Quantile Regression in the Presence of Monotone Missingness

August 14, 2013

Abstract

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 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 comes from the estimated conditional quantile regression and specification of conditional density of missing covariates given observed ones. However, they put more focus on the missing covariates rather 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 ignorable. However, because their imputation method is not derived from a joint distribution, the joint distribution with such conditionals may not exist. In addition, their approach does not allow for MNAR.

Yuan and Yin (2010) introduced a fully parametric Bayesian quantile regression approach for longitudinal data with nonignorable missing data. They used random effects to explain the within-subject correlation and applied multivariate normal priors on the random terms to match the traditional quantile regression check function with penalties, which can also shrink the subject-specific effect toward zero. However, the quantile regression coefficients are conditional on the random effects, which is not of interest if we are interested in interpreting regression coefficients unconditional on random effects. In addition, 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 (on Handling Missing Data in Clinical Trials; 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 nonignorable 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 allows non-ignorable missingness and sensitivity analysis.

The structure of this article is as follows. First, we introduce a quantile regression method to address monotone nonignorable 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.

Under monotone dropout, without loss of generality, denote $S_i \in \{1, 2, \ldots, J\}$ to be the follow up time, and $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iJ})^T$ to be the 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 $\boldsymbol{\gamma}_j = (\gamma_{j0}, \gamma_{j2}, \ldots, \gamma_{jp})^T$,

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

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

Let

$$\begin{aligned} \mathrm{p}_k(Y) &= \mathrm{p}(Y|S = k), \\ \mathrm{p}_{\geq k}(Y) &= \mathrm{p}(Y|S \geq k) \end{aligned}$$

be the densities of response $\boldsymbol{Y}$ given follow-up time $S = k$ and $S \geq k$. 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 is 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 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

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

where $\boldsymbol{\theta}_E$ indexes the parameters in an extrapolation distribution, $\boldsymbol{\theta}_{y,O}, \boldsymbol{\phi}$ indexes parameters in distributions of observed data.

We assume models within pattern to be multivariate normal distributions and specify a sequential model parametrization. Let the subscript i stand for subject i. We specify the conditional distributions as:

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

where $\boldsymbol{y}_{ij^-} = (y_{i1}, \ldots, y_{i(j-1)})^T$, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_J)$,$\boldsymbol{h}_j^{(k)} = (h_{j1}^{(k)}, \ldots, h_{jp}^{(k)})$, $\boldsymbol{x}_j$ is a $p \times 1$ covariate vector, $\boldsymbol{\beta}_{y,j-1}^{(k)} = \big(\beta_{y_1,j-1}^{(k)}, \ldots, \beta_{y_{j-1},j-1}^{(k)}\big)^T$ and $\sigma_j^{(k)}$ is conditional standard deviation of response component j. We specify the model as in (3) to , have multivariate normal distribution within patterns and to allow MAR to exist (Wang and Daniels, 2011). More details are presented in section 2.2.

In (3), Δ_{ij} are functions of $\tau, \boldsymbol{x}_{ij}, \boldsymbol{\alpha}_j, \gamma_j$ and other parameters and are determined by the marginal quantile regressions,

$$\tau = \Pr(Y_{ij} \leq \boldsymbol{x}_{ij}^T \gamma_j) = \sum_{k=1}^{J} \pi_k \Pr_k(Y_{ij} \leq \boldsymbol{x}_{ij}^T \gamma_j), \tag{4}$$

for $j = 1$ and

$$\begin{aligned} \tau = \Pr(Y_{ij} \leq \boldsymbol{x}_{ij}^T \gamma_j) &= \sum_{k=1}^{J} \pi_k \Pr_k(Y_{ij} \leq \boldsymbol{x}_{ij}^T \gamma_j) \\ &= \sum_{k=1}^{J} \pi_k \int \cdots \int \Pr_k(Y_{ij} \leq \boldsymbol{x}_{ij}^T \gamma_j | y_{i1}, \ldots, y_{i(j-1)}) \, \mathrm{p}_k(y_{i(j-1)} | y_{i1}, \ldots, y_{i(j-2)}) \\ &\quad \cdots \mathrm{p}_k(y_{i2} | y_{i1}) \, \mathrm{p}_k(y_{i1}) dy_{i(j-1)} \cdots dy_{i1}. \end{aligned} \tag{5}$$

for $j = 2, \ldots, J$. Computational details will be given in section 2.3.

The idea is to model the marginal quantile regressions directly, 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.

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

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

where $\boldsymbol{\beta}_1^{(k)} = (\beta_{11}^{(k)}, \ldots, \beta_{p1}^{(k)})^T$. Further details on these restrictions can be found in Appendix A.

2.2 Missing Data Mechanism and Sensitivity Analysis

In general, mixture models are not identified due to insufficient information provided by observed data. Specific forms of missingness are needed to 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 $j \geq 2$ and $k < j$:

$$\mathrm{p}_k(y_j | y_1, \ldots, y_{j-1}) = \mathrm{p}_{\geq j}(y_j | y_1, \ldots, y_{j-1}). \tag{6}$$

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

$$\begin{aligned} \log \sigma_j^{(k)} &= \log \sigma_j^{(\geq j)} + \delta_j^{(k)}, \\ \boldsymbol{\beta}_{y,j-1}^{(k)} &= \boldsymbol{\beta}_{y,j-1}^{(\geq j)} + \boldsymbol{\eta}_{j-1}^{(k)}, \end{aligned}$$

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

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.

For fully Bayesian inference, we can put priors on $(\boldsymbol{\xi}_s, \boldsymbol{\xi}_m)$ as :

$$p(\boldsymbol{\xi}_s, \boldsymbol{\xi}_m) = p(\boldsymbol{\xi}_s)p(\boldsymbol{\xi}_m),$$

where $\boldsymbol{\xi}_m = \left(\gamma_j, \boldsymbol{\beta}_{y,j-1}^{(\geq j)}, \boldsymbol{\alpha}_j^{(\geq j)}, \boldsymbol{\pi}\right)$, the identified parameters in the data distribution. If we assume MAR with no uncertainty, the prior of $\boldsymbol{\xi}_s$ is $\mathrm{p}(\boldsymbol{\xi}_s = \mathbf{0}) \equiv 1$. Sensitivity analysis can be executed by putting point mass priors on $\boldsymbol{\xi}_s$ to examine the effect of priors on the posterior inference about quantile regression coefficients γ_{ij}^{τ}. For example, if MAR is assumed with uncertainty, priors can be assigned as $\mathrm{E}(\boldsymbol{\xi}_s) = \boldsymbol{\xi}_{s0} = \mathbf{0}$ with $\mathrm{Var}(\boldsymbol{\xi}_s) \neq \mathbf{0}$. If we assume MNAR with no uncertainty, we can put priors satisfying $\mathrm{E}(\boldsymbol{\xi}_s) = \Delta_{\xi}$, where $\Delta_{\xi} \neq \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\xi}_s) = \mathbf{0}$. If MNAR is assumed with uncertainty, then priors could be $\mathrm{E}(\boldsymbol{\xi}_s) = \Delta_{\xi}$, where $\Delta_{\xi} \neq \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\xi}_s) \neq \mathbf{0}$.

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, 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 using an adaptive gradient descent algorithm in section 2.3.2. Finally, we present a Monte Carlo Markov Chain (MCMC) sampling algorithm for Bayesian inference in section 2.3.3.

