Bayesian Mixtures of Quantile Regressions

By Paul Thompson†[1] and Rana Moyeed[2]

[1]Department of Experimental Psychology, University of Oxford
[2]School of Computing, Electronics and Mathematics, Plymouth University

1. Introduction

The Bayesian quantile regression (BQR) methodology developed in Yu & Moyeed (2001) uses a parametric approach for BQR modelling based on polynomial regression.

In this paper we present a Bayesian alternative to the Frequentist approach of Tian et al. (2015) and Wu (2016).

Section 2 of the paper provides an introduction to the theoretical basis of Bayesian mixtures of quantile regression, and outlines the Markov chain Monte Carlo (MCMC) algorithm used to perform inference about our model parameters. Section 3 presents an application of our methodology to simulated data and a developmental psychology data set. The developmental psychology data used in this example relates to differentiation between normally developing and Dyslexic children. The data set consists of composites of reading and spelling measurements at age 8. Finally Section 4 is a short conclusion.

2. Bayesian Mixtures of Quantile Regression

In this section we define finite mixtures of quantile regressions following the approach of Tian et al. (2015). The aim of this paper is to fit mixtures of quantile regressions within a Bayesian Framework, so providing an alternative approach to Tian et al. (2015)'s technique. Despite Tian et al. (2015) and Wu (2016)'s techniques novelty using the EM algorothm, it is known that Bayesian approach to mdoellling mixture distributions is often favourable. Richardson & Green (1997, p.732) stated, "However, we strongly believe that the Bayesian paradigm is particularly suited to mixture analysis...". We begin with an overview of the finite mixture model and describe the extension to quantile regressions following Wu (2016) and Tian et al. (2015).

2.1. *Finite Gaussian mixture model*

The Gaussian mixture model with m components for the response variable Y_i given the covariates $\boldsymbol{x}_i, i = 1, 2, \ldots, n.$, has the conditional distribution as follows:

$$f(y_i|\boldsymbol{x},\boldsymbol{\beta},\boldsymbol{\sigma}) = \sum_{k=1}^{m} \phi_k f(y; \boldsymbol{x}^{\boldsymbol{T}} \boldsymbol{\beta}_{\boldsymbol{k}}, \sigma_k^2) \tag{2.1}$$

where ϕ_k is a vector of mixing proportions provided $\sum_{k=1}^{m} \phi_k = 1$, $\phi_k > 0$, $k = 1, 2, \ldots, m$, β_k are k vectors of model parameters, and σ_k are the k variance parameters. We also define y as observations from our response variable and x as a set of covariates. This model requires that the error terms are independent and identically distributed, and follow a normal distribution with mean zero and constant variance.

† paul.thompson@psy.ox.ac.uk

Following Richardson & Green (1997) and Gelman et al. (2003), we ensure identifiability by introducing a latent allocation variable $z_i, i = 1, 2, \ldots, n.$ that indicates the group membership for each i^{th} observation as this is initially unknown.

$$z_i = \begin{cases} 1 & \text{if the ith unit is drawn from the mth mixture component} \\ 0 & \text{Otherwise.} \end{cases}$$

The $z_i = (z_{i,1}, \ldots, z_{i,m})$ are independently drawn from the distributions

$$p(z_i = k) = \phi_k, \quad \text{for } k = 1, 2, \ldots, m. \tag{2.2}$$

then, the observations are drawn from their corresponding subpopulations, given the values of z_i,

$$y_i|z \sim f(\cdot|\theta_{z_k}), \quad \text{for } i = 1, 2, \ldots, n. \tag{2.3}$$

where $\theta = (\beta, \sigma, \phi)$ are the model parameters.

2.2. *Finite Mixtures of Quantile Regressions*

The Finite mixture model defined in the previous section can be extended to

2.3. *Bayesian Mixtures of Quantile regressions*

In this section we present a framework for Bayesian mixtures of quantile regressions.

