IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 13, Issue 3 Ver. III (May - June 2017), PP 04-09
www.iosrjournals.org

Inference with Normal-Jeffreys Prior Distributions in Quantile Regression

Rahim Alhamzawi[1], Intisar Ibrahim Allyas[2]
[1]*Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq*
[2]*College of Administration and Economics, Nawroz University*

Abstract: *Decades after its discussion in (Koenker and Bassett, 1978), quantile regression (QR) has been the topic of great practical applications in many areas: economics, ecology, biology and so on. In this paper, we present Bayesian quantile regression using two level prior distributions. Specifically, we assume that the prior distribution of each regression coefficient is a zero mean normal prior distribution with unknown variance. Then, we assign noninformative Jeffreys prior distributions for the variances assuming they are independent. A Gibbs sampler algorithm is developed for the posterior inference. The new method is illustrated via simulations and a real dataset.*

Keywords*: Bayesian, Jeffreys prior, Noninformative, MCMC.*

I. Introduction

Decades after its discussion in (Koenker and Bassett, 1978), QR has been the topic of great practical applications in many areas: economics, ecology, biology and so on (Cade and Noon, 2003). Suppose that we have a sample of observations $(x_1, y_1), \dots, (x_n, y_n)$ where x_i is a k × 1 vector of predictors. Then, the linear QR model for the τth quantile ($0 < \tau < 1$) is $y_i = x_i'\beta + \varepsilon_i$, where β is a k × 1 vector of regression coefficients and ε_i's are independent with τth quantile equal to zero. According to Koenker and Bassett (1978), QR estimation for β proceeds by

$$min \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta), \qquad (1)$$

where $\rho_\tau(.)$ is the empirical check function defined by $\rho_\tau(z) = z\{\tau - I(z < 0)\}$, and I(.) denotes the usual indicator function. This empirical check function is not differentiable at 0. Thus, a closed-form solution is not available for the QR coefficients vector β (Koenker and Bassett, 1978). However, the minimization of (1) can be achieved through an algorithm suggested by Koenker and D'Orey (1987). Alternatively, Koenker and Machado (1999) observed that minimizing the empirical loss function of Koenker and Bassett (1978) is closely related to maximising the likelihood of the Asymmetric Laplace Distribution (ALD) and consequently the vector β can be estimated through exploiting this link. Yu and Moyeed (2001) and Yu and Stander (2007) proposed a Bayesian formulation of QR using the ALD for the errors and sampling β from its posterior distribution using a random walk Metropolis- Hastings algorithm. Recently, Kozumi and Kobayashi (2011) developed a Gibbs sampler method to estimate the vector β. Specifically, by expressing the ALD as a location-scale mixture of normals and by data augmentation; they propose a Gibbs sampler algorithm which converges to the joint posterior distribution of all unknowns (parameters and latent variables). This approach has been used in a large number of studies (see for example, Li et al., 2010; Khare et al., 2011; Alhamzawi et al., 2011; Lubrano and Ndoye, 2012; Kurose and Omori, 2012; Alhamzawi and Yu, 2013; Wichitaksorn and Tsurumi, 2013; Hu et al., 2013; Benoit et al., 2013; Sriram et al., 2013; Alhamzawi, 2014; Huang and Chen, 2015; Hashem et al., 2015; Alhamzawi, 2015; Alshaybawee et al., 2016; Huang et al., 2016; Alhamzawi and Ali, 2017).

The density of the ALD for the error term (ε_i) is written explicitly as

$$f(\varepsilon_i) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau\left(\frac{\varepsilon_i}{\sigma}\right)\right\}. \qquad (2)$$

Under the above density, the joint distribution of $y = (y_1, \dots, y_n)$ given $X = (X_1, \dots, X_n)$ is

$$f(y|X, \beta, \sigma) = \left(\frac{\tau(1-\tau)}{\sigma}\right)^n \exp\left\{-\sum_{i=1}^{n} \rho_\tau\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right\}. \qquad (3)$$

Recent numerical studies (see for example, Yang et al., 2015; Sriram et al., 2013) support the utilize of ALD even though it may fail to exactly represent the actual underlying distribution for the errors.