2.3.1 Calculation of Δ

From equation (4) and (5), 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 $\boldsymbol{\xi} = (\boldsymbol{\xi}_m, \xi_s)$.

- Δ_{i1} : Expand equation (4):

$$\tau = \sum_{k=1}^{J} \pi_k \Phi \left(\frac{\boldsymbol{x}_{i1}^T \gamma_1 - \Delta_{i1} - \boldsymbol{x}_{i1}^T \boldsymbol{\beta}_1^{(k)}}{\sigma_1^{(k)}} \right),$$

  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.

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

  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:*

$$\int \Phi\left(\frac{x-b}{a}\right) d\Phi(x;\mu,\sigma) = \begin{cases} 1-\Phi\left(\frac{b-\mu}{\sigma}\Big/\sqrt{\frac{a^2}{\sigma^2}+1}\right) & a>0, \\ \Phi\left(\frac{b-\mu}{\sigma}\Big/\sqrt{\frac{a^2}{\sigma^2}+1}\right) & a<0, \end{cases} \tag{7}$$

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

Proof of 2.1 is in Appendix B.

To solve equation (5), we propose a recursive approach. For the first multiple integral in equation (5), apply lemma 2.1 once to obtain:

$$\begin{aligned}
\Pr(Y_j \leq \boldsymbol{x}^T\gamma_j | S=1) &= \int \dots \int \Pr(Y_j \leq \boldsymbol{x}^T\gamma_j | S=1, \boldsymbol{x}, Y_{j-1}, \dots, Y_1) \\
&\quad dF(Y_{j-1}|S=1, Y_{j-2}, \dots, Y_1) \cdots dF(Y_2|S=1, Y_1) dF(Y_1|S=1), \\
&= \int \dots \int \Phi\left(\frac{\boldsymbol{x}^T\gamma_j - \mu_{j|1,\dots,j-1}(y_{j-1})}{\sigma_{j|1,\dots,j-1}}\right) \\
&\quad dF(Y_{j-1}|S=1, Y_{j-2}, \dots, Y_1) \cdots dF(Y_2|S=1, Y_1) dF(Y_1|S=1), \\
&= \int \dots \int \Phi\left(\frac{y_{j-2} - b^*}{a^*}\right) dF(Y_{j-2}|S=1, Y_{j-3}, \dots, Y_1) \cdots dF(Y_1|S=1).
\end{aligned}$$

Then, by recursively applying lemma 2.1 $(j-1)$ times, each multiple integral in equation (5) can be simplified to single normal CDF. Thus it can be solved 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 $\boldsymbol{y}_i$ with follow-up time $S=k$ is

$$\begin{aligned}
L_i(\boldsymbol{\xi}|\boldsymbol{y}_i, S_i=k) &= \pi_k \, \mathrm{p}_k(y_k|y_1,\dots,y_{k-1}) \, \mathrm{p}_k(y_{k-1}|y_1,\dots,y_{k-2}) \cdots \mathrm{p}_k(y_1), \\
&= \pi_k \, \mathrm{p}_{\geq k}(y_k|y_1,\dots,y_{k-1}) \, \mathrm{p}_{\geq k-1}(y_{k-1}|y_1,\dots,y_{k-2}) \cdots \mathrm{p}_k(y_1),
\end{aligned} \tag{8}$$

where $\boldsymbol{y}_i = (y_1,\dots,y_k)$.

We use a 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 minimize the target function $J(\boldsymbol{\xi})$. Under an MAR assumption, we fix $\boldsymbol{\xi}_s = \boldsymbol{0}$, 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 352 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 computation.

We use the bootstrap (Efron, 1979; Efron and Tibshirani, 1993; Davison and Hinkley, 1997) to construct confidence interval and make inferences. For quantile regression models, we resample the subject and use bootstrap percentile interval to form confidence intervals.

A simple goodness-of-fit test can be done by examine normal QQ plot of the fitted residuals from the model. The visual test can help to diagnose if the parametric assumption of normal distributions is suitable for model.

2.3.3 Bayesian Framework

For Bayesian inference, we specify priors on the parameters $\boldsymbol{\xi}$ and use a block Gibbs sampling method to draw samples from the posterior distribution. Denote all the parameters (including sensitivity parameters) to sample as :

$$\boldsymbol{\xi}_m = \left\{ \gamma_1, \gamma_2, \ldots, \gamma_J, \boldsymbol{\beta}_{y,j-1}^{(\geq j)}, \boldsymbol{\alpha}_j^{(\geq j)} \right\} \text{ for } j = 1, \ldots, J, \boldsymbol{\xi}_s = \left\{ \boldsymbol{h}_j^{(k)}, \boldsymbol{\eta}_{j-1}^{(k)}, \delta_j^{(k)} \right\} \text{ for } k = 1, \ldots, j; 2 \leq j \leq J.$$

Comma separated parameters are marked to sample as a block. Updates of $\boldsymbol{\xi}_m$ require a Metropolis-Hasting algorithm, while $\boldsymbol{\xi}_s$ samples are drawn directly from priors as desired for missingness mechanism assumptions.

As mentioned in section 2.2, MAR or MNAR assumptions are implemented via specific priors. For example, if MAR is assumed with no uncertainty, then $\boldsymbol{\xi}_s = \mathbf{0}$ with probability 1. Details for updating parameters are:

- γ_1: Use Metropolis-Hasting algorithm.
  1. Draw (γ_1^c) candidates from candidate distribution;
  2. Based on the new candidate parameter $\boldsymbol{\xi}^c$, calculate candidate Δ_{i1}^c for each subject i as we described in section 2.3.1. If $S > 1$ for subject i, update candidate $\Delta_{ij}^c, j \geq 2$ as well since $\Delta_{ij}, j \geq 2$ depends on Δ_{i1}. (For $S = 1$, we only need to update Δ_{i1}^c);
  3. Plug in Δ_{i1}^c or ($\Delta_{i1}^c, \Delta_{ij}^c, j \geq 2$) in likelihood (8) to get candidate likelihood;
  4. Compute Metropolis-Hasting ratio, and accept the candidate value or keep the previous value.
- For the rest of the identifiable parameters, algorithms for updating the samples are all similar to γ_j.
- For sensitivity parameters, because we do not get any information from the data, we sample them from priors, which are specified based on assumptions about the missingness.

3 Simulation Study

In this section, we compared the performance of our proposed model in section 2.1 with the *rq* function 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*. Their imputation approach assumes ignorable missing data.

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 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:

$$
\begin{aligned}
Y_{i1}|R = 1 &\sim 2 + x_i + \epsilon_{i1}, \\
Y_{i1}|R = 0 &\sim -2 - x_i + \epsilon_{i1}, \\
Y_{i2}|R = 1, y_{i1} &\sim 1 - x_i - 1/2 y_{i1} + \epsilon_{i2},
\end{aligned}
$$

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

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

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

Under true MNAR assumption, we still used the same $\boldsymbol{\xi}_s = \mathbf{0}$ in scenario 2, mis-specifying the MDM. But we fixed $\boldsymbol{\xi}_s$ at the true value, assuming there was an intercept shift between distribution of $Y_{i2}|Y_{i1}, R = 1$ and $Y_{i2}|Y_{i1}, R = 0$ in scenario 3.