As our approach is Bayesian, we begin by defining the prior distribution for $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_k)^T$. We adopt a Dirichlet prior for $\boldsymbol{\phi}$ defined by,

$$\phi_k \sim Dir(\alpha_1, \ldots, \alpha_M) \tag{2.4}$$

$$\pi(\phi_k) = \frac{\Gamma(\sum_{k=1}^{M} \alpha_k)}{\prod_{k=1}^{M} \Gamma(\alpha_k)} \prod_{k=1}^{M} \phi_k^{\alpha_k - 1} \tag{2.5}$$

Next, the model parameters $\boldsymbol{\beta}$ are assigned a normal prior for each component,k and element, d in the $\boldsymbol{\beta_k} = (\beta_{1,k}, \ldots, \beta_{d,k})^T, k = 1, 2, ..., m$. The normal distribution prior for $\beta_{d,k}$ is defined,

$$\pi(\beta_{d,k}) = \frac{1}{\sqrt{2\pi}\sigma_{\beta_{d,k}}} \exp\left(-\frac{1}{2\sigma^2_{\beta_{d,k}}}(\boldsymbol{\beta}_k - \mu_{\beta_k})^2\right), \tag{2.6}$$

We next require a prior on the variance parameter σ_k which is constrained by a lower limit of zero. We use the Gamma distribution as our prior for σ_k which takes the form

$$\pi(\sigma_k) = \frac{\sigma_k^{a-1} \exp(-\sigma_k/b)}{\Gamma(a)b^a}, \tag{2.7}$$

in which Γ is the usual gamma function. The user is able to specify the hyperparameters a and b. Under this prior $\mathrm{E}[\sigma_k] = ab$ and $\mathrm{Var}[\sigma_k] = ab^2$, results that can be used to guide hyperparameter choice.

We can define our $p(z_{i,k}|\phi_k), k = 1, 2, \ldots, m$ following a multinomial distribution,

$$z_i \sim Multin(1, \phi_1, \ldots, \phi_m), \text{where} \sum_{k=1}^{M} z_i = 1, \tag{2.8}$$

$$\pi(z_{i,k}|\phi_k) = \frac{1!}{z_{i,1}! \cdots z_{i,m}!} \prod_{k=1}^{M} \phi_k^{z_k} \tag{2.9}$$

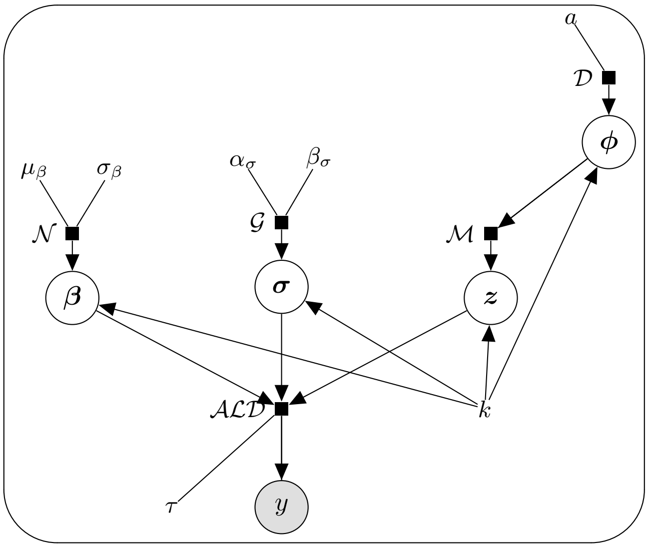

FIGURE 1. DAG for the mixture model of quantile regression presented in this paper

