[Table 4 about here.]

For weights at 18 months after baseline, we have similar results. Weights after 18 months still have a strong relationship with baseline weights. However, whites do not weigh significantly less than blacks at 18 months unlike at 6 months.

We also did a sensitivity analysis based on an assumption of missing not at random. Based on previous studies of pattern of weight regain after lifestyle treatment (Wadden et al., 2001; Perri et al., 2008), we assume that

$$E(Y_2 - Y_1 | R = 0) = 3.6\text{kg},$$

which corresponds to 0.3kg regain per month after finishing the initial 6-month program. Therefore, we incorporate the sensitivity parameters in $\chi(\boldsymbol{x}_{i2}, Y_{i1})$ in the distribution of $Y_2|Y_1, R = 0$ in (4):

$$\chi(\boldsymbol{x}_{ij}, y_{i1}) = 3.6\text{kg} + y_{i1}.$$

Table 4 presents the estimates and bootstrap percentile confidence intervals under the above MNAR mechanism. There are not large differences for estimates for Y_2 under MNAR vs MAR. This is partly due to the low proportion of missing data in this study.

We also checked the goodness of fit via QQ plots on the fitted residuals as described in section 2.3.3 for each quantile regression fit. Plots are presented in the Appendix. The QQ plots showed minimal evidence against the assumption that the residuals were normally distributed; thus we are confident with the conclusion of our quantile regression models.

5. Discussion

In this paper, we have developed a marginal quantile regression model for data with monotone missingness. We use a pattern mixture model to jointly model the full data response and missingness. Here we estimate marginal quantile regression coefficients instead of coefficients conditional on random effects as in Yuan and Yin (2010). In addition, our approach allows non-parallel quantile lines over different quantiles via the mixture distribution and allows for sensitivity analysis which is essential for the analysis of missing data (NAS 2010).

Our method allows the missingness to be non-ignorable. We illustrated how to find sensitivity parameters to allow different missing data mechanisms. The recursive integration algorithm simplifies computation and can be easily implemented even in high dimensions. Simulation studies demonstrates that our approach has smaller MSE than the traditional frequentist method *rq* function for most cases, especially for inferences of partial missing responses. And it has advantages over Bottai's approach for middle quantiles regression inference even when the distribution is mis-specified. However, our approach also shows some bias for extreme quantiles comparing to Bottai and Zhen (2013) if error is mis-specified.

Our model assumes a multivariate normal distribution for each component in the pattern mixture model, which might be too restrictive in some settings. Simulation results showed that the mis-specification of the error term did have an impact on the extreme quantile regression inferences. It is possible to replace it with a non-parametric model, for example, a Dirichlet process mixture or Pólya trees. However, computational algorithms would need to be developed for those settings and we are currently working on this. In addition, since

we assume multivariate normal distributions within patterns, which can easily quantify departures from MAR via differences in means and (co)-variances, we need the specification of the model with a sequential multivariate normal distribution within each pattern; otherwise MAR constraints do not exist (Wang and Daniels, 2011). The R package "qrmissing" which can be downloaded from the author's website.

Acknowledgments

This study is supported by NIH grant CA85295.

References

Bates, D., Mullen, K. M., Nash, J. C., and Varadhan, R. (2012). *minqa: Derivative-free optimization algorithms by quadratic approximation*. R package version 1.2.1.

Bottai, M. and Zhen, H. (2013). Multiple imputation based on conditional quantile estimation. *Epidemiology, Biostatistics and Public Health* **10,**.

Buchinsky, M. (1994). Changes in the US Wage Structure 1963-1987: Application of Quantile Regression. *Econometrica* **62,** pp. 405–458.

Buchinsky, M. (1998). Recent advances in quantile regression models: A practical guideline for empirical research. *The Journal of Human Resources* **33,** pp. 88–126.

Daniels, M. J. and Hogan, J. W. (2008). *Missing data in longitudinal studies*, volume 109 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL. Strategies for Bayesian modeling and sensitivity analysis.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*, volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. With 1 IBM-PC floppy disk (3.5 inch; HD).

