Quantile Regression with Monotone Dropout Missingness

April 27, 2013

1 Introduction

The structure of this article is as below: first, we introduce our quantile regression methods to deal with monotone dropout in bivariate normal mixed model in section 2. Then in section 3, we generalize our method to trivariate case. Section 4 describes our proposed method in generalized cases. Computation details are presented in section 2.3. And we also include a simulation study to demonstrate the performance of our algorithm in section 5.

2 Bivariate Normal Mixed Model with Monotone Dropout

In this section, we first introduce the bivariate scenario with monotone dropout, then describe our proposed quantile regression model in section 2.1. And we will give more details on deploying MAR and MNAR and computation in section 2.2 and section 2.3.

Let $\boldsymbol{Y} = (Y_1, Y_2)^T$ and $R = I\{Y_2 \text{ observed}\}$. Suppose we are interested in the τ-th marginalized quantile regression coefficients $\boldsymbol{\gamma} = (\gamma_1^\tau, \gamma_2^\tau)$, where

$$\gamma_1^\tau = (\gamma_{01}^\tau, \gamma_{11}^\tau),$$
$$\gamma_2^\tau = (\gamma_{02}^\tau, \gamma_{12}^\tau),$$

such that

$$\mathrm{p}(Y_{i1} \leq \gamma_{01}^\tau + x_i\gamma_{11}^\tau | x_i) = \tau, \tag{1}$$
$$\mathrm{p}(Y_{i2} \leq \gamma_{02}^\tau + x_i\gamma_{12}^\tau | x_i) = \tau. \tag{2}$$

Here we assume there is only one single covariate x_i for each subject i, which is constant across treatments or time points, while the quantile regression coefficients $\boldsymbol{\gamma}^\tau$ are time varying.

2.1 Model Setting

To adopt mixed model to deal with missingness, we assume $\boldsymbol{Y}$ follows bivariate normal distribution:

$$\boldsymbol{Y}|R = r \sim \mathrm{N}(\boldsymbol{\mu}^{(r)}, \Sigma^{(r)}), r = 0, 1,$$
$$R \sim \mathrm{Ber}(\phi).$$

Reparametrize the above model as

$$Y_{i1}|R=r,x_i \sim \mathrm{N}(\Delta_{i1}+\beta_{01}^{(r)}+x_i\beta_{11}^{(r)},\sigma_{11}^{(r)}), \tag{3}$$
$$Y_{i2}|R=r,x_i,y_{i1} \sim \mathrm{N}(\Delta_{i2}+\beta_{02}^{(r)}+x_i\beta_{12}^{(r)}+y_{i1}\beta_{22}^{(r)},\sigma_{2|1}^{(r)}). \tag{4}$$

where the quantity $\Delta_{it}, t=1,2$ is determined by other parameters in the model and can be solved from

$$\tau=\sum_{r=0}^{1}\mathrm{p}(Y_{i1}\leq\gamma_{01}^{\tau}+x_i\gamma_{11}^{\tau}|x_i,R=r)\,\mathrm{p}(R=r), \tag{5}$$
$$\tau=\sum_{r=0}^{1}\mathrm{p}(Y_{i2}\leq\gamma_{02}^{\tau}+x_i\gamma_{12}^{\tau}|x_i,R=r)\,\mathrm{p}(R=r). \tag{6}$$

Specific derivation can be found in section 2.3.

In this section we consider $\mathrm{p}(R=1|x_i)=\mathrm{p}(R=1)=\pi=1-\mathrm{p}(R=0)$, which means missingness does not depend on covariates (missingness depending on covariates could be a further topic in future research).

Meanwhile, for identifiability issue in equation (3) and (4) , we put constraints on the β parameters:

$$\beta_{01}^{(1)}+\beta_{01}^{(0)}=0, \tag{7}$$
$$\beta_{11}^{(1)}+\beta_{11}^{(0)}=0, \tag{8}$$
$$\beta_{02}^{(1)}+\beta_{02}^{(0)}=0, \tag{9}$$
$$\beta_{12}^{(1)}+\beta_{12}^{(0)}=0, \tag{10}$$
$$\beta_{22}^{(1)}+\beta_{22}^{(0)}=0. \tag{11}$$

2.2 Sensitivity Analysis

For identifiability issue in equation (4), we put constraints (9 - 11), thus we denote

$$\beta_{02}^{(1)}=-\beta_{02}^{(0)}, \tag{12}$$
$$\beta_{12}^{(1)}=-\beta_{12}^{(0)}, \tag{13}$$
$$\beta_{22}^{(2)}=-\beta_{22}^{(0)}, \tag{14}$$
$$\sigma_{2|1}^{(0)}=\lambda\sigma_{2|1}^{(1)}. \tag{15}$$

And when $R=0$, $Y_{i2}|R=0$ is not observed, $\boldsymbol{\xi}_s=(\beta_{02}^{(0)},\beta_{12}^{(0)},\beta_{22}^{(0)},\lambda)$ could be a group of sensitivity parameters. We illustrate how to deploy MAR and MNAR assumption from both frequentist way and Bayesian framework.

- **Frequentist way:**
  When $\boldsymbol{\xi}_s=\boldsymbol{\xi}_{s0}=(0,0,0,1)$, it yields $p(Y_{i2}|y_{i1},x_i,R=1)=p(Y_{i2}|y_{i1},x_i,R=0)$, where MAR condition satisfies. If $\boldsymbol{\xi}_s$ is fixed at $\boldsymbol{\xi}_s\neq\boldsymbol{\xi}_{s0}$, then $\mathrm{p}(R|Y_{i1},Y_{i2})\neq\mathrm{p}(R|Y_{i1})$, thus the missing mechanism is missing not at random.

- **Bayesian Framework:**

  We put priors on $(\boldsymbol{\xi}_s, \boldsymbol{\xi}_m)$ ($\boldsymbol{\xi}_m$ are identifiable parameters) as :

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

  If we assume MAR with no uncertainty, the prior of $\boldsymbol{\xi}_s$ is p$(\boldsymbol{\xi}_s = (0,0,0,1)) = 1$. Sensitivity analysis can be executed through putting a set of priors on $\boldsymbol{\xi}_s$ to check 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 $\mathrm{E}(\boldsymbol{\xi}_s) = (0,0,0,1)$ 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 $\Delta_{\xi} \neq (0,0,0,1)$ 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 $\Delta_{\xi} \neq (0,0,0,1)$ and $\mathrm{Var}(\boldsymbol{\xi}_s) \neq \mathbf{0}$.

2.3 Computation

In section 2.3.1 , we give details on how to calculate Δ_{it} in model (3, 4) for $t = 1,2$. Then we illustrate how to get maximum likelihood estimator using gradient descent algorithm in section 2.3.2. Last, we present the Bayesian sampling procedure in section 2.3.3.

2.3.1 Δ Calculation

First we illustrate how to calculate Δ_{i1}, Δ_{i2} given all the other parameters $\boldsymbol{\xi} = (\boldsymbol{\xi}_m, \xi_s) = (\gamma_{01}, \gamma_{11}, \gamma_{02}, \gamma_{12}, R, x_i, \beta_{01}^{(1)}, \beta_{11}^{(1)}, \beta_{02}^{(0)}, \beta_{12}^{(0)}, \beta_{22}^{(0)}, \sigma_{11}^{(1)}, \sigma_{11}^{(0)}, \sigma_{2|1}^{(1)}, \lambda, \phi)$.

- Δ_{i1} : Expand equation (5) with constraints (7, 8):

  $$\begin{aligned}\tau &= \mathrm{p}(R=1|x_i)\,\mathrm{p}(Y_{i1} \le \gamma_{01}^{\tau} + x_i\gamma_{11}^{\tau}|x_i, R=1) + \mathrm{p}(R=0|x_i)\,\mathrm{p}(Y_{i1} \le \gamma_{01}^{\tau} + x_i\gamma_{11}^{\tau}|x_i, R=0)\\ &= \pi F_{i1}^{(1)}(\gamma_{01}^{\tau} + x_i\gamma_{11}^{\tau}; \Delta_{i1}, \beta_{01}^{(1)}, \beta_{11}^{(1)}, x_i, \sigma_{11}^{(1)}) + (1-\pi)F_{i1}^{(0)}(\gamma_{01}^{\tau} + x_i\gamma_{11}^{\tau}; \Delta_{i1}, \beta_{01}^{(0)}, \beta_{11}^{(0)}, x_i, \sigma_{11}^{(0)}),\\ &= \pi\Phi\left(\frac{\gamma_{01} + x_i\gamma_{11} - \Delta_{i1} - \beta_{01}^{(1)} - x_i\beta_{11}^{(1)}}{\sigma_{11}^{(1)}}\right) + (1-\pi)\Phi\left(\frac{\gamma_{01} + x_i\gamma_{11} - \Delta_{i1} + \beta_{01}^{(1)} + x_i\beta_{11}^{(1)}}{\sigma_{11}^{(0)}}\right),\end{aligned} \tag{16}$$

  where Φ is the standard normal CDF. Because equation (16) is continuous and monotone on Δ_{i1}, it can be solved by standard numerical root-find method without much difficulty, for example, the bisection method.

