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]Department of Clinical and Health Psychology, 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 a straightforward interpretation of 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 the required integrals. We conduct simulations to examine the robustness of our approach to modeling assumptions and compare its performance to competing approaches. The model is applied to data from a recent 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, the approach does not naturally extend to missing data. Extensions of minimizing loss function include the use of regularization. Koenker (2004) adopted l_1 regularization methods to shrink a large number of individual effects to a common value for quantile regression models for longitudinal data. Li and Zhu (2008) considered L_1-norm (LASSO) regularized quantile regression. Wu and Liu (2009) demonstrated using a smoothly clipped absolute deviation model for variable selection in penalized quantile regression. In terms of Bayesian inference, both parametric and semiparametric Bayesian approaches have been

proposed in the literature (Yu and Moyeed, 2001; Walker and Mallick, 1999; Hanson and Johnson, 2002; Reich et al., 2010).

All 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 typical 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 (National Research Council, 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, denote $S_i \in \{1, 2, \ldots, J\}$ to be the number of observed $Y_{ij}^T s$ 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 $S = k$ 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 the number of observed responses, 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. As in Daniels and Hogan (2008), we denote $\boldsymbol{\xi}(\boldsymbol{\omega}) = (\boldsymbol{\xi}_m, \boldsymbol{\xi}_s)$ as a reparametrization of the parameters in the full data model above where $\boldsymbol{\xi}_s$ are sensitivity

parameters (not identified by the observed data) and $\boldsymbol{\xi}_m$ are identified by the observed data (just the parameters of the observed data distribution).

We assume a sequential finite (K) mixture of normals within each pattern. A random variable x follows a finite (K) mixture of normal distribution (MN), when

$$p(x|\boldsymbol{\theta}) = \sum_{i=1}^{K} \omega_i \phi_N(x; \mu_i, \sigma_i^2),$$

where $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{\omega})$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_K)$, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K)$, and $\phi_N(x; \mu, \sigma^2)$ stands for the PDF evaluated at x of a normal distribution with mean μ and variance σ^2.

We specify the model in this way such 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 in (1) hold. In particular, we specify the distributions conditional on $S = k$ as:

$$\begin{aligned}
&Y_{i1} = \Delta_{i1} + \boldsymbol{x_i^T \beta_1^{(k)}} + \epsilon_{i1}, k = 1, \ldots, J, \\
&Y_{ij}|\boldsymbol{Y}_{ij^-} = \begin{cases} \Delta_{ij} + \boldsymbol{y}_{ij^-}^T \boldsymbol{\beta}_{y,j-1} + \epsilon_{ij}, & k \geqslant j; \\ \chi(\boldsymbol{x}_i, \boldsymbol{y}_{ij^-}) + \epsilon_{ij}, & k < j; \end{cases}, \text{ for } 2 \leqslant j \leqslant J, \\
&S_i = k \sim \text{Multinomial}(1, \boldsymbol{\phi}),
\end{aligned} \tag{3}$$

where

$$\begin{aligned}
\epsilon_{ij}|\boldsymbol{\theta} &\overset{\text{i.i.d}}{\sim} \text{MN}(\boldsymbol{\theta}), \\
\boldsymbol{\theta} &= (\mu_l, \sigma_l, \omega_l), l = 1, \ldots, K,
\end{aligned} \tag{4}$$

with the marginal quantile regression constraints (1). And $\boldsymbol{\beta}_1^{(k)}$ is the shift of the coefficients for distribution of $Y_{i1}|S = k$, $\boldsymbol{y}_{ij^-} = (y_{i1}, \ldots, y_{i(j-1)})^T$ is the response history for subject i up to time point $(j-1)$, $\boldsymbol{\beta}_{y,j-1} = \left(\beta_{y_1,j-1}, \ldots, \beta_{y_{j-1},j-1}\right)^T$ are autoregressive coefficients, σ_j 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. Δ_{ij} are subject/time specific intercepts determined by the parameters in (1) and (3); more details are given in what follows. $\chi(\boldsymbol{x}_i, \boldsymbol{y}_{ij^-})$ is the mean of the unobserved data distribution and allows sensitivity