Under the model $y_i = x_i'\beta + \varepsilon_i$, it is assumed that only an unknown subset of the covariates is important in the regression, so that the variable selection problem is to identify this important subset of covariates. Many methods for selecting the important variables in linear regression models have been proposed over the recent years. Among these, two-level hierarchical Bayesian models have been shown to be effective in linear

regression in improving the prediction accuracy (see for example, Bae and Mallick, 2004; Alhamzawi, 2015; Benoit et al., 2013; Alhamzawi and Yu, 2015).

According to Kozumi and Kobayashi (2011), the joint density of $y|X, \beta, \sigma$ in Eq (3) is given by:

$$p(y|X,\beta,\sigma,z_i) \propto \left(\prod_{i=1}^{n}(\sigma z_i)^{-\frac{1}{2}}\right)\exp\left\{-\sum_{i=1}^{n}\frac{(y_i - x_i'\beta - \theta z_i)^2}{4\sigma z_i} - \frac{\tau(1-\tau)}{\sigma}z_i\right\}, \qquad (4)$$

where, z_i is a mixing variable and $\theta = 1 - 2\tau$. In this paper, we present a Bayesian regression model and specify prior distributions that favor sparseness in terms of number of predictors used. Our model involves a zero mean normal prior distributions for the unknown regression coefficients β with unknown variances. Then we assign noninformative Jeffreys prior distributions for the variances assuming they are independent.

II. Methods

2.1. Priors

In this paper, we assume that the prior distribution of each β_j is a zero mean normal prior distribution with unknown variance, i.e. $\beta_j \sim N(0, \sigma\lambda_j)$. Then, we assign a non informative Jeffreys prior for λ_j takes the form of $p(\lambda_j) \propto 1/\lambda_j$. We further assign a non informative Jeffreys prior for the scale parameter σ takes the form of $p(\sigma) \propto 1/\sigma$. To summarize, our Bayesian hierarchical formulation is provided below.

$$\begin{aligned} p(y_i|X, \beta, \sigma, z_i) &\sim N(xi\beta + \theta z_i, 2\sigma z_i), \\ p(\beta) &\sim \prod_{j=1}^{k} N(0, \sigma\lambda_j), \\ p(z_i) &\sim \text{Exponential}\left(\frac{\tau(1-\tau)}{\sigma}\right), \\ p(\lambda_j) &\sim 1/\lambda_j \\ p(\sigma) &\propto 1/\sigma. \end{aligned} \qquad (5)$$

2.2. Gibbs sampler for Bayesian inference

Let $z = (z_1, \cdots, z_n)$, $V = diag(1/(2z_1), \cdots, 1/(2z_n))$ and $\Lambda = diag(\lambda_1, \cdots, \lambda_k)$. From (5), we can obtain a tractable and efficient Gibbs sampler algorithm that works as follows:

• Sample β

The full conditional distribution of β is normal, i.e.

$$\beta|\sigma, z, \Lambda, y, X \sim N_k((X'VX + \Lambda^{-1})^{-1}X'V(y - \theta z_i), \sigma (X'VX + \Lambda^{-1})^{-1}),$$

• *Sample* z_i^{-1} *for* $i = 1, \cdots, n$.

The full conditional distribution of each z_i^{-1} *is an inverse Gaussian IG* (μ_i', λ_i'), *where* $\mu_i' = |y_{i-} x_i'\beta|^{-1}$ and $\lambda_i' = 1/2\sigma$.

• Sample λ_j for $j = 1, \cdots, k$.

The full conditional distribution of each λ_j is an inverse gamma with shape parameter 1/2 and rate parameter $\beta_j^2/(2\sigma)$.