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7,** 1–26.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York.

Hanson, T. and Johnson, W. O. (2002). Modeling regression error with a mixture of Polya trees. *J. Amer. Statist. Assoc.* **97,** 1020–1033.

He, X., Ng, P., and Portnoy, S. (1998). Bivariate quantile smoothing splines. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60,** 537–550.

Heagerty, P. J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics* **55,** pp. 688–698.

Hjort, N. L. and Petrone, S. (2007). Nonparametric quantile inference using Dirichlet processes. In *Advances in statistical modeling and inference*, volume 3 of *Ser. Biostat.*, pages 463–492. World Sci. Publ., Hackensack, NJ.

Hjort, N. L. and Walker, S. G. (2009). Quantile pyramids for Bayesian nonparametrics. *Ann. Statist.* **37,** 105–131.

Koenker, R. (2005). *Quantile regression*, volume 38 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.

Koenker, R. (2012). *quantreg: Quantile Regression*. R package version 4.91.

Koenker, R. and Bassett, Gilbert, J. (1978). Regression quantiles. *Econometrica* **46,** pp. 33–50.

Koenker, R. and Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *J. Amer. Statist. Assoc.* **94,** 1296–1310.

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association* **96,** pp. 1458–1468.

Kottas, A. and Krnjajić, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scand. J. Stat.* **36,** 297–319.

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal*

of the American Statistical Association **88,** 125–134.

Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81,** 471–483.

Molenberghs, G., Michiels, B., Kenward, M. G., and Diggle, P. J. (1998). Monotone missing data and pattern-mixture models. *Statist. Neerlandica* **52,** 153–161.

Perri, M. G., Limacher, M. C., Durning, P. E., Janicke, D. M., Lutes, L. D., Bobroff, L. B., Dale, M. S., Daniels, M. J., Radcliff, T. A., and Martin, A. D. (2008). Extended-care programs for weight management in rural communities: the treatment of obesity in underserved rural settings (tours) randomized trial. *Archives of internal medicine* **168,** 2347.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Reich, B. J., Bondell, H. D., and Wang, H. J. (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics* **11,** 337–352.

Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics* **59,** 829–836.

Roy, J. and Daniels, M. J. (2008). A general class of pattern mixture models for nonignorable dropout with many possible dropout times. *Biometrics* **64,** 538–545, 668.

Rubin, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *J. Amer. Statist. Assoc.* **72,** 538–543.

Wadden, T. A., Berkowitz, R. I., Sarwer, D. B., Prus-Wisniewski, R., and Steinberg, C. (2001). Benefits of lifestyle modification in the pharmacologic treatment of obesity: a randomized trial. *Archives of Internal Medicine* **161,** 218–227.

Walker, S. and Mallick, B. K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics* **55,** 477–483.

Wang, C. and Daniels, M. J. (2011). A note on MAR, identifying restrictions, model comparison, and sensitivity analysis in pattern mixture models with and without covariates for incomplete data. *Biometrics* **67,** 810–818.

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

Wei, Y., Pere, A., Koenker, R., and He, X. (2006). Quantile regression methods for reference growth charts. *Stat. Med.* **25,** 1369–1382.

Yu, K., Lu, Z., and Stander, J. (2003). Quantile regression: applications and current research areas. *The Statistician* **52,** 331–350.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statist. Probab. Lett.* **54,** 437–447.

Yuan, Y. and Yin, G. (2010). Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics* **66,** 105–114.

