

Bayesian estimation and variables selection for binary composite quantile regression

Taha Alshaybawee, Ahmad Naeem Flaih, Fadel Hamid Hadi Alhusseini

Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah

ABSTRACT

In this paper, Bayesian hierarchical model proposed to estimate the coefficients of the composite quantile regression model when the response variable is binary. For selecting variables in binary composite quantile regression lasso the adaptive lasso penalty is derived in a Bayesian framework. Simulation study and real data examples are used to examine the performance of the proposed methods compared to the other existing methods. We conclude that the proposed method is comparable.

Keywords: Bayesian regression, Composite quantile regression, Adoptive lasso, Prior distribution, Variable selection.

Corresponding Author:

Taha Alshaybawee

Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah,

Al Diwaniyah, Iraq

E-mail: taha.alshaybawee@qu.edu.iq

1. Introduction

Modelling the relationship between the average of a dependent variable Y with the set of explanatory variables X not always be convenient. In many application studies mean regression may be not appropriate to describe the behavior of outcome variable Y with covariates variable X . For example, the effect of demographic properties and maternal conduct on the weight of infant born was study by [1] in the United States. This study concerned in low birth weight for infant it is cause many health problems, these data analyze by standard mean regression, the conditional mean was not attractive approach for low tail distribution. Quantile regression (QReg) was proposed by [14] as an extension for standard mean regression to conditional different quantiles of a dependent variables. Quantile regression model is a capable of providing a complete information about different quantiles of the stochastic relationships between dependent and predictors variables. Recently, quantile regression has received much attention in theoretical and application study, it is applied in different field of study biology, medicine, survival analysis, financial and economics and environment for more detail see [26].

For any τ th quantile, ($0 < \tau < 1$), the τ th quantile regression can be denoted as $Q_{y_i|x_i}(\tau) = x'_i\beta_\tau$, where y_i is the response variable, x'_i is a k -dimensional vector, β_τ is a coefficient vector of quantile regression. To estimate the coefficient vector [14] proposed this equation

$$\sum_{i=1}^n \rho_\tau(y_i - x'_i\beta_\tau), \quad (1)$$

where $\rho_\tau(u) = \tau(1 - I(u < 0))$, $I(u < 0)$ is the indicator function this equation can be minimization by the algorithm proposed by [13]? Bayesian method employ to estimate quantile regression coefficient where the errors are independent asymmetric Laplace distribution [25].

In recent years, selection important subset of explanatory variables takes a lot of attention in the literature, many techniques suggest to get the important groups in the model for instance, lasso [24], SCAD [8], the elastic net method [29], adaptive Lasso [28]. Variables selection techniques has been used in quantile regression model, lasso penalty was applied to the mixed-effect QReg model or longitudinal data by [12], a solution path was introduced by [19] to L_1 -penalized for quantile regression model. Bayesian hierarchical model was proposed by [18], in lasso, group lasso and elastic net in quantile regression. In linear-mixed quantile regression A hierarchical Bayesian Lasso was used by [2] and was developed Bayesian adoptive lasso in quantile regression [3].

Binary quantile regression was developed for by [20,21], he was employed quantile regression in a classification and indicted to the drawbacks in the frequentist process as the difficulty optimization to estimate the parameters and the problem of computing confidence interval to the parameters. [16] studied models with binary response variable by quantile regression and was concluded this approach drive to good classification. Bayesian approach was adopted by [5] to avoid the drawback that mention above by setting some assumptions on the error term. [22] considered approach that proposed by [6] to evaluate the credit risk was modelled by binary quantile regression.

New method was proposed for estimating the coefficients in regression model called composite quantile regression (CQReg), and show the relative efficiency of these estimators is greater than 70% when compare with least square estimator regardless of the error distribution [30]. Composite quantile regression (CQReg) estimators are robust to the heavy tailed or outliers in the dependent variables and more efficient than a single quantile regression. For these characteristics we employ this approach in this study.

The novel in this paper, Bayesian hierarchical model proposed to estimate the coefficients of the composite quantile regression model when the response variable is binary. For selecting variables in binary composite quantile regression lasso and the adoptive lasso penalty is derived in a Bayesian framework.

2. Prior assumptions and hierarchical models

2.1. Bayesian Binary composite quantile regression model (BBCQReg)

Consider the following model

$$\begin{aligned} \check{y}_i &= b_\tau + x'_i\beta + \varepsilon_i, & i = 1, \dots, n, & y_i = h(\check{y}_i) \\ y_i &= \begin{cases} 1 & \text{if } \check{y}_i \geq 0 \\ 0 & \text{if } \check{y}_i < 0 \end{cases} \end{aligned} \tag{2}$$

Where y_i is an observed binary response variable determined by the unobserved scalar latent variable \check{y}_i , b_τ is the τ th quantile intercept parameter where $0 < \tau < 1$, x'_i is a p-dimensional vector of explanatory variables, β is a vector of coefficient, ε_i is the error term. For $(0 < \tau_1 < \tau_2 < \dots < \tau_q < 1)$, composite quantile regression parameters estimate by solving the following equation

$$\left(\hat{b}_{\tau_1}, \hat{b}_{\tau_2}, \dots, \hat{b}_{\tau_q}, \hat{\beta} \right) = \text{arg} \sum_{j=1}^q \left\{ \sum_{i=1}^n \rho_{\tau_j}(\check{y}_i - b_{\tau_j} - x'_i\beta) \right\}, \tag{3}$$

where $\rho_{\tau_j}(t) = t(\tau_j - I(t < 0))$, is the check function and $I(\cdot)$, is indicator function, $\tau_j = \frac{j}{q+1}$, $j = 1, 2, \dots, q$. Equation (3) is a mixture of the objective functions from different quantile models.

Quantile regression was adopted by many researchers to treat the models with binary response variable see for example [20,21,16,9,5,6]. Recently, composite quantile regression (CQReg) that proposed by [30] have been proven more efficient than the single quantile regression and more robust in the non-normal distribution for the error. To ameliorate the efficiency of CQReg a new estimation approach was proposed by [27], based on different weights of the components and they introduced a technique to estimate the optimal weight. Follow [27], weighted composite quantile regression proposed in this paper to model the binary response data, so Equation (3) rewrites as follow:

$$(\hat{b}_{\tau_1}, \hat{b}_{\tau_2}, \dots, \hat{b}_{\tau_q}, \hat{\beta}) = \arg \sum_{j=1}^q \left\{ \sum_{i=1}^n w_j \rho_{\tau_j}(\check{y}_i - b_{\tau_j} - x'_i \beta) \right\} \quad (4)$$

where the weight $0 \leq w_j \leq 1$, and $\sum_{j=1}^q w_j = 1$, for each component j th.

