

Bayesian group bridge composite quantile regression

Mayyadah Aljasimee¹, Rahim Alhamzawi²

^{1,2} Department of Statistics, University of Al-Qadisiyah,
¹Mayyadahjawad3@gmail.com , ²Rahim.alhamzawi@qu.edu.iq

ABSTRACT

Bayesian regularized composite quantile regression (CQR) method with group bridge penalty is adopted to conduct covariate selection and estimation in CQR. MCMC algorithm was improved for posterior inference employing a scale mixture of normals of the asymmetric Laplace distribution (ALD). The suggested algorithm uses priors for the coefficients of regression, which are scale mixtures of multivariate uniform distributions with a particular Gamma distribution as a mixing distribution. Simulation results and analyses of real data show that the suggested MCMC sampler has excellent mixing feature and outperforms the current approaches in terms of prediction accuracy and model selection.

Keywords: Bayesian inference, Composite quantile regression, Group bridge, MCMC, scale mixture of uniform.

1. Introduction

The normal linear regression model supposes that an outcomes vector $\mathbf{y} = (y_1, \dots, y_n)'$ can be written as

$$\mathbf{y} = b_0 \mathbf{1} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1)$$

where $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ is a $n \times p$ covariates matrix, b_0 is the intercept, $\mathbf{1}$ is an $n \times 1$ unit vector, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)'$ are independent, as well as ε_i has a Gaussian distribution having mean 0 and variance σ^2 . According to model (1), it's supposed that only an unfamiliar subset from covariates are effective in the regression; therefore, the issue of covariate selecting is to find this unfamiliar subset of covariates.

Traditional approaches to model selection based on the observed data log likelihood, comparing a set of candidate models include Mallows's C_p [1], Akaike information criterion [AIC; 2] and Bayesian information criterion [BIC; 3]. Among the new approaches that are based on regularization and selection operator involve the bridge regression [4], lasso [5], smoothly clipped absolute deviation [6], fused lasso [7], adaptive lasso [8], graphical lasso [9], dantzig selector [10], and matrix completion [11, 12], among others. These approaches are setup for selecting individual covariates. However, covariates are naturally grouped in many real studies. An important example appears in association studies, genes may form overlapping sets where each gene can be involved in multiple tracks [13]. For this and other situations, Yuan and Lin [9] suggested the group lasso penalty for choosing covariates groups by introducing a suitable expansion of the lasso penalty. Since Yuan and Lin [9], over the years, various group lasso methods have been improved for dealing with chosen groups of covariates [see for example, 14- 21].

Although covariate selection methods in standard mean regression models have been well developed, we frequently require to assess effects of covariates on outcome variable at various quantile levels. Koenker and Bassett [22] suggested quantile regression (QR) to overcome this

issue. Compared to standard mean regression, QR is more strong to data outliers than standard mean regression, and can provide a more clear picture of the relation between covariates and outcome of interest. However, for linear regression models, [Zou and Yuan \[23\]](#) indicated that QR may result in an arbitrarily tiny relative efficiency when compared with the standard mean regression. Since, QR at one quantile can provide more efficient estimators than QR at another quantile, [Zou and Yuan \[23\]](#) suggested a composite QR (CQR) approach to simultaneously study multiple QR models. They proved that, irrespective of the error distribution, the relative efficiency of the CQR estimator is higher than 70% when compared to the mean regression estimator. Recently, when p is finite, CQR has been employed in covariate selection methods; for example see, [Zou and Yuan \[23\]](#), [Bradic et al. \[24\]](#) and [Jiang et al. \[25\]](#). In this paper, we suggest a Bayesian framework to combine CQR and group bridge penalty together to perform model selection and estimation of coefficients simultaneously.

We introduce the CQR with the group bridge penalty in Section 2. We also outline the Bayesian sampler algorithm for CQR. Section 3 is where we run examples of simulation to investigate the performance of the suggested approach, and we explain our approach employing the prostate cancer data in Section 4. Finally, in Section 5, we conclude with a summarized discussion.