- Δ_{i2} : First we introduce a lemma:

  Lemma 2.1. *An integral of a normal CDF over a non-standard normal distribution can be simplified to a closed form in terms of another normal CDF:*

  $$\int \Phi\left(\frac{x-b}{a}\right) d\Phi(x;\mu,\sigma) = \begin{cases} 1 - \Phi\left(\frac{b-\mu}{\sigma} \Big/ \sqrt{\frac{a^2}{\sigma^2}+1}\right) & a > 0,\\ \Phi\left(\frac{b-\mu}{\sigma} \Big/ \sqrt{\frac{a^2}{\sigma^2}+1}\right) & a < 0,\end{cases} \tag{17}$$

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

Proof of lemma can be seen in appendix A.

Hence, expand equation (6) with constraints (9 to 11) similarly to get

$$
\begin{aligned}
\tau &= \mathrm{p}(R=1)\,\mathrm{p}(Y_{i2} \leq \gamma_{02}^{\tau} + x_i\gamma_{12}^{\tau}|x_i, R=1) + \mathrm{p}(R=0)\,\mathrm{p}(Y_{i2} \leq \gamma_{02}^{\tau} + x_i\gamma_{12}^{\tau}|x_i, R=0) \\
&= \pi \int \mathrm{p}(y_{i2} \leq \gamma_{02}^{\tau} + x_i\gamma_{12}^{\tau}|x_i, y_{i1}, R=1) dF_{i1}^{(1)}(y_{i1}; \Delta_{i1}, x_i, \beta_{*1}^{(1)}, \sigma_{11}^{(1)}) \\
&\quad + (1-\pi) \int \mathrm{p}(y_{i2} \leq \gamma_{02}^{\tau} + x_i\gamma_{12}^{\tau}|x_i, y_{i1}, R=0) dF_{i1}^{(0)}(y_{i1}; \Delta_{i1}, x_i, \beta_{*1}^{(0)}, \sigma_{11}^{(0)}) \\
&= \pi \int F_{i2|1}^{(1)}(\gamma_{02}^{\tau} + x_i\gamma_{12}^{\tau}; \Delta_{i2}, \beta_{*2}^{(1)}, \sigma_{2|1}^{(1)}, x_i, y_{i1}) dF_{i1}^{(1)}(y_{i1}; \Delta_{i1}, x_i, \beta_{*1}^{(1)}, \sigma_{11}^{(1)}) \\
&\quad + (1-\pi) \int F_{i2|1}^{(0)}(\gamma_{02}^{\tau} + x_i\gamma_{12}^{\tau}; \Delta_{i2}, \beta_{*2}^{(0)}, \sigma_{2|1}^{(0)}, x_i, y_{i1}) dF_{i1}^{(0)}(y_{i1}; \Delta_{i1}, x_i, \beta_{*1}^{(0)}, \sigma_{11}^{(0)}), \\
&= \pi \int \Phi\left(\frac{\gamma_{02} + x_i\gamma_{12} - \Delta_{i2} + \beta_{02}^{(0)} + x_i\beta_{12}^{(0)} + y_{i1}\beta_{22}^{(0)}}{\sigma_{2|1}^{(1)}}\right) dF_{i1}^{(1)}(y_{i1}; \Delta_{i1}, x_i, \beta_{*1}^{(1)}, \sigma_{11}^{(1)}) \\
&\quad + (1-\pi) \int \Phi\left(\frac{\gamma_{02} + x_i\gamma_{12} - \Delta_{i2} - \beta_{02}^{(0)} - x_i\beta_{12}^{(0)} - y_{i1}\beta_{22}^{(0)}}{\lambda\sigma_{2|1}^{(1)}}\right) dF_{i1}^{(0)}(y_{i1}; \Delta_{i1}, x_i, \beta_{*1}^{(0)}, \sigma_{11}^{(0)}).
\end{aligned}
\tag{18}
$$

By notations in (12 to 15) and lemma (17), the integral on the right hand side of above equation (18) can be simplified as

$$
\begin{aligned}
&\int \Phi\left(\frac{\gamma_{02} + x_i\gamma_{12} - \Delta_{i2} + \beta_{02}^{(0)} + x_i\beta_{12}^{(0)} + y_{i1}\beta_{22}^{(0)}}{\sigma_{2|1}^{(1)}}\right) dF_{i1}^{(1)}(y_{i1}; \Delta_{i1}, x_i, \beta_{*1}^{(1)}, \sigma_{11}^{(1)}) \\
&= \int \Phi\left(\frac{y_{i1} - (\Delta_{i2} - \gamma_{02} - x\gamma_{12} - \beta_{02}^{(0)} - x\beta_{12}^{(0)})/\beta_{22}^{(0)}}{\sigma_{2|1}/\beta_{22}^{(0)}}\right) dF_{i1}^{(1)}(y_{i1}; \Delta_{i1} + \beta_{01}^{(1)} + x\beta_{11}^{(1)}, \sigma_{11}^{(1)}) \\
&= \begin{cases}
1 - \Phi\left(\frac{(\Delta_{i2} - \gamma_{02} - x\gamma_{12} - \beta_{02}^{(0)} - x\beta_{12}^{(0)})/\beta_{22}^{(0)} - (\Delta_{i1} + \beta_{01}^{(1)} + x\beta_{11}^{(1)})}{\sigma_{11}^{(1)}} \Big/ \sqrt{\frac{\sigma_{2|1}^2}{\beta_{22}^{(0)2}\sigma_{11}^{(1)2}} + 1}\right) & \beta_{22}^{(0)} > 0 \\
\Phi\left(\frac{(\Delta_{i2} - \gamma_{02} - x\gamma_{12} - \beta_{02}^{(0)} - x\beta_{12}^{(0)})/\beta_{22}^{(0)} - (\Delta_{i1} + \beta_{01}^{(1)} + x\beta_{11}^{(1)})}{\sigma_{11}^{(1)}} \Big/ \sqrt{\frac{\sigma_{2|1}^2}{\beta_{22}^{(0)2}\sigma_{11}^{(1)2}} + 1}\right) & \beta_{22}^{(0)} < 0
\end{cases}
\end{aligned}
$$

and

$$
\begin{aligned}
&\int \Phi\left(\frac{\gamma_{02}+x_i\gamma_{12}-\Delta_{i2}-\beta_{02}^{(0)}-x_i\beta_{12}^{(0)}-y_{i1}\beta_{22}^{(0)}}{\sigma_{2|1}^{(0)}}\right) dF_{i1}^{(0)}(y_{i1};\Delta_{i1},x_i,\beta_{*1}^{(0)},\sigma_{11}^{(0)})\\
&=\int \Phi\left(\frac{y_{i1}-(\gamma_{02}+x\gamma_{12}-\Delta_{i2}-\beta_{02}^{(0)}-x\beta_{12}^{(0)})/\beta_{22}^{(0)}}{\lambda\sigma_{2|1}/(-\beta_{22}^{(0)})}\right) dF_{i1}^{(0)}(y_{i1};\Delta_{i1}-\beta_{01}^{(1)}-x\beta_{11}^{(1)},\sigma_{11}^{(0)})\\
&=\begin{cases}\Phi\left(\frac{(\gamma_{02}+x\gamma_{12}-\Delta_{i2}-\beta_{02}^{(0)}-x\beta_{12}^{(0)})/\beta_{22}^{(0)}-(\Delta_{i1}-\beta_{01}^{(1)}-x\beta_{11}^{(1)})}{\sigma_{11}^{(0)}}\Big/\sqrt{\frac{\lambda^2\sigma_{2|1}^2}{\beta_{22}^{(0)2}\sigma_{11}^{(0)2}}+1}\right) & \beta_{22}^{(0)}>0\\
1-\Phi\left(\frac{(\gamma_{02}+x\gamma_{12}-\Delta_{i2}-\beta_{02}^{(0)}-x\beta_{12}^{(0)})/\beta_{22}^{(0)}-(\Delta_{i1}-\beta_{01}^{(1)}-x\beta_{11}^{(1)})}{\sigma_{11}^{(0)}}\Big/\sqrt{\frac{\lambda^2\sigma_{2|1}^2}{\beta_{22}^{(0)2}\sigma_{11}^{(0)2}}+1}\right) & \beta_{22}^{(0)}<0\end{cases}
\end{aligned}
$$

Therefore, Δ_{i2} can be solved through simplified equation (18) for each subject i.

For computation of higher order Δ_{ij} for $j >= 3$, we propose two approaches: (for convenience, we use Δ_j to save typing for subject i and denote S as the follow-up time.)

1. **Assume first order relationship:** We assume

$$
\mathrm{p}(Y_j|S,x,Y_{j-1},\ldots,Y_1)=\mathrm{p}(Y_j|S,x,Y_{j-1}) \tag{19}
$$

After obtaining Δ_{j-1}

$$
\begin{aligned}
\mathrm{p}(Y_j\le x\gamma|S=s)&=\int\ldots\int \mathrm{p}(Y_j\le x\gamma|S=s,x,Y_{j-1},\ldots,Y_1)dF(Y_{j-1}|S=s,Y_{j-2},\ldots,Y_1)\\
&\quad\cdots dF(Y_2|S=s,Y_1)\\
&=\int \mathrm{p}(Y_j\le x\gamma|S=s,x,Y_{j-1})dF(Y_{j-1}|S=s,Y_{j-2})
\end{aligned}
$$

Thus, only one integral is needed. Furthermore, by lemma (17), we can simplify above integral by closed form in terms of normal CDF.