The latent variables $\check{y}_1, \check{y}_2, \dots, \check{y}_n$, come from an asymmetric Laplace distribution with parameters $ALD(\mu = b_{\tau} + x'_i \beta, \sigma = 1, \tau)$, for identification reasons σ set at unity, for more detail, see ([5, 15, 25]),

$$p(\check{y}_i | y_i, x_i, b_{\tau}, \beta, \tau) = \tau(1 - \tau) \exp(-\rho_{\tau}(\check{y}_i - b_{\tau} - x'_i \beta)).$$

So the joint distribution function of $\check{y} = (\check{y}_1, \check{y}_2, \dots, \check{y}_n)$ given $X = (x'_1, x'_2, \dots, x'_n)'$ is:

$$l = \prod_{j=1}^q \prod_{i=1}^n \tau(1 - \tau) \exp \exp \left(-w_j \rho_{\tau_j}(\check{y}_i - b_{\tau_j} - x'_i \beta) \right) \quad (5)$$

Maximization the likelihood function (5) is equivalent to Minimization the loss function (3), Equation (5) is difficult to solve directly, follow f [10] cluster assignment matrix K with the elements

$$k_{ij} = \begin{cases} 1 & \text{when } i\text{th belong to } j\text{th cluster} \\ 0 & \text{when } i\text{th does not belong to } j\text{th cluster} \end{cases}$$

Where k_{ij} Treat as missing value, the likelihood function will be as follows:

$$\prod_{j=1}^q \prod_{i=1}^n [w_j p(\check{y}_i | y_i, x_i, b_{\tau}, \beta, \tau)]^{k_{ij}} \quad (6)$$

To facilitate of calculations by using an MCMC algorithm a mixed representation of asymmetric Laplace distribution have been used, the error term can be written as a mixture of standard normal distribution and standard exponential see [17], suppose that $u \sim \exp(\frac{1}{\tau(1-\tau)})$ and $v \sim N(0,1)$. Therefore, the error term in (2) can be written as $\varepsilon = \theta u + \sqrt{\varphi uv}$, where $\theta = (1 - 2\tau)$ and $\varphi = 2$. Using MCMC algorithm with a mixed representation provided by [11] lead to converge at a geometric rate. Then the conditional distribution of \check{y}_i will be as follows:

$$p(\check{y}_i | y_i, X, b, \beta, u_i, w, K) = \exp \left(- \sum_{j=1}^q \sum_{i=1}^n \frac{k_{ij}}{4u_i} (\check{y}_i - b_{\tau_j} - x'_i \beta - \theta u_i)^2 \right) \prod_{i=1}^n (4\pi u_i)^{-\frac{k_{ij}}{2}} \quad (7)$$

The fully conditional distribution of \check{y}_i as we see in equation (7) is a mixture of two truncated normal distribution and can write as follow:

$$\check{y}_i|y_i, X, b, \beta, u_i \sim \begin{cases} N(b_{\tau_j} + x'_i\beta + \theta u_i, 2u_i) I(\check{y}_i > 0) & \text{if } y_i = 1 \\ N(b_{\tau_j} + x'_i\beta + \theta u_i, 2u_i) I(\check{y}_i < 0) & \text{if } y_i = 0 \end{cases} \quad (8)$$

The prior distribution for the coefficients $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ set as normal distribution $\pi(\beta) \sim N(0, 100)$, follow [10] Dirichlet prior distribution consider for the weight vector $w = (w_1, w_2, \dots, w_q)$, so the prior distribution is:

$$\pi(w) = \text{Dirichlet}(\delta_1, \delta_2, \dots, \delta_q)$$

The Bayesian hierarchical model for binary composite quantile regression will write as follows:

$$\begin{aligned} \check{y}|y, x, b_{\tau}, \beta, u &\sim N(b_{\tau_j} + x'_i\beta + \theta u_i, 2u_i) \\ \pi(\beta) &\sim N(0, 100) \\ p(u_i) &\sim \prod_{i=1}^n \tau(1 - \tau) \exp(-\tau(1 - \tau)u_i) \\ p(v_i) &\sim \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v_i^2\right) \end{aligned} \quad (9)$$

2.2. Bayesian Lasso penalty for binary composite quantile regression (BLBCQReg)

In this section, the lasso penalty for binary composite quantile regression was considered, coefficient can be estimated by the following

$$\sum_{j=1}^q \sum_{i=1}^n [w_j \rho_{\tau_j}(\check{y}_i - b_{\tau_j} - x'_i\beta)] + \lambda \sum_{k=1}^p |\beta_k| \quad (10)$$

Where the second term is l_1 the norm penalty for β , $\lambda > 0$, is the Lagrange multiplier. Laplace distribution put as a prior for β ,

$$\pi(\beta_k|\lambda, \sigma) = \frac{\sigma\lambda}{2} \exp\{-\sigma\lambda|\beta_k|\} \quad (11)$$

For any $b \geq 0$, Equation (11) can be expressed as follows [4]:

$$\frac{b}{2} \exp \exp(-b|z|) = \int_0^\infty \frac{1}{\sqrt{2\pi s}} \exp \exp\left(-\frac{z^2}{2s}\right) \frac{b^2}{2} \exp \exp\left(-\frac{b^2}{2}s\right) ds.$$

Let $\gamma = \sigma\lambda$, then the prior distribution for the coefficients β will be can be shown as follow

$$\pi(\beta_k|\gamma) = \prod_{k=1}^p \int_0^\infty \frac{1}{\sqrt{2\pi s_k}} \exp \exp \left(-\frac{\beta_k}{2s_k} \right) \frac{\gamma^2}{2} \exp \left(-\frac{\gamma^2}{2} s_k \right) ds_k. \quad (12)$$

Following [18], we considered gamma distribution as a prior for the parameter γ^2 , the Bayesian hierarchical model for binary composite quantile regression will be

$$\begin{aligned} \check{y}|y, x, b_\tau, \beta, u &\sim N(b_{\tau_j} + x'_i\beta + \theta u_i, 2u_i) \\ p(u_i) &\sim \prod_{i=1}^n \tau(1 - \tau) \exp \exp (-\tau(1 - \tau)u_i) \end{aligned} \quad (13)$$

$$p(v_i) \sim \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2}v_i^2 \right)$$

$$\pi(\beta, s|\gamma^2) = \prod_{k=1}^p \int_0^\infty \frac{1}{\sqrt{2\pi s_k}} \exp \exp \left(-\frac{\beta_k}{2s_k} \right) \frac{\gamma^2}{2} \exp \exp \left(-\frac{\gamma^2}{2} s_k \right) ds_k.$$

$$p(\gamma^2) = (\gamma^2)^{a-1} \exp (-b\gamma^2)$$

where $s = (s_1, s_2, \dots, s_p)$, $a, b > 0$, are the hyperparameters.