2. Methods

2.1. QR

QR [\[22\]](#) has acquired growing popularity since it makes few assumptions about the error distribution. For the θ th quantile ($0 < \theta < 1$), the linear QR model is $y = b_\theta + X\beta + \epsilon$, where $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ are independent, and their θ th quantiles equal to zero. The θ th QR model takes the form of

$$Q_{y_i}(x_i) = b_\theta + x_i'\beta, \quad (2)$$

where b_θ is the quantile intercept. The regression parameters b_θ and β are estimated by minimizing [\[22\]](#):

$$b_\theta, \beta \min \sum_{i=1}^n \rho_\theta(y_i - b_\theta - x_i'\beta), \quad (3)$$

where $\rho_\theta(w) = w\theta - wI(w \leq 0)$ denotes the quantile check (loss) function and $I(\cdot)$ denotes the indicator function. The ALD provides a possible parametric correlation between the minimization issue in (3) and the maximum likelihood theorem [\[26,27\]](#). The ALD density function for the response y is

$$f(\mu, \sigma) = \frac{\theta(1-\theta)}{\sigma} \exp\left\{-\frac{\rho_\theta(y-\mu)}{\sigma}\right\}, \quad (4)$$

where σ is the scale parameter and μ is the location parameter. [Yu and Moyeed \[27\]](#) introduced a Bayesian framework for QR employing the ALD for the errors, and the MCMC Metropolis-Hastings sampling algorithm is utilized to (approximately) draw β from it's conditional distribution. [Kozumi and Kobayashi \[28\]](#) improved an efficient Gibbs sampling algorithm for Bayesian QR by assuming that the random variable $\epsilon_i = (1 - 2\theta)w_i + \sqrt{2\sigma w_i}z_i$ follows the ALD, where w_i and z_i have an exponential distribution having scale parameter $(\theta(1 - \theta)/\sigma)$ and a standard normal distribution, respectively [see, [29- 33](#)]. As the conditional distribution of y_i given w_i is normal having mean $b_\theta + x_i'\beta + (1 - 2\theta)w_i$ and variance $2\sigma w_i$, the density of y_i is given by

$$p(y_i|x_i, \beta, b_\theta, w_i, \sigma) = \frac{1}{\sqrt{4\pi\sigma w_i}} \exp\left\{-\frac{(y_i - b_\theta - x_i'\beta - (1 - 2\theta)w_i)^2}{4\sigma w_i}\right\}. \quad (5)$$

2.2. CQR

CQR [23] has acquired growing popularity as it can combine information of numerous quantiles simultaneously to get a group of good estimations. Denote $0 < \theta_1 < \theta_2 < \dots < \theta_K < 1$, where $\theta_k = k / (K + 1)$. The CQR estimators of $\mathbf{b}_\theta = (b_{\theta_1}, \dots, b_{\theta_K})$ and β can be estimated by minimizing

$$(\hat{b}_\theta, \hat{\beta}) = b_\theta, \beta \min \sum_{i=1}^n \{ \sum_{k=1}^K \rho_{\theta_k}(y_i - b_{\theta_k} - x_i' \beta) \}, \quad (6)$$

Huang and Chen [34] and Alhamzawi [35] proposed Bayesian formulations for CQR using the ALD for the errors. Under these formulations, the joint distribution of \mathbf{y} is

$$p(X, \beta, b_\theta, w, \sigma) = \prod_{k=1}^K \prod_{i=1}^n \left(\frac{1}{\sqrt{4\pi\sigma w_{ik}}} \right) \exp \left\{ - \frac{(y_i - b_{\theta_k} - x_i' \beta - \xi_k w_{ik})^2}{4\sigma w_{ik}} \right\}, \quad (7)$$

where $w = (w_1, \dots, w_K)$, $w_k = (w_{1k}, \dots, w_{nk})$ and $\xi_k = 1 - 2\theta_k$.