• Sample σ

The full conditional distribution of σ is an inverse gamma with shape parameter $\frac{3}{2}n + k$ and rate parameter $\frac{1}{2}[(y - X\beta - \theta z)'V(y - X\beta - \theta z) + \sum_{j=1}^{k}\frac{\beta_j^2}{\lambda_j}]$.

III. Simulation Studies

In this section, we carry out simulation studies to study the performance of the proposed method with comparison to some Bayesian and non-Bayesian approaches. The methods in the comparison include:

• The standard QR (referred to as "QR").
• Bayesian Lasso QR (referred to as "BLQR").
• Bayesian elastic net QR (referred to as "BENQR").
• The proposed method (referred to as "BQR").

The data in the simulations are simulated by $y_i = x_i'\beta + \varepsilon_i$. Predictors were generated independently from a multivariate normal distribution N(0, Σ), where the *(i, j)*th element of Σ is $0.75^{|i-j|}$. We consider the following simulation studies:

1. $\beta = (2, 1, 0, 0, 2, 0, 0, 0)$, which corresponds to the sparse case.
2. $\beta = (2, 0, 0, 0, 0, 0, 0, 0)$, which corresponds to the very sparse case.
3. $\beta = (0.90, 0.90, 0.90, 0.90, 0.90, 0.90, 0.90, 0.90)$, which corresponds to the dense case.

Within each simulation study, three different choices for the distribution of the error are considered: the standard normal distribution N(0, 1), a $t_{(3)}$ distribution, and a $\chi^2_{(3)}$ distribution. We centered the response variable to have

mean zero, while the covariates have been standardized. For each simulation study and each choice of the error term, we run 100 replications. In each simulation study, we simulate a training set with 20 observations and a testing set with 200 observations. Methods are evaluated based on MMAD=median($\sum_{i=1}^{200}(|x_i'\hat{\beta} - x_i'\beta|)$.
The results are summarized in Tables 1, 2 and 3. From the tables, it can be seen that the proposed method BQR performs reasonably well outperforming the others (QR, BLQR and BENQR). We can see that the proposed method tends to produce smaller MMAD compared with other methods. From the tables 1, 2 and 3, we can also see that the Bayesian approaches yield similar performance and outperform the frequentist QR.
The results are summarized in Tables 1, 2 and 3. From the tables, it can be seen that the proposed method BQR performs reasonably well outperforming the others (QR, BLQR and BENQR). We can see that the proposed method tends to produce smaller MMAD compared with other methods. From the tables 1, 2 and 3, we can also see that the Bayesian approaches yield similar performance and outperform the frequentist QR.

Table 1: MMADs for Simulation 1. In the parentheses are standard deviations of the MADs.

Error	Methods	τ = 0.50		τ = 0.75		τ = 0.95	
ε ~ N(0,1)	BQR	0.473	(0.121)	0.502	(0.121)	0.635	(0.406)
	BLQR	0.463	(0.124)	0.535	(0.168)	0.664	(0.428)
	BENQR	0.450	(0.127)	0.539	(0.167)	0.654	(0.416)
	QR	0.686	(0.220)	0.727	(0.225)	0.839	(0.492)
ε~t(3)	BQR	0.583	(0.346)	0.617	(0.311)	0.671	(0.439)
	BLQR	0.646	(0.386)	0.665	(0.356)	0.701	(0.444)
	BENQR	0.648	(0.377)	0.654	(0.349)	0.717	(0.441)
	QR	0.791	(0.459)	0.816	(0.421)	0.898	(0.553)
ε ~χ_3^2	BQR	0.666	(0.436)	0.736	(0.412)	0.824	(0.724)
	BLQR	0.758	(0.441)	0.799	(0.448)	0.848	(0.728)
	BENQR	0.751	(0.435)	0.797	(0.443)	0.854	(0.734)
	QR	0.970	(0.551)	1.045	(0.623)	1.089	(1.243)

Table 2: MMADs for Simulation 2. In the parentheses are standard deviations of the MADs.