Appendix

Goodness of Fit Check for Tours Data

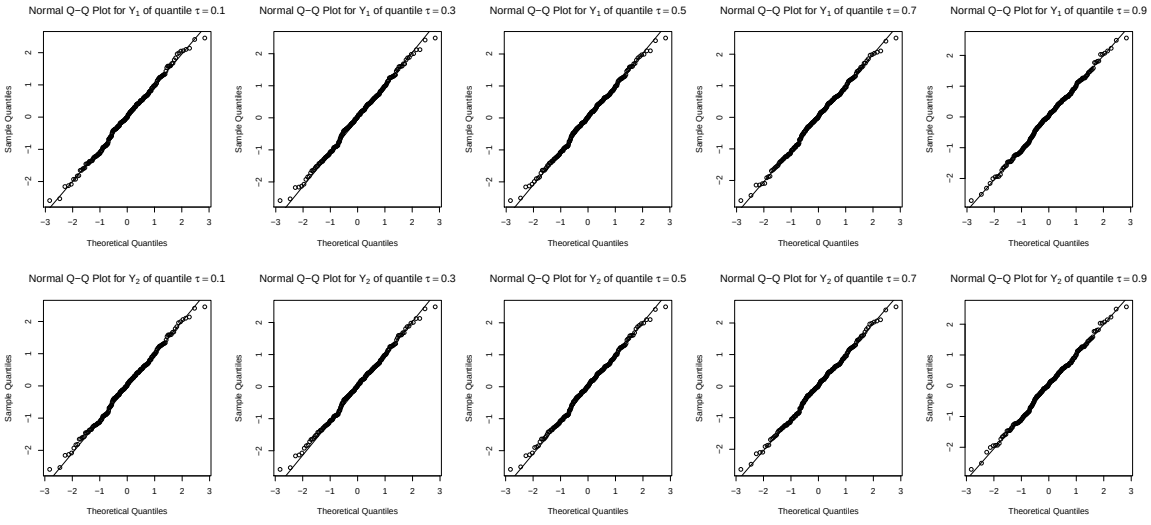

Table 1

Scenario 1: MSE(MCSE) for coefficients estimates of quantiles 0.1, 0.3, 0.5, 0.7, 0.9 under MAR assumptions. $(\gamma_{01}, \gamma_{11})$ *are quantile regression coefficients for* Y_{i1}*, and* $(\gamma_{02}, \gamma_{12})$ *are coefficients for* Y_{i2}*. MM stands for our proposed method, RQ stands for the 'rq' function in R package 'quantreg', and BZ stands for Bottai's approach.*