2.3. CQR with the group bridge penalty

Assume that the covariates are collected into G groups so that $x_i = (x_{i1}', \dots, x_{iG}')$, $\beta = (\beta_1', \dots, \beta_G')$, β_g is the m_g -dimensional coefficient vector of the g th group covariates x_{ig} , $\sum_{g=1}^G m_g = p$ and $G < p$. In this paper, we define the following group bridge regularized CQR:

$$(\hat{b}_\theta, \hat{\beta}) = b_\theta, \beta \min \sum_{i=1}^n \{ \sum_{k=1}^K \rho_{\theta_k}(y_i - b_{\theta_k} - x_i' \beta) \} + \sum_{g=1}^G \lambda_g \|\beta_g\|_1^\alpha, \quad (8)$$

where $\|\beta_g\|_1$ is the L_1 norm of β_g , $\lambda_g > 0$, $g = 1, \dots, G$ are the group-specific shrinkage parameters and $\alpha > 0$ denotes the concavity parameter. The bridge parameter α does covariate selection when $\alpha \in (0, 1]$, and shrinks the coefficients of regression when $\alpha > 1$. From a Bayesian point of view, one may define the following group bridge prior on the coefficients [36-38]:

$$p(\alpha, \lambda_1, \dots, \lambda_G) \propto \prod_{g=1}^G \exp(-\lambda_g \|\beta_g\|_1^\alpha). \quad (9)$$

If we remove the dependence on the group index g , the prior for a group bridge may be written as follows

$$p(\beta) = \frac{\lambda^\alpha \Gamma(p+1)}{2^p \Gamma(\frac{p}{\alpha}+1)} \exp(-\lambda \|\beta\|_1^\alpha). \quad (10)$$

If we put the group bridge prior (9) on β and assume the errors ϵ_i is from the ALD (4), the conditional distribution of β is

$$p(X, \beta, b_{\theta}, w, \sigma) \propto \exp \left\{ - \sum_{i=1}^n \sum_{k=1}^K \frac{(y_i - b_{\theta_k} - x_i' \beta - \xi_k w_{ik})^2}{4\sigma w_{ik}} - \sum_{g=1}^G \lambda_g \|\beta_g\|_1^\alpha \right\}. \quad (11)$$

So minimizing the group bridge regularized CQR (8) is equivalent to maximizing the composite likelihood (11). Mallick and Yi [38] show that the group bridge prior may be expressed as a scale mixture of multivariate uniform (SMU) distribution, the mixing density is a specific Gamma distribution, in other words, $\beta|u \sim \text{Multivariate Uniform}(A)$, where $A = \{\beta \in R^q : \|\beta_g\|_1^\alpha < u\}$, $u > 0$ and $u \sim \text{Gamma}(\frac{q}{\alpha} + 1, \lambda)$. Putting Beta prior on α and Gamma priors on λ_g and σ_k , the Bayesian hierarchical model for CQR with group bridge penalty (8) is as follows

$$y_i = \prod_{k=1}^K (b_{\theta_k} + x_i' \beta + \xi_k w_{ik} + \sqrt{2\sigma w_{ik}} z_i), i = 1, \dots, n,$$

$$w|\sigma \sim \prod_{k=1}^K \prod_{i=1}^n \frac{\theta_k(1 - \theta_k)}{\sigma} \exp\left(-\frac{\theta_k(1 - \theta_k)}{\sigma} w_{ik}\right),$$

$$z \sim \prod_{i=1}^n N(0,1),$$

$$\beta_g|u_g, \alpha \sim \text{Multivariate Uniform}(\Omega_g) \text{ independently for } g = 1, \dots, G,$$

$$\text{where } \Omega_g = \{\beta_g \in R^{m_g} : \|\beta_g\|_1^\alpha < u_g\},$$

$$u_1, \dots, u_G | \lambda_1, \dots, \lambda_G, \alpha \sim \prod_{g=1}^G \text{Gamma}\left(\frac{m_g}{\alpha} + 1, \lambda_g\right),$$

$$\lambda_1, \dots, \lambda_G \sim \prod_{g=1}^G \text{Gamma}(a, b), \quad (12)$$

$$\alpha \sim \text{Beta}(c, d),$$

$$\sigma \sim \text{Gamma}(r, \delta),$$

where $u = (u_1, \dots, u_G)$, and $\lambda = (\lambda_1, \dots, \lambda_G)$. It's clear that the full conditional posteriors may be obtained by employing easy algebra for the prior description and the parameters of interest $(b_{\theta}, \beta, \sigma, w, u, \lambda, \alpha)$ can be sampled as listed in Figure 1.