2. **Recursive Computation:**

From equation (17), we can find after integral, the kernel part is still a normal cdf, but with other coefficients. So recursive simplification can be applied. Here we take trivariate case as an example:

For convenience, we only calculate $\mathrm{p}(Y_3\le \boldsymbol{x}\gamma|S=1)$, omit superscript and denote $\beta_{ij}^{(1)}$ as β_{ij} to save typing. The model we have is

$$
\begin{aligned}
Y_3|S=1,Y_2,Y_1&\sim \mathrm{N}(\Delta_3+\mu_3+y_2\beta_{23}+y_1\beta_{13},\tau_3),\\
Y_2|S=1,Y_1&\sim \mathrm{N}(\Delta_2+\mu_2+y_1\beta_{12},\tau_2),\\
Y_1|S=1&\sim \mathrm{N}(\Delta_1+\mu_1,\sigma).
\end{aligned}
$$

Assume $\beta_{23}, \beta_{13}, \beta_{12} > 0$, then

$$\begin{aligned}
\mathrm{p}(Y_3 \leq x\gamma | S = 1) &= \int\int \mathrm{p}(Y_3 \leq x\gamma | S = 1, y_2, y_1) f(y_2 | S = 1, y_1) f(y_1 | S = 1) dy_2 dy_1 \\
&= \int \left[\int \Phi\left(\frac{x\gamma - \Delta_3 - \mu_3 - y_2\beta_{23} - y_1\beta_{13}}{\tau_3} \right) f(y_2; \Delta_2 + \mu_2 + y_1\beta_{12}, \tau_2) dy_2 \right] \\
&\quad f(y_1 | S = 1) dy_1 \\
&= \int \left[\int \Phi\left(\frac{y_2 - (x\gamma - \Delta_3 - \mu_3 - y_1\beta_{13})/\beta_{23}}{-\tau_3/\beta_{23}} \right) f(y_2; \Delta_2 + \mu_2 + y_1\beta_{12}, \tau_2) dy_2 \right] \\
&\quad f(y_1 | S = 1) dy_1 \\
\text{by (45)} \quad &= \int \Phi\left(\frac{(x\gamma - \Delta_3 - \mu_3 - y_1\beta_{13})/\beta_{23} - \Delta_2 - \mu_2 - y_1\beta_{12}}{\tau_2\sqrt{\tau_3^2/\beta_{23}^2/\tau_2^2 + 1}} \right) f(y_1; \Delta_1 + \mu_1, \sigma) dy_1 \\
&= \int \Phi\left(\frac{y_1 - b^*}{a^*} \right) f(y_1; \mu^*, \sigma) dy_1.
\end{aligned}$$

where again the last equation has closed form from equation (17).

2.3.2 Maximum Likelihood Estimation

Denote the parameters $\boldsymbol{\xi} = (\boldsymbol{\xi}_m, \boldsymbol{\xi}_s)$, where $\boldsymbol{\xi}_m = (\gamma_{0i}^\tau, \gamma_{1i}^\tau, \beta_{11}^{(1)}, \beta_{11}^{(1)}, \sigma_{11}^{(1)}, \sigma_{11}^{(0)} \sigma_{2|1}^{(1)}, \phi), i = 1, 2$ are the identifiable parameters and $\boldsymbol{\xi}_s = (\beta_{02}^{(0)}, \beta_{12}^{(0)}, \beta_{22}^{(0)}, \lambda)$ are sensitivity parameters.

The contributed likelihood for observation $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})$ of subject i with missing indicator R is

$$L_i(Y_{i1}, Y_{i2}, R = 1 | \boldsymbol{\xi}, x_i) = \mathrm{p}(R = 1 | \phi)\, \mathrm{p}(Y_{i1} | R = 1, \gamma_{*1}^\tau, \beta_{*1}^{(1)}, \sigma_{11}^{(1)}, x_i)\, \mathrm{p}(Y_{i2} | R = 1, Y_{i1}, \gamma_{*2}^\tau, x_i, \beta_{*2}^{(0)}, \sigma_{2|1}^{(1)}) \tag{20}$$

$$L_i(Y_{i1}, R = 0 | \boldsymbol{\xi}, x_i) = \mathrm{p}(R = 0 | \phi)\, \mathrm{p}(Y_{i1} | R = 0, x_i, \gamma_{*1}^\tau, \beta_{*1}^{(1)}, \sigma_{11}^{(0)}) \tag{21}$$

One of the methods to get maximum likelihood estimates is by gradient descent algorithm. Denote $J(\boldsymbol{\xi}_m) = -\log L = -\log \sum_{i=1}^n L_i$. Then to maximize likelihood (20) and (21) is equivalent to minimize the target function $J(\boldsymbol{\xi}_m)$. Under MAR assumption, we fix $\boldsymbol{\xi}_s = (0, 0, 0, 1)$, while under MNAR assumption, one example of $\boldsymbol{\xi}_s$ is $(1, 0, 0, 1)$, assuming there is an intercept shift between the conditional distribution of $Y_{i2} | Y_{i1}, R$.

Maximization can be reached by the following procedures:

1. Initialize $\boldsymbol{\xi}$
2. calculate $\partial J(\boldsymbol{\xi})/\partial \xi_j$ for all j,
3. update $\boldsymbol{\xi}$ by $\xi_j = \xi_j - \alpha \partial J(\boldsymbol{\xi})/\partial \xi_j$ for all j
4. evaluate new $J(\boldsymbol{\xi})$
5. if the amount of descent of $J(\boldsymbol{\xi})$ is great than certain number, say 10^{-3}, then go back to step 2 and repeat. Otherwise, algorithm converges.

In previous algorithm, α controls the speed of convergence. If α is too large, $J(\boldsymbol{\xi})$ may be floating around the minimum and diverge. If α is too small, the convergence could be slow. But fixing at a small enough α, convergence is ensured, at least to a local minimum.

We can use numerical approximation to calculate $\partial J(\boldsymbol{\xi})/\partial \xi_j$ in algorithm step 2 by

$$\frac{\partial J(\boldsymbol{\xi})}{\partial \xi_1} \approx \frac{J(\xi_1+\epsilon, \xi_2, \ldots) - J(\xi_1-\epsilon, \xi_2, \ldots)}{2\epsilon},$$

for $j = 1$.

2.3.3 Bayesian Framework

Under Bayesian framework, we are going to put priors on the parameters $\boldsymbol{\xi}$ and make exact inference through posterior samples.

We can use the block Gibbs sampling method to draw sample from posterior distribution. Denote all the parameters (including sensitivity parameters) to sample as : (TODO: specifying priors)

$$\boldsymbol{\xi} = \left((\gamma_{01}, \gamma_{11}), (\gamma_{02}, \gamma_{12}), (\beta_{01}^{(1)}, \beta_{11}^{(1)}), (\beta_{02}^{(0)}, \beta_{12}^{(0)}, \beta_{22}^{(0)}), \sigma_{11}^{(1)}, \sigma_{11}^{(0)}, \sigma_{2|1}^{(1)}, \lambda, \phi \right).$$

Bracketed parameters are marked to sample as a block in block Gibbs sampling. All the procedures require Metropolis-Hasting algorithm to update samples since the likelihood is extremely complicated.

As mentioned in section 2.2, MAR or MNAR assumptions are adopted using different priors. For example, if MAR is assumed with no uncertainty, then $\beta_{02}^{(0)} = \beta_{12}^{(0)} = \beta_{22}^{(0)} = 0$ and $\lambda = 1$ with probability 1. Details for updating parameters are:

- $(\gamma_{01}, \gamma_{11})$: using Metropolis-Hasting algorithm.
  1. Draw $(\gamma_{01}^c, \gamma_{11}^c)$ candidates from candidate distribution
  2. Based on the new candidate parameter $\boldsymbol{\xi}^c$, calculate candidate Δ_{i1}^c from equation (16) for each subject i as we described in section 2.3.1. If $R = 1$ for corresponding subject i, update candidate Δ_{i2}^c as well. (For $R = 0$, only need to update Δ_{i1}^c).
  3. Plug in Δ_{i1}^c or $(\Delta_{i1}^c, \Delta_{i2}^c)$ in likelihood (21) or likelihood (20) for $R = 0$ or $R = 1$ to get candidate likelihood.
  4. Obtain Metropolis-Hasting ratio, move the chain and repeat.
- $(\gamma_{02}, \gamma_{12})$: similar algorithm but only update Δ_{i2} for subjects with $R = 1$.
- $(\beta_{01}^{(1)}, \beta_{11}^{(1)})$: similar to $(\gamma_{01}, \gamma_{11})$.
- For the rest of the parameters, algorithms for updating the samples are all similar to the first one.