analysis by varying assumptions on χ; for computational reasons we assume that χ is linear in y_{ij^-}. For example, here we specify

$$\chi(\boldsymbol{x}_i, \boldsymbol{y}_{ij^-}) = \Delta_{ij} + \boldsymbol{y}_{ij^-}^T \boldsymbol{\beta}_{y,j-1} + \boldsymbol{x_i^T h^{(k)}}, \tag{5}$$

where $\boldsymbol{h}^{(k)}$ is a set of sensitivity parameters and Δ_{ij} depends on x_i. Comparing (3) and (5), the mean of the unidentified distribution, $\mathrm{p}_k(y_{ij}|\boldsymbol{y}_{ij^-})$ for $k < j$ is characterized by an intercept shift from identified observed data distribution, $\mathrm{p}_k(y_{ij}|\boldsymbol{y}_{ij^-})$ for $k \geqslant j$. More details about sensitivity parameters are given in section 2.2.

For (standard) identifiability of the distribution of the observed data, we use the following restrictions (without loss of generality),

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

Also in order to not confound the marginal quantile regression parameters, we put the following constraint on the parameters $\boldsymbol{\theta}$ in the mixture of normals distribution,

$$\sum_{l=1}^{K} \omega_l \mu_l = 0.$$

In (3) and (5), Δ_{ij} are functions of $\tau, \boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{h}, \boldsymbol{\theta}, \boldsymbol{\gamma}_j, \boldsymbol{\phi}$ and are determined by the marginal quantile regressions,

$$\tau = \Pr(Y_{ij} \leqslant \boldsymbol{x}_i^T \boldsymbol{\gamma}_j) = \sum_{k=1}^{J} \phi_k \Pr_k(Y_{ij} \leqslant \boldsymbol{x}_i^T \boldsymbol{\gamma}_j) \text{ for } j = 1, \tag{6}$$

and

$$\begin{aligned} \tau = \Pr(Y_{ij} \leqslant \boldsymbol{x}_i^T \boldsymbol{\gamma}_j) &= \sum_{k=1}^{J} \phi_k \Pr_k(Y_{ij} \leqslant \boldsymbol{x}_i^T \boldsymbol{\gamma}_j) \\ &= \sum_{k=1}^{J} \phi_k \int \cdots \int \Pr_k(Y_{ij} \leqslant \boldsymbol{x}_i^T \boldsymbol{\gamma}_j | \boldsymbol{y}_{ij^-}) \mathrm{p}_k(y_{i(j-1)} | \boldsymbol{y}_{i(j-1)^-}) \\ &\quad \cdots \mathrm{p}_k(y_{i2}|y_{i1}) \mathrm{p}_k(y_{i1}) dy_{i(j-1)} \cdots dy_{i1}. \text{ for } j = 2, \ldots, J. \end{aligned} \tag{7}$$

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

finite mixture of normals distribution makes model flexible, which allows heavy tails and skewness for each pattern. 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 previous approaches.

2.2 *Missing Data Mechanism and Sensitivity Analysis*

Mixture models as specified in Section 2.1 are not identified by the observed data (as stated in the previous subsection). 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 $(\boldsymbol{h^{(1)},\ldots,h^{(J-1)}})$ can not be identified from the observed data. Then $\boldsymbol{\xi}_s = (\boldsymbol{h}^{(k)})$ is a set of sensitivity parameters (Daniels and Hogan, 2008) as mentioned earlier.

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 (6) and (7), Δ_{ij} depends on subject-specific covariates $\boldsymbol{x}_i$, 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 (6) with (3) and (4):

$$\tau = \sum_{k=1}^{J} \phi_k \left(\sum_{l=1}^{K} \omega_l \Phi \left(\frac{\boldsymbol{x}_i^T \boldsymbol{\gamma}_1 - \Delta_{i1} - \boldsymbol{x_i^T} \boldsymbol{\beta_1^{(k)}} - \mu_l}{\sigma_l} \right) \right).$$

  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 (7), we propose a recursive approach.