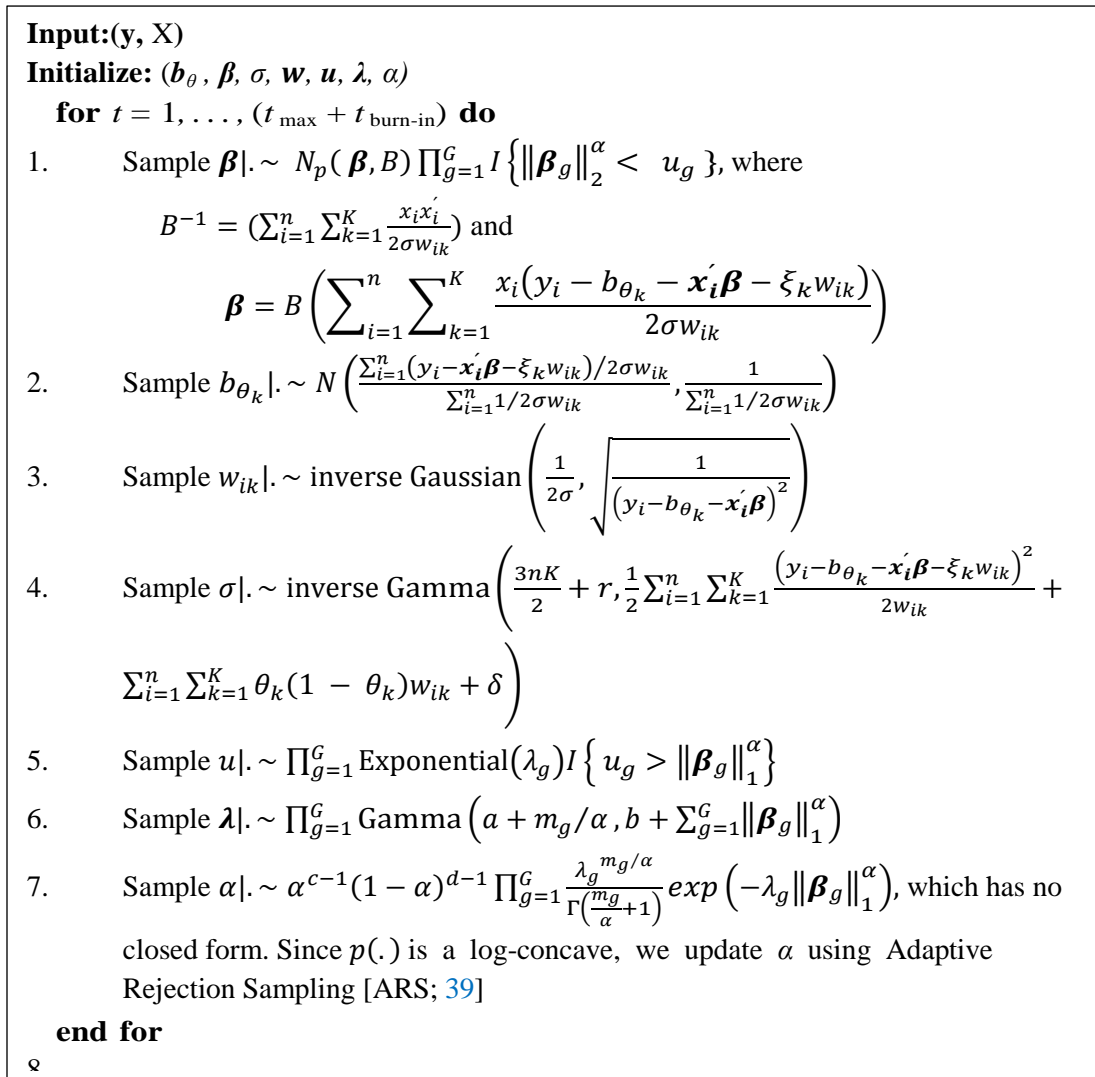


Figure 1: MCMC sampling for the Bayesian group bridge CQR.

3. Simulation Studies

Here, we use simulations of Monte Carlo to illustrate the performance of Bayesian group Bridge CQR (BgBCQR) with comparison to Bayesian group bridge regression [BgBR, 40], group bridge regression [gBR, 16] and group Lasso regression [gLR, 9]. The Bayesian estimations are posterior means employing 20,000 draws of the MCMC algorithm following burn-in the first 10,000 draws. For our approach, we set $a = 1, b = 0.1, r = 10, \delta = 10, c = 0.1,$ and $d = 0.1$.