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

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

where γ_{ij} is the true value for quantile regression coefficient, $\hat{\gamma}_{ij}^{(k)}$ 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

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

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. In each scenario, our proposed method and Bottai's have smaller MSE than *rq* function in all cases for normal error. And no matter if we mis-specified or used correct sensitivity parameters, our approach has smaller MSE than *BZ*.

As to the heavier tail distributions, t_3 and Laplace distribution, our approach shows better performance in middle quantiles and lose to *rq* for extreme quantiles for observed data Y_1. Nevertheless, our algorithm provides larger gains over *rq* function for each marginal quantile for the second component Y_2, which are missing for some observations. No matter what missing data mechanism (MAR or MNAR), what assumption we use in our approach (misspecification or correct specified), our method shows advantages over *rq* function, especially for Y_2, because *quantreg* does not consider the missingness mechanism. 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, however, shows great advantage over *rq* function for missing data because its imputing method for missing responses. It also has smaller MSE than ours on extreme quantiles regression when distribution has heavy tail. However, our approach has advantages on middle quantiles (30% - 70%) for marginal inference on missing responses regardless using mis-specification or correct sensitivity parameters. And we can see more gains over *BZ* in the quantile regression slope estimates for Y_2.

We also proposed a simple goodness of fit tool to check our parametric model on model fit. We used the QQ plot on fitted residuals in model (3) to check the normality assumption on the error term. One sample goodness of fit diagnostic from the simulation shows that when our error assumption is correct (normal), the QQ plot reflects the fitted residuals follow exact normal distribution. However, when we misspecified the error distribution, the proposed diagnostic method did suggest heavier tail error than normal, and this also explains why our approach has some disadvantages on the extreme quantiles regression when error is not normal.

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 ones for Y_{i2}. MM stands for our proposed method, and RQ stands for the ’rq’ function in R package ’quantreg’.

	MAR 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}	0.05 (0.01)	0.10 (0.01)	0.10 (0.01)	0.05 (0.01)	0.09 (0.02)	0.09 (0.02)	0.20 (0.03)	1.15 (0.13)	1.15 (0.13)	0.04 (0.02)	0.09 (0.01)	0.09 (0.01)	0.04 (0.01)	0.06 (0.01)	0.06 (0.01)
γ_{11}	0.02 (0.00)	0.07 (0.01)	0.07 (0.01)	0.02 (0.00)	0.07 (0.01)	0.07 (0.01)	0.90 (0.04)	2.41 (0.21)	2.41 (0.21)	0.03 (0.01)	0.07 (0.01)	0.07 (0.01)	0.03 (0.00)	0.05 (0.01)	0.05 (0.01)
γ_{02}	0.04 (0.01)	0.31 (0.04)	0.13 (0.02)	0.05 (0.01)	0.68 (0.05)	0.11 (0.02)	0.07 (0.01)	1.07 (0.06)	0.18 (0.04)	0.14 (0.02)	1.51 (0.06)	0.21 (0.03)	0.19 (0.03)	2.35 (0.11)	0.25 (0.05)
γ_{12}	0.03 (0.00)	0.11 (0.01)	0.09 (0.01)	0.03 (0.00)	0.06 (0.01)	0.08 (0.01)	0.06 (0.01)	0.29 (0.02)	0.16 (0.03)	0.07 (0.01)	0.96 (0.05)	0.13 (0.02)	0.06 (0.01)	1.05 (0.06)	0.11 (0.02)
	MAR 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 (0.04)	0.14 (0.03)	0.14 (0.03)	0.15 (0.03)	0.14 (0.03)	0.14 (0.03)	0.12 (0.02)	1.29 (0.15)	1.29 (0.15)	0.10 (0.02)	0.11 (0.02)	0.11 (0.02)	0.20 (0.03)	0.15 (0.03)	0.15 (0.03)
γ_{11}	0.07 (0.01)	0.09 (0.02)	0.09 (0.02)	0.06 (0.01)	0.09 (0.02)	0.09 (0.02)	0.49 (0.06)	2.00 (0.21)	2.00 (0.21)	0.06 (0.01)	0.08 (0.02)	0.08 (0.02)	0.06 (0.01)	0.10 (0.01)	0.10 (0.01)
γ_{02}	0.26 (0.05)	0.62 (0.09)	0.18 (0.04)	0.12 (0.03)	0.54 (0.06)	0.11 (0.02)	0.14 (0.04)	1.09 (0.06)	0.20 (0.03)	0.25 (0.05)	1.84 (0.09)	0.18 (0.04)	0.54 (0.12)	2.58 (0.16)	0.48 (0.08)
γ_{12}	0.08 (0.02)	0.19 (0.05)	0.12 (0.03)	0.09 (0.02)	0.07 (0.02)	0.08 (0.02)	0.12 (0.02)	0.28 (0.02)	0.27 (0.05)	0.17 (0.03)	0.86 (0.05)	0.18 (0.03)	0.18 (0.03)	0.93 (0.10)	0.19 (0.04)
	MAR 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}	2.43 (0.17)	1.91 (0.23)	1.91 (0.23)	0.19 (0.03)	0.21 (0.04)	0.21 (0.04)	0.13 (0.02)	0.72 (0.09)	0.72 (0.09)	0.17 (0.03)	0.20 (0.04)	0.20 (0.04)	2.26 (0.18)	1.49 (0.20)	1.49 (0.20)
γ_{11}	0.19 (0.03)	0.52 (0.09)	0.52 (0.09)	0.21 (0.03)	0.21 (0.03)	0.21 (0.03)	0.12 (0.02)	1.06 (0.14)	1.06 (0.14)	0.17 (0.03)	0.20 (0.05)	0.20 (0.05)	0.16 (0.03)	0.35 (0.06)	0.35 (0.06)
γ_{02}	3.06 (0.22)	5.17 (0.46)	2.52 (0.28)	0.54 (0.07)	1.31 (0.17)	0.35 (0.04)	0.24 (0.04)	1.08 (0.09)	0.32 (0.07)	0.63 (0.08)	1.03 (0.12)	0.52 (0.09)	3.35 (0.27)	0.86 (0.24)	3.45 (0.37)
γ_{12}	0.17 (0.03)	0.62 (0.12)	0.48 (0.07)	0.21 (0.04)	0.22 (0.04)	0.19 (0.03)	0.14 (0.03)	0.35 (0.05)	0.29 (0.04)	0.20 (0.04)	1.21 (0.15)	0.26 (0.04)	0.17 (0.04)	1.70 (0.23)	0.55 (0.09)

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 ones for Y_{i2}. MM stands for our proposed method, and RQ stands for the 'rq' function in R package 'quantreg'.