Normal	0.1			0.3			0.5			0.7			0.9		
	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ
γ_{01}	0.06(0.01)	0.08(0.02)	0.08(0.02)	0.09(0.06)	0.09(0.03)	0.09(0.03)	0.23(0.04)	1.13(0.15)	1.13(0.15)	0.05(0.01)	0.07(0.02)	0.07(0.02)	0.05(0.01)	0.06(0.01)	0.06(0.01)
γ_{11}	0.04(0.01)	0.07(0.01)	0.07(0.01)	0.04(0.02)	0.07(0.02)	0.07(0.02)	0.95(0.04)	2.87(0.20)	2.87(0.20)	0.02(0.01)	0.06(0.01)	0.06(0.01)	0.04(0.01)	0.05(0.01)	0.05(0.01)
γ_{02}	0.08(0.01)	0.32(0.05)	0.09(0.02)	0.07(0.02)	0.59(0.05)	0.11(0.02)	0.09(0.02)	0.96(0.06)	0.14(0.03)	0.18(0.02)	1.47(0.08)	0.20(0.03)	0.45(0.05)	2.40(0.11)	0.26(0.04)
γ_{12}	0.05(0.01)	0.11(0.02)	0.08(0.01)	0.06(0.01)	0.08(0.01)	0.09(0.02)	0.07(0.01)	0.34(0.03)	0.20(0.04)	0.10(0.02)	1.00(0.06)	0.13(0.02)	0.11(0.02)	1.07(0.07)	0.12(0.02)
T_3	0.1			0.3			0.5			0.7			0.9		
	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ
γ_{01}	0.21(0.05)	0.12(0.03)	0.12(0.03)	0.14(0.03)	0.11(0.02)	0.11(0.02)	0.13(0.05)	1.35(0.14)	1.35(0.14)	0.12(0.04)	0.10(0.02)	0.10(0.02)	0.16(0.05)	0.12(0.03)	0.12(0.03)
γ_{11}	0.11(0.03)	0.10(0.02)	0.10(0.02)	0.09(0.02)	0.08(0.02)	0.08(0.02)	0.37(0.05)	1.96(0.20)	1.96(0.20)	0.07(0.02)	0.07(0.01)	0.07(0.01)	0.10(0.02)	0.12(0.02)	0.12(0.02)
γ_{02}	0.20(0.19)	0.48(0.10)	0.13(0.11)	0.18(0.13)	0.53(0.05)	0.10(0.03)	0.21(0.07)	1.03(0.05)	0.20(0.04)	0.25(0.07)	1.74(0.09)	0.25(0.05)	0.37(0.07)	2.36(0.18)	0.49(0.11)
γ_{12}	0.09(0.02)	0.19(0.04)	0.09(0.02)	0.09(0.02)	0.06(0.01)	0.06(0.01)	0.09(0.03)	0.30(0.03)	0.20(0.04)	0.16(0.04)	0.96(0.06)	0.15(0.03)	0.16(0.04)	1.14(0.11)	0.16(0.02)
Laplace	0.1			0.3			0.5			0.7			0.9		
	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ
γ_{01}	2.34(0.26)	1.80(0.22)	1.80(0.22)	0.19(0.03)	0.22(0.04)	0.22(0.04)	0.17(0.03)	0.69(0.09)	0.69(0.09)	0.23(0.05)	0.21(0.05)	0.21(0.05)	1.77(0.20)	1.23(0.20)	1.23(0.20)
γ_{11}	0.22(0.04)	0.46(0.06)	0.46(0.06)	0.20(0.04)	0.19(0.04)	0.19(0.04)	0.14(0.02)	0.93(0.11)	0.93(0.11)	0.16(0.03)	0.24(0.04)	0.24(0.04)	0.28(0.05)	0.46(0.09)	0.46(0.09)
γ_{02}	2.94(0.21)	4.67(0.55)	1.89(0.27)	0.49(0.06)	1.28(0.15)	0.26(0.05)	0.24(0.04)	1.07(0.08)	0.20(0.03)	0.59(0.09)	1.04(0.09)	0.39(0.07)	2.82(0.30)	1.10(0.18)	2.87(0.37)
γ_{12}	0.29(0.04)	1.08(0.15)	0.45(0.09)	0.25(0.04)	0.23(0.05)	0.16(0.04)	0.21(0.03)	0.37(0.06)	0.18(0.04)	0.30(0.04)	1.14(0.12)	0.26(0.06)	0.34(0.05)	1.54(0.21)	0.58(0.14)

Table 2

Scenario 2: MSE(MCSE) for coefficients estimates of quantiles 0.1, 0.3, 0.5, 0.7, 0.9 under MNAR scenario. In this scenario, we adopted MAR assumption for our approach and thus misspecified the MDM. $(\gamma_{01}, \gamma_{11})$ are quantile regression coefficients for Y_{i1}, and $(\gamma_{02}, \gamma_{12})$ are coefficients for Y_{i2}. MM stands for our proposed method, RQ stands for the 'rq' function in R package 'quantreg', and BZ stands for Bottai's approach.