We generate data using the following real model

$$y = X\beta + \varepsilon$$

In each generated data, we consider three different choices for the error distribution: $N(0,9), t(3)$ distribution having (3) freedom degrees, and $\chi_{(3)}^2$ distribution having (3) freedom degrees. Additionally, we run 100 replications. In each replication, we simulate a training set of 20 observations and a testing set of 200 observations.

Example 1 [41]. In this example, the rows of the design matrix X are provided by $(I(S_1 = 0), I(S_1 = 1), I(S_1 = 2), \dots, I(S_5 = 0), I(S_5 = 1), I(S_5 = 2))$, where the latent variables $S = (S_1, \dots, S_5)'$ are simulated independently from $N(0, \Sigma)$ with the (i, j) th element of Σ is $\rho^{|i-j|}$ and $\rho = 0.5$. Each latent variable S_j for $j = 1, \dots, 5$ is trichotomized as zero, one or two, depending on whether it's less than $F^{-1}(1/3)$, between $F^{-1}(1/3)$ and $F^{-1}(2/3)$, or greater than $F^{-1}(2/3)$, where F^{-1} is the quantile function to standard normal distribution. We set the regression coefficients vector as $\beta = ((-1.2, 1.8, 0), (0, 0, 0), (0.5, 1, 0), (0, 0, 0), (1, 1, 0))$. Thus, the regression parameters in a group may be either all zero, all nonzero or partly. We use $\{n_T, n_P\} = \{20, 400\}$, $\{50, 400\}$ and $\{100, 400\}$ respectively, to simulate datasets, where n_T stands for the number of the observations in the training set, while n_P stands for the number of the observations in the testing set. The experimental outcomes are presented in Table 1. Here, in terms of prediction accuracy, our suggested approach outperforms current Bayesian and non-Bayesian approaches.

Example 2 (High Correlation Example). The setup for this example is identical to the first, excepting we set $\rho = 0.95$. The experimental outcomes are presented in Table 2. Here also, in terms of prediction accuracy, our suggested approach outperforms the other methods.

Example 3. The setup for this example is identical to the first, excepting we set the coefficients of regression vector as $\beta = ((0.5, 1, 1.5, 2, 2.5), (2, 2, 2, 2, 2), (0, 0, 0, 0, 0))$. Thus, in each group, the regression parameters are either all nonzero or all zero. The experimental outcomes are shown in Table 3. Again, we may observe that in terms of prediction accuracy, our proposed approach outperforms the other approaches.

Overall, the simulations show that all of the Bayesian approaches have the same accuracy of the prediction in most of the cases, so often outperform their frequentist counterparts in terms of prediction accuracy all over a wide range of scenarios.

Table 1: Median of mean absolute deviations (MMAD) with the standard deviations of MAD (SD) for Example 1. The bold numbers of MMAD stands for the least MMAD in each category.

Method	n_T	Error					
		$N(0, 9)$		$t(3)$		$\chi^2_{(3)}$	
		MMAD	SD	MMAD	SD	MMAD	SD
gLR	20	1.4278	1.4021	1.4325	1.4690	1.7025	1.6536
gBR	20	1.4166	1.4722	1.6533	1.8253	1.9837	2.3613
BgBR	20	1.3728	1.2083	1.5241	1.3422	1.6572	1.5344
BgBCQR	20	1.3213	1.4082	1.5221	1.3314	1.5267	1.4362
gLR	50	1.4099	1.5504	1.3722	1.4797	1.4359	1.4685
gBR	50	1.5313	2.1991	1.4850	2.0498	1.5083	2.0028
BgBR	50	1.3121	1.4467	1.2901	1.3788	1.3283	1.3936
BgBCQR	50	1.2614	1.1231	1.1751	1.1238	1.3781	1.5865
gLR	100	1.2543	1.4199	1.2347	1.3747	1.2459	1.3555
gBR	100	1.3013	1.9240	1.2365	1.8437	1.2331	1.7848
BgBR	100	1.1841	1.3446	1.1328	1.3026	1.1281	1.2801
BgBCQR	100	1.0021	1.5278	1.1206	1.4711	1.1061	1.4311
gLR	200	1.1197	1.3299	1.0750	1.2976	1.0859	1.2699
gBR	200	1.0892	1.7387	1.0194	1.6856	1.0292	1.6394
BgBR	200	1.0148	1.2540	0.9735	1.2209	0.9844	1.1984
BgBCQR	200	0.9893	1.3966	0.9059	1.3601	0.9137	1.3264

Table 2: MMAD with the standard deviations of MAD (SD) for Example 2. The bold numbers of MMAD stands for the least MMAD in each category.