	MNAR 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}	0.03 (0.00)	0.09 (0.01)	0.09 (0.01)	0.04 (0.00)	0.09 (0.02)	0.09 (0.02)	0.23 (0.03)	1.27 (0.14)	1.27 (0.14)	0.05 (0.01)	0.10 (0.02)	0.10 (0.02)	0.05 (0.01)	0.11 (0.02)	0.11 (0.02)
γ_{11}	0.03 (0.00)	0.07 (0.01)	0.07 (0.01)	0.03 (0.00)	0.07 (0.01)	0.07 (0.01)	0.90 (0.04)	2.64 (0.21)	2.64 (0.21)	0.04 (0.01)	0.09 (0.02)	0.09 (0.02)	0.04 (0.01)	0.08 (0.01)	0.08 (0.01)
γ_{02}	0.06 (0.01)	0.40 (0.06)	0.12 (0.02)	0.08 (0.01)	0.89 (0.05)	0.13 (0.02)	1.10 (0.07)	4.13 (0.11)	1.17 (0.09)	3.77 (0.16)	9.91 (0.19)	3.76 (0.18)	4.28 (0.20)	12.45 (0.26)	4.56 (0.25)
γ_{12}	0.04 (0.01)	0.13 (0.02)	0.09 (0.01)	0.05 (0.01)	0.08 (0.01)	0.10 (0.02)	0.08 (0.01)	0.30 (0.03)	0.21 (0.03)	0.09 (0.01)	1.04 (0.05)	0.12 (0.02)	0.07 (0.01)	1.11 (0.07)	0.11 (0.02)

	MNAR 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}	0.24 (0.29)	0.16 (0.07)	0.16 (0.07)	0.13 (0.02)	0.08 (0.02)	0.08 (0.02)	0.14 (0.02)	1.12 (0.15)	1.12 (0.15)	0.13 (0.02)	0.12 (0.02)	0.12 (0.02)	0.29 (0.29)	0.15 (0.03)	0.15 (0.03)
γ_{11}	0.08 (0.02)	0.12 (0.03)	0.12 (0.03)	0.08 (0.02)	0.06 (0.02)	0.06 (0.02)	0.36 (0.04)	1.91 (0.20)	1.91 (0.20)	0.07 (0.04)	0.08 (0.02)	0.08 (0.02)	0.07 (0.03)	0.12 (0.02)	0.12 (0.02)
γ_{02}	0.26 (0.08)	0.58 (0.07)	0.14 (0.02)	0.17 (0.04)	0.77 (0.06)	0.13 (0.02)	1.20 (0.09)	4.14 (0.12)	1.36 (0.10)	3.46 (0.19)	10.17 (0.23)	3.79 (0.19)	3.04 (0.25)	12.31 (0.41)	3.78 (0.28)
γ_{12}	0.06 (0.02)	0.18 (0.03)	0.10 (0.02)	0.09 (0.02)	0.07 (0.01)	0.08 (0.01)	0.08 (0.02)	0.25 (0.02)	0.18 (0.04)	0.13 (0.03)	0.99 (0.06)	0.08 (0.01)	0.13 (0.04)	0.95 (0.10)	0.12 (0.03)

	MNAR 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}	2.31 (0.18)	1.42 (0.16)	1.42 (0.16)	0.26 (0.03)	0.35 (0.05)	0.35 (0.05)	0.20 (0.03)	0.87 (0.10)	0.87 (0.10)	0.24 (0.04)	0.33 (0.06)	0.33 (0.06)	2.48 (0.18)	1.95 (0.20)	1.95 (0.20)
γ_{11}	0.21 (0.02)	0.40 (0.06)	0.40 (0.06)	0.27 (0.03)	0.31 (0.04)	0.31 (0.04)	0.16 (0.02)	0.92 (0.13)	0.92 (0.13)	0.23 (0.03)	0.26 (0.04)	0.26 (0.04)	0.18 (0.03)	0.46 (0.07)	0.46 (0.07)
γ_{02}	3.95 (0.27)	6.29 (0.63)	3.33 (0.40)	0.99 (0.10)	1.69 (0.14)	0.47 (0.06)	1.55 (0.13)	4.08 (0.18)	1.66 (0.13)	2.34 (0.17)	8.09 (0.33)	3.12 (0.22)	0.77 (0.10)	4.90 (0.45)	1.44 (0.18)
γ_{12}	0.25 (0.04)	0.83 (0.13)	0.63 (0.11)	0.33 (0.05)	0.21 (0.03)	0.20 (0.03)	0.22 (0.03)	0.41 (0.04)	0.31 (0.04)	0.34 (0.04)	1.20 (0.12)	0.37 (0.05)	0.29 (0.04)	1.87 (0.23)	0.48 (0.06)

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 ones for Y_{i2}. MM stands for our proposed method, and RQ stands for the 'rq' function in R package 'quantreg'.

	MNAR 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}	0.04 (0.01)	0.10 (0.01)	0.10 (0.01)	0.07 (0.02)	0.10 (0.01)	0.10 (0.01)	0.28 (0.04)	1.41 (0.17)	1.41 (0.17)	0.08 (0.01)	0.18 (0.03)	0.18 (0.03)	0.06 (0.01)	0.11 (0.02)	0.11 (0.02)
γ_{11}	0.03 (0.00)	0.07 (0.01)	0.07 (0.01)	0.04 (0.01)	0.09 (0.01)	0.09 (0.01)	0.81 (0.04)	2.73 (0.22)	2.73 (0.22)	0.04 (0.01)	0.11 (0.02)	0.11 (0.02)	0.04 (0.01)	0.08 (0.01)	0.08 (0.01)
γ_{02}	0.06 (0.01)	0.32 (0.03)	0.13 (0.02)	0.08 (0.01)	0.82 (0.05)	0.16 (0.02)	0.24 (0.04)	4.21 (0.12)	1.10 (0.10)	0.26 (0.04)	9.96 (0.19)	3.53 (0.19)	0.30 (0.04)	12.60 (0.27)	4.28 (0.24)
γ_{12}	0.04 (0.00)	0.09 (0.01)	0.10 (0.01)	0.04 (0.00)	0.07 (0.01)	0.13 (0.03)	0.21 (0.03)	0.32 (0.03)	0.25 (0.04)	0.10 (0.01)	1.11 (0.06)	0.24 (0.03)	0.10 (0.01)	1.13 (0.07)	0.18 (0.03)

	MNAR 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}	0.29 (0.05)	0.14 (0.02)	0.14 (0.02)	0.16 (0.03)	0.18 (0.04)	0.18 (0.04)	0.18 (0.03)	1.14 (0.14)	1.14 (0.14)	0.14 (0.03)	0.13 (0.02)	0.13 (0.02)	0.38 (0.06)	0.20 (0.03)	0.20 (0.03)
γ_{11}	0.08 (0.01)	0.11 (0.02)	0.11 (0.02)	0.07 (0.01)	0.11 (0.02)	0.11 (0.02)	0.46 (0.04)	2.15 (0.19)	2.15 (0.19)	0.08 (0.01)	0.11 (0.02)	0.11 (0.02)	0.07 (0.01)	0.16 (0.02)	0.16 (0.02)
γ_{02}	0.41 (0.07)	0.69 (0.12)	0.17 (0.03)	0.14 (0.03)	0.74 (0.05)	0.15 (0.02)	0.36 (0.06)	4.19 (0.10)	1.32 (0.10)	0.65 (0.10)	10.29 (0.20)	3.79 (0.21)	1.11 (0.21)	12.08 (0.41)	4.13 (0.30)
γ_{12}	0.07 (0.01)	0.33 (0.05)	0.14 (0.02)	0.08 (0.01)	0.06 (0.01)	0.10 (0.02)	0.25 (0.06)	0.27 (0.02)	0.40 (0.07)	0.26 (0.06)	1.02 (0.06)	0.22 (0.03)	0.25 (0.06)	1.10 (0.09)	0.23 (0.04)