2.3. Bayesian adaptive Lasso penalty for binary composite quantile regression (BALBCQReg)

Adaptive lasso penalty [28] was extended for a lasso approach that proposed by [24]. Adaptive lasso has been proved by [28] yields a consistent estimate and enjoys oracles properties [8] while in high dimensional. Binary composite quantile regression with adaptive lasso penalty estimates by solving the following

$$\sum_{j=1}^q \sum_{i=1}^n \left[w_j \rho_{\tau_j}(\check{y}_i - b_{\tau_j} - x'_i\beta) \right] + \sum_{k=1}^p \lambda_k |\beta_k| \quad (14)$$

where $\lambda_k > 0$, λ_k are weighted penalty parameters, the second term $\sum_{k=1}^p \lambda_k |\beta_k|$ is the adaptive weighted penalty for parameter selection. Follow [3], Laplace prior distribution was setting for the parameters β_k and the form is

$$\pi(\beta_k|\lambda_k, \sigma) = \frac{\sigma^{\frac{1}{2}}}{2\lambda_k} \exp \exp \left(-\frac{\sigma^{\frac{1}{2}}|\beta_k|}{\lambda_k} \right) \quad (15)$$

Equation (15) can be expressed as a mixture of exponential and normal mix function [4]. For any $b > 0$, then

$$\frac{b}{2} \exp \exp (-b|z|) = \int_0^\infty \frac{1}{\sqrt{2\pi s}} \exp \exp \left(-\frac{z^2}{2s} \right) \frac{b^2}{2} \exp \exp \left(-\frac{b^2}{2} s \right) ds.$$

Let $\eta_k = \frac{\sigma^{\frac{1}{2}}}{\lambda_k}$, so the prior distribution for β_k can be rewritten as follows:

$$\pi(\beta_k|\eta_k) = \prod_{k=1}^p \int_0^\infty \frac{1}{\sqrt{2\pi r_k}} \exp \exp \left(-\frac{\beta_k^2}{2r_k} \right) \frac{\eta_k^2}{2} \exp \exp \left(-\frac{\eta_k^2}{2} r_k \right) dr_k. \quad (16)$$

For the parameters η_k^2 Gamma distribution set as a prior distribution [3].

The Bayesian hierarchical model for binary composite quantile regression with an adaptive lasso penalty can be shown as follows:

$$\begin{aligned}
 \check{y}|y, x, b_\tau, \beta, u &\sim N(b_{\tau_j} + x'_i\beta + \theta u_i, 2u_i) \\
 p(u_i) &\sim \prod_{i=1}^n \tau(1 - \tau) \exp \exp(-\tau(1 - \tau)u_i) \\
 p(v_i) &\sim \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}v_i^2) \\
 \pi(\beta, r_k | \eta_k^2) &= \prod_{k=1}^p \int_0^\infty \frac{1}{\sqrt{2\pi r_k}} \exp \exp\left(-\frac{\beta_k^2}{2r_k}\right) \frac{\eta_k^2}{2} \exp\left(-\frac{\eta_k^2}{2}r_k\right) dr_k. \\
 p(\eta_k^2) &= \prod_{k=1}^p (\eta_k^2)^{c-1} \exp(-d\eta_k^2)
 \end{aligned}
 \tag{17}$$

where, $b > 0$ are the hyperparameters.

Under the hierarchical Bayesian models (9), (13) and (17), the Gibbs sampler algorithm is used to sample and update the parameters. The full conditional distributions for the three methods above are derivative in Appendix to get the posterior distribution.

3. Simulation scenarios

In this section, simulation scenarios will be considered to investigating our proposed methods compared to some other existing methods; binary regression quantiles (BRQ) which is proposed by [20] and Bayesian lasso binary quantile regression (BBRQL) that method was proposed by [6]. Four quantiles are used $\tau = (0.20, 0.40, 0.60, 0.80)$. For each simulation scenario, random error ε_i are generated from four different distributions: Normal distribution with mean (0) and variance (1), $\varepsilon_i \sim N(0,1)$ mixture of normal distributions $N(1,1)$ and $N(-1,1)$, student distribution with (3) degree of freedom, $\varepsilon_i \sim t_{(3)}$ and Laplace distribution with location parameter (0) and scale parameter (1), $\varepsilon_i \sim Lap(0,1)$. In each simulation scenario, the algorithm run 13000 iterations and the first 3000 were takeout as burn in. The consideration methods are evaluated based on the root mean square error (RMSE) and mean absolute error (MAE). In our study, we will used three simulation scenarios, the computations were done by using R package.

3.1. First simulation

In the first simulation scenario, very sparse model will be used within the model:

$$\check{y}_i = 5x_{ij} + \varepsilon_i \quad \text{where } i = 1, 2, \dots, 100 \text{ and } j = 1, \dots, 9.$$

The covariate variables are generated from the standard Uniform (0,1), where the true parameters of independent variables are $\beta = (0, 5, 0, 0, 0, 0, 0, 0, 0)$.

The (RMSE) and (MAE) are listed in Table 1. We can readily saw across all the quantile levels, our proposed methods (BBCQReg, BLBCQReg and BALBCQReg) have high performance compared to the BRQ and BBRQL.

In general, Table (1) show that the (RMSE) and (MAE) for the proposed methods BBCQReg, BLBCQReg and BALBCQReg are smaller than that for existing methods.

Table 1. RMSE and MAE for the first simulation scenario

| Methods under study | Error distributions | | | |
|---------------------|-----------------------------|--|------------------------------|-------------------------------|
| | $\varepsilon_i \sim N(0,1)$ | $\varepsilon_i \sim \text{Normal mix}$ | $\varepsilon_i \sim t_{(3)}$ | $\varepsilon_i \sim Lap(0,1)$ |
| | | | | |

| | | | | |
|-----------------------|---------------|---------------|---------------|---------------|
| $BRQ_{\tau_1=0.20}$ | 0.754 (0.147) | 0.842 (0.212) | 0.645 (0.142) | 0.724 (0.154) |
| $BRQ_{\tau_2=0.40}$ | 0.864 (0.154) | 0.987 (0.217) | 0.749 (0.356) | 0.784 (0.128) |
| $BRQ_{\tau_3=0.60}$ | 0.971 (0.318) | 0.871 (0.259) | 0.741 (0.445) | 0.687 (0.154) |
| $BRQ_{\tau_4=0.80}$ | 1.245 (0.341) | 1.545 (0.231) | 1.754 (0.421) | 1.457 (0.215) |
| $BBRQL_{\tau_1=0.20}$ | 0.647 (0.218) | 0.782 (0.317) | 0.974 (0.151) | 1.245 (0.417) |
| $BBRQL_{\tau_2=0.40}$ | 0.451 (0.412) | 0.841 (0.265) | 0.872 (0.124) | 0.148 (0.421) |
| $BBRQL_{\tau_3=0.60}$ | 0.874 (0.145) | 0.947 (0.248) | 0.974 (0.254) | 0.879 (0.457) |
| $BBRQL_{\tau_4=0.80}$ | 1.974 (0.287) | 1.674 (0.574) | 1.784 (0.697) | 0.975 (0.647) |
| BBCQReg | 0.412 (0.107) | 0.347 (0.119) | 0.417 (0.207) | 0.292 (0.097) |
| BLBCQReg | 0.318 (0.108) | 0.292 (0.088) | 0.642 (0.467) | 0.321 (0.247) |
| BALBCQReg | 0.487 (0.058) | 0.347 (0.217) | 0.528 (0.092) | 0.541 (0.124) |

In the parentheses are belong to MAE.

3.2. Second simulation

In the second simulation scenario, we will used dense case within the true our model:

$$\check{y}_i = 0.85x_{i1} + 0.85x_{i2} + 0.85x_{i3} + 0.85x_{i4} + 0.85x_{i5} + 0.85x_{i6} + 0.85x_{i7} + 0.85x_{i8} + \varepsilon_i \quad \text{where } i = 1 \dots 100$$

The independent variables also are generated from the standard Uniform (0,1). where the true parameters of independent variables, including the intercept, are $\beta = 0, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85$.