The final step in our Bayesian approach is to define the likelihood of the data $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ given $\phi_k, \boldsymbol{\beta}_k, \sigma_k, \boldsymbol{z}_k$ in accordance with the BQR approach of Yu & Moyeed (2001) and mixture model reparameterization in Gelman et al. (2003):

$$L(\boldsymbol{y}|\phi, \boldsymbol{\beta}, \sigma, z) = \prod_{i=1}^{n} \prod_{k=1}^{m} \left(\phi_k \frac{\tau(1-\tau)}{\sigma_k} \exp\left\{ -\rho_\tau \left(\frac{y_i - \boldsymbol{x^T} \boldsymbol{\beta_k}}{\sigma_k} \right) \right\} \right)^{z_{i,k}} \tag{2.10}$$

where τ is the quantile of interest, $0 < \tau < 1$, and ρ_τ is the standard loss function

$$\rho_\tau(u) = u(\tau - I(u < 0)) \tag{2.11}$$

in which I is the usual indicator function. Combining $\pi(\phi_k)$, $\pi(\boldsymbol{\beta})$, $\pi(\sigma_k)$ and $L(\boldsymbol{y}|\phi, \boldsymbol{\beta}, \sigma, z)$, we can write the posterior density function of ϕ, $\boldsymbol{\beta}$, σ and z as

$$\pi(\phi, \boldsymbol{\beta}, \sigma, z|\boldsymbol{y}) \propto L(\boldsymbol{y}|\phi, \boldsymbol{\beta}, \sigma, z)\pi(\boldsymbol{\phi})\pi(\sigma)\pi(\boldsymbol{\beta})\pi(z|\phi). \tag{2.12}$$

Figure 1 shows the Directed Acyclic Graph (DAG) for the proposed mixture model. ....

2.3.1. *Sampling from Posterior distribution of ϕ*

We first consider the mixing parameters $\phi_k, k = 1, 2, \ldots, m$. These parameters can be sampled directly from their posterior $\pi(\phi_k|z_{i,k})$ as the Dirichlet prior $\pi(\phi_k)$ is conjugate and $\pi(z|\phi)$ follows a multinomial distribution; hence, we have a Dirichlet posterior distribution. The Dirichlet posterior is

$$\pi(\phi_k|z_{i,k}) = \left(\frac{\Gamma(\sum_{k=1}^{M} \alpha_k)}{\prod_{k=1}^{M} \Gamma(\alpha_k)} \prod_{k=1}^{M} \phi_k^{\alpha_{k-1}} \right) \left(\frac{1!}{z_{i,1}! \cdots z_{i,m}!} \prod_{k=1}^{M} \phi_k^{z_k} \right) \tag{2.13}$$

$$\pi(\phi_k|z_{i,k}) = \prod_{k=1}^{M} \phi_k^{\alpha_{k-1} + \sum_{z_i \in D} z_i^{(k)}} \tag{2.14}$$

2.3.2. *Sampling from Posterior distribution of β, σ, and z using MCMC*

The posterior distributions for the remaining parameters are more complex and cannot be evaluated directly, so we employ MCMC methods via the Metropolis-Hastings algorithm to simulate realizations of our parameters from the posterior density. Our algorithm can be summarized as follows:

(i) Specify the number of components, $k = 1, \ldots, m$.

(ii) Draw a sample from $\pi(\phi_k|z_{i,k})$, which is $\phi_k^{(j)}$, where $j = 1, \ldots, r$, refers to the iteration number in the Metropolis-Hastings algorithm. As in the previous section, we can sample directly from the posterior without requiring MCMC.

Updating β

(iii) Assign initial values $\boldsymbol{\beta}_k^{(0)}$, $\sigma_k^{(0)}$ and $z_k^{(0)}$ to $\boldsymbol{\beta_k}$, σ_k and z. We set $\boldsymbol{\beta}_k^{(0)}$to be vector of zeros, $\sigma_k^{(0)}$ to be a vector of 1s, and $z_k^{(0)}$ as a matrix of zeros.