Error	Methods	τ = 0.50		τ = 0.75		τ = 0.95	
ε ~ N(0,1)	BQR	0.485	(0.163)	0.422	(0.107)	0.446	(0.108)
	BLQR	0.476	(0.158)	0.476	(0.144)	0.467	(0.167)
	BENQR	0.474	(0.155)	0.471	(0.144)	0.448	(0.169)
	QR	0.864	(0.295)	0.899	(0.313)	0.857	(0.310)
ε~t(3)	BQR	0.411	(0.155)	0.467	(0.108)	0.397	(0.122)
	BLQR	0.445	(0.165)	0.478	(0.167)	0.487	(0.184)
	BENQR	0.441	(0.167)	0.474	(0.169)	0.478	(0.184)
	QR	0.861	(0.355)	0.908	(0.408)	0.919	(0.427)
ε ~χ_3^2	BQR	0.476	(0.236)	0.530	(0.181)	0.573	(0.276)
	BLQR	0.512	(0.247)	0.532	(0.263)	0.576	(0.325)
	BENQR	0.505	(0.247)	0.537	(0.265)	0.579	(0.324)
	QR	1.014	(0.478)	1.060	(0.580)	1.107	(0.670)

Table 3: MMADs for Simulation 3. In the parentheses are standard deviations of the MADs.

Error	Methods	τ = 0.50		τ = 0.75		τ = 0.95	
ε ~ N(0,1)	BQR	0.515	(0.194)	0.547	(0.134)	0.674	(0.196)
	BLQR	0.520	(0.188)	0.541	(0.195)	0.699	(0.253)
	BENQR	0.525	(0.197)	0.544	(0.196)	0.718	(0.255)
	QR	0.778	(0.405)	0.745	(0.344)	0.893	(0.431)
ε~t(3)	BQR	0.655	(0.200)	0.721	(0.234)	0.733	(0.199)
	BLQR	0.699	(0.255)	0.733	(0.262)	0.772	(0.284)
	BENQR	0.696	(0.255)	0.735	(0.261)	0.782	(0.283)
	QR	0.980	(0.431)	1.026	(0.460)	1.150	(0.603)

$\varepsilon \sim \chi_3^2$	BQR	0.810	(0.216)	0.814	(0.342)	0.887	(0.410)
	BLQR	0.826	(0.291)	0.856	(0.336)	0.894	(0.424)
	BENQR	0.837	(0.291)	0.868	(0.335)	0.901	(0.418)
	QR	1.282	(0.688)	1.329	(0.823)	1.426	(0.966)

3.1. Body dimensions data

Here, we consider a real dataset to investigate the performance of the proposed method. We use the dataset on body dimensions (Heinz et al., 2003) where there are 507 observations and 24 predictors. We use a sub sample of 9 predictors from this data set. The data is available in the R Package Brq (Alhamzawi, 2017). The response of interest is the weight in kilogram. The 9 predictors are: Gender, Age in years (Age), Height in cm (Height), Biacromial diameter in cm (BiacSk), Biiliac diameter in cm (BiilSk), Bitrochanteric diameter in cm (BitrSk), Chest depth in cm (CheDeSk), Chest diameter in cm (CheDiSk), Elbow diameter in cm (ElbowSk) and Wrist diameter in cm (WristSk). We assume a QR model between the weight and the 9 regressors. We assume that the response variable to be centered to have mean 0, while the predictors have been standardized. We divided the data into a training set with 107 observations and a test set with 400 observations. The histograms of the body dimensions data predictors based on posterior samples of 11,000 iterations are illustrated in Figure 1. These plots reveal that the conditional posterior distributions are the desired univariate normals. The Trace plot of the Gibbs sampler is shown in Figure 2 for this data set predictors. We can observe that for this benchmark dataset the samples traverse the posterior space very fast. Table 4 presents the mean squared prediction errors (MSE) based on a test set with 400 observations for $\tau \in \{0.50, 0.75, 0.95\}$. We can see that the proposed method tends to produce smaller MMAD than the other methods.

