Simulation Study for Penalized Bayesian Elastic Net Quantile Regression

Muntadher Hashim Mnati Almusaedi   Dr. Ahmad Naeem Flaih

stat.post32@qu.edu.iq   ahmad.flaih@qu.edu.iq

University of Al-Qadisiyah

Abstract

Bayesian regression analysis has great importance in recent years, especially in the Regularization method, Such as ridge, Lasso, adaptive lasso, elastic net methods, where choosing the prior distribution of the interested parameter is the main idea in the Bayesian regression analysis. By penalizing the Bayesian regression model, the variance of the estimators are reduced notable and the bias is getting smaller. The tradeoff between the bias and variance of the penalized Bayesian regression estimator consequently produce more interpretable model with more prediction accuracy. In this paper, we proposed new hierarchical model for the Bayesian quantile regression by employing the scale mixture of normals mixing with truncated gamma distribution that stated by (Li and Lin, 2010) as Laplace prior distribution. Therefore, new Gibbs sampling algorithms are introduced. A comparison has made with classical quantile regression model and with lasso quantile regression model by conducting simulations studies. Our model is comparable and gives better results.

Keywords: Bayesian analysis, quantile regression, elastic net, Gibbs sampler.

1. Introduction

Regression analysis concerning in the relationship between the response variable Y and one or more predictor variables X. However, regression analysis can be used for find the regression model that offers more prediction accuracy and more interpretability. Additionally, regression analysis provides variable selection procedure. Linear regression model is a statistical tool to estimate the mean of the response variable (y) by using the information from the predictor variables. The Ordinary Least Squares (OLS) estimators are BLUE. It is well known that the estimation methods of regression coefficients produce reliable estimators with tradeoff between the variance and bias, (Kirkland, 2014) as well as the model explainability. Meanwhile, the OLS offers biased and inconsistent (inflated variance) estimators when the collinearity problem present in the data, and when the number of predictors p greater or near the sample size n. To address the drawbacks of least squares estimates quality. Briefly, the regularization procedure is a tradeoff between the variance and bias of estimator. The regularization regression methods are used to overcome the lack of least squares method in case of p>n (many predictors) or in the presence of collinearity, but it is taken that produces biased estimators with the reduction of the variance (James et al., 2013). The ridge method proposed by (Hoerl and Kennard, 1970) adding the L_1-norm constrain to residuals sum of squares (RSS) term to overcome the collinearly or p>n problem, but ridge parameters estimates will not set to zero (not sparse). (Tibshirani, 1996), Suggested the lasso (Least absolute shrinkage and selection operator) method which is works under the same circumstances of ridge method but with adding L_2 – norm constrain to RSS term. The lasso method has ability to set the coefficient estimates equal to zero, that is mean the lasso method has the ability to remove the irrelevant predictor variables and consequently produce more interpretable model. Also, the Elastic Net (EN) is another regularization regression method proposed by (Zou and Hastie, 2005) which adding the ridge and lasso to the RSS term, EN method deal with many relevant predictors that have highly pairwise correlation and EN oftentimes outperforms the lasso (Osborne et al., 2000). The combined penalties method, such as, the elastic net considered two penalty functions $\mathrm{L}_1 - \mathrm{norm}$ and $\mathrm{L}_2 - \mathrm{norm}$, that is, the lasso and ridge penalty function added to residual sum of squares, the elastic net was proposed by (Zou and Hastie, 2005) to combine the ridge and lasso functions to deal with the grouping effect when there are strong pairwise correlations between groups of predictor variables, the elastic net estimator is defined as follows,

$$\hat{\beta}_{EN} = argmin\|\boldsymbol{y} - \boldsymbol{x}_i'\boldsymbol{\beta}\|^2 + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|^2,$$

Where λ_1 and $\lambda_2 \geq 0$ are the regularization parameters λ_1 and λ_2 that controls the amount of shrinkage that forced on the regression parameters. The elastic net works well with high correlated predictor variables.

Many of times in practice we find out that the data exhibits the violation of the linear model assumptions or the researchers are interested in modelling other quantities rather than the mean of the response variable $E(y|x)$, Such as the median, and other quantiles (Chatterjee and Hadi, 2013).