Normal	0.1			0.3			0.5			0.7			0.9		
	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ
γ_{01}	0.08(0.01)	0.10(0.01)	0.10(0.01)	0.08(0.02)	0.14(0.02)	0.14(0.02)	0.30(0.03)	1.35(0.15)	1.35(0.15)	0.10(0.04)	0.13(0.04)	0.13(0.04)	0.06(0.01)	0.08(0.01)	0.08(0.01)
γ_{11}	0.05(0.01)	0.08(0.01)	0.08(0.01)	0.03(0.00)	0.11(0.02)	0.11(0.02)	1.00(0.04)	2.87(0.23)	2.87(0.23)	0.04(0.01)	0.09(0.02)	0.09(0.02)	0.04(0.00)	0.06(0.01)	0.06(0.01)
γ_{02}	0.15(0.03)	0.42(0.05)	0.14(0.02)	0.10(0.01)	0.85(0.06)	0.15(0.02)	1.30(0.08)	4.05(0.12)	1.31(0.09)	4.88(0.13)	9.95(0.19)	3.71(0.17)	6.89(0.20)	12.67(0.26)	4.43(0.24)
γ_{12}	0.08(0.01)	0.13(0.02)	0.09(0.01)	0.08(0.01)	0.08(0.01)	0.09(0.01)	0.05(0.01)	0.31(0.02)	0.24(0.05)	0.12(0.02)	1.03(0.05)	0.15(0.02)	0.10(0.01)	1.06(0.07)	0.11(0.02)
T_3	0.1			0.3			0.5			0.7			0.9		
	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ
γ_{01}	0.21(0.05)	0.15(0.03)	0.15(0.03)	0.14(0.03)	0.11(0.02)	0.11(0.02)	0.14(0.03)	0.96(0.13)	0.96(0.13)	0.17(0.03)	0.12(0.02)	0.12(0.02)	0.37(0.25)	0.16(0.03)	0.16(0.03)
γ_{11}	0.10(0.03)	0.11(0.03)	0.11(0.03)	0.07(0.01)	0.08(0.03)	0.08(0.03)	0.43(0.05)	1.92(0.21)	1.92(0.21)	0.09(0.01)	0.09(0.02)	0.09(0.02)	0.10(0.02)	0.11(0.02)	0.11(0.02)
γ_{02}	0.31(0.08)	0.64(0.15)	0.17(0.04)	0.15(0.04)	0.74(0.06)	0.10(0.03)	1.18(0.09)	4.14(0.11)	1.23(0.10)	3.97(0.18)	10.19(0.19)	3.59(0.20)	4.18(0.24)	11.28(0.43)	3.57(0.30)
γ_{12}	0.10(0.04)	0.23(0.05)	0.12(0.02)	0.12(0.02)	0.05(0.01)	0.06(0.02)	0.09(0.02)	0.26(0.02)	0.24(0.05)	0.19(0.05)	1.01(0.05)	0.17(0.03)	0.21(0.06)	1.26(0.12)	0.19(0.04)
Laplace	0.1			0.3			0.5			0.7			0.9		
	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ
γ_{01}	2.69(0.24)	2.22(0.23)	2.22(0.23)	0.34(0.05)	0.37(0.06)	0.37(0.06)	0.20(0.03)	0.96(0.12)	0.96(0.12)	0.23(0.03)	0.30(0.05)	0.30(0.05)	2.62(0.23)	2.03(0.23)	2.03(0.23)
γ_{11}	0.39(0.05)	0.55(0.08)	0.55(0.08)	0.33(0.04)	0.37(0.07)	0.37(0.07)	0.16(0.02)	1.15(0.14)	1.15(0.14)	0.25(0.03)	0.27(0.06)	0.27(0.06)	0.23(0.03)	0.52(0.07)	0.52(0.07)
γ_{02}	4.47(0.37)	7.17(0.77)	3.45(0.45)	1.04(0.12)	1.78(0.16)	0.58(0.08)	1.57(0.15)	4.15(0.20)	1.52(0.12)	2.49(0.20)	7.98(0.31)	2.80(0.21)	0.90(0.10)	5.05(0.45)	1.32(0.18)
γ_{12}	0.53(0.08)	1.07(0.20)	0.72(0.13)	0.38(0.05)	0.27(0.04)	0.27(0.05)	0.20(0.03)	0.48(0.06)	0.29(0.04)	0.34(0.04)	1.22(0.10)	0.36(0.04)	0.30(0.03)	1.71(0.24)	0.55(0.08)