In Table 2, we show the summary of the RMSE and MAE for the five methods under study. From this table we can see that clearly our three proposed methods BBCQReg, BLBCQReg and BALBCQReg are get the smallest values of RMSE and MAE during all error distribution.

Table 2. RMSE and MAE for the second simulation scenario.

| Methods under study | Error distributions | | | |
|-----------------------|-----------------------------|--|------------------------------|--------------------------------------|
| | $\varepsilon_i \sim N(0,1)$ | $\varepsilon_i \sim \text{Normal mix}$ | $\varepsilon_i \sim t_{(3)}$ | $\varepsilon_i \sim \text{Lap}(0,1)$ |
| $BRQ_{\tau_1=0.20}$ | 0.865 (0.215) | 0.926 (0.219) | 0.743 (0.211) | 0.682 (0.197) |
| $BRQ_{\tau_2=0.40}$ | 0.925 (0.129) | 0.869 (0.186) | 0.795 (0.254) | 0.659 (0.156) |
| $BRQ_{\tau_3=0.60}$ | 0.828 (0.398) | 0.936 (0.359) | 0.854 (0.523) | 1.225 (0.524) |
| $BRQ_{\tau_4=0.80}$ | 1.678 (0.354) | 1.892 (0.245) | 1.828(0.502) | 1.871 (0.547) |
| $BBRQL_{\tau_1=0.20}$ | 0.781 (0.326) | 0.745 (0.492) | 0.624 (0.125) | 0.598 (0.438) |
| $BBRQL_{\tau_2=0.40}$ | 0.645 (0.641) | 0.648 (0.128) | 0.846 (0.065) | 0.648 (0.154) |
| $BBRQL_{\tau_3=0.60}$ | 0.924 (0.105) | 0.862 (0.135) | 0.792 (0.121) | 0.754 (0.214) |
| $BBRQL_{\tau_4=0.80}$ | 2.542 (0.412) | 1.524 (0.321) | 1.421 (0.654) | 1.135 (0.266) |
| BBCQReg | 0.354 (0.057) | 0.451 (0.110) | 0.531 (0.198) | 0.311 (0.158) |
| BLBCQReg | 0.298 (0.175) | 0.304 (0.091) | 0.521 (0.350) | 0.214 (0.114) |
| BALBCQReg | 0.385 (0.124) | 0.218 (0.017) | 0.124 (0.063) | 0.411 (0.151) |

In the parentheses are belong to MAE

3.3. Third simulation

In the third simulation scenario, we will used sparse case with in the true our model:

$$\check{y}_i = 3x_{i1} + x_{i2} + x_{i5} + \varepsilon_i \quad \text{where } i = 1 \dots 100$$

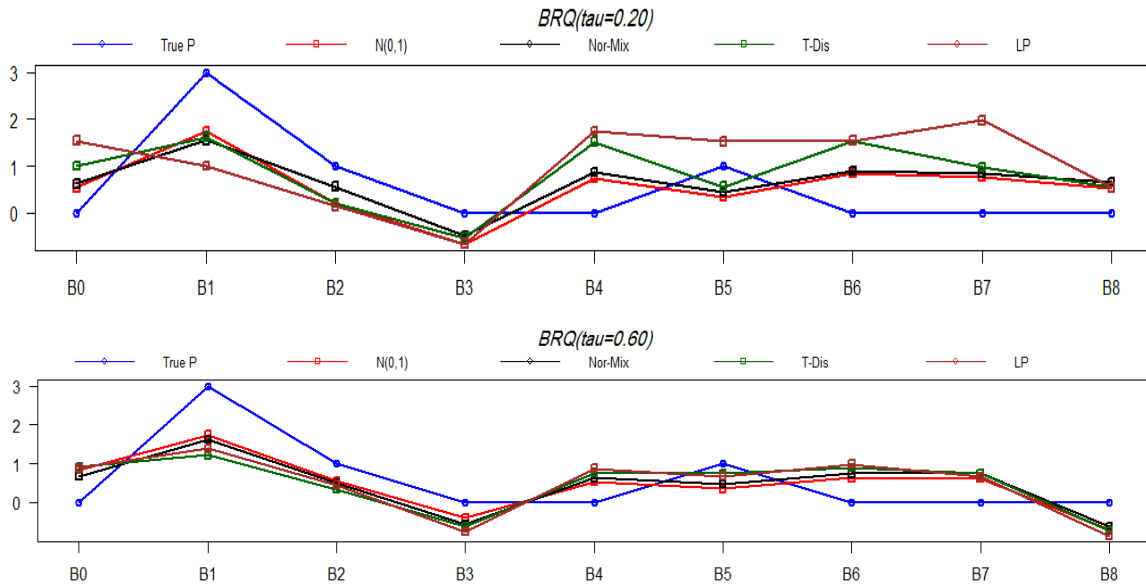
Similar to second Simulation the independent variables also are generated from the standard Uniform (0,1). Where the true parameters of independent variables, including the intercept, are $\beta = 0,3,1,0,0,1,0,0,0$.