It is well known that the quantile regression required no assumptions to impose on the residual term (Koenker and Bassett, 1978). Quantile regression can be applied in many different fields such as, econometrics, ecology, biology, survival analysis and many other fields of sciences. The quantile regression model is

$$\boldsymbol{y_i} = \boldsymbol{x_i'\beta(\gamma)} + \boldsymbol{\epsilon_i(\gamma)}, \qquad (\mathbf{1})$$

Where β (γ) can be estimated by minimizing the RSS, that is

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = \boldsymbol{argmin}\sum_{i=1}^{n}\boldsymbol{\rho_\gamma}\,(\boldsymbol{y_i} - \boldsymbol{x_i'\beta}), \qquad (\mathbf{2})$$

Here $\rho_\gamma\,(\cdot)$ the quantile loss function (Koenker and Bassett, 1978) and defined as the following piecewise function,

$$\boldsymbol{\rho_\gamma}\,(\boldsymbol{\epsilon}) = \frac{|\boldsymbol{\epsilon}| + (\mathbf{2}\boldsymbol{\gamma} - \mathbf{1})\boldsymbol{\epsilon}}{\mathbf{2}}\;;\quad \boldsymbol{\epsilon} = \boldsymbol{y_i} - \boldsymbol{x_i'\beta} \qquad (\mathbf{3})$$

(Ghosh, 2007) introduced new method of regularization of the elastic net that is called adaptive elastic net where the estimator have desirable properties of adaptive lasso method and elastic net method. (Alshaybowee et al., 2016) introduced the Bayesian elastic net in the single index quantile regression model as a method to address the high dimensionally in data with the nonparametric regression model. (Lee et al., 2016) presented the elastic net shrinkage method to overcome the dimensionality problem in the data that have high correlation between the predictor variables with group selections. (Jiratchayut and Bumrungsup, 2015) studied the adaptive elastic net with different adaptive weight along with least squares estimators weights. They showed in the simulation example that the adaptive elastic net weights estimator performs better in terms of estimation accuracy and variable selection procedure. (Feng, 2011) developed Bayesian MCMC algorithm for estimating the quantile linear regression parameters under two proposed Bayesian quantile model methods, the estimators are efficient compared with some existing regression methods. (Al-hamzawi, 2013) proposed some extensions on the Bayesian quantile regression through the prior distribution that allows the full conditional conjugate prior. (Al-hamzawi, 2016) proposed the Bayesian Tobit quantile regression model under the gamma prior for the regression coefficients with the elastic net penalty function. (Li et.al, 2010) studied the regularization regression method, such as, Lasso, elastic net, and group lasso with Bayesian analysis of the quantile regression.

(Li and Lin, 2010) proposed new prior distribution for the elastic net under the Bayesian analysis of the linear regression to avoid the double shrinkage problem in the elastic net penalty function, the prior form of $\pi(\beta|\sigma^2)$ is proportional to

$$\mathbf{c}(\lambda_1, \lambda_2, \sigma^2) \prod_{j=1}^{p} \int_{1}^{\infty} \sqrt{\frac{t}{t-1}} \exp\left\{-\frac{\beta_j^2}{2}\left(\frac{\lambda_2}{\sigma^2}\frac{t}{t-1}\right)\right\} t^{\frac{-1}{2}} \exp\left(-\frac{1}{2\sigma^2}\frac{\lambda_1^2}{4\lambda_2}\right) dt \quad (4)$$

In this paper new hierarchical model and new Gibbs sampler algorithm have been proposed for the quantile regression improving the prediction accuracy of the proposed model.

2. The model hierarchy and prior distributions

Based on the quantile regression model (1.1) and the prior density (1.4), we have the following Bayesian elastic net quantile model hierarchy representation

$$y_i^* = x'_i \beta_p,$$

$$y_i^* | \beta, V_i \sim N\left(y_i^*; x_i' \beta_p + \ \delta V_i, 2V_i \right),$$

$$V_i \sim \text{Exp}\left(V_i ; \frac{1}{p} (1 - p) \right),$$

$$\beta_j | \tau, \sigma^2 \sim \prod_{j=1}^{p} N\left(0, \left(\frac{\lambda_2}{\sigma^2} \frac{\tau_j}{\tau_j - 1} \right)^{-1} \right),$$

$$\tau | \sigma^2 \sim \prod_{j=1}^{p} \text{Truncated Gamma} \left(\frac{1}{2}, \frac{8\lambda_2 \sigma^2}{\lambda_1^2} \right), \tau \in (1, 0)$$

$$\sigma^2 \sim \frac{1}{\sigma^2}, \qquad (5)$$

3. Posterior Distributions with Full Conditional Model.

Supposing that all priors for the different parameters are independent, we can write down the full conditional distribution as follows.