	MNAR 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}	2.66 (0.22)	2.42 (0.30)	2.42 (0.30)	0.30 (0.05)	0.41 (0.06)	0.41 (0.06)	0.21 (0.03)	0.87 (0.12)	0.87 (0.12)	0.24 (0.03)	0.37 (0.07)	0.37 (0.07)	2.61 (0.18)	1.82 (0.26)	1.82 (0.26)
γ_{11}	0.25 (0.03)	0.69 (0.08)	0.69 (0.08)	0.28 (0.04)	0.37 (0.06)	0.37 (0.06)	0.19 (0.03)	1.00 (0.13)	1.00 (0.13)	0.24 (0.03)	0.32 (0.07)	0.32 (0.07)	0.20 (0.03)	0.58 (0.09)	0.58 (0.09)
γ_{02}	2.80 (0.21)	6.44 (0.59)	3.12 (0.34)	0.38 (0.05)	1.86 (0.15)	0.55 (0.07)	0.43 (0.06)	4.19 (0.19)	1.63 (0.15)	0.57 (0.08)	8.43 (0.34)	2.91 (0.23)	3.18 (0.33)	4.44 (0.43)	1.60 (0.24)
γ_{12}	0.23 (0.04)	1.05 (0.14)	0.67 (0.08)	0.24 (0.04)	0.27 (0.04)	0.25 (0.04)	0.23 (0.04)	0.46 (0.06)	0.31 (0.05)	0.35 (0.06)	1.17 (0.11)	0.34 (0.05)	0.34 (0.05)	1.75 (0.22)	0.77 (0.09)

4 Real Data Analysis

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 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) after the trial. Here, we are interested in how the distribution of weights at six months and eighteenth months change with covariates. The regressors of interest include AGE, RACE (black and white) and weights 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). All weights were scaled by 1/100 for computation stability.

We fitted regression models for bivariate responses $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})$ for quantiles (10%, 30%, 50%, 70%, 90%). We ran 100 bootstrap samples to obtain 95% confidence intervals.

Estimates under MAR 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 basically reflected from intercept, though only the intercept of 90% quantile regression is significant. The main contribution of the weights at six months or eighteen months is the baseline weight. Coefficients of baseline weight show strong linearity relationship between baseline weight and weights after months. Confidence intervals of coefficients for baseline weight are statistically significant away from 1, which is another sign that the six-month weight management program is effective.

For weights at 18 months after baseline, we have similar conclusions. Although estimates are similar, we can still see that participants at 18 months regain weights after the 6-month program for all quantiles. Intercept increased for all quantiles except for 10%. Whites still weigh less than blacks for all quantiles, but the magnitude is 1kg smaller than that at 6th months. All confidence intervals of baseline weight include 1 now.

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

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

which means 0.3kg regain per month after finishing the initial 6-month program. Therefore, we have sensitivity parameters in distribution of $Y_2 | Y_1, R = 0$ in a restriction as:

$$\Delta_{i2} + \boldsymbol{x}_{i2}^T \boldsymbol{h}_2^{(1)} + E(y_{i1} | R = 0)(\beta_{y,1}^{(1)} + \eta_1^{(1)} - 1) = 3.6\text{kg}.$$

Table 5 presents the estimates and bootstrap percentile confidence intervals under the above MNAR mechanism. There is no big difference for estimates of Y_1, Y_2 due to the missing data mechanism assumption change.

Table 4: Estimated marginal quantile regression coefficients with 95% bootstrap percentile confidence interval for weight of participants at 6 and 18 months. Weight measurement is scaled by 1/100. Missing data mechanism assumption is MAR.

	Intercept	Age.centered.	White	BaseWeight
Weight at 6 months				
10%	-0.06 (-0.10, 0.00)	0.00 (-0.00, 0.01)	-0.04 (-0.06, -0.02)	0.92 (0.87, 0.97)
30%	-0.02 (-0.07, 0.03)	0.00 (-0.00, 0.01)	-0.04 (-0.06, -0.03)	0.92 (0.88, 0.97)
50%	0.00 (-0.04, 0.06)	0.00 (-0.00, 0.01)	-0.04 (-0.06, -0.03)	0.93 (0.88, 0.97)
70%	0.03 (-0.02, 0.08)	0.00 (-0.00, 0.01)	-0.04 (-0.06, -0.03)	0.93 (0.88, 0.98)
90%	0.05 (0.01, 0.11)	0.00 (-0.00, 0.01)	-0.04 (-0.06, -0.03)	0.94 (0.89, 0.99)
Weight at 18 months				
10%	-0.08 (-0.16, 0.02)	-0.00 (-0.01, 0.01)	-0.03 (-0.06, -0.01)	0.92 (0.82, 1.00)
30%	-0.02 (-0.10, 0.08)	-0.00 (-0.01, 0.01)	-0.03 (-0.06, -0.01)	0.92 (0.82, 1.00)
50%	0.03 (-0.05, 0.13)	-0.00 (-0.01, 0.01)	-0.03 (-0.06, -0.01)	0.92 (0.82, 1.00)
70%	0.06 (-0.01, 0.17)	-0.00 (-0.01, 0.01)	-0.03 (-0.06, -0.01)	0.93 (0.83, 1.01)
90%	0.12 (0.04, 0.20)	0.00 (-0.01, 0.01)	-0.03 (-0.06, -0.01)	0.93 (0.84, 1.01)

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 estimates 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 (on Handling Missing Data in Clinical Trials; National Research Council, 2010).

Our method allows the missingness to be non-ignorable. We illustrated how to put informative priors for Bayesian inference and how to find sensitivity parameters to allow different missing data mechanisms in general. 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 appraoch for middle quantiles regression inference when error is heavy tailed. We also demonstrated sensitivity analysis and how to allow non-ignorable missingness assumptions.

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 on the error term did have impact on the extreme quantile regres-

Table 5: Estimated marginal quantile regression coefficients with 95% confidence interval for weight of participants at 6 and 18 months under MNAR assumption. Weight measurement is scaled by 1/100.

	Intercept	Age(centered)	White	BaseWeight
Weight at 6 months				
10%	-0.06 (-0.10, 0.00)	0.00 (-0.00, 0.01)	-0.04 (-0.06, -0.02)	0.92 (0.86, 0.97)
30%	-0.02 (-0.06, 0.03)	0.00 (-0.00, 0.01)	-0.04 (-0.05, -0.03)	0.92 (0.87, 0.97)
50%	0.00 (-0.04, 0.05)	0.00 (-0.00, 0.01)	-0.04 (-0.06, -0.03)	0.93 (0.87, 0.97)
70%	0.03 (-0.02, 0.07)	0.00 (-0.00, 0.01)	-0.04 (-0.06, -0.03)	0.93 (0.88, 0.98)
90%	0.05 (0.01, 0.11)	0.00 (-0.00, 0.01)	-0.04 (-0.06, -0.03)	0.95 (0.89, 0.99)
Weight at 18 months				
10%	-0.09 (-0.15, 0.01)	-0.00 (-0.01, 0.01)	-0.03 (-0.06, 0.00)	0.92 (0.80, 0.99)
30%	-0.02 (-0.08, 0.06)	-0.00 (-0.01, 0.01)	-0.03 (-0.06, -0.01)	0.91 (0.82, 0.98)
50%	0.04 (-0.04, 0.11)	-0.00 (-0.01, 0.01)	-0.04 (-0.07, -0.00)	0.91 (0.84, 0.98)
70%	0.08 (0.01, 0.16)	0.00 (-0.01, 0.01)	-0.04 (-0.07, -0.01)	0.91 (0.82, 0.98)
90%	0.15 (0.07, 0.22)	0.00 (-0.01, 0.01)	-0.04 (-0.07, -0.01)	0.90 (0.82, 0.98)

sion inferences. It is possible to replace that with a semi-parametric model, for example, the Dirichlet process mixture or Pólya tree. Meanwhile, even though we use a multivariate normal distributions within patterns, which can easily departures from MAR via differences in means and (co)-variances, we still need strong assumptions for sequential multivariate normal distribution within each pattern; otherwise MAR constraints do not exist (Wang and Daniels, 2011).