Method	n_T	Error					
		$N(0, 9)$		$t(3)$		$\chi^2_{(3)}$	
		MMAD	SD	MMAD	SD	MMAD	SD
gLR	20	1.2157	1.0958	1.4045	1.5105	1.6392	3.4381
gBR	20	1.1169	1.6627	1.3276	2.2563	1.5980	3.8811
BgBR	20	1.2516	1.0141	1.3050	1.2258	1.4053	1.4456
BgBCQR	20	1.1087	1.3581	1.2732	1.7005	1.4008	2.1010
gLR	50	1.3368	3.0464	1.3128	2.8754	1.3578	2.6771
gBR	50	1.2352	3.4627	1.1757	3.2631	1.2064	3.0724
BgBR	50	1.1389	1.3623	1.1205	1.4771	1.1517	1.4497
BgBCQR	50	1.1123	1.9312	1.0463	1.9399	1.0749	1.8490
gLR	100	1.2226	2.5129	1.1945	2.3731	1.2195	2.2664
gBR	100	1.0464	2.8894	1.0008	2.7298	1.0234	2.6102
BgBR	100	1.0228	1.3897	0.9612	1.3315	0.9882	1.3033
BgBCQR	100	0.9636	1.7549	0.8931	1.6710	0.9162	1.6120
gLR	200	1.1040	2.1808	1.0616	2.0965	1.0670	2.0238
gBR	200	0.8993	2.5071	0.8571	2.4100	0.8449	2.3241
BgBR	200	0.8784	1.2708	0.8372	1.2350	0.8332	1.2030
BgBCQR	200	0.8304	1.5592	0.7751	1.5073	0.7689	1.4594

4. Real Data Analyses

In this section, we implement the suggested approach for the standard datasets, namely the data of prostate cancer [42]. This dataset has been utilized for illustration in previous regularization papers. In this dataset, the logarithm of prostate-specified antigen is the outcome of interest. Here is a list describing briefly the response variable and 8 covariates.

We compare the mean squared prediction errors (MMSE) for Prostate data analyses in Table 4, which shows that our suggested approach outperforms both the existing Bayesian and non-Bayesian approaches in terms of prediction accuracy.

lcavol	Log(volume of cancer)
lweight	Log(weight of the prostate)
age	Age
lbph	Log(The quantity of benign prostatic hyperplasia)
svi	Invasion of seminal vesicles
lcp	Log(capsular breakthrough)
gleason	The Gleason result
pgg45	The rate of Gleason results is four or five
lpsa	Log(prostatic specified antigen)

Table 3: MMAD with the standard deviations of MAD (SD) for Example 3. The bold numbers of MMAD stands for the least MMAD in each category.

Method	n_T	Error					
		$N(0, 9)$		$t(3)$		$\chi^2_{(3)}$	
		MMAD	SD	MMAD	SD	MMAD	SD
gLR	20	0.9023	1.3315	0.9852	1.1570	0.9991	1.9758
gBR	20	1.2054	2.0352	1.3890	1.9173	1.5677	3.5526
BgBR	20	0.8993	0.9335	1.0620	0.9569	1.1418	1.4970
BgBCQR	20	1.0882	1.4828	1.1909	1.4525	1.3595	1.8149
gLR	50	0.8698	1.7703	0.8549	1.6444	0.8512	1.5493
gBR	50	1.2195	3.1602	1.1014	2.8997	1.1144	2.7275
BgBR	50	0.9995	1.3676	0.9356	1.2845	0.9457	1.2530
BgBCQR	50	1.1121	1.6576	1.0524	1.5576	1.0498	1.4973
gLR	100	0.8051	1.4692	0.7990	1.4078	0.8051	1.3559
gBR	100	0.9750	2.5649	0.9377	2.4384	0.9430	2.3289
BgBR	100	0.8438	1.1981	0.7990	1.1597	0.8272	1.1301
BgBCQR	100	0.8306	1.4276	0.7750	1.3702	0.8164	1.3255
gLR	200	0.7708	1.3108	0.7582	1.2707	0.7685	1.2389
gBR	200	0.8369	2.2343	0.7950	2.1482	0.7827	2.0760
BgBR	200	0.7502	1.0969	0.7218	1.0677	0.7216	1.0481
BgBCQR	200	0.7493	1.2841	0.7159	1.2453	0.7111	1.2131