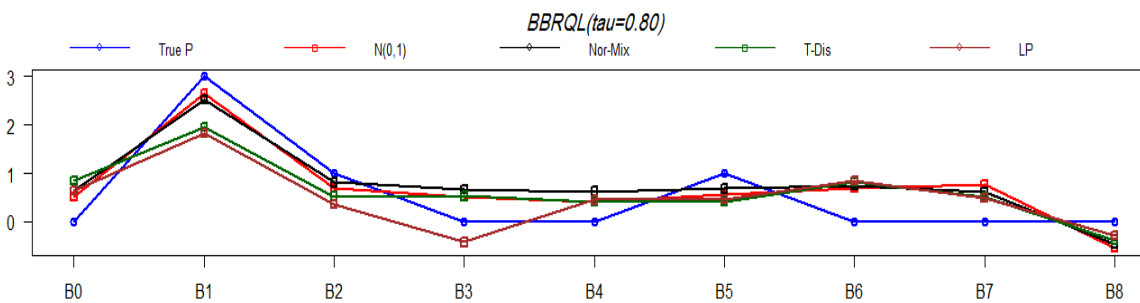
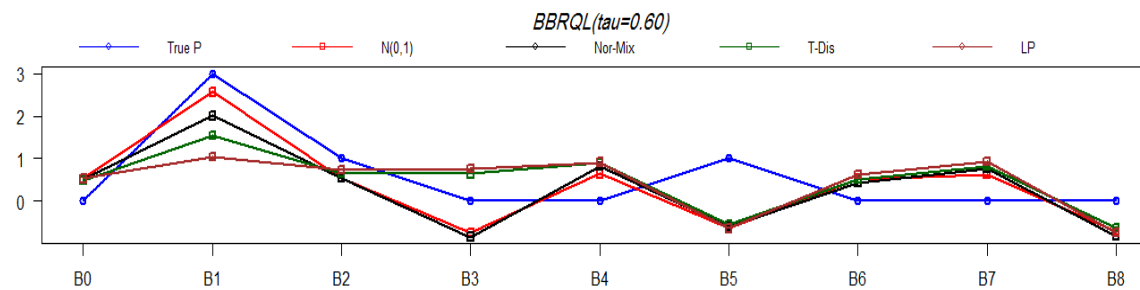
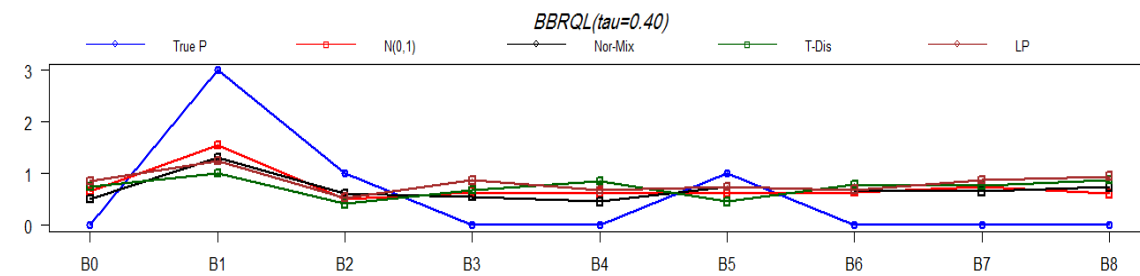
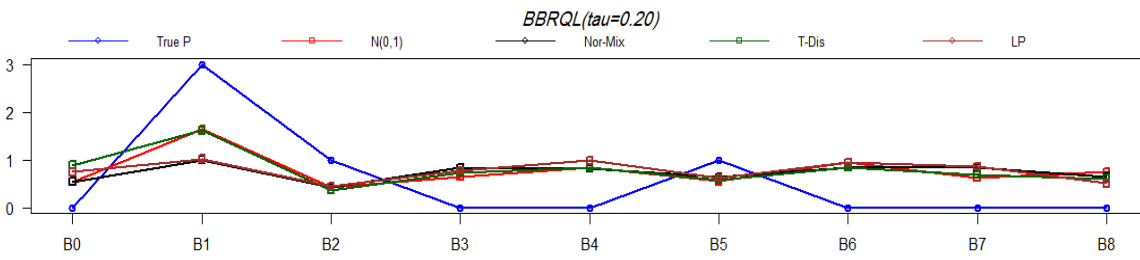
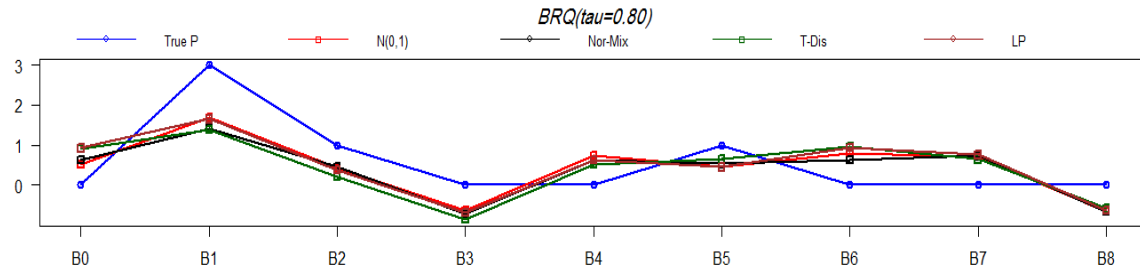
Table 3. RMSE and MAE for the third simulation scenario.

| Methods under study | Error distributions | | | |
|-----------------------|-----------------------------|--|------------------------------|--------------------------------------|
| | $\varepsilon_i \sim N(0,1)$ | $\varepsilon_i \sim \text{Normal mixture}$ | $\varepsilon_i \sim t_{(3)}$ | $\varepsilon_i \sim \text{Lap}(0,1)$ |
| $BRQ_{\tau_1=0.20}$ | 0.954 (0.124) | 0.857 (0.195) | 0.695 (0.247) | 0.726 (0.215) |
| $BRQ_{\tau_2=0.40}$ | 0.845 (0.254) | 0.695 (0.251) | 0.647 (0.154) | 0.652 (0.224) |
| $BRQ_{\tau_3=0.60}$ | 0.765 (0.259) | 0.854 (0.135) | 0.625 (0.185) | 0.684 (0.128) |
| $BRQ_{\tau_4=0.80}$ | 1.845 (0.121) | 1.913 (0.182) | 1.772(0.231) | 1.647 (0.315) |
| $BBRQL_{\tau_1=0.20}$ | 0.645 (0.254) | 0.628 (0.205) | 0.692 (0.254) | 0.855 (0.358) |
| $BBRQL_{\tau_2=0.40}$ | 0.851 (0.234) | 0.705 (0.206) | 0.877 (0.301) | 0.705 (0.242) |
| $BBRQL_{\tau_3=0.60}$ | 0.953 (0.181) | 0.765 (0.104) | 0.854 (0.314) | 0.824 (0.247) |
| $BBRQL_{\tau_4=0.80}$ | 1.664 (0.193) | 1.547 (0.405) | 1.625 (0.431) | 1.685 (0.371) |
| BBCQReg | 0.465 (0.102) | 0.354 (0.152) | 0.365 (0.112) | 0.282 (0.106) |
| BLBCQReg | 0.354 (0.205) | 0.216 (0.103) | 0.342 (0.116) | 0.354 (0.214) |
| BALBCQReg | 0.265 (0.084) | 0.312 (0.117) | 0.225 (0.098) | 0.257 (0.101) |

In the parentheses are belong to MAE.

As same as Table 1 and Table 2, Table 3 show the values of RMSE and MAE of all methods proposed and existing. This table showing that proposed methods BBCQReg, BLBCQReg and BALBCQReg are better than the other existing methods BRQ and BBRQL. Whereas the results showing that the RMSE and MAE for the proposed methods BBCQReg, BLBCQReg and BALBCQReg are smaller than that for the other methods existing method (BRQ and BBRQL). In this paper, we will use another procedure for the evaluating the methods under study is parameters estimations in direct way. Figure 1 shows plot of parameter estimations in the second simulation. From the Figure 1, it can see that our proposed methods are estimated the parameters very close to the true parameters compared with other methods. As the result, our proposed methods have high efficiency compared with previous methods.