$$y_i^* / V_i, \beta \sim N\left(x_i' \beta + \delta V_i, 2V_i\right)$$

Where $i = 1, 2, \dots, n$

Following (Alhamzawi, 2016) and (Li and Lin, 2010) and conditioning ony^*, V_i, β the posterior distribution of β is

$$\pi\left(\beta / y^*, \sigma^2, \tau \right) \propto \pi\left(y^* / \beta, \sigma^2, \tau \right) \pi\left(\beta / \sigma^2 \right)$$

$$\propto \exp\left\{ -\frac{1}{2\sigma^2} (y^* - x'\beta)'(y^* - x'\beta) \right\} \exp\left\{ -\frac{1}{2\sigma^2} \beta' D_\tau \beta \right\}$$

$$= -\frac{1}{2\sigma^2} \left(\beta - c^{-1} x' y^* \right)' c \left(\beta - c^{-1} x' y^* \right)$$

Then β distribution is the multivariable normal with mean $c^{-1} x' y^*$ and variance $\sigma^2 c^{-1}$;

$$\beta / y, \sigma^2, \tau, \sim N\left(c^{-1} x'^{y^*}, \sigma^2 c^{-1} \right) \qquad (6)$$

The second variable σ^2 distributed as follows

$$(\frac{1}{\sigma^2})^{\frac{n}{2}+p+1}\{\Gamma_u\left(\frac{1}{2},\frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\}^{-p}exp[-\frac{1}{2\sigma^2}\left\{\left((y^* - x'\beta)'(y^* - x'\beta) + \lambda_2\sum_{j=1}^{p}\frac{\tau_j}{\tau_j - 1}\beta_j^2\right.\right.$$

$$\left.\left. + \frac{\lambda^2}{4\lambda_2}\sum_{j=1}^{p}\tau_j\right\}\right] \qquad (7)$$

The third variable $(\tau - 1_P)$ distributed as

$$(\tau_j - 1_p)^{-1}/y, \sigma^2, \beta \sim IG\left(\mu = \sqrt{\lambda_1}/(2\lambda_2\,|\beta_j|),\ \lambda = \frac{\lambda_1}{4\lambda_2\sigma^2}\right) \qquad (8)$$

4. The Gibbs Sampling From the Full Conditional Distribution

We will use the Markov Chain Monte Carlo (MCMC) special algorithm that is called Gibbs sampling to implement the hierarchical model (1.5). The Gibbs sample generates (samples) random variables indirectly from the full conditional distributions of a parameter fixed all the other parameters (Evans, 2012). The conditional posterior densities of each parameter will be generate for the elastic net quantile regression by using the following algorithms:

1- Updating y_i^* from the following full conditional distribution

$$y_i^*/V_i, \beta \sim N\,(x_i'\beta + \delta V_i, 2V_i)$$

Where $i = 1, 2, \ldots, n$.

2- Updating $\beta/y, \sigma^2, \tau$ from the full conditional posterior density which following the multivariate normal distribution (1.6) with mean $C^{-1}X'Y^*$ and variance $\sigma^2 C^{-1}$, where

$$C = x'x + \lambda_2\,(D_\tau)\,;\ D_\tau = diag\,(\frac{\tau_1}{\tau_1 - 1}, \ldots, \frac{\tau_p}{\tau_p - 1}).$$

3- Updating V_i^{-1} ; $i = 1,2, \ldots, n$ from the full conditional posterior distribution of V_i^{-1} which is follows Inverse Gaussian (μ', λ') see (Alhamzawi, 2016), where

$$\mu' = \sqrt{\frac{1}{(y_i^* - x_i'\beta)^2}} \quad and \quad \lambda' = \frac{1}{2},$$

(Chhikarn and Folks, 1988) stated the inverse Gaussian density is:

$$\mathrm{f}\,(\boldsymbol{x}/\boldsymbol{\lambda}',\boldsymbol{\mu}') = \sqrt{\frac{\boldsymbol{\lambda}'}{2\pi \boldsymbol{x}^3}}\,\boldsymbol{exp}\left\{\frac{-\boldsymbol{\lambda}'(\boldsymbol{x}-\boldsymbol{\mu}')^2}{2(\boldsymbol{\mu}')^2\boldsymbol{x}}\right\};\boldsymbol{x} > \mathbf{0}$$

4- Updating $(\tau_j - 1)^{-1}/y\,,\sigma^2\,,\beta$ from the full conditional inverse Gaussian distribution (Chhikarn and Folks, 1988)

$$\boldsymbol{With}\;\boldsymbol{\mu} = \frac{\sqrt{\boldsymbol{\lambda}_1}}{(2\boldsymbol{\lambda}_2|\boldsymbol{\beta}_j|)}\;\boldsymbol{and}\;\boldsymbol{\lambda} = \frac{\boldsymbol{\lambda}_1}{4\boldsymbol{\lambda}_2\boldsymbol{\sigma}^2}\;;\;\boldsymbol{j} = \mathbf{1},\mathbf{2},\dots,\boldsymbol{p}$$