Table 3

Scenario 3: MSE(MCSE) for coefficients estimates of quantiles 0.1, 0.3, 0.5, 0.7, 0.9 under MNAR scenario. In this scenario, we used the correct sensitivity parameters for our approach. $(\gamma_{01}, \gamma_{11})$ *are quantile regression coefficients for* Y_{i1}*, and* $(\gamma_{02}, \gamma_{12})$ *are coefficients for* Y_{i2}*. MM stands for our proposed method, RQ stands for the 'rq' function in R package 'quantreg', and BZ stands for Bottai's approach.*

Normal	0.1			0.3			0.5			0.7			0.9		
	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ
γ_{01}	0.13(0.03)	0.11(0.01)	0.11(0.01)	0.13(0.02)	0.16(0.02)	0.16(0.02)	0.37(0.04)	1.15(0.12)	1.15(0.12)	0.08(0.01)	0.13(0.02)	0.13(0.02)	0.06(0.01)	0.07(0.01)	0.07(0.01)
γ_{11}	0.07(0.01)	0.08(0.01)	0.08(0.01)	0.05(0.01)	0.13(0.02)	0.13(0.02)	0.94(0.05)	2.48(0.20)	2.48(0.20)	0.04(0.01)	0.09(0.02)	0.09(0.02)	0.04(0.01)	0.05(0.01)	0.05(0.01)
γ_{02}	0.13(0.02)	0.48(0.05)	0.14(0.02)	0.15(0.03)	0.92(0.05)	0.15(0.02)	0.37(0.05)	4.25(0.11)	1.31(0.09)	0.64(0.06)	10.20(0.17)	3.70(0.16)	0.97(0.08)	12.95(0.26)	4.27(0.22)
γ_{12}	0.07(0.01)	0.09(0.02)	0.09(0.01)	0.08(0.01)	0.07(0.01)	0.10(0.02)	0.14(0.02)	0.28(0.03)	0.23(0.05)	0.10(0.02)	0.97(0.05)	0.14(0.02)	0.11(0.02)	1.04(0.07)	0.12(0.02)
T_3	0.1			0.3			0.5			0.7			0.9		
	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ
γ_{01}	0.36(0.08)	0.19(0.03)	0.19(0.03)	0.28(0.04)	0.14(0.02)	0.14(0.02)	0.17(0.03)	1.11(0.12)	1.11(0.12)	0.19(0.03)	0.15(0.02)	0.15(0.02)	0.25(0.05)	0.17(0.03)	0.17(0.03)
γ_{11}	0.16(0.03)	0.14(0.03)	0.14(0.03)	0.16(0.04)	0.11(0.02)	0.11(0.02)	0.56(0.05)	2.00(0.20)	2.00(0.20)	0.10(0.01)	0.11(0.02)	0.11(0.02)	0.11(0.02)	0.12(0.01)	0.12(0.01)
γ_{02}	0.36(0.08)	0.83(0.12)	0.29(0.05)	0.26(0.04)	0.73(0.05)	0.15(0.02)	0.33(0.05)	3.88(0.11)	1.24(0.09)	0.77(0.08)	9.80(0.21)	3.72(0.20)	0.67(0.08)	10.92(0.40)	3.76(0.26)
γ_{12}	0.16(0.03)	0.27(0.04)	0.16(0.03)	0.11(0.02)	0.08(0.01)	0.10(0.02)	0.19(0.03)	0.35(0.03)	0.27(0.04)	0.30(0.06)	1.19(0.07)	0.22(0.03)	0.32(0.06)	1.52(0.12)	0.24(0.03)
Laplace	0.1			0.3			0.5			0.7			0.9		
	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ	MM	RQ	BZ
γ_{01}	2.16(0.20)	1.82(0.21)	1.82(0.21)	0.29(0.04)	0.43(0.07)	0.43(0.07)	0.22(0.03)	0.94(0.12)	0.94(0.12)	0.28(0.04)	0.33(0.04)	0.33(0.04)	2.49(0.24)	2.06(0.26)	2.06(0.26)
γ_{11}	0.37(0.05)	0.53(0.07)	0.53(0.07)	0.21(0.03)	0.29(0.04)	0.29(0.04)	0.15(0.02)	1.10(0.14)	1.10(0.14)	0.28(0.04)	0.43(0.07)	0.43(0.07)	0.35(0.06)	0.59(0.08)	0.59(0.08)
γ_{02}	2.42(0.26)	5.97(0.57)	3.34(0.38)	0.38(0.05)	1.48(0.14)	0.55(0.07)	0.44(0.05)	3.89(0.18)	1.50(0.13)	0.56(0.08)	7.58(0.35)	2.81(0.24)	1.48(0.18)	4.43(0.39)	1.37(0.21)
γ_{12}	0.38(0.06)	0.96(0.12)	0.76(0.10)	0.31(0.05)	0.30(0.05)	0.27(0.04)	0.25(0.04)	0.48(0.05)	0.34(0.06)	0.30(0.05)	1.35(0.14)	0.39(0.07)	0.56(0.11)	1.89(0.27)	0.56(0.08)