3 Trivariate Normal Mixed Model under Ignorability with Monotone Dropout

Let $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3})$ and follow-up time $S = \{1, 2, 3\}$. Assuming covariates x_i are constant and marginalized quantile regression coefficients to be time varying $\gamma = (\gamma_1^\tau, \gamma_2^\tau, \gamma_3^\tau)$:

$$\begin{aligned}
\gamma_1^\tau &= (\gamma_{01}^\tau, \gamma_{11}^\tau), \\
\gamma_2^\tau &= (\gamma_{02}^\tau, \gamma_{12}^\tau), \\
\gamma_3^\tau &= (\gamma_{03}^\tau, \gamma_{13}^\tau).
\end{aligned}$$

Distribution of Y_{i1} follows :

$$\begin{aligned}
Y_{i1}|S = 1, x_i &\sim \mathrm{N}(\Delta_{i1} + \beta_{01}^{(1)} + x_i\beta_{11}^{(1)}, \sigma^{(1)}), && (22)\\
Y_{i1}|S = 2, x_i &\sim \mathrm{N}(\Delta_{i1} + \beta_{01}^{(2)} + x_i\beta_{11}^{(2)}, \sigma^{(2)}), && (23)\\
Y_{i1}|S = 3, x_i &\sim \mathrm{N}(\Delta_{i1} + \beta_{01}^{(3)} + x_i\beta_{11}^{(3)}, \sigma^{(3)}). && (24)\\
& && (25)
\end{aligned}$$

Note, for identifiability issue, we require constraints:

$$\begin{aligned}
\beta_{01}^{(1)} + \beta_{01}^{(2)} + \beta_{01}^{(3)} &= 0, && (26)\\
\beta_{11}^{(1)} + \beta_{11}^{(2)} + \beta_{11}^{(3)} &= 0. && (27)
\end{aligned}$$

Distribution of $Y_{i2}|Y_{i1}$:

$$\begin{aligned}
*Y_{i2}|y_{i1}, S = 1, x_i &\sim \mathrm{N}(\Delta_{i2} + \beta_{02}^{(1)} + x_i\beta_{12}^{(1)} + y_{i1}\beta_{22}^{(1)}, \tau_2^{(1)}), && (28)\\
Y_{i2}|y_{i1}, S \geq 2, x_i &\sim \mathrm{N}(\Delta_{i2} + \beta_{02}^{(\geq 2)} + x_i\beta_{12}^{(\geq 2)} + y_{i1}\beta_{22}^{(\geq 2)}, \tau_2^{(\geq 2)}). && (29)
\end{aligned}$$

$*$ means the observation Y_{i2} is not observed when the follow-up time is 1. Due to the same reason of identifiability, we need to apply constraints as:

$$\begin{aligned}
\beta_{02}^{(1)} + \beta_{02}^{(\geq 2)} &= 0, && (30)\\
\beta_{12}^{(1)} + \beta_{12}^{(\geq 2)} &= 0, && (31)\\
\beta_{22}^{(1)} + \beta_{22}^{(\geq 2)} &= 0. && (32)\\
& && (33)
\end{aligned}$$

Distribution of $Y_{i3}|Y_{i2}, Y_{i1}$:

$$\begin{aligned}
*Y_{i3}|y_{i2}, y_{i1}, S = 1, x_i &\sim \mathrm{N}(\Delta_{i3} + \beta_{03}^{(1)} + x_i\beta_{13}^{(1)} + y_{i1}\beta_{23}^{(1)} + y_{i2}\beta_{33}^{(1)}, \tau_3^{(1)}), && (34)\\
*Y_{i3}|y_{i2}, y_{i1}, S = 2, x_i &\sim \mathrm{N}(\Delta_{i3} + \beta_{03}^{(2)} + x_i\beta_{13}^{(2)} + y_{i1}\beta_{23}^{(2)} + y_{i2}\beta_{33}^{(2)}, \tau_3^{(2)}), && (35)\\
Y_{i3}|y_{i2}, y_{i1}, S = 3, x_i &\sim \mathrm{N}(\Delta_{i3} + \beta_{03}^{(3)} + x_i\beta_{13}^{(3)} + y_{i1}\beta_{23}^{(3)} + y_{i2}\beta_{33}^{(3)}, \tau_3^{(3)}). && (36)
\end{aligned}$$

$*$ means the observation Y_{i3} is not observed when the follow-up time is 1 or 2. Constraints are:

$$\beta_{03}^{(1)} + \beta_{03}^{(2)} + \beta_{03}^{(3)} = 0, \tag{37}$$
$$\beta_{13}^{(1)} + \beta_{13}^{(2)} + \beta_{13}^{(3)} = 0, \tag{38}$$
$$\beta_{23}^{(1)} + \beta_{23}^{(2)} + \beta_{23}^{(3)} = 0, \tag{39}$$
$$\beta_{33}^{(1)} + \beta_{33}^{(2)} + \beta_{33}^{(3)} = 0. \tag{40}$$

Sensitivity parameters are :

$$\boldsymbol{\xi}_{s2} = (\beta_{02}^{(1)}, \beta_{12}^{(1)}, \beta_{22}^{(1)}, \tau_2^{(1)}),$$
$$\boldsymbol{\xi}_{s3} = (\beta_{03}^{(1)}, \beta_{03}^{(2)}, \beta_{13}^{(1,2)}, \beta_{23}^{(1,2)}, \beta_{33}^{(1,2)}, \tau_2^{(1,2)})$$

While $\boldsymbol{\xi}_{s2} = (0, 0, 0, \tau_2^{(\geq 2)})$ and $\boldsymbol{\xi}_{s3} = (0, 0, 0, 0, 0, 0, 0, 0, \tau_3^{(3)}, \tau_3^{(3)})$ yield MAR.

4 TODO: Generalized multivariate quantile regression with monotone missingness

- Generalized J
- GLM: not only normal distribution, but all the exponential family as well
- Pattern mixture: group the patter
- $\exp(\lambda)$: consider MNAR for more general case, not only $\lambda\sigma_{2|1}^{(1)}$

4.1 General Case

We extend our mixture model to general monotone dropout scenario. Suppose the follow up time $S \in \{1, 2, \ldots, J\}$, denote response vector for subject i: $\boldsymbol{y} = (y_1, y_2, \ldots, y_J)$ and in the following context, we omit subject subscript i to save typing. In principle, $y_{ij}, \Delta_{ij}, x_{ij}$ can be changing with subject.

First we introduce some notations. Let

$$\mathrm{p}_k(y) = \mathrm{p}(y|S = k)$$
$$\mathrm{p}_{\geq k}(y) = \mathrm{p}(y|S \geq k).$$

be the density of response y given follow-up time $S = k$ and $S \geq k$. And $\Pr_k$ be the corresponding probability given $S = k$.

For follow-up time S, we assume:

$$\Pr(S = k) = \pi_k,$$
$$\sum_{k=1}^{J} \pi_k = 1.$$

Table 1: Table of conditional distribution for $J = 4$.

	$j=1$	$j=2$	$j=3$	$j=4$
$S=1$	$\mathrm{p}_1(y_1)$	$\mathrm{p}_1(y_2\|y_1)$	$\mathrm{p}_1(y_3\|y_1,y_2)$	$\mathrm{p}_1(y_4\|y_1,y_2,y_3)$
$S=2$	$\mathrm{p}_2(y_1)$	$\mathrm{p}_{\geq 2}(y_2\|y_1)$	$\mathrm{p}_2(y_3\|y_1,y_2)$	$\mathrm{p}_2(y_4\|y_1,y_2,y_3)$
$S=3$	$\mathrm{p}_3(y_1)$	$\mathrm{p}_{\geq 2}(y_2\|y_1)$	$\mathrm{p}_{\geq 3}(y_3\|y_1,y_2)$	$\mathrm{p}_3(y_4\|y_1,y_2,y_3)$
$S=4$	$\mathrm{p}_4(y_1)$	$\mathrm{p}_{\geq 2}(y_2\|y_1)$	$\mathrm{p}_{\geq 3}(y_3\|y_1,y_2)$	$\mathrm{p}_4(y_4\|y_1,y_2,y_3)$

For Y_1, suppose

$$\begin{aligned}
&\Pr(Y_1 \leq X\gamma_1) = \tau,\\
&\mathrm{p}_k(y_1) = \mathrm{p}(Y_1|S=k) \sim \mathrm{N}(\Delta_1 + \beta_{01}^{(k)} + \beta_{x1}^{(k)}x, \sigma_1^{(k)}), k = 1, \ldots, J,\\
&\sum_{k=1}^{J} \beta_{i1}^{(k)} = 0, i = 0, x.
\end{aligned}$$

The Δ_1 can be determined by

$$\tau = \Pr(Y_1 \leq X\gamma_1) = \sum_{k=1}^{J} \pi_k \Pr_k(Y_1 \leq X\gamma_1)$$

For $Y_j, j \geq 2$, we assume

$$\begin{aligned}
&\Pr(Y_j \leq X\gamma_j) = \tau,\\
&\mathrm{p}_k(y_j|y_1,\ldots,y_{j-1}) = \begin{cases} \mathrm{N}(\Delta_j + \beta_{0j}^{(k)} + \beta_{xj}^{(k)}x + \beta_{1j}^{(k)}y_1 \cdots + \beta_{(j-1)j}^{(k)}y_{j-1}, \sigma_j^{(k)}) & k < j;\\ \mathrm{N}(\Delta_j + \beta_{0j}^{(\geq k)} + \beta_{xj}^{(\geq k)}x + \beta_{1j}^{(\geq k)}y_1 \cdots + \beta_{(j-1)j}^{(\geq k)}y_{j-1}, \sigma_j^{(\geq k)}) & k \geq j; \end{cases}\\
&\sum_{k=1}^{j-1} \beta_{ij}^{(k)} + \beta_{ij}^{(\geq j)} = 0, i = 0, x.
\end{aligned}$$

By Molenberghs et al. (1998) theorem (MAR for discrete-time pattern mixture models under monotone dropout), MAR holds if and only if, for each $j \geq 2$ and $k < j$:

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

So MAR condition is satisfied by fixing $\beta_{ij}^{(k)} = \beta_{ij}^{(\geq j)} = 0$ for $i = 0, x; 1 \leq k < j$, and $\beta_{ij}^{(k)} = \beta_{ij}^{(\geq j)}$ for $i = 1, \ldots, j-1; 1 \leq k < j$.