(iv) The elements of the vector of βs is updated individually. Hence, we generate a candidate value $\beta^\star_{d,k}$ from the normal distribution for each element of $\beta_{d,k} = (\beta_{1,k}, \ldots, \beta_{D,k}), d = 1, 2, ..., D$.

$$\beta^\star_{d,k}|\beta^{(j-1)}_{d,k} \sim N(\beta^{(j-1)}_{d,k}, \Sigma) \tag{2.15}$$

with mean $\beta^{(j-1)}_{d,k}$ and variance $\sigma^2_{\beta_{d,k}}$, where $d = 1, \ldots, D$ is the number of β parameters. The constant $\sigma^2_{\beta_{d,k}}$ is specified by the user.

(v) We then calculate the acceptance probability of a move from $\beta^{(j-1)}_{d,k}$ to $\beta^\star_{d,k}$ which takes the form:

$$\begin{aligned}
\alpha(\beta^{(j-1)}_{d,k}, \beta^\star_{d,k}) &= \min\left\{1, \frac{\pi(\beta^\star_{d,k}, \phi^{(j)}_k, \sigma^{(j-1)}_k, \boldsymbol{z}^{(j-1)}_k|\boldsymbol{y})q(\beta^{(j-1)}_{d,k}|\beta^\star_{d,k})}{\pi(\beta^{(j-1)}_{d,k}, \phi^{(j)}_k, \sigma^{(j-1)}_k, \boldsymbol{z}^{(j-1)}_k|\boldsymbol{y})q(\beta^\star_{d,k}|\beta^{(j-1)}_{d,k})}\right\} \\
&= \min\left\{1, \frac{L(\boldsymbol{y}|\beta^\star_{d,k}, \phi^{(j)}_k, \sigma^{(j-1)}_k, \boldsymbol{z}^{(j-1)}_k)\pi(\beta^\star_{d,k})q(\beta^{(j-1)}_{d,k}|\beta^\star_{d,k})}{L(\boldsymbol{y}|\beta^{(j-1)}_{d,k}, \phi^{(j)}_k, \sigma^{(j-1)}_k, \boldsymbol{z}^{(j-1)}_k)\pi(\beta^{(j-1)}_{d,k})q(\beta^\star_{d,k}|\beta^{(j-1)}_{d,k})}\right\}
\end{aligned} \tag{2.16}$$

where q is the probability density function of the normal specified in (2.15). In fact, because of the symmetry of the normal density, q disappears from (B 2).

(vi) A random variable u is simulated from a uniform distribution $U(0, 1)$. If $u \leq \alpha(\beta^{(j-1)}_{d,k}, \beta^\star_{d,k})$, then $\beta^\star_{d,k}$ is accepted by setting $\beta^{(j)}_{d,k} = \beta^\star_{d,k}$, otherwise the chain does not move and $\beta^{(j)}_{d,k} = \beta^{(j-1)}_{d,k}$.

Updating σ

(vii) We now generate a candidate $\sigma^\star_k$ from the log-normal distribution as follows:

$$\mu^\star_k \sim N(\log(\sigma^{j-1}), \sigma^2_\lambda) \tag{2.17}$$

$$\sigma^\star_k = \exp(\mu^\star_k) \tag{2.18}$$

where the normal distribution has mean $\log(\sigma^{(j-1)}_k)$ and variance σ^2_λ. The variance σ^2_λ

can be specified by the user.

(viii) We then calculate the acceptance probability of a move from $\sigma_k^{(j-1)}$ to $\sigma_k^\star$ which takes the form:

$$\begin{aligned}\alpha(\sigma_k^{(j-1)},\sigma_k^\star) &= \min\left\{1, \frac{\pi(\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^\star,\boldsymbol{z}_k^{(j-1)}|\boldsymbol{y})q(\sigma_k^{(j-1)}|\sigma_k^\star)}{\pi(\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^{(j-1)},\boldsymbol{z}_k^{(j-1)}|\boldsymbol{y})q(\sigma_k^\star|\sigma_k^{(j-1)})}\right\}\\ &= \min\left\{1, \frac{L(\boldsymbol{y}|\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^\star,\boldsymbol{z}_k^{(j-1)})\pi(\sigma_k^\star)q(\sigma_k^{(j-1)}|\sigma_k^\star)}{L(\boldsymbol{y}|\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^{(j-1)},\boldsymbol{z}_k^{(j-1)})\pi(\sigma_k^{(j-1)})q(\sigma_k^\star|\sigma_k^{(j-1)})}\right\} \quad (2.19)\end{aligned}$$

where q is the log-normal probability density function specified in (2.17)and (2.18).