Expand equation (7) with with (3) and (4):

$$\begin{aligned}\tau &= \sum_{k=1}^{J} \Pr_k(Y_{ij} \leqslant \boldsymbol{x}_i^T \boldsymbol{\gamma}_j) \\ &= \sum_{k=1}^{J} \left(\sum_{l=1}^{K} \omega_l \Pr_k(Y_{ij} \leqslant \boldsymbol{x}_i^T \boldsymbol{\gamma}_j; \boldsymbol{\theta}_l) \right).\end{aligned}$$

For the first multiple integral in equation (7), apply lemma 2.1 once to obtain:

$$\begin{aligned}\Pr_1(Y_{ij} \leqslant \boldsymbol{x}_i^T \boldsymbol{\gamma}_j; \boldsymbol{\theta}_l) = \int \dots \int \Pr(Y_{ij} \leqslant \boldsymbol{x}_i^T \boldsymbol{\gamma}_j | S = 1, \boldsymbol{x}_i, \boldsymbol{Y}_{ij^-}, \boldsymbol{\theta}_l) \\ dF(Y_{i(j-1)} | S = 1, \boldsymbol{x}_i, \boldsymbol{Y}_{i(j-1)^-}) \cdots dF(Y_{i1} | S = 1, \boldsymbol{x}_i), \\ = \int \dots \int \Phi \left(\frac{\boldsymbol{x}_i^T \boldsymbol{\gamma}_j - \chi(\boldsymbol{x}_i, \boldsymbol{y}_{ij^-}) - \mu_l}{\sigma_l} \right) \\ dF(Y_{i(j-1)} | S = 1, \boldsymbol{x}_i, \boldsymbol{Y}_{i(j-1)^-}) \cdots dF(Y_{i1} | S = 1, \boldsymbol{x}_i),\end{aligned}$$

while

$$\begin{aligned}\int \Phi \left(\frac{\boldsymbol{x}_i^T \boldsymbol{\gamma}_j - \chi(\boldsymbol{x}_i, \boldsymbol{y}_{ij^-}) - \mu_l}{\sigma_l} \right) dF(Y_{i(j-1)} | S = 1, \boldsymbol{x}_i, \boldsymbol{Y}_{i(j-1)^-}) \\ = \sum_{m=1}^{K} \omega_m \int \Phi \left(\frac{Y_{i(j-2)} - b_m^*}{a_m^*} \right) dF(Y_{i(j-2)} | S = 1, \boldsymbol{x}_i, \boldsymbol{Y}_{i(j-2)^-}, \boldsymbol{\theta}_m)\end{aligned}$$

where

$$a_m^* = \sqrt{\sigma_l^2 / \beta_{y_{j-1},j-1}^2 + \sigma_m^2} \Big/ \left(-\frac{\beta_{y_{j-2},j-1}}{\beta_{y_{j-1},j-1}} - \beta_{y_{j-2},j-2} \right),$$

$$b_m^* = \frac{(\boldsymbol{x}_i^T \boldsymbol{\gamma}_j - \Delta_{ij} - \boldsymbol{x_i^T h^{(k)}} - \mu_l)/\beta_{y_{j-1},j-1} - (\Delta_{i(j-1)} + \sum_{k=1}^{j-3} Y_{i(k)} \beta_{y_k,j-2} + \boldsymbol{x_i^T h^{(k)}} + \mu_m)}{\frac{\beta_{y_{j-2},j-1}}{\beta_{y_{j-1},j-1}} + \beta_{y_{j-2},j-2}},$$

when $\chi(\boldsymbol{x}_i, \boldsymbol{y}_{ij^-})$ takes the linear form given in (5) with $\beta_{y_{j-1},j-1} > 0$. Similar results hold as long as $\chi(\cdot, \cdot)$ is linear in $\boldsymbol{y}_{ij^-}$.