Similarly, Δ_j can also be deterministic by

$$\begin{aligned}
\tau &= \Pr(Y_j \leq X\gamma_j) = \sum_{k=1}^{J} \pi_k \Pr_k(Y_j \leq X\gamma_j)\\
&= \sum_{k=1}^{J} \pi_k \int \cdots \int \Pr_k(Y_j \leq X\gamma_j|y_1,\ldots,y_{j-1})\, \mathrm{p}_k(y_{j-1}|y_1,\ldots,y_{j-2}) \cdots \mathrm{p}_k(y_2|y_1)\, \mathrm{p}_k(y_1) dy_{j-1} \cdots dy_1.
\end{aligned}$$

The contributed observed likelihood for $\boldsymbol{y} = (y_1, \ldots, y_k)$ given follow-up time $S = k$ under MAR assumption is

$$\begin{aligned}
\mathrm{p}(\boldsymbol{y}, S = k) &= \pi_k \, \mathrm{p}_k(y_k|y_1, \ldots, y_{k-1}) \, \mathrm{p}_k(y_{k-1}|y_1, \ldots, y_{k-2}) \cdots \mathrm{p}_k(y_1) \\
&= \pi_k \, \mathrm{p}_{\geq k}(y_k|y_1, \ldots, y_{k-1}) \, \mathrm{p}_{\geq k-1}(y_{k-1}|y_1, \ldots, y_{k-2}) \cdots \mathrm{p}_k(y_1)
\end{aligned}$$

5 Simulation

We proposed a simulation study to test performance of our algorithm. The simulation study included 1000 data sets. Each dataset consists 200 bivariate observations $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})$ for $i = 1, \ldots, 200$. Y_{i1} was always observed, while some of Y_{i2} were missing with probability 0.5. Covariates x were sampled from uniform (0,2). We sampled $\boldsymbol{Y}_i$ from:

$$\begin{aligned}
&\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \Big| R = 1 \sim N\left(\begin{pmatrix} 1+x \\ 1-x \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right), \\
&Y_{i1}|R = 0 \sim \mathrm{N}(-1-x, 1), \\
&\mathrm{p}(R = 1) = 0.5.
\end{aligned}$$

We conducted simulation study under two different assumptions: MAR and MNAR. Under MAR assumption, we fixed sensitivity parameter $\boldsymbol{\xi}_s$ in section 2.2 at $(0, 0, 0, 1)$, while under MNAR assumption, we fixed $\boldsymbol{\xi}_s = (1, 0, 0, 1)$, assuming there was an intercept shift between distribution of $Y_{i2}|Y_{i1}, R = 1$ and $Y_{i2}|Y_{i1}, R = 0$.

For each dataset in both scenario, we fit quantile regression for quantiles $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$.

Algorithms were evaluated by mean squared error (MSE), which is defined by :

$$\mathrm{MSE}(\gamma_{ij}) = \frac{1}{1000} \sum_{k=1}^{1000} \left(\hat{\gamma}_{ij}^{(k)} - \gamma_{ij} \right)^2, \tag{41}$$

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

We also compared our results with those using 'rq' function in R package 'quantreg'. We fit models for each component using 'rq' function, and obtained estimates for each quantile.

Simulation results show estimates from our algorithm are closer to the true value for all quantiles from 0.1 to 0.9. Table 2 and 3 list the MSE for coefficients estimates of quantile 0.1, 0.3, 0.5, 0.7, 0.9 under MAR and MNAR assumptions. Even for extreme quantiles ($\tau = 0.1$ and $\tau = 0.9$), our algorithm behave as good as for non-extreme quantile ($\tau = 0.3, 0.5, 0.7$) in terms of MSE. Furthermore, 'rq' function did not consider the missing mechanism, so under MNAR assumption, 'quantreg' method led to tremendous MSE and our proposed method were much closer to the true value.

6 Discussion

7 TODO

correlation within subject

Table 2: Simulation result: 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 ones for Y_{i2}. MM stands for our proposed method, and RQ stands for the 'rq' function in R package 'quantreg'.

	MAR									
	0.1		0.3		0.5		0.7		0.9	
	MM	RQ	MM	RQ	MM	RQ	MM	RQ	MM	RQ
γ_{01}	0.05	0.09	0.04	0.10	0.03	0.24	0.04	0.10	0.05	0.10
γ_{11}	0.03	0.07	0.02	0.08	0.58	0.74	0.03	0.08	0.03	0.07
γ_{02}	0.04	0.12	0.05	0.07	0.04	0.06	0.05	0.07	0.05	0.11
γ_{12}	0.03	0.09	0.03	0.05	0.03	0.05	0.03	0.05	0.03	0.09

Table 3: Simulation result: MSE for coefficients estimates of quantiles 0.1, 0.3, 0.5, 0.7, 0.9 under MNAR 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'.

	MNAR									
	0.1		0.3		0.5		0.7		0.9	
	MM	RQ	MM	RQ	MM	RQ	MM	RQ	MM	RQ
γ_{01}	0.04	0.09	0.04	0.10	0.03	0.24	0.04	0.10	0.04	0.10
γ_{11}	0.03	0.07	0.02	0.08	0.64	0.74	0.03	0.08	0.03	0.07
γ_{02}	0.04	0.30	0.05	0.52	0.07	1.06	0.05	1.79	0.05	2.59
γ_{12}	0.03	0.09	0.03	0.05	0.03	0.05	0.03	0.05	0.03	0.09

A Proof of Lemma (17)

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

$$\begin{aligned} I(a,b) &= -\frac{1}{\sqrt{a^2+1}} \int_b^{\infty} \phi\left(\frac{s}{\sqrt{a^2+1}}\right) ds \\ &= \int_{b/\sqrt{a^2+1}}^{\infty} \phi(t)dt \\ &= 1 - \Phi(b/\sqrt{a^2+1}) \end{aligned} \tag{42}$$

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

$$\begin{aligned} I(a,b) &= \int_{-\infty}^{b/\sqrt{a^2+1}} \phi(t)dt \\ &= \Phi(b/\sqrt{a^2+1}) \end{aligned} \tag{43}$$

- If integrating over non-standard normal distribution:

$$\begin{aligned} \int \Phi(x) d\Phi(x;\mu,\sigma) &= \int \Phi(x) \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) dx \\ &= \int \Phi(\sigma t + \mu) \phi(t) dt \\ \text{by equation (42):} \quad &= 1 - \Phi(-\mu/\sigma/\sqrt{1/\sigma^2+1}) \end{aligned}$$

- If integrating a non standard cdf over a non-standard normal distribution:

$$\begin{aligned} \int \Phi\left(\frac{x-b}{a}\right) d\Phi(x;\mu,\sigma) &= \int \Phi\left(\frac{x-b}{a}\right) \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) dx \\ &= \int \Phi\left(\frac{\sigma y + \mu - b}{a}\right) \phi(y) dy \\ &= 1 - \Phi\left(\frac{b-\mu}{\sigma} / \sqrt{\frac{a^2}{\sigma^2}+1}\right) \end{aligned} \tag{44}$$

If $a < 0$,

$$\int \Phi\left(\frac{x-b}{a}\right) d\Phi(x;\mu,\sigma) = \Phi\left(\frac{b-\mu}{\sigma} / \sqrt{\frac{a^2}{\sigma^2}+1}\right) \tag{45}$$

B Adaptive Gradient Descent Algorithm

After using the adaptive gradient descent algorithm (Riedmiller and Braun, 1993), the convergence goes faster and consistent. The algorithm is :

```
alphamax = 1
alphamin = 0.000001
etap = 1.2
etam = 0.5
      if (pp(i) * ppp(i) > 0) then
         alpha(i) = min(alpha(i)*etap, alphamax)
         param(i) = param(i) - alpha(i)*pp(i)/abs(pp(i))
      else if (pp(i)*ppp(i) < 0) then
         alpha(i) = max(alpha(i)*etam, alphamin)
         param(i) = param(i) - alpha(i)*pp(i)/abs(pp(i))
         pp(i) = 0
      else if (ppp(i)*pp(i) .eq. 0) then
         if (pp(i) .eq. 0.d0) then
            param(i) = param(i)
         else
            param(i) = param(i) - alpha(i)*pp(i)/abs(pp(i))
         end if
      end if
```