(ix) A random variable u is simulated from a uniform distribution $U(0,1)$. If $u \leq \alpha(\sigma_k^{(j-1)},\sigma_k^\star)$, then $\sigma_k^\star$ is accepted by setting $\sigma_k^{(j)} = \sigma_k^\star$, otherwise the chain does not move and $\sigma_k^{(j)} = \sigma_k^{(j-1)}$.

(x) We now increment k by 1, and repeat steps (ii)-(xi) until $k = m$, then proceed to step (xi).

Updating z

(xi) $\boldsymbol{z}$ is a n by m matrix of zeros and ones, where for any i, $\sum_{k=1}^m z_{i,k} = 1$, i.e. each row should contain only one non-zero value and all other elements are zero. We generate a candidate vector each $\boldsymbol{z}_i^\star$ using a two-step procedure. First, we draw a random integer w from a uniform distribution $w \sim Unif(1,m)$. This random value w can be thought of as a location parameter within the $\boldsymbol{z}_i$ vector for the non-zero value,i.e. 1. The candidate vector $\boldsymbol{z}_i^\star$ is then simply a m-length vector with a 1 imputed at the location w.

(xii) We then calculate the acceptance probability of a move from $\boldsymbol{z}_i^{(j-1)}$ to $\boldsymbol{z}_i^\star$ which takes the form:

$$\begin{aligned}\alpha(\boldsymbol{z}_i^{(j-1)},\boldsymbol{z}_i^\star) &= \min\left\{1, \frac{\pi(\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^{(j)},\boldsymbol{z}_i^\star|\boldsymbol{y})q(\boldsymbol{z}_i^{(j-1)}|\boldsymbol{z}_i^\star)}{\pi(\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^{(j)},\boldsymbol{z}_i^{(j-1)}|\boldsymbol{y})q(\boldsymbol{z}_i^\star|\boldsymbol{z}_i^{(j-1)})}\right\}\\ &= \min\left\{1, \frac{L(\boldsymbol{y}|\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^j,\boldsymbol{z}_i^\star)\pi(\boldsymbol{z}_i^\star|\phi)\pi(\phi)q(\boldsymbol{z}_i^{(j-1)}|\boldsymbol{z}_i^\star)}{L(\boldsymbol{y}|\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^{(j)},\boldsymbol{z}_i^{(j-1)})\pi(\boldsymbol{z}_i^{(j-1)}|\phi)\pi(\phi)q(\boldsymbol{z}_i^\star|\boldsymbol{z}_i^{(j-1)})}\right\}\\ &= \min\left\{1, \frac{L(\boldsymbol{y}|\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^j,\boldsymbol{z}_i^\star)\pi(\boldsymbol{z}_i^\star|\phi)q(\boldsymbol{z}_i^{(j-1)}|\boldsymbol{z}_i^\star)}{L(\boldsymbol{y}|\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^{(j)},\boldsymbol{z}_i^{(j-1)})\pi(\boldsymbol{z}_i^{(j-1)}|\phi)q(\boldsymbol{z}_i^\star|\boldsymbol{z}_i^{(j-1)})}\right\} \quad (2.20)\end{aligned}$$

where q is the Uniform probability density function.

(xiii) A random variable u is simulated from a uniform distribution $U(0,1)$. If $u \leq \alpha(\boldsymbol{z}_i^{(j-1)},\boldsymbol{z}_i^\star)$, then $\boldsymbol{z}_i^\star$ is accepted by setting $\boldsymbol{z}_i^{(j)} = \boldsymbol{z}_i^\star$, otherwise the chain does not move and $\boldsymbol{z}_i^{(j)} = \boldsymbol{z}_i^{(j-1)}$.