Table 4: Body dimensions data analysis: Mean squared prediction errors (MSE) based on a test set with 400 observations.

τ	Method	QR	BLQR	BENQR	BQR
0.50	MSE	21.36089	21.33972	21.44317	21.29310
0.75	MSE	21.44281	21.41752	21.53110	21.39221
0.95	MSE	21.77739	21.92418	21.66234	21.56287

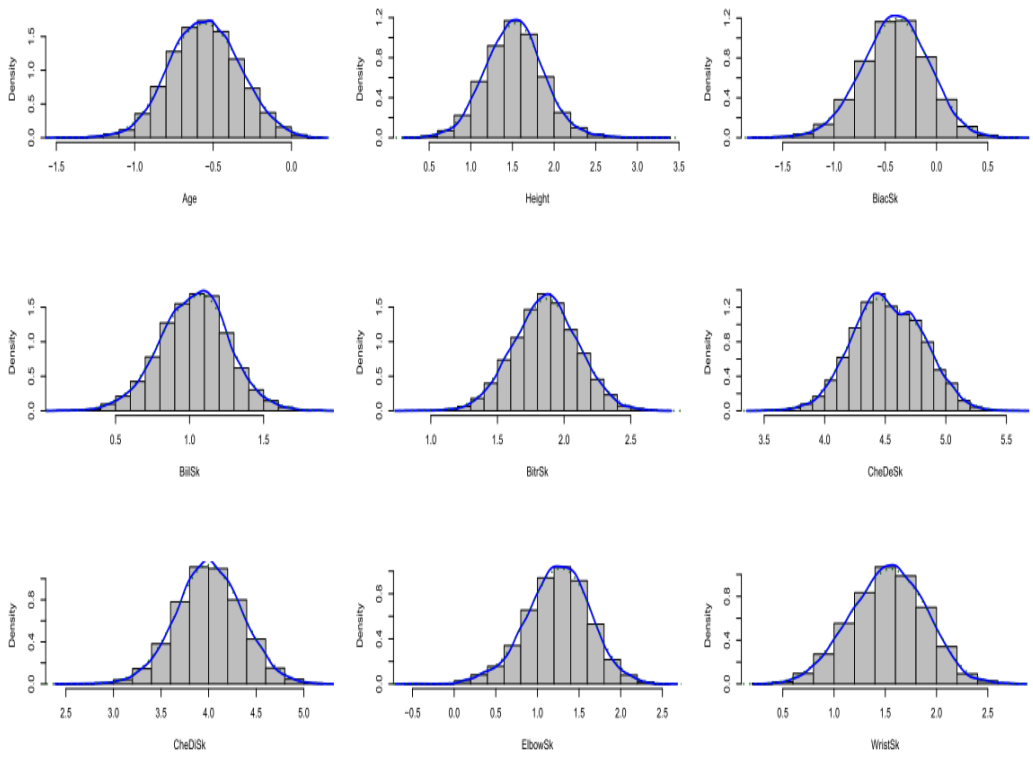

Figure 1: Histograms based on posterior samples of body dimensions data.