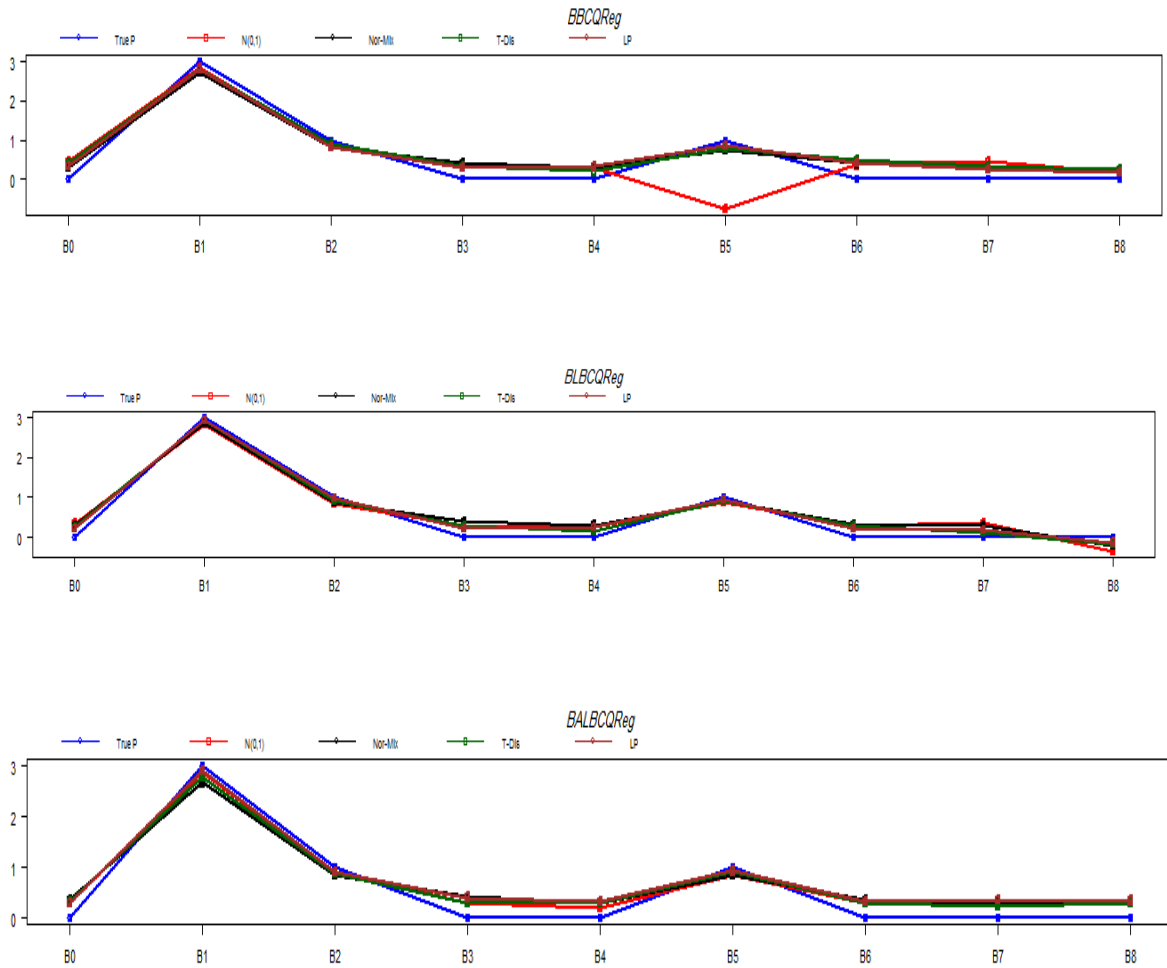


Figure 1. Plot of parameter estimations in in direct way for the second simulation to the methods under study.

4. Real data

To clarification our proposed methods for (BBCQReg, BLBCQReg and BALBCQReg) and compare with *BRQ* and *BBRQL* approaches, the data of Pima Indians have been considered, these data set available in the caret package in R programs, from [23]. The Pima Indians Diabetes data consists of (532) observations of which 200 are test positive observations and 332 are test positive observations. The important part in the study of Pima Indians is in achieving the relationship between diabetic according to WHO criteria, (*diabetes*) and eight independent variables. The seven independent variables are X_1 : Number of pregnancies code by (npreg), X_2 : Plasma glucose concentration in an oral glucose tolerance test code by (glu), X_3 : diastolic blood pressure (mm Hg) code by (bp), X_4 : triceps skin fold thickness (mm) code by (skin), X_5 : body mass index (weight in kg/(height in m)²) code by (bmi), X_6 : diabetes pedigree function code by (ped) and X_7 : age in years code by (age).

As same as Section three, in this section we compare our proposed methods with the previous two existing methods are assessed based on the mean squared error (MSE).

Table 4. MSE for the methods under comparison

| Methods | $t = 0.16$ | $t = 0.33$ | $t = 0.50$ | $t = 0.66$ | $t = 0.83$ | MSE | MSE | MSE |
|-----------|------------|------------|------------|------------|------------|-------|-------|-------|
| | MSE | MSE | MSE | MSE | MSE | | | |
| BRQ | 12.674 | 12.435 | 11.102 | 11.009 | 11.003 | | | |
| BBRQL | 11.414 | 11.341 | 10.652 | 10.766 | 10.112 | | | |
| BBCQReg | | | | | | 8.734 | | |
| BLBCQReg | | | | | | | 8.104 | |
| BALBCQReg | | | | | | | | 6.485 |

From the results are listed in Table 4, the MSE of our proposed methods Bayesian Binary composite quantile regression model, Bayesian lasso penalty for binary composite quantile regression and Bayesian adaptive lasso penalty for binary composite quantile regression are 8.734, 8.104 and 6.485 respectively. Where, the MSE generated by our proposed methods are much smaller than MSE generated by BRQ and BBRQL methods via all quantile levels. So, our methods have performance better than previous methods (BRQ and BBRQL).

5. Conclusion

In this paper, Bayesian binary composite quantile regression approach is proposed for estimating the model. In binary composite quantile regression lasso and the adoptive lasso penalty are considered for selecting variables in a Bayesian framework. We developed a Bayesian hierarchical model for the lasso and adaptive lasso penalty methods, whereas Gibbs sampler algorithm was adopted for posterior inference.

Simulation examples and real data were considered to compare our proposed methods BBCQReg, BLBCQReg and BALBCQReg with other methods, BRQ and BBRQL with different quantiles. Based on the simulation scenarios the RMSE and MAE of our proposed methods BBCQReg, BLBCQReg and BALBCQReg are smaller than the RMSE and MAE of the other existing methods BRQ and BBRQL.

The numerical study signify that the proposed methods offers substantial improvement over the other two methods. So that, we concluded that our proposed methods perform better than the other two existing methods.

References

- [1] J. Abrevaya, "The effects of demographics and maternal behavior on the distribution of birth outcomes," *Economic applications of quantile regression* (pp. 247-257). Physica, Heidelberg, 2002.
- [2] R. Alhamzawi, and K. Yu, "Bayesian Lasso-mixed quantile regression." *Journal of Statistical Computation and Simulation*, vol. 84, no. 4, p. 868-880. 2014.
- [3] R. Alhamzawi, K. Yu, and D. F. Benoit, "Bayesian adaptive Lasso quantile regression." *Statistical Modelling*, vol. 12, no. 3, p. 279-297, 2012.
- [4] D. F. Andrews, and C. L. Mallows, "Scale mixtures of normal distributions." *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no.1, p. 99-102, 1974.
- [5] D. F. Benoit, and D. Van den Poel, "Binary quantile regression: A Bayesian approach based on the asymmetric Laplace distribution." *Journal of Applied Econometrics*, vol. 27, no.7, p. 1174-1188, 2012.
- [6] D. F. Benoit, R. Alhamzawi and K. Yu, "Bayesian lasso binary quantile regression." *Computational Statistics*, vol. 28, no. 6, p. 2861-2873, 2013.
- [7] S. A. Campbell, and Y. Yuan, "Zero singularities of codimension two and three in delay differential equations." *Nonlinearity*, vol. 21, no. 11, p. 2671-2691, 2008.
- [8] J. Fan, and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties." *Journal of the American statistical Association*, vol. 96, no.456, p. 1348-1360, 2001.