(xiv) We now increment i by 1, and repeat steps (xi)-(xiii) until $i = n$ iterations.

(xv) We now increment j by 1, and repeat steps (ii)-(xiv+) for a total of r iterations.

A considerable advantage of the Bayesian approach is that we can calculate associated credible intervals to provide an idea of the associated posterior uncertainty. In the next section an example of this methodology applied to the data set introduced in Section 1 is presented.

3. Application to Developmental Psychology Data

4. Conclusions

In this paper we have presented methodology to mixtures of quantile regressions within a Bayesian framework....

5. Acknowledgement

The first author is funded by Wellcome Trust Programme Grant number 082498/Z/07/Z. The funders had no role in the design of the methods, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

BROOKS, S.P. & GELMAN, A. 1998, General methods for monitoring convergence of iterative simulations, *Journal of Computational and Graphical Statistics*, **7**, part 4, page 434–455.

GAMERMAN, D. 1997, *Markov Chain Monte Carlo: Stocastic Simulation for Bayesian Inference*, Chapman and Hall, London.

GELMAN, A., CARLIN, J.B, STERN, H.S, & RUBIN, D.B. 2003, *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton.

KASS, R.E, CALIN, B.P, GELMAN, A. & NEAL, R.M. 1998, MCMC in practice: A roundtable discussion, *The American Statistician: Statistical Practice*, **52**, part 2, page 93–100.

KOENKER, R. 2005, *Quantile Regression*, Cambridge University Press, New York.

RAO, R.C. 1973 *Linear Statistical Inference and its Applications (2nd Ed.)*, John Wiley & Sons, New York.

RICHARDSON, S., & GREEN, P. 1997, On Bayesian analysis of mixtures with an unknown number of Components, *J. R. Stat. Soc: Series B*, **59**, part 4, page 731–792.

TIAN, Y, TANG, M & TIAN, M. 2015, A class of finite mixture of quantile regressions with its applications, *Journal of Applied Statistics*, page 1–13, doi: 10.1080/02664763.2015.1094035.

VENABLES, W.N & RIPLEY, B.D. 2002, *Modern Applied Statistics with S*, Springer-Verlag, New York.

YU, K., & MOYEED, R. 2001, Bayesian quantile regression, *Statistics and Probability Letters*, **54**, page 437–447.

YU, K., LU, Z., & STANDER, J. 2003, Quantile regression: application and current research areas, *J. R. Stat. Soc: The Statistician*, **52**, part 3, page 331–350.

WU, Q & YAO, W. 2016, Mixtures of quantile regressions, *Computational Statistics and Data Analysis*, **93**, page 162–176.

Appendix A. Trace plots

Figure # provides time series plots for a sample of the parameters ###. We include these plots as a justification of the convergence of ###.

Appendix B. Full posterior and Acceptance functions

B.1. *Full posterior*

$$\pi(\phi,\boldsymbol{\beta},\sigma,\boldsymbol{z}|\boldsymbol{y}) = \prod_{i=1}^{n}\prod_{k=1}^{m}\left(\phi_k\frac{\tau(1-\tau)}{\sigma_k}\exp\left\{-\rho_\tau\left(\frac{y_i-\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta_k}}{\sigma_k}\right)\right\}\right)^{z_{i,k}}\left(\frac{1}{\sqrt{2\pi}\sigma_{\beta_{d,k}}}\exp\left\{-\frac{1}{2\sigma^2_{\beta_{d,k}}}(\boldsymbol{\beta}_k-\mu_{\beta_k})^2\right\}\right)$$
$$\cdot\left(\frac{\sigma_k^{a-1}\exp\left(-\sigma_k/b\right)}{\Gamma(a)b^a}\right)\left(\frac{1!}{z_{i,1}!\cdots z_{i,m}!}\prod_{k=1}^{M}\phi_k^{z_k}\right) \quad \text{(B 1)}$$