Table 4: MMSE for Prostate data analyses.

Method	MMSE
gLR	0.48
gBR	0.48
BgBR	0.47
BgBCQR	0.45

5. Discussion

We have introduced a Bayesian analysis of group bridge CQR and employing a scale mixture of normals of the ALD, we have proposed Gibbs sampler algorithm for posterior inference. The suggested algorithm uses priors for the coefficients of regression, which are scale mixtures of multivariate uniform distributions with a particular Gamma distribution as a mixing distribution. The suggested algorithm is active in regularization under a variety of scenarios, as demonstrated by simulation examples. We have as well illustrated the advantages of the new method on prostate data example. Thus, both the simulation and the data of prostate cancer reveal great support for the employment of Bayesian group bridge CQR.

References

- [1] Mallows CL, "Some comments on C_p ", *Technometrics* 15 : 661–675, (1973).
- [2] H. Akaike, "Information theory and an extension of maximum likelihood principle", *In Second International Symposium on Information Theory*, B. N. Petrov and F. Csaki (eds), Budapest: Akademiai Kiado, (1973).
- [3] G. Schwarz, (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.
- [4] L. E. Frank, and J. H. Friedman, "A statistical view of some chemometrics regression tools", *Technometrics* 35 (2), 109–135, (1993).
- [5] R. Tibshirani, "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288, (1996).
- [6] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties", *Journal of the American statistical Association* 96 (456), 1348–1360, (2001).
- [7] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Knight Sparsity and smoothness via the fused lasso". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (1), 91–108, (2005).
- [8] H. Zou "The adaptive lasso and its oracle properties", *Journal of the American statistical association* 101 (476), 1418–1429, (2006).
- [9] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables", *Journal of the Royal Statistical Society, Series B* 68, 49–67, (2006).
- [10] E. Candès and T. Tao, "The dantzig selector: Statistical estimation when p is much larger than n ", *The Annals of Statistics*, 2313–2351, (2007).
- [11] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion", *IEEE Transactions on Information Theory* 56 (5), 2053–2080, (2010).
- [12] R. Mazumder, J. H. Friedman, and T. Hastie, "Sparsenet: Coordinate descent with nonconvex penalties", *Journal of the American Statistical Association* 106 (495), 1125–1138, (2011).
- [13] L. Jacob, G. Obozinski, and J.-P. Vert "Group lasso with overlap and graph lasso", *In Proceedings of the 26th annual international conference on machine learning*, pp. 433–440, (2009).
- [14] P. Breheny, "The group exponential lasso for bi-level variable selection", *Biometrics* 71 (3), 731–740, (2015).
- [15] J. Huang, P. Breheny, and S. Ma. "A selective review of group selection in high-dimensional models", *Statistical science: a review journal of the Institute of Mathematical Statistics* 27 (4), (2012).
- [16] J. Huang, S. Ma, H. Xie, and C.-H. Zhang, "A group bridge approach for variable selection", *Biometrika* 96 (2), 339–355, (2009).
- [17] L. Meier, S. Van De Geer, and P. Bühlmann, "The group lasso for logistic regression", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70 (1), 53–71, (2008).
- [18] C. Park, and Y. J. Yoon Bridge regression: "adaptivity and group selection", *Journal of Statistical Planning and Inference* 141 (11), 3506–3519, (2011).
- [19] W. Qian, Y. Yang, and H. Zou "Tweedie's compound poisson model with grouped elastic net", *Journal of Computational and Graphical Statistics* 25 (2), 606–625, (2016).
- [20] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse-group lasso", *Journal of Computational and Graphical Statistics* 22 (2), 231–245, (2013).
- [21] N. Simon and R. Tibshirani, "Standardization and the group lasso penalty", *Statistica Sinica* 22 (3), 983–1001, (2012).