5- Updating $\sigma^2/y\,,\beta,\tau$ by using the acceptance-rejection algorithm that depends on the incomplete gamma functions;

$$(\boldsymbol{\sigma}^2) \le \frac{\boldsymbol{\Gamma}_a\,\boldsymbol{\Gamma}_{\frac{1}{2}}^{-p}}{\boldsymbol{b}^a}\,\boldsymbol{h}(\boldsymbol{\sigma}^2);$$

$$\boldsymbol{Where}\;\boldsymbol{a} = \frac{\boldsymbol{n}}{\mathbf{2}} + \boldsymbol{p}\,,\boldsymbol{b} = \frac{\mathbf{1}}{\mathbf{2}}\left[\|\boldsymbol{y}^* - \boldsymbol{x}'\boldsymbol{\beta}\| + \boldsymbol{\lambda}_2\sum_{j=1}^{p}\frac{\boldsymbol{\tau}_j}{\boldsymbol{\tau}_j - \mathbf{1}}\,\boldsymbol{\beta}_j^2 + \frac{\boldsymbol{\lambda}_1^2}{4\boldsymbol{\lambda}_2}\sum_{j=1}^{p}\boldsymbol{\tau}_j\right].$$

And h (.) is the inverse gamma (a, b).

5- Simulation Analysis

In this section simulation study will be conducted to show the behavior of our proposed model, Bayesian elastic net quantile regression (Benqr) using R packag and compared with different other models; the classic quantile regression model (cqr) by implementing R package, and the lasso quantile regression model (lqr) by implementing R package. Our comparison is based on the parameters estimates of the different models under different quantiles $(\tau = 0.25\,,\tau = 0.50\,,\tau = 0.75\,,\text{and}\;\tau = 0.95)^*$. Also, we used the median mean absolute deviation (mmad) criterion,

$$\boldsymbol{mmad} = \boldsymbol{median}\left[\boldsymbol{mean}\left|\boldsymbol{x}'\widehat{\boldsymbol{\beta}} - \boldsymbol{x}'\boldsymbol{\beta}^{true}\right|\right].$$

The mmad and the standard deviation (sd) are used to measure the performance of prediction accuracy for different model. The Gibbs sampler algorithm have been used with 10000 iterations to generate the stability of the posterior distribution of the interested parameter, the first 1000 iterations have burned in we generated the observations of $x_1, \ldots, x_9$ predictor variables from$N_{n=9}(0\ ,\Sigma)$, where the matrix $\Sigma_{ij} = \rho^{|i-j|}$ with three distribution distributions of the (iid) errors. For each simulation study, we run 400 simulations.

1. Simulation Example

In this simulation, we supposed that the true vector of parameter $\beta = (\ 0,3,0,0,0,0,0,0,0)^t$ with error terms followed $\epsilon_i \sim N(\mu = 0, \sigma^2 = 1), \epsilon_i \sim Nnormal\ mixture$. Also, we generated the observation of $x_1, \ldots, x_9$ predictor variables based on $N_{n=9}(0\ ,\Sigma)$, where Σ is the var-cov matrix defined as$\sum ij = 0.5^{|i-j|}$. As well as, we simulated $y_i = 3x_2$.

True para	0	3	0	0	0	0	0	0	0
Our method 0.25	-0.53433	3.05755	-0.04878	0.15044	0.05801	0.02518	0.05821	0.09411	-0.11741
rq lasso0.25	-0.48510	3.11706	-0.10233	0.22113	0.03017	0.06392	-0.00889	0.05554	-0.13766
rq method0.25	-0.31952	3.26606	-0.22838	0.30980	0.02538	0.36293	-0.07736	-0.12678	0.01154
Our method 0. 50	0.00348	2.61704	0.02319	0.05258	0.19002	0.13365	-0.03415	-0.13002	-0.12188
rq lasso0. 50	0.01652	2.71332	-0.00693	0.05948	0.16909	0.13757	-0.01339	-0.14076	-0.12048
rq method0. 50	-0.06342	2.80605	-0.42396	0.40156	0.04339	0.65369	-0.16830	-0.24411	0.17263
Our method 0. 75	0.59039	2.64516	0.33294	0.08248	-0.0428	-0.06223	-0.07704	0.00512	-0.19677
rq lasso 0. 75	0.54035	2.88732	0.25855	0.01500	-0.0327	-0.18471	-0.08628	0.10899	-0.18795
rq method 0. 75	0.60376	3.15254	-0.15802	-0.3066	0.31255	0.15051	0.08138	0.64905	-0.18132
Our method 0. 99	2.16136	2.82956	0.06402	-0.02756	-0.0494	-0.05186	-0.00629	-0.09735	0.01993
rq lasso 0. 99	1.47320	3.27608	0.06881	-0.04868	0.02646	0.09077	-0.17528	-0.22168	-0.01531
rq method 0. 99	0.80679	3.50108	-0.05175	-0.36564	0.16453	0.31269	-0.13776	-0.18034	-0.02319