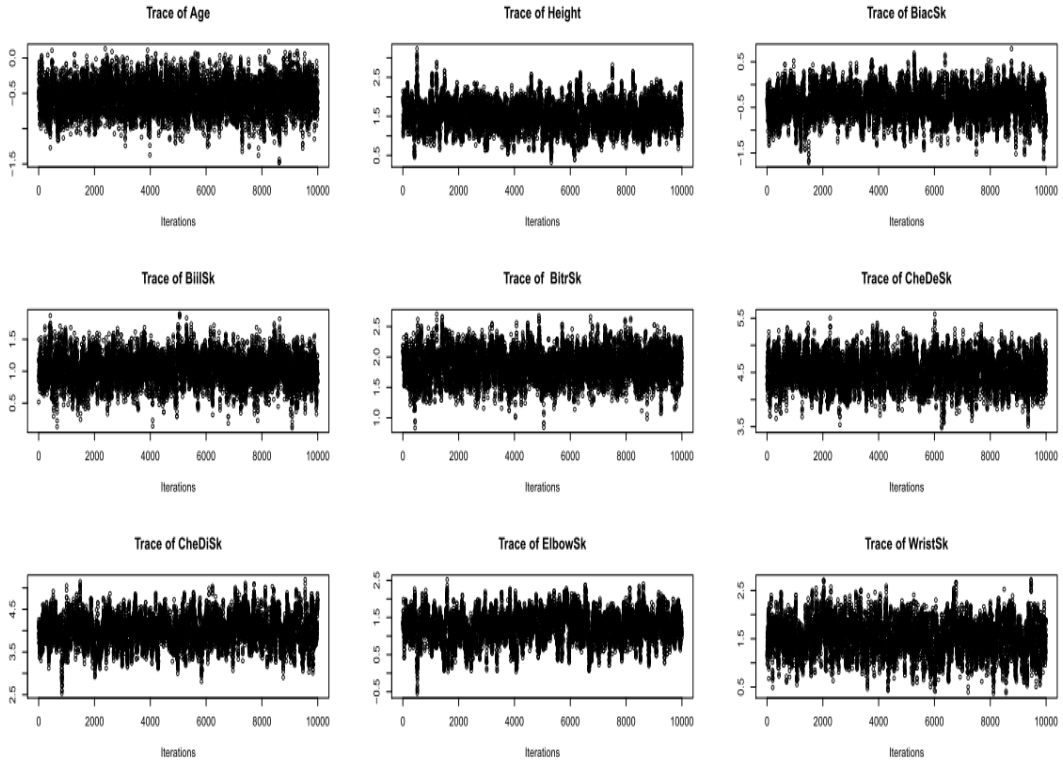

Figure 2: Trace plots based on posterior samples of body dimensions data.

IV. Conclusion

We have proposed the Bayesian quantile regression using normal-Jeffreys prior distributions for the regression coefficients. Specifically, we assume that the prior distribution of each regression coefficient is a zero mean normal prior distribution with unknown variance. Then, we assign noninformative Jeffreys prior distributions for the variances assuming they are independent. We developed a new algorithm for Bayesian sampler from the posteriors. The proposed approach is then illustrated via simulations and a real dataset. Results show that the proposed approach performs very well.