-
- [22] R. Koenker and G. Bassett, "Regression quantiles", *Econometrica: Journal of the Econometric Society*, 33–50, (1978).
- [23] H. Zou and M. Yuan "Composite quantile regression and the oracle model selection theory", *The Annals of Statistics* 36 (3), 1108–1126, (2008).
- [24] J. Bradic, J. Fan, and W. Wang, "Penalized composite quasi-likelihood for ultrahigh dimensional variable selection", *Journal of the Royal Statistical Society, Ser. B* 73, 325–349, (2011).
- [25] X. Jiang, J. Jiang, and X. Song, "Oracle model selection for nonlinear models based on weighted composite quantile regression", *Statistica Sinica*, 1479–1506, (2012).
- [26] R. Koenker and J. A. Machado, "Goodness of fit and related inference processes for quantile regression", *Journal of the American Statistical association* 94 (448), 1296–1310, (1999).
- [27] K. Yu and R. A. Moyeed, "Bayesian quantile regression", *Statistics & Probability Letters* 54, 437–4471, (2001).
- [28] H. Kozumi and G. Kobayashi, "Gibbs sampling methods for Bayesian quantile regression", *Journal of statistical computation and simulation* 81 (11), 1565–1578, (2011).
- [29] R. Alhamzawi and K. Yu, "Conjugate priors and variable selection for Bayesian quantile regression", *Computational Statistics & Data Analysis*, in press, (2012).
- [30] T. Alshaybawee, H. Midi, & R. Alhamzawi, "Bayesian elastic net single index quantile regression", *Journal of Applied Statistics*, 44(5), 853-871, (2017).
- [31] R. Alhamzawi, & H. T. M. Ali, "The Bayesian elastic net regression", *Communications in Statistics-Simulation and Computation*, 47(4), 1168-1178, (2018).
- [32] R. Alhamzawi, A. Alhamzawi, , & H. T. M. Ali, "New Gibbs sampling methods for Bayesian regularized quantile regression", *Computers in biology and medicine*, 110, 52-65. (2019).
- [33] R. Alhamzawi, & H. Taha Mohammad Ali, "A new Gibbs sampler for Bayesian lasso", *Communications in Statistics-Simulation and Computation*, 49(7), 1855-1871, (2020).
- [34] H. Huang, and Z. Chen, "Bayesian composite quantile regression" *Journal of Statistical Computation and Simulation* 85 (18), 3744–3754, (2015).
- [35] R. Alhamzawi, "Bayesian analysis of composite quantile regression". *Statistics in Biosciences* 8 (2), 358–373, (2016).
- [36] E. Gómez-Sánchez-Manzano, M. Gómez-Villegas, and J. Marín, "Multivariate exponential power distributions as mixtures of normal distributions with bayesian applications", *Communications in Statistics—Theory and Methods* 37 (6), 972–985, (2008).
- [37] M. A. Gómez-Villegas, E. Gómez-Sánchez-Manzano, P. Maín, and H. Navarro, "The effect of non-normality in the power exponential distributions", In *Modern mathematical tools and techniques in capturing complexity*, pp. 119–129. Springer, (2011).
- [38] H. Mallick & N. Yi, "Bayesian group bridge for bi-level variable selection", *Computational statistics & data analysis*, 110, 115-133, (2017).
- [39] W. R. Gilks, "Derivative-free adaptive rejection sampling for gibbs sampling", *Bayesian statistics* 4 (2), 641–649, (1992).
- [40] H. Mallick and N. Yi, "Bayesian bridge regression", *Journal of applied statistics* 45 (6), 988–1008, (2018).
- [41] Q. Li, R. Xi, and N. Lin, "Bayesian regularized quantile regression", *Bayesian Analysis* 5, 533–556, (2010).
- [42] T. Stamey, J. Kabalin, J. McNeal, I. Johnstone, F. Freiha, E. Redwine, and N. Yang, "Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate, II: Radical prostatectomy treated patients", *Journal of Urology* 141, 1076–1083, (1989).