6 Acknowledgments

References

Douglas Bates, Katharine M. Mullen, John C. Nash, and Ravi Varadhan. *minqa: Derivative-free optimization algorithms by quadratic approximation*, 2012. URL http://CRAN.R-project.org/package=minqa. R package version 1.2.1.

Matteo Bottai and Huiling Zhen. Multiple imputation based on conditional quantile estimation. *Epidemiology, Biostatistics and Public Health*, 10(1), 2013. ISSN 2282-0930. doi: 10.2427/8758. URL http://ebph.it/article/view/8758.

Moshe Buchinsky. Changes in the US Wage Structure 1963-1987: Application of Quantile Regression. *Econometrica*, 62(2):pp. 405–458, 1994. ISSN 00129682. URL http://www.jstor.org/stable/2951618.

Moshe Buchinsky. Recent advances in quantile regression models: A practical guideline for empirical research. *The Journal of Human Resources*, 33(1):pp. 88–126, 1998. ISSN 0022166X. URL http://www.jstor.org/stable/146316.

Michael J. Daniels and Joseph W. Hogan. *Missing data in longitudinal studies*, volume 109 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL, 2008. ISBN 978-1-58488-609-9; 1-58488-609-9. doi: 10.1201/9781420011180. URL http://dx.doi.org/10.1201/9781420011180. Strategies for Bayesian modeling and sensitivity analysis.

A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*, volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1997. ISBN 0-521-57391-2. With 1 IBM-PC floppy disk (3.5 inch; HD).

B. Efron. Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7(1):1–26, 1979. ISSN 0090-5364. URL http://links.jstor.org/sici?sici=0090-5364(197901)7:1<1:BMALAT>2.0.CO;2-6&origin=MSN.

Bradley Efron and Robert J. Tibshirani. *An introduction to the bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1993. ISBN 0-412-04231-2.

Timothy Hanson and Wesley O. Johnson. Modeling regression error with a mixture of Polya trees. *J. Amer. Statist. Assoc.*, 97(460):1020–1033, 2002. ISSN 0162-1459. doi: 10.1198/016214502388618843. URL http://dx.doi.org/10.1198/016214502388618843.

Xuming He, Pin Ng, and Stephen Portnoy. Bivariate quantile smoothing splines. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(3):537–550, 1998. ISSN 1369-7412. doi: 10.1111/1467-9868.00138. URL http://dx.doi.org/10.1111/1467-9868.00138.

Patrick J. Heagerty. Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*, 55(3):pp. 688–698, 1999. ISSN 0006341X. URL http://www.jstor.org/stable/2533591.

Nils Lid Hjort and Sonia Petrone. 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, 2007.

Nils Lid Hjort and Stephen G. Walker. Quantile pyramids for Bayesian nonparametrics. *Ann. Statist.*, 37(1):105–131, 2009. ISSN 0090-5364. doi: 10.1214/07-AOS553. URL http://dx.doi.org/10.1214/07-AOS553.

Roger Koenker. *Quantile regression*, volume 38 of *Econometric Society Monographs*. Cambridge University Press, Cambridge, 2005. ISBN 978-0-521-60827-5; 0-521-60827-9. doi: 10.1017/CBO9780511754098. URL http://dx.doi.org/10.1017/CBO9780511754098.

Roger Koenker. *quantreg: Quantile Regression*, 2012. URL http://CRAN.R-project.org/package=quantreg. R package version 4.91.

Roger Koenker and Jr. Bassett, Gilbert. Regression quantiles. *Econometrica*, 46(1):pp. 33–50, 1978. ISSN 00129682. URL http://www.jstor.org/stable/1913643.

Roger Koenker and José A. F. Machado. Goodness of fit and related inference processes for quantile regression. *J. Amer. Statist. Assoc.*, 94(448):1296–1310, 1999. ISSN 0162-1459. doi: 10.2307/2669943. URL http://dx.doi.org/10.2307/2669943.

Athanasios Kottas and Alan E. Gelfand. Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, 96(456):pp. 1458–1468, 2001. ISSN 01621459. URL http://www.jstor.org/stable/3085913.

Athanasios Kottas and Milovan Krnjajić. Bayesian semiparametric modelling in quantile regression. *Scand. J. Stat.*, 36(2):297–319, 2009. ISSN 0303-6898. doi: 10.1111/j.1467-9469.2008.00626.x. URL http://dx.doi.org/10.1111/j.1467-9469.2008.00626.x.

Roderick J. A. Little. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134, 1993. doi: 10.1080/01621459.1993.10594302. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1993.10594302.

Roderick J. A. Little. A class of pattern-mixture models for normal incomplete data. *Biometrika*, 81(3):471–483, 1994. ISSN 0006-3444. doi: 10.1093/biomet/81.3.471. URL http://dx.doi.org/10.1093/biomet/81.3.471.

G. Molenberghs, B. Michiels, M. G. Kenward, and P. J. Diggle. Monotone missing data and pattern-mixture models. *Statist. Neerlandica*, 52(2):153–161, 1998. ISSN 0039-0402. doi: 10.1111/1467-9574.00075. URL http://dx.doi.org/10.1111/1467-9574.00075.

Panel on Handling Missing Data in Clinical Trials; National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. The National Academies Press, 2010. ISBN 9780309158145. URL http://www.nap.edu/openbook.php?record_id=12955.

Michael G Perri, Marian C Limacher, Patricia E Durning, David M Janicke, Lesley D Lutes, Linda B Bobroff, Martha Sue Dale, Michael J Daniels, Tiffany A Radcliff, and A Daniel Martin. 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(21):2347, 2008.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/. ISBN 3-900051-07-0.

Brian J. Reich, Howard D. Bondell, and Huixia J. Wang. Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics*, 11(2):337–352, 2010. doi: 10.1093/biostatistics/kxp049. URL http://biostatistics.oxfordjournals.org/content/11/2/337.abstract.

Martin Riedmiller and Heinrich Braun. A direct adaptive method for faster backpropagation learning: the rprop algorithm. In *Neural Networks, 1993., IEEE International Conference on*, pages 586–591 vol.1. IEEE, 1993. doi: 10.1109/ICNN.1993.298623.

Jason Roy. Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics*, 59(4):829–836, 2003. ISSN 0006-341X. doi: 10.1111/j.0006-341X.2003.00097.x. URL http://dx.doi.org/10.1111/j.0006-341X.2003.00097.x.

Jason Roy and Michael J. Daniels. A general class of pattern mixture models for nonignorable dropout with many possible dropout times. *Biometrics*, 64(2):538–545, 668, 2008. ISSN 0006-341X. doi: 10.1111/j.1541-0420.2007.00884.x. URL http://dx.doi.org/10.1111/j.1541-0420.2007.00884.x.