-
- [9] J. L. Horowitz, "A smoothed maximum score estimator for the binary response model." *Econometrica: Journal of the Econometric Society*, vol. 60, no. 3, p. 505-531, 1992.
- [10] H. Huang, and Z. Chen, "Bayesian composite quantile regression." *Journal of Statistical Computation and Simulation*, vol. 85, no. 18, p. 3744-3754, 2015.
- [11] K. Khare, and J. P. Hobert, "Geometric ergodicity of the Gibbs sampler for Bayesian quantile regression." *Journal of Multivariate Analysis*, vol. 112, p. 108-116, 2012.
- [12] R. Koenker, "Quantile regression for longitudinal data." *Journal of Multivariate Analysis*, vol. 91, no. 1, p. 74-89, 2004.
- [13] R. Koenker, and V. D'Orey, "Algorithm AS 229: Computing regression quantiles." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 36, no. 3, p. 383-393, 1987.
- [14] R. Koenker, and JR. Bassett, "Regression quantiles." *Econometrica: journal of the Econometric Society*, vol. 46, no. 1, p. 33-50, 1978.
- [15] R. Koenker, and J. A. Machado, "Goodness of fit and related inference processes for quantile regression." *Journal of the American Statistical Association*, vol. 94, no. 448, p. 1296-1310, 1999.
- [16] G. Kordas, "Smoothed binary regression quantiles." *Journal of Applied Econometrics*, vol. 21, no. 3, p. 387-407, 2006.
- [17] H. Kozumi, & G. Kobayashi, "Gibbs sampling methods for Bayesian quantile regression." *Journal of statistical computation and simulation*, vol. 81, no. 11, p. 1565-1578, 2011.
- [18] Q. Li, R. Xi, and N. Lin, "Bayesian regularized quantile regression." *Bayesian Analysis*, vol. 5, no. 3, p. 533-556, 2010.
- [19] Y. Li, and J. Zhu, " L_1 -norm quantile regression." *Journal of Computational and Graphical Statistics*, vol. 17, no.1, p.163-185, 2008.
- [20] C. F. Manski, "Maximum score estimation of the stochastic utility model of choice." *Journal of econometrics*, vol. 3, no. 3, 205-228, 1975.
- [21] C. F. Manski, "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator." *Journal of econometrics*, vol. 27, no. 3, p. 313-333, 1985.
- [22] V. L. Miguéis, D. F. Benoit, and D. Van den Poel, "Enhanced decision support in credit scoring using Bayesian binary quantile regression." *Journal of the Operational Research Society*, vol. 64, no. 9, p. 1374-1383, 2013.
- [23] J. W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, and R. S. Johannes, "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus." *Proceedings of the Annual Symposium on Computer Application in Medical Care*, p. 261-265, Nov. 1988.
- [24] R. Tibshirani, "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, p. 267-288., 1996.
- [25] K. Yu, and R. A. Moyeed, "Bayesian quantile regression." *Statistics & Probability Letters*, vol. 54, no. 4, p. 437-447, 2001.
- [26] K. Yu, Z. Lu, & J. Stander, "Quantile regression: applications and current research areas." *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 52, no. 3, p. 331-350, 2003.
- [27] Z. Zhao, and Z. Xiao, "Efficient regressions via optimally combining quantile information." *Econometric theory*, vol. 30, no. 6, p. 1272-1314, 2014.
- [28] H. Zou, "The adaptive lasso and its oracle properties." *Journal of the American statistical association*, vol. 101, no. 476, p. 1418-1429, 2006.
- [29] H. Zou, and T. Hastie, "Regularization and variable selection via the elastic net." *Journal of the royal statistical society: series B (statistical methodology)*, vol. 67, no. 2, p. 301-320, 2005.
-

[30] H. Zou, and M. Yuan, “Composite quantile regression and the oracle model selection theory.” *The Annals of Statistics*, vol. 36, no. 3, 1108-1126, 2008.

Appendix

I. Gibbs sampler Bayesian binary composite quantile regression

The conditional distribution $f(\beta_k | \check{y}_i, X, \beta_{-k}, b_\tau, u_i, w, k)$, where β_{-k} is the parameters vector excepting β_k , is normal

$$\begin{aligned} f(\beta_k | \check{y}_i, X, \beta_{-k}, b, u, w, K) &\propto f(\check{y}_i | X, \beta, b, u, w, K) \times \pi(\beta) \\ &\propto \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^n \frac{(50x'_{ik} + u_i)}{100u_i} \beta_k^2 - 2 \sum_{i=1}^n \frac{50\check{y}_i^* x'_{ik}}{100u_i} \beta_k \right] \right\} \\ &\propto N \left[\frac{\sum_{i=1}^n \left(\frac{50k_{ij}\check{y}_i^* x_{ik}}{100u_i} \right)}{\sum_{i=1}^n \left(\frac{(50k_{ij}x'_{ik} + u_i)}{100u_i} \right)}, \frac{1}{\sum_{i=1}^n \left(\frac{(50k_{ij}x'_{ik} + u_i)}{100u_i} \right)} \right] \\ \check{y}_i^* &= \check{y}_i - b_{\tau_j} - x'_{i,-k}\beta_{-k} - \theta u_i \end{aligned}$$

The conditional distribution $f(u_i | \check{y}_i, X, \beta, b, u_{-i}, w, K)$, where u_{-i} is the variable u excepting the component u_i is given by

$$\begin{aligned} f(u_i | \check{y}_i, X, \beta, b, u_{-i}, w, K) &\propto f(\check{y}_i | X, \beta, b, u, w, K) \times \pi(u_i | \tau) \\ &\propto (u_i)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \sum_{j=1}^q k_{ij} \left[\left(\frac{(\check{y}_i - b_{\tau_j} - x'_{i,-k}\beta_{-k})^2}{2} \right) u_i^{-1} + \left(\frac{\theta^2}{2} + 2\tau(1 - \tau) \right) u_i \right] \right\} \end{aligned}$$

So the conditional distribution of u_i is generalized inverse Gaussian $GIG(u_i, A, B)$ where $A = \frac{(\check{y}_i - b_{\tau_j} - x'_{i,-k}\beta_{-k})^2}{2}$ and $B = \frac{\theta^2}{2} + 2\tau(1 - \tau)$.

The conditional distribution $f(b_{\tau_k} | \check{y}_i, X, \beta, b_{-\tau_k}, u, w, K)$ is a normal distribution

$$\begin{aligned} f(b_{\tau_k} | \check{y}_i, X, \beta, b_{-\tau_k}, u, w, K) &\propto f(\check{y}_i | X, \beta, b, u, w, K) \\ &\propto \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^n \frac{k_{ij}}{2u_i} b_{\tau_k}^2 - 2 \sum_{i=1}^n \frac{k_{ij}\check{y}_i^{**}}{2u_i} b_{\tau_k} \right] \right\} \\ &\propto N \left(\frac{\sum_{i=1}^n \left(\frac{k_{ij}\check{y}_i^{**}}{2u_i} \right)}{\sum_{i=1}^n \left(\frac{k_{ij}}{2u_i} \right)}, \frac{1}{\sum_{i=1}^n \left(\frac{k_{ij}}{2u_i} \right)} \right) \\ \check{y}_i^{**} &= \check{y}_i - x'_i\beta - \theta u_i \end{aligned}$$

Follow [10] the conditional distribution $f(w|\check{y}_i, X, \beta, b, u, K)$ is given as follow

$$f(w|X, \check{y}_i, \beta, b_{\tau_k}, u_i, k) \propto \prod_{j=1}^q w_j^{n_j + \delta_j}$$

$$\propto \text{Dirichlet}(n_1 + \delta_1, \dots, n_q + \delta_q)$$

Where n_j is the summation of the objects k_{ij} in the j th cluster, i.e. $\sum_{i=1}^n k_{ij}$.