B.2. *β posterior and acceptance function*

$$\alpha(\beta_{d,k}^{(j-1)},\beta_{d,k}^{\star}) = \min\left\{1,\frac{L(\boldsymbol{y}|\beta_{d,k}^{\star},\phi_k^{(j)},\sigma_k^{(j-1)},\boldsymbol{z}_k^{(j-1)})\pi(\beta_{d,k}^{\star})q(\beta_{d,k}^{(j-1)}|\beta_{d,k}^{\star})}{L(\boldsymbol{y}|\beta_{d,k}^{(j-1)},\phi_k^{(j)},\sigma_k^{(j-1)},\boldsymbol{z}_k^{(j-1)})\pi(\beta_{d,k}^{(j-1)})q(\beta_{d,k}^{\star}|\beta_{d,k}^{(j-1)})}\right\}$$
$$= \min\left\{1,\frac{\prod_{i=1}^{n}\prod_{k=1}^{m}\left(\phi_k\frac{\tau(1-\tau)}{\sigma_k}\exp\left\{-\rho_\tau\left(\frac{y_i-\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta_k^\star}}{\sigma_k}\right)\right\}\right)^{z_{i,k}}\left(\frac{1}{\sqrt{2\pi}\sigma_{\beta_{d,k}}}\exp\left\{-\frac{1}{2\sigma^2_{\beta_{d,k}}}(\boldsymbol{\beta}_k^\star-\mu_{\beta_k})^2\right\}\right)}{\prod_{i=1}^{n}\prod_{k=1}^{m}\left(\phi_k\frac{\tau(1-\tau)}{\sigma_k}\exp\left\{-\rho_\tau\left(\frac{y_i-\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta}_k^{(j-1)}}{\sigma_k}\right)\right\}\right)^{z_{i,k}}\left(\frac{1}{\sqrt{2\pi}\sigma_{\beta_{d,k}}}\exp\left\{-\frac{1}{2\sigma^2_{\beta_{d,k}}}(\boldsymbol{\beta}_k^{(j-1)}-\mu_{\beta_k})^2\right\}\right)}\right\}$$
$$= \min\left\{1,\exp\left(\left[\sum_{i=1}^{n}\sum_{k=1}^{m}(z_{i,k})\rho_\tau\left(\frac{y_i-\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta_k^\star}}{\sigma_k}\right)(\boldsymbol{\beta}_k^\star-\mu_{\beta_k})^2\right]-\left[\sum_{i=1}^{n}\sum_{k=1}^{m}(z_{i,k})\rho_\tau\left(\frac{y_i-\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta_k^\star}}{\sigma_k}\right)(\boldsymbol{\beta}_k^{(j-1)}-\mu_{\beta_k})^2\right]\right)\right\}$$

B.3. *σ posterior and acceptance function*

$$\alpha(\sigma_k^{(j-1)},\sigma_k^{\star}) = \min\left\{1,\frac{L(\boldsymbol{y}|\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^{\star},\boldsymbol{z}_k^{(j-1)})\pi(\sigma_k^{\star})q(\sigma_k^{(j-1)}|\sigma_k^{\star})}{L(\boldsymbol{y}|\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^{(j-1)},\boldsymbol{z}_k^{(j-1)})\pi(\sigma_k^{(j-1)})q(\sigma_k^{\star}|\sigma_k^{(j-1)})}\right\}$$
$$= \min\left\{1,\frac{\prod_{i=1}^{n}\prod_{k=1}^{m}\left(\phi_k\frac{\tau(1-\tau)}{\sigma_k^\star}\exp\left\{-\rho_\tau\left(\frac{y_i-\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta_k}}{\sigma_k^\star}\right)\right\}\right)^{z_{i,k}}\left(\frac{\sigma_k^{(\star)a-1}\exp\left(-\sigma_k^\star/b\right)}{\Gamma(a)b^a}\right)}{\prod_{i=1}^{n}\prod_{k=1}^{m}\left(\phi_k\frac{\tau(1-\tau)}{\sigma_k^{(j-1)}}\exp\left\{-\rho_\tau\left(\frac{y_i-\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta_k}}{\sigma_k^{(j-1)}}\right)\right\}\right)^{z_{i,k}}\left(\frac{\sigma_k^{(\star)a-1}\exp\left(-\sigma_k^{(j-1)}/b\right)}{\Gamma(a)b^a}\right)}\right\}$$