Donald B. Rubin. Formalizing subjective notions about the effect of nonrespondents in sample surveys. *J. Amer. Statist. Assoc.*, 72(359):538–543, 1977. ISSN 0162-1459.

Thomas A Wadden, Robert I Berkowitz, David B Sarwer, Richard Prus-Wisniewski, and Carrie Steinberg. Benefits of lifestyle modification in the pharmacologic treatment of obesity: a randomized trial. *Archives of Internal Medicine*, 161(2):218–227, 2001. doi: 10.1001/archinte.161.2.218.

Stephen Walker and Bani K. Mallick. A Bayesian semiparametric accelerated failure time model. *Biometrics*, 55(2):477–483, 1999. ISSN 0006-341X. doi: 10.1111/j.0006-341X.1999.00477.x. URL http://dx.doi.org/10.1111/j.0006-341X.1999.00477.x.

Chenguang Wang and Michael J. Daniels. A note on MAR, identifying restrictions, model comparison, and sensitivity analysis in pattern mixture models with and without covariates for incomplete data. *Biometrics*, 67(3):810–818, 2011. ISSN 0006-341X. doi: 10.1111/j.1541-0420.2011.01565.x. URL http://dx.doi.org/10.1111/j.1541-0420.2011.01565.x.

Ying Wei, Anneli Pere, Roger Koenker, and Xuming He. Quantile regression methods for reference growth charts. *Stat. Med.*, 25(8):1369–1382, 2006. ISSN 0277-6715. doi: 10.1002/sim.2271. URL http://dx.doi.org/10.1002/sim.2271.

Ying Wei, Yanyuan Ma, and Raymond J. Carroll. Multiple imputation in quantile regression. *Biometrika*, 99(2):423–438, 2012. ISSN 0006-3444. doi: 10.1093/biomet/ass007. URL http://dx.doi.org/10.1093/biomet/ass007.

Keming Yu and Rana A. Moyeed. Bayesian quantile regression. *Statist. Probab. Lett.*, 54(4): 437–447, 2001. ISSN 0167-7152. doi: 10.1016/S0167-7152(01)00124-9. URL http://dx.doi.org/10.1016/S0167-7152(01)00124-9.

Keming Yu, Zudi Lu, and Julian Stander. Quantile regression: applications and current research areas. *The Statistician*, 52(3):331–350, 2003. ISSN 0039-0526. doi: 10.1111/1467-9884.00363. URL http://dx.doi.org/10.1111/1467-9884.00363.

Ying Yuan and Guosheng Yin. Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics*, 66(1):105–114, 2010. ISSN 0006-341X. doi: 10.1111/j.1541-0420.2009.01269.x. URL http://dx.doi.org/10.1111/j.1541-0420.2009.01269.x.

A Identifiability

First suppose y is univariate and there are two patterns $R = 1$ and $R = 0$.

Before going forward to quantile regression, first we consider identifiability problem in mean regression.

Consider a pattern mixture model:

$$\begin{aligned} Y|R = 1 &\sim N(\Delta + \mu_1, \sigma_1), \\ Y|R = 0 &\sim N(\Delta + \mu_0, \sigma_0), \\ \Pr(R = 1) &= \pi, \\ E(y) &= \theta. \end{aligned} \tag{9}$$

Thus by iterated expectation, we have

$$\begin{aligned} \theta &= \Delta + \mu_1\pi + \mu_0(1 - \pi), \\ \Delta &= \theta - \pi\mu_1 - (1 - \pi)\mu_0. \end{aligned}$$

We can see Δ is determined by θ, μ_1, μ_0. Plugging in (9), we have

$$\begin{aligned} Y|R = 1 &\sim N(\theta + (1 - \pi)\mu_1 - (1 - \pi)\mu_0, \sigma_1), \\ Y|R = 0 &\sim N(\theta - \pi\mu_1 + \pi\mu_0, \sigma_0). \end{aligned}$$

Denote $\xi_1 = (\theta, \mu_1, \mu_0)$, and if $\xi_2 = (\theta, \mu_1 + c, \mu_0 + c)$, both groups of parameters lead to the same distribution of $\mathrm{p}(y, R) = \mathrm{p}(y|R)\,\mathrm{p}(R)$. Therefore, ξ is not identifiable. If we put constraints on μ_1 and μ_0, for example $\mu_0 = 0$, then

$$\begin{aligned} Y|R = 1 &\sim N(\theta + \mu_1, \sigma_1), \\ Y|R = 0 &\sim N(\theta, \sigma_0). \end{aligned}$$

Thus $\boldsymbol{\xi} = (\theta, \mu_1)$ is identifiable. If $\xi_2 \neq \xi_1$, then $\mathrm{p}_2(y, R) \neq \mathrm{p}_1(y, R)$.

Secondly, we consider quantile regression for a pattern mixture model:

$$\begin{aligned}
Y|R = 1 &\sim N(\Delta + \mu_1, \sigma_1), \\
Y|R = 0 &\sim N(\Delta + \mu_0, \sigma_0), \\
\Pr(R = 1) &= \pi, \\
\mathrm{p}(Y \leq \theta) &= \tau,
\end{aligned}$$

where θ is the quantile estimate of interest. We again show $\boldsymbol{\xi} = (\theta, \mu_1, \mu_0)$ is not identifiable.

Again by iterated expectations, we have

$$\tau = \pi\Phi\left(\frac{\theta - \Delta - \mu_1}{\sigma_1}\right) + (1 - \pi)\Phi\left(\frac{\theta - \Delta - \mu_0}{\sigma_0}\right).$$

Thus Δ is again determined by the other parameters:

$$\Delta = h(\theta, \mu_1, \mu_0, \sigma_1, \sigma_0, \pi, \tau).$$

To show $\boldsymbol{\xi} = (\theta, \mu_1, \mu_0, \sigma_1, \sigma_0)$ is not identifiable, we need to find $\boldsymbol{\xi}' \neq \boldsymbol{\xi}$, such that $\mathrm{p}(y|R) = \mathrm{p}'(y|R)$. If the last equation holds, then we must have $\sigma_1' = \sigma_1, \sigma_0' = \sigma_0$, thus we still need to find θ', μ_1', μ_0' such that

$$\begin{aligned}
h(\boldsymbol{\xi}) + \mu_1 &= h(\boldsymbol{\xi}') + \mu_1', \\
h(\boldsymbol{\xi}) + \mu_0 &= h(\boldsymbol{\xi}') + \mu_0'.
\end{aligned}$$

By substracting previous equations, we have $\mu_1' - \mu_0' = \mu_1 - \mu_0$. Denote $\mu_1' = \mu_1 + \delta$ and $\mu_0' = \mu_0 + \delta$, and let $\theta' = \theta$ such that

$$\Delta' = h(\theta', \mu_1, \mu_0, \sigma_1, \sigma_0, \delta) = h(\boldsymbol{\xi}) - \delta = \Delta - \delta.$$

Then the new parameter $\boldsymbol{\xi}'$ yields the same distribution as $\boldsymbol{\xi}$. Therefore $\boldsymbol{\xi}$ is not identifiable.

If we use a constraint, for example $\mu_1 = -\mu_0$, then $\mathrm{p}(y|R; \boldsymbol{\xi}) = \mathrm{p}(y|R; \boldsymbol{\xi}')$ yields $\boldsymbol{\xi} = \boldsymbol{\xi}'$.