Then, by recursively applying lemma 2.1 $(j-1)$ times, each multiple integral in equation (7) 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

$$\begin{aligned} L_i(\boldsymbol{\xi}|\boldsymbol{y}_i, S_i = k) &= \phi_k \mathrm{p}_k(y_{ik}|y_{i1}, \ldots, y_{i(k-1)})\mathrm{p}_k(y_{i(k-1)}|y_{i1}, \ldots, y_{i(k-2)}) \cdots \mathrm{p}_k(y_{i1}) \\ &= \phi_k \mathrm{p}_{\geqslant k}(y_{ik}|y_{i1}, \ldots, y_{i(k-1)})\mathrm{p}_{\geqslant k-1}(y_{i(k-1)}|y_{i1}, \ldots, y_{i(k-2)}) \cdots \mathrm{p}_k(y_{i1}), \end{aligned} \tag{8}$$

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

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 and Tibshirani, 1993) to construct confidence interval and make inferences. We resample subjects and use bootstrap percentile intervals to form confidence intervals.

3. Simulation Study

In this section, we compare the performance of our proposed model with the rq function (denoted as RQ) in quantreg R package (Koenker, 2012) and Bottai's algorithm (Bottai

and Zhen, 2013) (denoted 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 missingness, but does not allow MNAR.

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:

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

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

$$Y_{i2}|R = 1, Y_{i1} \sim 1 - x_i - 0.5Y_{i1} + \epsilon_{i2},$$

where $\epsilon_{i1}, \epsilon_{i2} \overset{\text{i.i.d}}{\sim} \text{N}(0,1)$, t_3, LP(rate = 1), and $0.5N(-2,1) + 0.5N(2,1)$ 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 $Y_{i2}|R = 0, Y_{i1} \sim 3 -$

$x_i - 0.5Y_{i1} + \epsilon_{i2}$. Therefore, there is a shift of 2 in the intercept between $\mathrm{p}(Y_2|R=1,Y_1)$ and $\mathrm{p}(Y_2|R=0,Y_1)$.

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),

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

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}).

The true values for quantile regression coefficients are obtained through the following procedures:

- Generate $x_1, \ldots, x_{100}$ evenly spaced in the range of x.
- For each x_i, compute the τ-th conditional quantile of $Y_i|x_i$, denoted as $\hat{y}_i^{\tau}$.
- Regress $\hat{y}_i^{\tau}$ on x_i to compute the regression coefficients (the τ-th quantile regression coefficients).

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 BZ are closer to the true value for all quantiles from 0.1 to 0.9. The smallest MSE is in red.

Under normal errors and scenario 1, the proposed method (M1 and M2) has the lowest MSE for all regression coefficients. However, there is bias for the estimates of the coefficients for the first component of the responses when MDM or error distribution is mis-specified

comparing to BZ. Even with mis-specification of MDM or error distribution, our method still shows better performance in most middle quantiles. Meanwhile, 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 MCAR 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.

For the heavier tail distributions, t_3 and Laplace distribution, our approach (M2) still shows best performance in most quantiles. BZ does much better than rq function for missing data because it imputes missing responses under MAR.

For errors with skewness (mixture of two normals), our approach (M2) succeeded to capture the shape of the error. Therefore, M2 provides better performance in most cases.

It can also be shown that models adopting mixture of normals (M2) perform better than models using errors of single normal (M1). The mixture of normals allows heavy tail and skewness.

[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 4.2 kg 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.]

We calculated the BIC for median regression models with $K = 1, 2, 3, 4, 5$. From table 5, we have strong evidence that $K = 1$ is the best model when using mixture of normals.

[Table 5 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, the effect of gender is slightly less than that for 6 month. Whites weigh 3.5 kg less than blacks at 18 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 specify $\chi(\boldsymbol{x}_i, Y_{i1})$ as

$$\chi(\boldsymbol{x}_i, y_{i1}) = 3.6 + y_{i1},$$

where $h_0^{(1)} = 3.6 + y_{i1} - (\Delta_{ij} + y_{i1}\beta_{y,1})$. Table 4 presents the estimates and bootstrap percentile confidence intervals under the above MNAR mechanism. There are not large differences from the estimates for Y_2 under MNAR vs MAR. This is partly due to the low proportion of missing data in this study.

5. Discussion

In this chapter, 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 (National Research Council, 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 demonstrate that our approach has smaller MSE than the traditional frequentist method rq function for missing component, especially for inferences in the median regression. And it has advantages over BZ for most quantile regression inference even when the tails of the distribution are mis-specified.