C trivariate case

Suppose distribution of $Y_{i2}|Y_{i1}$:

$$Y_{i2}|y_{i1}, S = 2, x_i \sim \mathrm{N}(\Delta_{i2} + \beta_{02}^{(2)} + x_i\beta_{12}^{(2)} + y_{i1}\beta_{22}^{(2)}, \tau_2^{(2)}),$$
$$Y_{i2}|y_{i1}, S = 3, x_i \sim \mathrm{N}(\Delta_{i2} + \beta_{02}^{(3)} + x_i\beta_{12}^{(3)} + y_{i1}\beta_{22}^{(3)}, \tau_2^{(3)}),$$

To identify $y_2|y_1, S = 1$, denote it in general by the mixture:

$$\mathrm{p}_1(y_2|y_1) = \Delta\, \mathrm{p}_2(y_2|y_1) + (1 - \Delta)\, \mathrm{p}_3(y_2|y_1)$$

where Δ is not identifiable from the data and defined as a sensitivity parameter. If Δ satisfies:

$$\Delta(y_1) = \frac{\pi_2\, \mathrm{p}_2(y_1)}{\pi_2\, \mathrm{p}_2(y_1) + \pi_3\, \mathrm{p}_3(y_1)}$$

then

$$\mathrm{p}_1(y_2|y_1) = \mathrm{p}_{\geq 2}(y_2|y_1) = \frac{\pi_2\, \mathrm{p}_2(y_1)\, \mathrm{p}_2(y_2|y_1) + \pi_3\, \mathrm{p}_3(y_2|y_1)\, \mathrm{p}_3(y_1)}{\pi_2\, \mathrm{p}_2(y_1) + \pi_3\, \mathrm{p}_3(y_1)}$$

which yields MAR condition.

To solve Δ_2,

$$\begin{aligned}\tau &= \mathrm{p}(y_2 \le x\gamma)\\ &= \pi_1\,\mathrm{p}_1(y_2 \le x\gamma) + \pi_2\,\mathrm{p}_2(y_2 \le x\gamma) + \pi_3\,\mathrm{p}_3(y_2 \le x\gamma)\end{aligned}$$

where the last two terms on the right hand side are easy to get as we mentioned before, it is an integral of a normal cdf over a normal pdf . For the first term:

$$\begin{aligned}\mathrm{p}_1(y_2 \le x\gamma) &= \int \mathrm{p}_1(y_2 \le x\gamma|y_1)dF^{(1)}(y_1|S=1)\\ &= \int \left(\frac{\pi_2\,\mathrm{p}_2(y_1)\,\mathrm{p}_2(y_2 \le x\gamma|y_1) + \pi_3\,\mathrm{p}_3(y_2 \le x\gamma|y_1)\,\mathrm{p}_3(y_1)}{\pi_2\,\mathrm{p}_2(y_1) + \pi_3\,\mathrm{p}_3(y_1)}\right) dF^{(1)}(y_1|S=1)\end{aligned}$$

where the integral is not closed form, but if using numerical method (Laplace approximation), it might be feasible.

D True Values for Quantile Regression Coefficients

Recall if y is distributed from:

$$\begin{aligned}y &= \beta_0 + \beta_x x + (\alpha_0 + \alpha_x x)\epsilon\\ \epsilon &\sim N(0,1)\end{aligned}$$

then for τ - quantile, the quantile regression coefficient is $(\gamma_0, \gamma_1) = (\beta_0 + \alpha_0 F_\epsilon(\tau), \beta_x + \alpha_x \Phi_\epsilon(\tau))$ such that

$$\Pr(y \le \gamma_0 + \gamma_1 x) = \tau.$$

Now if y is distributed from:

$$y \sim \beta_0 + \beta_x x + \exp(\alpha_0 + \alpha_x x) N(0,1)$$

then by Taylor expansion, $\exp(\alpha_0 + \alpha_x x) \approx 1 + \alpha_0 + \alpha_x x$, and by previous result, the corresponding quantile regression coefficient can be approximated as $(\gamma_0, \gamma_1) = (\beta_0 + (1 + \alpha_0)\Phi_\epsilon(\tau), \beta_x + \alpha_x \Phi_\epsilon(\tau))$

Now consider mixture of distributions:

$$\begin{aligned}y|R = 1 &\sim N(\beta_{01} + \beta_{11}x, \sigma)\\ y|R = 0 &\sim N(\beta_{00} + \beta_{10}x, \sigma)\\ \mathrm{p}(R = 1) &= \pi\end{aligned}$$

or

$$\begin{aligned}y|R = 1 &\sim N(\beta_{01} + \beta_{11}x, \exp(\alpha_{01} + \alpha_{11}x))\\ y|R = 0 &\sim N(\beta_{00} + \beta_{10}x, \exp(\alpha_{00} + \alpha_{10}x))\\ \mathrm{p}(R = 1) &= \pi\end{aligned}$$

Here R stands for mixture normal indicator, not for missing indicator.

If distribution of $y|R = 1$ and $y|R = 0$ are almost *separable*, then we can obtain quantile lines in closed form and those quantiles given x are linear form of covariates x.

E Heterogeneous Model

In this section we illustrate how to allow heterogeneous variance changing with covariates in different patterns. First we apply the method in bivariate case. Instead of assuming constant variance over covariates in model (3) and (4), we apply another heterogeneity parameter $\alpha_{xi}^{(R)}$ to capture the varying standard deviation for subject i in pattern R.

The reason why we include heterogeneity in the model is that we want to capture the slope change of quantile lines not only by mixture of distributions, but also by the heterogeneity of variance (as we did in previous study).

Thus model (3) and (4) becomes:

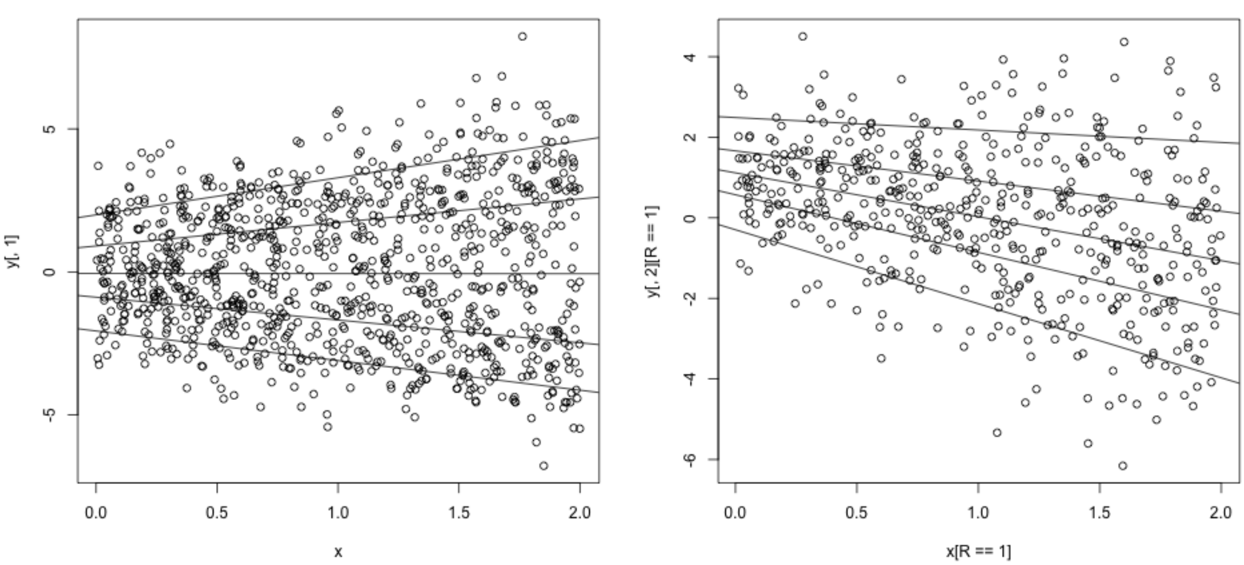
$$
\begin{aligned}
Y_{i1}|R = r, x_i &\sim \mathrm{N}(\Delta_{i1} + \beta_{01}^{(r)} + x_i\beta_{x1}^{(r)}, \exp(\alpha_{01}^{(r)} + \alpha_{x1}^{(r)}x_1)),\\
Y_{i2}|R = r, x_i, y_{i1} &\sim \mathrm{N}(\Delta_{i2} + \beta_{02}^{(r)} + x_i\beta_{x2}^{(r)} + y_{i1}\beta_{12}^{(r)}, \exp(\alpha_{02}^{(r)} + \alpha_{x2}^{(r)}x_1)).
\end{aligned}
$$

To deploy the missing mechanism, we can expand previous notation for sensitivity parameters in section 2.2. Let

$$
\begin{aligned}
\alpha_{02}^{(0)} &= \lambda_0\alpha_{02}^{(1)}\\
\alpha_{x2}^{(0)} &= \lambda_x\alpha_{x2}^{(1)}
\end{aligned}
$$

Here $\boldsymbol{\xi}_\lambda = (\lambda_0, \lambda_x)$ can be additive parameters too. Combined with discussion in section 2.2, when $\boldsymbol{\xi}_s = (\beta_{02}^{(0)}, \beta_{x2}^{(0)}, h) = (0, 0, 0)$ and $\boldsymbol{\xi}_\lambda = (1, 1)$, MAR condition holds.