$$= \min\left\{1,\exp\left(\sum_{i=1}^{n}\sum_{k=1}^{m}(a-z_{i,k}-1)\log(\sigma_k^\star)-z_{i,k}\left(\rho_\tau\left\{\frac{y_i-\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta_k}}{\sigma_k^\star}\right\}\right)-\frac{\sigma_k^\star}{b}\right.\right.$$
$$\left.\left.-\sum_{i=1}^{n}\sum_{k=1}^{m}(a-z_{i,k}-1)\log(\sigma_k^{(j-1)})-z_{i,k}\left(\rho_\tau\left\{\frac{y_i-\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta_k}}{\sigma_k^{(j-1)}}\right\}\right)-\frac{\sigma_k^{(j-1)}}{b}\right)\right\} \quad \text{(B 2)}$$

B.4. *$\boldsymbol{z}_i$ posterior and acceptance function*

$$\alpha(\boldsymbol{z}_i^{(j-1)},\boldsymbol{z}_i^{\star}) = \min\left\{1,\frac{L(\boldsymbol{y}|\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^{j},\boldsymbol{z}_i^{\star})\pi(\boldsymbol{z}_i^{\star}|\phi)q(\boldsymbol{z}_i^{(j-1)}|\boldsymbol{z}_i^{\star})}{L(\boldsymbol{y}|\beta_{d,k}^{(j)},\phi_k^{(j)},\sigma_k^{(j)},\boldsymbol{z}_i^{(j-1)})\pi(\boldsymbol{z}_i^{(j-1)}|\phi)q(\boldsymbol{z}_i^{\star}|\boldsymbol{z}_i^{(j-1)})}\right\}$$
$$= \min\left\{1,\frac{\prod_{i=1}^{n}\prod_{k=1}^{m}\left(\phi_k\frac{\tau(1-\tau)}{\sigma_k}\exp\left\{-\rho_\tau\left(\frac{y_i-\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta_k}}{\sigma_k}\right)\right\}\right)^{z_{i,k}^\star}\left(\frac{\Gamma(2)}{\prod_{k=1}^{m}\Gamma(z_{i,k}^\star+1)}\prod_{k=1}^{M}\phi_k^{z_{i,k}^\star}\right)}{\prod_{i=1}^{n}\prod_{k=1}^{m}\left(\phi_k\frac{\tau(1-\tau)}{\sigma_k}\exp\left\{-\rho_\tau\left(\frac{y_i-\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta_k}}{\sigma_k}\right)\right\}\right)^{z_{i,k}^{(j-1)}}\left(\frac{\Gamma(2)}{\prod_{k=1}^{m}\Gamma(z_{i,k}^{(j-1)}+1)}\prod_{k=1}^{M}\phi_k^{z_{i,k}^{(j-1)}}\right)}\right\}$$
$$= \min\left\{1,\exp\left(\sum_{i=1}^{n}\sum_{k=1}^{m}(z_{i,k}^\star-z_{i,k}^{(j-1)})\log\left[\left(\phi_k\frac{\tau(1-\tau)}{\sigma_k}\exp\left\{-\rho_\tau\left(\frac{y_i-\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta_k}}{\sigma_k}\right)\right\}\right)\left(\frac{\sum_{k=1}^{m}\Gamma(z_{i,k}^{(j-1)}+1)}{\sum_{k=1}^{m}\Gamma(z_{i,k}^\star+1)}\sum_{k=1}^{M}\phi_k\right)\right]\right)\right\}$$