Our model assumes a multivariate mixture of normals distribution for each component in the pattern mixture model, which can deal with heavy tail and skewness. 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; we are currently working on this. Our approach is implemented with the R package "qrmissing" which can be downloaded from the author's website: `https://github.com/liuminzhao/qrmissing`.

Supplementary Materials

Web Figure 1 referenced in Section 3 is available with this paper at the Biometrics website on Wiley Online Library.

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.

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.

Koenker, R. (2004). Quantile regression for longitudinal data. *J. Multivariate Anal.* **91,** 74–89.

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.

Li, Y. and Zhu, J. (2008). L_1-norm quantile regression. *J. Comput. Graph. Statist.* **17,** 163–185.

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.

National Research Council (2010). *The Prevention and Treatment of Missing Data in Clinical Trials*. The National Academies Press.

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.

Wu, Y. and Liu, Y. (2009). Variable selection in quantile regression. *Statist. Sinica* **19,** 801–817.

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.

Table 1

Scenario 1: MSE 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}*. M1 stands for our proposed method with* $K = 1$*, M2 stands for proposed model with* $K = 2$*, RQ stands for the 'rq' function in R package 'quantreg', and BZ stands for Bottai's approach. The titles for sub-columns indicate models with four errors distributed from: Normal(N), t distribution with degrees of freedom 3(*t_3*), Laplace distribution(LP) and mixture of two normals(Mix).*

	N				t_3				LP				Mix			
	M1	M2	RQ	BZ	M1	M2	RQ	BZ	M1	M2	RQ	BZ	M1	M2	RQ	BZ
10%																
γ_{01}	0.05	0.06	0.06	0.06	0.16	0.08	0.16	0.16	2.12	1.18	1.22	1.22	0.28	0.08	0.16	0.16
γ_{11}	0.03	0.04	0.05	0.05	0.11	0.05	0.12	0.12	0.25	0.16	0.41	0.41	0.16	0.05	0.12	0.12
γ_{02}	0.10	0.08	0.37	0.10	0.19	0.08	0.53	0.17	3.26	3.32	6.58	2.57	0.25	0.09	0.86	0.39
γ_{12}	0.05	0.04	0.09	0.07	0.10	0.06	0.24	0.10	0.33	0.19	0.99	0.51	0.18	0.10	0.28	0.36
30%																
γ_{01}	0.06	0.05	0.09	0.09	0.12	0.07	0.10	0.10	0.21	0.17	0.30	0.30	0.41	0.14	0.44	0.44
γ_{11}	0.03	0.03	0.07	0.07	0.07	0.04	0.07	0.07	0.18	0.16	0.28	0.28	0.16	0.04	0.53	0.53
γ_{02}	0.06	0.06	0.60	0.08	0.15	0.08	0.44	0.13	0.62	0.36	1.24	0.30	0.32	0.08	1.59	0.32
γ_{12}	0.05	0.05	0.05	0.04	0.10	0.04	0.08	0.08	0.25	0.16	0.22	0.15	0.14	0.05	0.48	0.26
50%																
γ_{01}	0.22	0.17	1.17	1.17	0.09	0.12	1.01	1.01	0.16	0.17	0.80	0.80	0.10	0.03	0.18	0.18
γ_{11}	0.99	0.90	2.58	2.58	0.46	0.48	1.70	1.70	0.12	0.19	0.88	0.88	0.19	0.11	0.36	0.36
γ_{02}	0.11	0.12	1.15	0.13	0.16	0.12	1.03	0.21	0.34	0.17	1.05	0.27	0.22	0.08	1.88	0.44
γ_{12}	0.07	0.06	0.24	0.19	0.08	0.06	0.31	0.21	0.18	0.14	0.38	0.20	0.13	0.05	0.51	0.28
70%																
γ_{01}	0.05	0.06	0.10	0.10	0.17	0.07	0.10	0.10	0.25	0.18	0.32	0.32	0.39	0.16	0.84	0.84
γ_{11}	0.03	0.03	0.06	0.06	0.09	0.04	0.08	0.08	0.24	0.13	0.24	0.24	0.14	0.04	0.68	0.68
γ_{02}	0.12	0.16	1.64	0.17	0.26	0.20	1.82	0.28	0.59	0.39	0.84	0.55	0.26	0.08	1.96	0.56
γ_{12}	0.11	0.09	0.90	0.13	0.15	0.10	0.88	0.17	0.24	0.22	1.21	0.28	0.15	0.06	0.47	0.31
90%																
γ_{01}	0.05	0.05	0.08	0.08	0.21	0.07	0.13	0.13	2.36	1.21	1.80	1.80	0.38	0.13	0.20	0.20
γ_{11}	0.04	0.03	0.06	0.06	0.11	0.06	0.11	0.11	0.19	0.11	0.44	0.44	0.14	0.08	0.15	0.15
γ_{02}	0.37	0.48	2.49	0.23	0.36	0.28	2.63	0.64	2.79	2.95	1.44	3.38	0.41	0.27	2.13	0.51
γ_{12}	0.12	0.10	1.02	0.10	0.14	0.09	1.05	0.20	0.27	0.28	1.91	0.63	0.18	0.08	1.00	0.23