Now consider the situation with covariates. Suppose the model is

$$\begin{aligned}
Y|R = 1, x &\sim N(\Delta + \mu_1 + \beta_{x1}x, \sigma_1), \\
Y|R = 0, x &\sim N(\Delta - \mu_1 + \beta_{x0}x, \sigma_0), \\
\Pr(R = 1) &= \pi, \\
\mathrm{p}(Y \leq \gamma_0 + \gamma_1 x) &= \tau.
\end{aligned}$$

Δ can still be determined by

$$\Delta = h(x, \gamma_0, \gamma_1, \mu_1, \beta_{x1}, \beta_{x0}, \sigma_1, \sigma_0, \pi, \tau).$$

We want to show the parameter $\boldsymbol{\xi} = (\gamma_0, \gamma_1, \mu_1, \beta_{x1}, \beta_{x0}, \sigma_1, \sigma_0, \pi)$ is not identifiable by finding $\boldsymbol{\xi}' \neq \boldsymbol{\xi}$, but $\mathrm{p}(y|R; \boldsymbol{\xi}) = \mathrm{p}(y|R; \boldsymbol{\xi}')$. If the last equation holds, we have $\sigma_1' = \sigma_1, \sigma_0' = \sigma_0$, and to equate the two means, we have

$$\begin{aligned}
\Delta + \mu_1 + \beta_{x1}x &= \Delta' + \mu_1' + \beta_{x1}'x, \\
\Delta - \mu_1 + \beta_{x0}x &= \Delta' - \mu_1' + \beta_{x0}'x.
\end{aligned}$$

By substracting the two equations, we have

$$2\mu_1 + (\beta_{x1} - \beta_{x0})x = 2\mu_1^{'} + (\beta_{x1}^{'} - \beta_{x0}^{'})x,$$

which holds for all x. Thus $\mu_1 = \mu_1^{'}$ and $(\beta_{x1} - \beta_{x0}) = (\beta_{x1}^{'} - \beta_{x0}^{'})$. Then let

$$\beta_{x1}^{'} = \beta_{x1} + \delta,$$
$$\beta_{x0}^{'} = \beta_{x0} + \delta,$$

and keep all the other parameters in $\xi^{'}$ the same. We can still have the same distribution of $y|R;\xi$ but with different ξ. Therefore, ξ is not identifiable One solution is to restrict $\beta_{x1} = -\beta_{x0}$ or $\beta_{x1} = 0$.

Now consider the bivariate (y_1, y_2) case, and we focus on the identifiability issue especially $y_2|y_1$. Suppose the model is

$$Y_2|y_1, x, R = 1 \sim N(\Delta + \mu_1 + x\beta_{x1} + \beta_{11}y_1, \sigma_1),$$
$$Y_2|y_1, x, R = 0 \sim N(\Delta - \mu_1 - x\beta_{x1} + \beta_{10}y_1, \sigma_0).$$

Here R stands for two different patterns, and missingness is not considered.

Regarding the identifiability of β_{11} and β_{10}, assume there exists $\beta_{11}^{'}$ and $\beta_{10}^{'}$, such that

$$\Delta + \mu_1 + x\beta_x + \beta_{11}y_1 = \Delta^{'} + \mu_1^{'} + x\beta_x^{'} + \beta_{11}^{'}y_1,$$
$$\Delta - \mu_1 - x\beta_x + \beta_{10}y_1 = \Delta^{'} - \mu_1^{'} - x\beta_x^{'} + \beta_{10}^{'}y_1.$$

By substracting two equations, we have $\mu_1 = \mu_1^{'}$ and $\beta_x = \beta_x^{'}$. Since Δ is determined by integrating out y_1, such that matching the two sides of the above equation for coefficient of y_1, we must have $\beta_{11} = \beta_{11}^{'}$ and $\beta_{10} = \beta_{10}^{'}$, therefore, ξ is identifiable.

For identifiability issue with the heterogeneous model described in section 2.1, it is easy to show there is no trouble with the heterogeneity parameters α, analogous to the linear model case. For the other parameters, it can be found similar to the above development.

B Proof of Lemma 2.1

- Denote
$$I(a,b) = \int \Phi\left(\frac{x-b}{a}\right)\phi(x)dx,$$
where Φ is the standard normal cdf and ϕ is the standard normal pdf and $a > 0$.
$$\begin{aligned}\frac{\partial I(a,b)}{\partial b} &= -\frac{1}{a}\int \phi\left(\frac{x-b}{a}\right)\phi(x)dx \\ &= -\frac{1}{\sqrt{2\pi}\sqrt{a^2+1}}\exp\left(-\frac{b^2}{2(a^2+1)}\right) \\ &= -\frac{1}{\sqrt{a^2+1}}\phi\left(\frac{b}{\sqrt{a^2+1}}\right).\end{aligned}$$

Since $I(a,\infty)=0$,

$$\begin{aligned} I(a,b) &= -\frac{1}{\sqrt{a^2+1}}\int_b^\infty \phi\left(\frac{s}{\sqrt{a^2+1}}\right)ds \\ &= \int_{b/\sqrt{a^2+1}}^\infty \phi(t)dt \\ &= 1-\Phi(b/\sqrt{a^2+1}). \end{aligned} \tag{10}$$

For $a<0$,

$$\begin{aligned} \frac{\partial I(a,b)}{\partial b} &= -\frac{1}{a}\int \phi\left(\frac{x-b}{a}\right)\phi(x)dx \\ &= -\frac{sgn(a)}{\sqrt{2\pi}\sqrt{a^2+1}}\exp\left(-\frac{b^2}{2(a^2+1)}\right) \\ &= -\frac{sgn(a)}{\sqrt{a^2+1}}\phi\left(\frac{b}{\sqrt{a^2+1}}\right). \end{aligned}$$

Since $I(a,-\infty)=0$:

$$\begin{aligned} I(a,b) &= \int_{-\infty}^{b/\sqrt{a^2+1}} \phi(t)dt \\ &= \Phi(b/\sqrt{a^2+1}). \end{aligned} \tag{11}$$

- For integrating over a normal distribution with mean μ and standard deviation σ:

$$\begin{aligned} \int \Phi(x)d\Phi(x;\mu,\sigma) &= \int \Phi(x)\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)dx \\ &= \int \Phi(\sigma t+\mu)\phi(t)dt \\ &= 1-\Phi(-\mu/\sigma/\sqrt{1/\sigma^2+1}). \end{aligned}$$

The last equation holds by (10)

- For integrating a $\mathrm{N}(b,a)$ CDF over another normal distribution ($\mathrm{N}(\mu,\sigma)$):

$$\begin{aligned} \int \Phi\left(\frac{x-b}{a}\right)d\Phi(x;\mu,\sigma) &= \int \Phi\left(\frac{x-b}{a}\right)\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)dx \\ &= \int \Phi\left(\frac{\sigma y+\mu-b}{a}\right)\phi(y)dy \\ &= 1-\Phi\left(\frac{b-\mu}{\sigma}/\sqrt{\frac{a^2}{\sigma^2}+1}\right). \end{aligned} \tag{12}$$

If $a<0$,

$$\int \Phi\left(\frac{x-b}{a}\right)d\Phi(x;\mu,\sigma) = \Phi\left(\frac{b-\mu}{\sigma}/\sqrt{\frac{a^2}{\sigma^2}+1}\right). \tag{13}$$