Table 4
Estimated marginal quantile regression coefficients with 95% bootstrap percentile confidence interval for weight of participants at 6 and 18 months.

	Intercept	Age	White	BaseWeight
6 months				
10%	-6.05 (-10.88, 2.67)	0.34 (-0.25, 0.85)	-3.86 (-5.75, -2.43)	0.92 (0.85, 0.97)
30%	-2.56 (-7.67, 3.66)	0.33 (-0.25, 0.84)	-3.90 (-5.43, -2.54)	0.92 (0.86, 0.97)
50%	-0.25 (-5.29, 5.60)	0.31 (-0.25, 0.85)	-4.04 (-5.57, -2.55)	0.93 (0.87, 0.98)
70%	1.79 (-3.27, 7.81)	0.35 (-0.22, 0.86)	-4.11 (-5.67, -2.68)	0.93 (0.87, 0.98)
90%	4.81 (-0.05, 11.32)	0.40 (-0.20, 0.94)	-4.07 (-5.68, -2.68)	0.94 (0.88, 0.99)
18 months(MAR)				
10%	-17.65 (-31.75, 21.41)	-0.73 (-1.99, 0.39)	-0.12 (-10.60, 2.96)	1.01 (0.63, 1.14)
30%	-18.26 (-28.27, 9.88)	-0.74 (-2.01, 0.32)	1.07 (-8.93, 3.67)	1.07 (0.79, 1.17)
50%	-12.72 (-24.20, 10.45)	-0.73 (-2.01, 0.30)	1.04 (-6.14, 3.94)	1.06 (0.83, 1.17)
70%	-9.12 (-19.69, 14.38)	-0.73 (-2.00, 0.31)	1.18 (-5.18, 3.92)	1.06 (0.84, 1.16)
90%	-3.90 (-12.65, 19.61)	-0.75 (-1.98, 0.36)	1.24 (-4.19, 3.76)	1.08 (0.85, 1.16)
18 months(MNAR)				
10%	-20.51 (-30.97, 25.00)	-0.69 (-2.23, 0.47)	0.24 (-10.19, 3.04)	1.04 (0.62, 1.14)
30%	-18.04 (-27.14, 8.73)	-0.74 (-2.04, 0.49)	1.08 (-9.22, 3.94)	1.07 (0.83, 1.16)
50%	-12.19 (-22.52, 8.79)	-0.73 (-2.06, 0.38)	1.05 (-6.36, 4.17)	1.06 (0.86, 1.16)
70%	-7.89 (-17.63, 12.26)	-0.73 (-1.95, 0.32)	1.17 (-4.43, 4.20)	1.06 (0.87, 1.16)
90%	-3.11 (-8.60, 21.70)	-0.73 (-2.02, 0.38)	1.68 (-3.90, 4.05)	1.10 (0.87, 1.15)