Table 2

Scenario 2: MSE 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}. M1 stands for our proposed method with $K = 1$, M2 stands for proposed model with $K = 2$, RQ stands for the 'rq' function in R package 'quantreg', and BZ stands for Bottai's approach. The titles for sub-columns indicate models with four errors distributed from: Normal(N), t distribution with degrees of freedom 3(t_3), Laplace distribution(LP) and mixture of two normals(Mix).

	N				t_3				LP				Mix			
	M1	M2	RQ	BZ	M1	M2	RQ	BZ	M1	M2	RQ	BZ	M1	M2	RQ	BZ
10%																
γ_{01}	0.10	0.07	0.09	0.09	0.31	0.12	0.20	0.20	2.38	1.19	1.78	1.78	0.34	0.21	0.20	0.20
γ_{11}	0.06	0.05	0.07	0.07	0.10	0.08	0.13	0.13	0.29	0.23	0.57	0.57	0.15	0.13	0.17	0.17
γ_{02}	0.11	0.10	0.47	0.14	0.49	0.13	1.02	0.36	4.64	3.79	7.07	3.51	0.36	0.21	1.28	0.50
γ_{12}	0.06	0.06	0.10	0.10	0.20	0.09	0.31	0.18	0.57	0.36	0.95	0.62	0.27	0.23	0.37	0.42
30%																
γ_{01}	0.14	0.11	0.16	0.16	0.18	0.08	0.16	0.16	0.26	0.20	0.36	0.36	0.62	0.22	0.81	0.81
γ_{11}	0.08	0.07	0.13	0.13	0.08	0.06	0.12	0.12	0.26	0.18	0.34	0.34	0.16	0.06	0.76	0.76
γ_{02}	0.09	0.07	0.83	0.13	0.33	0.10	0.88	0.20	1.01	0.54	1.95	0.60	0.31	0.34	3.06	0.68
γ_{12}	0.08	0.09	0.07	0.10	0.15	0.11	0.07	0.11	0.48	0.27	0.21	0.23	0.21	0.10	0.51	0.38
50%																
γ_{01}	0.23	0.23	1.68	1.68	0.12	0.17	1.30	1.30	0.18	0.16	0.95	0.95	0.14	0.04	0.21	0.21
γ_{11}	0.93	0.90	3.15	3.15	0.47	0.56	2.19	2.19	0.15	0.26	1.15	1.15	0.16	0.13	0.45	0.45
γ_{02}	1.44	1.47	4.14	1.29	1.37	1.14	4.25	1.45	1.58	1.26	4.36	1.76	1.02	0.98	4.57	1.34
γ_{12}	0.09	0.07	0.30	0.24	0.15	0.09	0.26	0.32	0.28	0.19	0.38	0.41	0.21	0.07	0.54	0.41
70%																
γ_{01}	0.07	0.07	0.14	0.14	0.23	0.07	0.13	0.13	0.23	0.23	0.39	0.39	0.52	0.27	1.09	1.09
γ_{11}	0.04	0.03	0.11	0.11	0.10	0.05	0.11	0.11	0.28	0.21	0.38	0.38	0.18	0.04	0.81	0.81
γ_{02}	4.77	4.87	10.24	3.94	3.97	4.08	10.28	3.57	2.48	2.63	8.20	2.77	1.98	1.35	4.93	1.85
γ_{12}	0.14	0.15	1.01	0.17	0.25	0.12	0.98	0.21	0.31	0.31	1.28	0.51	0.25	0.08	0.72	0.43
90%																
γ_{01}	0.06	0.06	0.08	0.08	0.26	0.08	0.12	0.12	2.30	1.18	2.09	2.09	0.51	0.15	0.27	0.27
γ_{11}	0.04	0.04	0.07	0.07	0.16	0.07	0.10	0.10	0.25	0.22	0.65	0.65	0.18	0.09	0.20	0.20
γ_{02}	6.57	6.72	12.96	4.79	4.41	4.35	11.82	3.33	1.04	0.78	4.10	1.50	2.65	2.39	6.46	2.04
γ_{12}	0.12	0.12	1.03	0.17	0.26	0.11	1.12	0.26	0.43	0.32	2.06	0.79	0.26	0.14	1.01	0.45