Bibliography

[1] Alhamzawi, R. (2014). Bayesian elastic net tobit quantile regression. Communications in Statistics-Simulation and Computation (just-accepted), 00-00.
[2] Alhamzawi, R. (2015). Model selection in quantile regression models. Journal of Applied Statistics 42 (2), 445-458.
[3] Alhamzawi, R. (2017). Brq: Bayesian Analysis of Quantile Regression Models. R package version 2.0.
[4] Alhamzawi, R. and H. T. M. Ali (2017). Bayesian quantile regression for ordinal longitudinal data. Journal of Applied Statistics, 1-14.
[5] Alhamzawi, R. and K. Yu (2013). Conjugate priors and variable selection for bayesian quantile regression. Computational Statistics & Data Analysis 64, 209-219.
[6] Alhamzawi, R. and K. Yu (2015). Bayesian tobit quantile regression using g-prior distribution with ridge parameter. Journal of Statistical Computation and Simulation 85 (14), 2903-2918.
[7] Alhamzawi, R., K. Yu, V. Vinciotti, and A. Tucker (2011). Prior elicitation for mixed quantile regression with an allometric model. Environmetrics 22 (7), 911-920.
[8] Alshaybawee, T., H. Midi, and R. Alhamzawi (2016). Bayesian elastic net single index quantile regression. Journal of Applied Statistics, 1-19.
[9] Bae, K. and B. K. Mallick (2004). Gene selection using a two-level hierarchical bayesian model. Bioinformatics 20 (18), 3423-3430.
[10] Benoit, D. F., R. Alhamzawi, and K. Yu (2013). Bayesian lasso binary quantile regression. Computational Statistics 28 (6), 2861-2873.
[11] Cade, B. S. and B. R. Noon (2003). A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment 1 (8), 412-420.
[12] Hashem, H., V. Vinciotti, R. Alhamzawi, and K. Yu (2015). Quantile regression with group lasso for classification. Advances in Data Analysis and Classification, 1-16.
[13] Heinz, G., L. J. Peterson, R. W. Johnson, and C. J. Kerk (2003). Exploring relationships in body dimensions. Journal of Statistics Education 11(2).
[14] Hu, Y., R. B. Gramacy, and H. Lian (2013). Bayesian quantile regression for single-index models. Statistics and Computing 23 (4), 437-454.
[15] Huang, H. and Z. Chen (2015). Bayesian composite quantile regression. Journal of Statistical Computation and Simulation 85 (18), 3744-3754.
[16] Huang, Y., J. Chen, and X. Lu (2016). Bayesian approach to nonlinear mixed-effects quantile regression models for longitudinal data with non-normality and left-censoring. Journal of Advanced Statistics 1 (3), 109.
[17] Khare, K., J. P. Hobert, et al. (2011). A spectral analytic comparison of trace-class data augmentation algorithms and their sandwich variants. The Annals of Statistics 39 (5), 2585-2606.

[18] Koenker, R. and G. Bassett (1978). Regression quantiles. Econometrica: journal of the Econometric Society, 33-50.
[19] Koenker, R. and J. A. Machado (1999). Goodness of fit and related inference processes for quantile regression. Journal of the american statistical association 94 (448), 1296-1310.
[20] Koenker, R. W. and V. D'Orey (1987). Algorithm as 229: Computing regression quantiles. Journal of the Royal Statistical Society. Series C (Applied Statistics) 36 (3), 383-393.
[21] Kozumi, H. and G. Kobayashi (2011). Gibbs sampling methods for bayesian quantile regression. Journal of statistical computation and simulation 81 (11), 1565-1578.
[22] Kurose, Y. and Y. Omori (2012). Bayesian analysis of time-varying quantiles using a smoothing spline. Journal of the Japan Statistical Society 42 (1), 23-46.
[23] Li, Q., R. Xi, N. Lin, et al. (2010). Bayesian regularized quantile regression. Bayesian Analysis 5 (3), 533-556.
[24] Lubrano, M. and A. A. J. Ndoye (2012). Bayesian unconditional quantile regression: An analysis of recent expansions in wage structure and earnings inequality in the us 1992-2009.
[25] Sriram, K., R. Ramamoorthi, P. Ghosh, et al. (2013). Posterior consistency of bayesian quantile regression based on the misspecified asymmetric laplace density. Bayesian Analysis 8 (2), 479-504.
[26] Wichitaksorn, N. and H. Tsurumi (2013). Comparison of mcmc algorithms for the estimation of tobit model with non-normal error: The case of asymmetric laplace distribution. Computational Statistics & Data Analysis 67, 226-235.
[27] Yang, Y., H. J. Wang, and X. He (2015). Posterior inference in bayesian quantile regression with asymmetric laplace likelihood. International Statistical Review .
[28] Yu, K. and R. A. Moyeed (2001). Bayesian quantile regression. Statistics & Probability Letters 54 (4), 437-447.
[29] Yu, K. and J. Stander (2007). Bayesian analysis of a tobit quantile regression model. Journal of Econometrics 137 (1), 260-276.