The plot shows our estimated quantile lines for a small toy example:



F Questions

- How to get true quantile regression estimates?
- Even not complete separable, those quantile lines are still linear.

G Correction

One correction about bivariate normal mixture model with monotone dropout, no constraints should be put on $\beta_{22}^{(r)}$ in equation (4), thus constraint (11) is not necessary and to apply missing mechanism, we can denote

$$\beta_{22}^{(1)} = \beta_{22}^{(0)} + h$$

where h is sensitivity parameter. $h = 0$ leads to MAR.

For general case, I have corrected them in section 4.1.

H Simulation(MAR)

In this section, we test performance of our proposed heterogeneous model in section E combined with adaptive gradient descent algorithm in section B, compared with *rq* function in *quantreg* R package.

The simulation study included 1000 data sets. Each data set consists 200 bivariate observations $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})$ for $i = 1, \ldots, 200$. Y_{i1} was always observed, while some of Y_{i2} were missing with probability 0.5. Covariates x were sampled from uniform (0,2). We sampled $\boldsymbol{Y}_i$ from:

$$\begin{aligned}
Y_{i1}|R = 1 &\sim N(2 + x, 1 + 0.5x) \\
Y_{i2}|R = 1, y_{i1} &\sim N(1 - x - 1/2y_{i1}, 1) \\
Y_{i1}|R = 0 &\sim N(-2 - x, 1 + 0.5x) \\
Y_{i2}|R = 0, y_{i1} &\sim N(1 - x - 1/2y_{i1}, 1) \\
\mathrm{p}(R = 1) &= 0.5.
\end{aligned}$$

$Y_{i2}|R = 0, y_{i1}$ is not observed. And in this setting, we assume distribution of $Y_{i2}|R = 0, y_{i1}$ is equal to $Y_{i2}|R = 1, Y_{i1}$, thus MAR condition satisfies.

By integrating $Y_{i1}|R$ out of $Y_{i2}|R, y_{i1}$, we have

$$\begin{aligned}
Y_{i2}|R = 1 &\sim N(-3x/2, 5/4 + x/8) \\
Y_{i2}|R = 0 &\sim N(2 - x/2, 5/4 + x/8)
\end{aligned}$$

To apply MAR assumption, we fix sensitivity parameter $\boldsymbol{\xi}_s = (\beta_{02}^{(0)}, \beta_{x2}^{(0)}, h, \lambda_0, \lambda_x) = (0, 0, 0, 1, 1)$ as discussed in section E for our proposed model. For *rq* function from *quantreg* R package, because only $Y_{i2}|R = 1$ is observed, quantile regression for Y_{i2} can only be fit from the information of $Y_{i2}|R = 1$ vs x.

For each dataset in both scenario, we fit quantile regression for quantiles $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$.

Algorithms were evaluated by mean squared error (MSE), which is defined by :

$$\mathrm{MSE}(\gamma_{ij}) = \frac{1}{1000} \sum_{k=1}^{1000} \left(\hat{\gamma}_{ij}^{(k)} - \gamma_{ij} \right)^2,$$

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

Table 4: Simulation result: 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 ones for Y_{i2}. MM stands for our proposed method, and RQ stands for the 'rq' function in R package 'quantreg'.

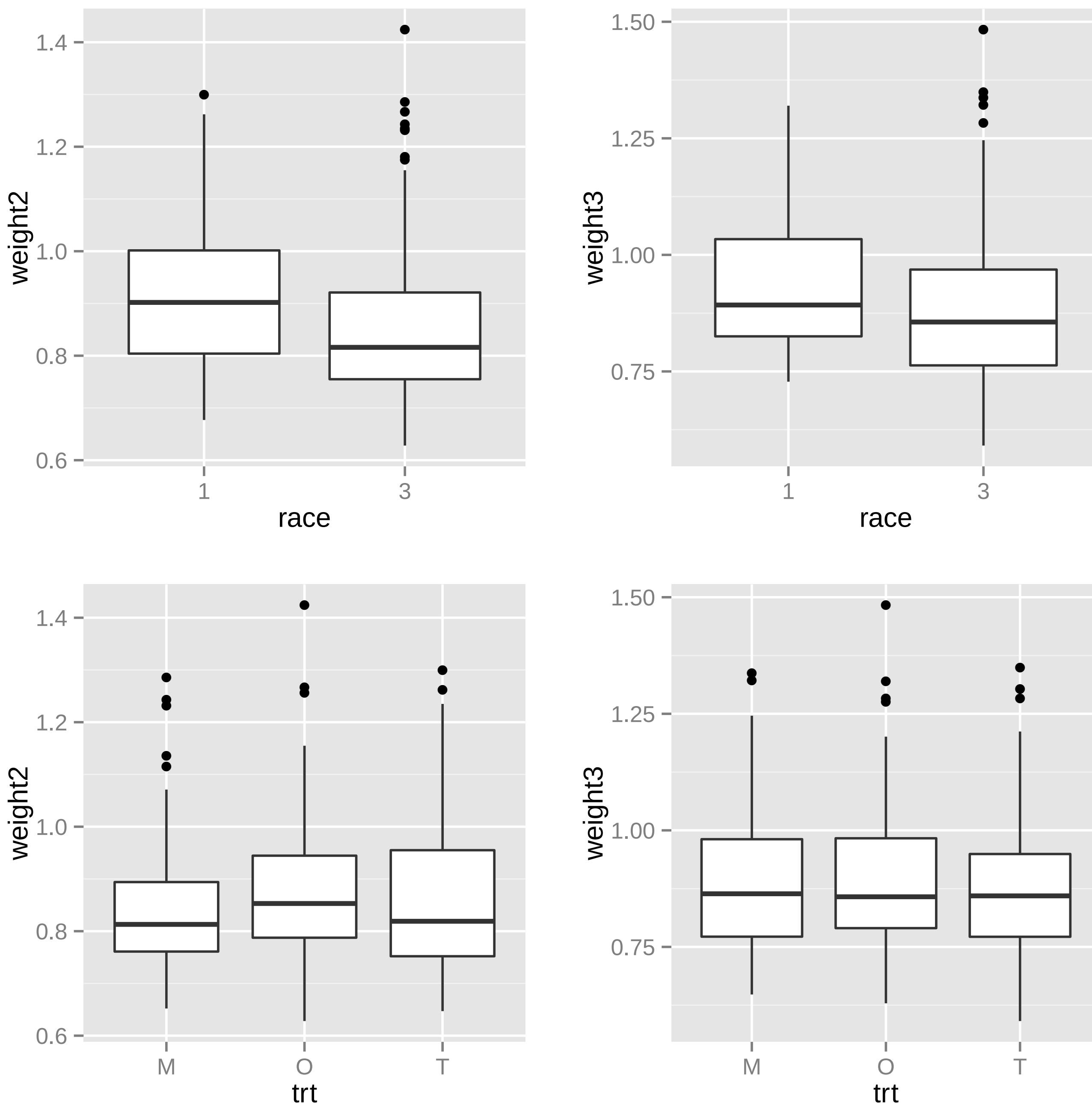
	MAR									
	0.1		0.3		0.5		0.7		0.9	
	MM	RQ	MM	RQ	MM	RQ	MM	RQ	MM	RQ
γ_{01}	0.09	0.15	0.12	0.19	0.11	1.08	0.16	0.19	0.10	0.15
γ_{11}	0.09	0.15	0.07	0.19	0.14	1.19	0.08	0.20	0.10	0.15
γ_{02}	0.08	0.27	0.07	0.59	0.06	1.08	0.12	1.75	0.24	2.92
γ_{12}	0.06	0.17	0.05	0.13	0.06	0.33	0.07	0.75	0.09	0.96

Mean squared errors are shown in table 4. Results show that our proposed method has smaller MSE than *rq* function in all cases. Furthermore, when Y_{i2} are missing at random, our method shows tremendous advantage over *rq* method, because *quantreg* does not consider the missing mechanism.

I Real Data Analysis

Here is the analysis for *tours* data. *Weight2* stands for weight at 6 months after the baseline measure, and *weight3* stands for the one at 18 months after the baseline. There were three treatments and two main races in this study (Treatment M, Treatment O and Treatment T; Race 1 and Race 3). Weights at 6th month were always observed and some weights at 18th month were missing (211 observed out of 224 , 94%). All weights are scaled by 1/100.

Here is the boxplot of weights vs treatments and races.



Quantile regression models were fitted for responses *weight2* and *weight3* together ($\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})$) for quantiles (10%, 30%, 50%, 70%, 90%). Covariates are treatments and races, and we assume their effects are additive. Treatment M and Race 1 are baseline references. We fit 1,000 bootstraps to obtain the 95% confidence intervals.

Estimates are presented in table 5. For *weight2*, quantile estimates show there is no significant difference for three treatment group through all the quantiles, because all 95 % confidence intervals include 0. However, when comparing weights from two races, weights at 6th month from race 3 are generally lower than ones from race 1. Estimates of race 3 effect to weights quantiles 10% up to 70% are all significantly away from zero (negative). However, for top weights of two races (90% quantile), the difference is not significant.