Table 3

Scenario 3: MSE 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}. M1 stands for our proposed method with $K = 1$, M2 stands for proposed model with $K = 2$, RQ stands for the 'rq' function in R package 'quantreg', and BZ stands for Bottai's approach. The titles for sub-columns indicate models with four errors distributed from: Normal(N), t distribution with degrees of freedom 3(t_3), Laplace distribution(LP) and mixture of two normals(Mix).

	N				t_3				LP				Mix			
	M1	M2	RQ	BZ	M1	M2	RQ	BZ	M1	M2	RQ	BZ	M1	M2	RQ	BZ
10%																
γ_{01}	0.10	0.07	0.10	0.10	0.30	0.11	0.16	0.16	2.43	1.29	2.06	2.06	0.46	0.21	0.34	0.34
γ_{11}	0.06	0.04	0.06	0.06	0.14	0.06	0.14	0.14	0.35	0.25	0.44	0.44	0.17	0.10	0.25	0.25
γ_{02}	0.17	0.12	0.50	0.14	0.30	0.13	0.68	0.17	2.73	2.74	7.59	3.24	0.52	0.26	1.07	0.42
γ_{12}	0.10	0.06	0.14	0.10	0.15	0.09	0.23	0.14	0.43	0.35	1.12	0.63	0.27	0.11	0.28	0.29
30%																
γ_{01}	0.27	0.13	0.14	0.14	0.22	0.10	0.16	0.16	0.32	0.25	0.37	0.37	0.48	0.24	1.32	1.32
γ_{11}	0.06	0.06	0.08	0.08	0.12	0.06	0.16	0.16	0.24	0.17	0.32	0.32	0.15	0.10	0.83	0.83
γ_{02}	0.19	0.09	0.84	0.11	0.27	0.14	0.82	0.14	0.46	0.26	1.97	0.54	0.42	0.34	2.68	0.65
γ_{12}	0.10	0.06	0.07	0.08	0.12	0.07	0.08	0.09	0.35	0.24	0.32	0.25	0.24	0.12	0.49	0.35
50%																
γ_{01}	0.23	0.30	1.33	1.33	0.13	0.22	1.22	1.22	0.23	0.23	0.81	0.81	0.15	0.05	0.21	0.21
γ_{11}	0.87	0.93	2.44	2.44	0.56	0.57	2.33	2.33	0.17	0.22	0.85	0.85	0.18	0.15	0.43	0.43
γ_{02}	0.44	0.35	4.16	1.28	0.21	0.25	4.05	1.22	0.45	0.30	4.27	1.42	0.34	0.11	4.18	1.34
γ_{12}	0.16	0.13	0.30	0.21	0.25	0.14	0.29	0.27	0.26	0.26	0.45	0.38	0.20	0.09	0.60	0.38
70%																
γ_{01}	0.09	0.09	0.12	0.12	0.16	0.08	0.16	0.16	0.28	0.29	0.33	0.33	0.61	0.30	0.72	0.72
γ_{11}	0.08	0.04	0.08	0.08	0.10	0.05	0.12	0.12	0.25	0.18	0.29	0.29	0.18	0.05	0.55	0.55
γ_{02}	0.79	0.77	10.33	3.99	0.66	0.49	10.31	3.69	0.59	0.45	7.74	2.73	0.44	0.18	4.88	1.66
γ_{12}	0.14	0.12	0.98	0.16	0.20	0.10	0.96	0.23	0.36	0.35	1.35	0.47	0.35	0.14	0.76	0.35
90%																
γ_{01}	0.07	0.06	0.10	0.10	0.17	0.10	0.19	0.19	2.62	1.39	2.37	2.37	0.30	0.15	0.23	0.23
γ_{11}	0.05	0.04	0.07	0.07	0.13	0.05	0.13	0.13	0.32	0.21	0.61	0.61	0.19	0.11	0.19	0.19
γ_{02}	0.93	0.99	13.00	4.87	0.68	0.49	11.49	4.23	1.52	1.55	4.27	1.34	0.70	0.29	6.85	2.02
γ_{12}	0.14	0.13	0.99	0.13	0.23	0.10	1.20	0.23	0.56	0.32	2.58	0.99	0.40	0.20	0.93	0.36