The conditional distribution $f(k_i|X, \check{y}_i, \beta, b, u, w, k_{-i})$ can be shown as ([10])

$$f(k_i|X, \check{y}_i, \beta, b, u, w, k_{-i})$$

$$\propto \prod_{j=1}^q \left\{ w_j \exp \left[-\frac{1}{4u_i} (\check{y}_i - b_{\tau_j} - x_i' \beta - \theta u_i)^2 \right] \right\}^{k_{ij}}$$

$$\propto \text{Multinomial}(1, \hat{p}_1, \dots, \hat{p}_q)$$

Where

$$\hat{p}_j = \frac{w_j \exp \left[-\frac{1}{4u_i} (\check{y}_i - b_{\tau_j} - x_i' \beta - \theta u_i)^2 \right]}{\sum_{j=1}^q w_j \exp \left[-\frac{1}{4u_i} (\check{y}_i - b_{\tau_j} - x_i' \beta - \theta u_i)^2 \right]}$$

II. Gibbs sampler Bayesian lasso binary composite quantile regression:

The conditional distribution of u, b, w and K will be as same as that in Appendix I. The conditional distribution of γ^2 is a Gamma distribution

$$f(\gamma^2|X, \check{y}_i, \beta, b, u, s, w, K) \propto f(s|\gamma^2) \times \pi(\gamma^2)$$

$$\propto (\gamma^2)^{p+a-1} \exp \left\{ -\left(\frac{1}{2} \sum_{k=1}^p s_k + b \right) \gamma^2 \right\}$$

$$\propto \text{Gamma}(p + a, \frac{1}{2} \sum_{k=1}^p s_k + b)$$

The conditional distribution $f(\beta_k|X, \check{y}_i, \beta_{-k}, b, u, s, \gamma^2, w, K)$, where β_{-k} is a parameters vector excepting β_k is a normal distribution

$$f(\beta_k|X, \check{y}_i, \beta_{-k}, b, u, s, \gamma^2, w, K) \propto f(\check{y}_i|X, \beta, b, u, s, \gamma^2, w, K) \times \pi(\beta_k|s_k)$$

$$\propto \exp \left\{ -\frac{1}{2} \left[\left(\sum_{j=1}^q \sum_{i=1}^n \frac{k_{ij} x_{ik}^2}{2u_i} + \frac{1}{s_k} \right) \beta_k^2 - 2 \sum_{j=1}^q \sum_{i=1}^n \frac{k_{ij} x_{ik} \check{y}_i}{2u_i} \beta_k \right] \right\}$$

$$\propto N \left[\frac{\sum_{j=1}^q \sum_{i=1}^n \left(\frac{k_{ij} x_{ik} \check{y}_i}{2u_i} \right)}{\sum_{j=1}^q \sum_{i=1}^n \left(\frac{k_{ij} x_{ik}^2}{2u_i} \right) + \left(\frac{1}{s_k} \right)}, \frac{1}{\sum_{j=1}^q \sum_{i=1}^n \left(\frac{k_{ij} x_{ik}^2}{2u_i} \right) + \left(\frac{1}{s_k} \right)} \right]$$

$$\check{y}_i = \check{y}_i - b_{\tau_j} - x'_{i,-k} \beta_{-k} - \theta u_i$$

The conditional distribution $f(s_k | X, \check{y}_i, \beta, b, u, s_{-k}, \gamma^2, w, K)$, where s_{-k} is the variable s excepting the component s_k is generalized inverse Gaussian

$$\begin{aligned} f(s_k | X, \check{y}_i, \beta, b, u, s_{-k}, \gamma^2, w, K) &\propto \pi(s_k | \gamma^2) \times \pi(\beta_k | \gamma^2) \\ &\propto s_k^{-\frac{1}{2}} \exp \exp \left\{ -\frac{1}{2} (\beta_k^2 s_k^{-1} + \gamma^2 s_k) \right\} \\ &\propto GIG(s_k, \beta_k^2, \gamma^2) \end{aligned}$$

III. Gibbs sampler Bayesian adaptive lasso binary composite quantile regression

As same as we mentioned in Appendix II the conditional distribution for u, b, w and K will not be change. The conditional distribution of η_k^2 is a Gamma distribution and derivative as follows

$$\begin{aligned} f(\eta_k^2 | X, \check{y}_i, \beta, b, u_i, r, \eta_{-k}^2, w, k) &\propto f(r | \eta_k^2) \times \pi(\eta_k^2 | c, d) \\ &\propto (\eta_k^2)^c \exp \exp \left(-\left(\frac{r_k}{2} + d \right) \eta_k^2 \right) \\ &\propto \text{Gamma} \left(c + 1, \left(\frac{r_k}{2} + d \right) \right) \end{aligned}$$

The conditional distribution $f(r_k | X, \check{y}_i, \beta, b, u, r_{-k}, \eta_k^2, w, k)$, where r_{-k} is the variable r excepting the component r_k is a generalized inverse Gamma distribution and given by

$$\begin{aligned} f(r_k | X, \check{y}_i, \beta, b, u_i, r_{-k}, \eta_k^2, w, k) &\propto f(r_k | \eta_k^2) \times \pi(\beta_k | \eta_k^2) \\ &\propto (r_k)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\beta_k^2 r_k^{-1} + \eta_k^2 r_k) \right\} \\ &\propto GIG(r_k, \beta_k^2, \eta_k^2) \end{aligned}$$

The conditional distribution of β_k is a normal distribution and shown as follow:

$$f(\beta_k | X, \check{y}_i, \beta_{-k}, b, u, r, \eta_k^2, w, K) \propto f(\check{y}_i | X, \beta, b, u, r, \eta_k^2, w, K) \times \pi(\beta_k | r_k)$$

$$\propto \exp \left\{ -\frac{1}{2} \left[\sum_{j=1}^q \sum_{i=1}^n \frac{k_{ij} x_{ik}^2}{2u_i} + \frac{1}{r_k} \right] \beta_k^2 - 2 \sum_{j=1}^q \sum_{i=1}^n \frac{k_{ij} x_{ik} \check{y}_i^*}{2u_i} \beta_k \right\}$$

$$\propto N \left[\frac{\sum_{j=1}^q \sum_{i=1}^n \left(\frac{k_{ij} x_{ik} \check{y}_i^*}{2u_i} \right)}{\sum_{j=1}^q \sum_{i=1}^n \left(\frac{k_{ij} x_{ik}^2}{2u_i} \right) + \left(\frac{1}{r_k} \right)}, \frac{1}{\sum_{j=1}^q \sum_{i=1}^n \left(\frac{k_{ij} x_{ik}^2}{2u_i} \right) + \left(\frac{1}{r_k} \right)} \right]$$

$$\check{y}_i^* = \check{y}_i - b_{\tau_j} - x'_{i,-k} \beta_{-k} - \theta u_i$$