For weights at 18th month (weight3), we have similar conclusions. Confidence intervals

Table 5: Marginal quantile regression coefficients for Weight2

	Intercept	Trt.O	Trt.T	Race.3
Weight2				
$\tau = 0.1$	0.80 (0.70, 0.86)	0.01 (-0.04, 0.07)	-0.01 (-0.06, 0.06)	-0.13 (-0.19, -0.04)
$\tau = 0.3$	0.83 (0.79, 0.92)	0.04 (-0.02, 0.07)	0.02 (-0.04, 0.05)	-0.07 (-0.16, -0.03)
$\tau = 0.5$	0.85 (0.82, 0.98)	0.05 (-0.03, 0.09)	0.04 (-0.06, 0.07)	-0.03 (-0.14, 0.00)
$\tau = 0.7$	0.95 (0.89, 1.03)	0.03 (-0.02, 0.10)	0.02 (-0.04, 0.08)	-0.04 (-0.12, 0.00)
$\tau = 0.9$	0.98 (0.94, 1.11)	0.07 (-0.02, 0.14)	0.06 (-0.02, 0.14)	-0.01 (-0.10, 0.05)
Weight3				
$\tau = 0.1$	0.78 (0.38, 0.84)	-0.01 (-0.07, 0.06)	-0.04 (-0.10, 0.04)	-0.13 (-0.18, -0.02)
$\tau = 0.3$	0.82 (0.78, 0.93)	0.01 (-0.04, 0.06)	-0.01 (-0.07, 0.05)	-0.06 (-0.16, -0.01)
$\tau = 0.5$	0.88 (0.84, 1.00)	0.02 (-0.06, 0.06)	0.02 (-0.08, 0.06)	-0.03 (-0.13, 0.02)
$\tau = 0.7$	0.99 (0.92, 1.07)	0.00 (-0.05, 0.08)	-0.00 (-0.07, 0.06)	-0.04 (-0.12, 0.01)
$\tau = 0.9$	1.02 (0.98, 1.16)	0.05 (-0.06, 0.11)	0.04 (-0.05, 0.12)	0.00 (-0.10, 0.06)

of treatment effects on weight3 for all quantiles (10% up to 90%) include zero. But after 18 months, weights of patients from race 3 are significantly lower than ones from race 1 only for lower quantiles (10% to 30%). They are not significantly different for quantiles (50% to 90%).

J Identifiability

First consider univariate case with two patterns. 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:

$$\begin{aligned} y|R = 1 &\sim N(\Delta + R_1, \sigma_1) \\ y|R = 0 &\sim N(\Delta + R_0, \sigma_0) \\ \Pr(R = 1) &= \pi \\ E(y) &= \theta \end{aligned}$$

Thus by iterated expectation, we have

$$\begin{aligned} \theta &= \Delta + R_1\pi + R_0(1 - \pi) \\ \Delta &= \theta - \pi R_1 - (1 - \pi)R_0 \end{aligned}$$

We can see Δ is deterministic by θ, R_1, R_0. If plugged in likelihood, we have

$$\begin{aligned} y|R = 1 &\sim N(\theta + (1-\pi)R_1 - (1-\pi)R_0, \sigma_1) \\ y|R = 0 &\sim N(\theta - \pi R_1 + \pi R_0, \sigma_0) \end{aligned}$$

Denote $\xi_1 = (\theta, R_1, R_0)$, and if $\xi_2 = (\theta, R_1 + 1, R_0 + 1)$, both 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 R_1 and R_0, for example $R_0 = -R_1$, then

$$\begin{aligned} y|R = 1 &\sim N(\theta + 2(1-\pi)R_1, \sigma_1) \\ y|R = 0 &\sim N(\theta - 2\pi R_1, \sigma_0) \end{aligned}$$

thus it 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 pattern mixture model:

$$\begin{aligned} y|R = 1 &\sim N(\Delta + R_1, \sigma_1) \\ y|R = 0 &\sim N(\Delta + R_0, \sigma_0) \\ \mathrm{Pr}(R = 1) &= \pi \\ \mathrm{p}(y \leq \theta) &= \tau \end{aligned}$$

where θ is the quantile estimate of interest and we does not include covariates so far. We will show $\boldsymbol{\xi} = (\theta, R_1, R_0)$ is not identifiable.

Again by iterated expectation, we have

$$\tau = \pi\Phi\left(\frac{\theta - \Delta - R_1}{\sigma_1}\right) + (1-\pi)\Phi\left(\frac{\theta - \Delta - R_0}{\sigma_0}\right)$$

thus Δ is again deterministic by other parameters:

$$\Delta = h(\theta, R_1, R_0, \sigma_1, \sigma_0, \pi, \tau)$$

To show $\boldsymbol{\xi} = (\theta, R_1, R_0, \sigma_1, \sigma_0)$ 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', R_1', R_0' such that

$$\begin{aligned} h(\boldsymbol{\xi}) + R_1 &= h(\boldsymbol{\xi}') + R_1' \\ h(\boldsymbol{\xi}) + R_0 &= h(\boldsymbol{\xi}') + R_0' \end{aligned}$$

By substracting previous equations, we have $R_1' - R_0' = R_1 - R_0$, thus denote $R_1' = R_1 + \delta$ and $R_0' = R_0 + \delta$, and let $\theta' = \theta$ such that

$$\Delta' = h(\theta', R_1, R_0, \sigma_1, \sigma_0, \delta) = h(\boldsymbol{\xi}) - \delta = \Delta - \delta$$

then the new parameter $\boldsymbol{\xi}'$ yields the same distribution with one from $\boldsymbol{\xi}$. Therefore $\boldsymbol{\xi}$ is not identifiable.

Instead, if we put constraint, for example $R_1 = -R_0$, then by calculation, $\mathrm{p}(y|R; \boldsymbol{\xi}) = \mathrm{p}(y|R; \boldsymbol{\xi}')$ yields $\boldsymbol{\xi} = \boldsymbol{\xi}'$.

Now consider the case with covariates. Suppose the model is

$$
\begin{aligned}
y|R = 1, x &\sim N(\Delta + R_1 + \beta_{x1}x, \sigma_1) \\
y|R = 0, x &\sim N(\Delta - R_1 + \beta_{x0}x, \sigma_0) \\
\Pr(R = 1) &= \pi \\
\mathrm{p}(y \leq \gamma_0 + \gamma_1 x) &= \tau
\end{aligned}
$$

Still Δ can be determined by

$$\Delta = h(x, \gamma_0, \gamma_1, R_1, \beta_{x1}, \beta_{x0}, \sigma_1, \sigma_0, \pi, \tau)$$

We want to show parameter $\boldsymbol{\xi} = (\gamma_0, \gamma_1, R_1, \beta_{x1}, \beta_{x0}, \sigma_1, \sigma_0, \pi)$ is not identifiable by finding $\boldsymbol{\xi}^{'} \neq \boldsymbol{\xi}$, but $\mathrm{p}(y|R;\boldsymbol{\xi}) = \mathrm{p}(y|R;\boldsymbol{\xi}^{'})$. Still if the last equation holds, first we have $\sigma_1^{'} = \sigma_1, \sigma_0^{'} = \sigma_0$, then to equate the two means, we have

$$
\begin{aligned}
\Delta + R_1 + \beta_{x1}x &= \Delta^{'} + R_1^{'} + \beta_{x1}^{'}x \\
\Delta - R_1 + \beta_{x0}x &= \Delta^{'} - R_1^{'} + \beta_{x0}^{'}x
\end{aligned}
$$

By substracting the two equations, we have

$$2R_1 + (\beta_{x1} - \beta_{x0})x = 2R_1^{'} + (\beta_{x1}^{'} - \beta_{x0}^{'})x$$

which holds for all x. Thus $R_1 = R_1^{'}$ and $(\beta_{x1} - \beta_{x0}) = (\beta_{x1}^{'} - \beta_{x0}^{'})$. Then let

$$
\begin{aligned}
\beta_{x1}^{'} &= \beta_{x1} + \delta \\
\beta_{x0}^{'} &= \beta_{x0} + \delta
\end{aligned}
$$

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

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

$$
\begin{aligned}
y_2|y_1, x, R = 1 &\sim N(\Delta + R_1 + x\beta_{x1} + \beta_{11}y_1, \sigma_1) \\
y_2|y_1, x, R = 0 &\sim N(\Delta - R_1 - x\beta_{x1} + \beta_{10}y_1, \sigma_0)
\end{aligned}
$$

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

we are wondering if β_{11} and β_{10} are identifiable, say if there exists $\beta_{11}^{'}$ and $\beta_{10}^{'}$, such that

$$
\begin{aligned}
\Delta + R_1 + x\beta_x + \beta_{11}y_1 &= \Delta^{'} + R_1^{'} + x\beta_x^{'} + \beta_{11}^{'}y_1 \\
\Delta - R_1 - x\beta_x + \beta_{10}y_1 &= \Delta^{'} - R_1^{'} - x\beta_x^{'} + \beta_{10}^{'}y_1
\end{aligned}
$$

still by substracting two equations, we have $R_1 = R_1^{'}$ and $\beta_x = \beta_x^{'}$. Considering 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, $\boldsymbol{\xi}$ is identifiable.

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