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%	-5.2(-10.7, 0.3)	0.3(-0.2,0.9)	-4.2(-5.6,-2.7)	0.9(0.9,1.0)
30%	-1.4(-7.0, 4.0)	0.3(-0.2,0.9)	-4.2(-5.6,-2.7)	0.9(0.9,1.0)
50%	1.0(-4.6, 6.3)	0.3(-0.2,0.9)	-4.2(-5.5,-2.7)	0.9(0.9,1.0)
70%	3.5(-2.3, 8.8)	0.3(-0.2,0.9)	-4.2(-5.5,-2.7)	0.9(0.9,1.0)
90%	7.0(1.4, 12.4)	0.3(-0.2,0.9)	-4.2(-5.5,-2.7)	0.9(0.9,1.0)
18 months(MAR)				
10%	-6.7(-14.1, 1.6)	-0.1(-1.0,0.9)	-3.5(-5.9,-1.2)	0.9(0.8,1.0)
30%	-0.1(-7.7, 8.2)	-0.1(-1.0,0.9)	-3.5(-5.9,-1.2)	0.9(0.8,1.0)
50%	4.0(-3.6, 12.2)	-0.0(-1.0,0.9)	-3.5(-5.8,-1.2)	0.9(0.8,1.0)
70%	8.4(0.5, 16.6)	-0.1(-1.0,0.9)	-3.5(-5.8,-1.2)	0.9(0.8,1.0)
90%	14.5(6.6, 23.1)	-0.1(-1.0,0.9)	-3.4(-5.9,-1.2)	0.9(0.8,1.0)
18 months(MNAR)				
10%	-6.5(-14.1,2.1)	-0.0(-1.0,0.8)	-3.5(-5.8,-1.0)	0.9(0.8,1.0)
30%	-0.2(-7.7,8.1)	-0.1(-1.0,0.8)	-3.5(-5.8,-1.0)	0.9(0.8,1.0)
50%	4.0(-3.7,12.3)	-0.1(-1.0,0.8)	-3.5(-5.8,-1.0)	0.9(0.8,1.0)
70%	8.4(0.5,16.8)	-0.0(-1.0,0.8)	-3.5(-5.8,-1.0)	0.9(0.8,1.0)
90%	14.8(6.5,23.2)	-0.1(-1.0,0.8)	-3.5(-5.8,-1.0)	0.9(0.8,1.0)

Table 5
BIC: The BIC for median regression models with $K = 1, 2, 3, 4, 5$.

	$K = 1$	$K = 2$	$K = 3$	$K = 4$	$K = 5$
BIC	878	889	894	912	928