Table 1. Parameter estimates of simulation 1 with $\epsilon_i \sim N(0,1)$.

True para	0	3	0	0	0	0	0	0	0
Our method 0.25	-0.94167	3.08580	0.18950	0.12903	-0.0174	0.23762	-0.09033	0.24558	-0.11196
rq lasso0.25	-0.91259	2.87556	0.11570	0.15341	0.05266	0.18572	-0.13328	0.53398	-0.28134
rq method0.25	-0.91319	2.52562	0.08169	0.28278	0.16490	-0.35038	-1.25653	1.04882	-0.64224
Our method 0. 50	-0.08201	2.35569	-0.03954	0.24500	0.05604	-0.11054	0.01792	-0.14199	0.14266
rq lasso0. 50	-0.10540	2.58759	-0.09529	0.26693	0.05386	-0.06312	0.05668	-0.10855	0.13835
rq method0. 50	-0.11045	2.72309	-0.17261	0.15492	0.10169	-0.03924	-0.13531	0.00210	0.01763
Our method 0. 75	0.77623	2.68634	0.06926	0.02502	-0.0233	-0.17077	-0.16754	-0.07847	0.15773
rq lasso 0. 75	0.77059	3.00004	0.16754	-0.02458	-0.0834	-0.09377	-0.22678	-0.08948	0.21058
rq method 0. 75	0.70621	3.37876	0.28050	-0.04396	-0.2131	0.55459	-0.44564	0.07573	0.30717
Our method 0. 99	1.19079	2.80197	-0.16346	0.10390	0.53281	0.15548	0.30575	-0.14864	0.32483
rq lasso 0. 99	1.94265	2.60253	0.01807	-0.05214	0.29374	0.18550	0.02513	-0.12150	0.05225
rq method 0. 99	2.78453	1.96509	0.20717	0.11471	0.42179	0.24586	0.23482	-0.08939	-0.07131

Table 2. Parameter estimates of simulation 1 with $\epsilon_i \sim$Normal mixture.

Errors distribution				
The methods	Quantile level	$\epsilon_i \sim N(0,1)$	$\epsilon_i \sim$ Normal mixture	$\epsilon_i \sim \chi_3^2$
Benqr	0.25	0.3617(0.37434)	0.6509 (0.84568)	0.352(0.33332)
Lqr	0.25	0.4428 (0.46550)	0.6617 (0.82850)	0.387(0.38830)
Cqr	0.25	0.5911 (0.59670)	1.0422 (1.11788)	0.532(0.54044)
Benqr	0.50	0.4394 (0.41602)	0.4890 (0.66236)	0.2731(0.28762)
Lqr	0.50	0.4642 (0.38906)	0.6096 (0.57216)	0.2897(0.32632)
Cqr	0.50	0.5975 (0.55410)	0.8125 (0.85222)	0.5209(0.48554)
Benqr	0.75	0.4075 (0.43762)	0.3674 (0.46434)	0.3045(0.3296)
Lqr	0.75	0.4465 (0.42018)	0.5239 (0.55014)	0.3564(0.37862)
Cqr	0.75	0.7371 (0.75410)	0.8570 (0.89084)	0.4747(0.50552)
Benqr	0.99	0.5442 (0.54374)	0.6967 (0.90858)	0.7749(0.73352)
Lqr	0.99	0.8628 (0.94924)	0.9078 (0.91796)	1.2781(1.31062)
Cqr	0.99	1.5671 (1.60992)	1.3995 (1.45826)	1.8570(1.83770)

Table 3. MMAD and S.D. for simulation example 1

From table 1- table 3 for the previous simulation example, obviously the parameter estimates of the proposed model ((Benqr) are comparable with (cqr) and (lqr), also from the values of the criterions mmad and SD it can be observed that the proposed model were relatively less than these results of classic quantile regression (cqr) and the lasso quantile regression (lqr) models and yields the best values of mmad and SD in the most of the simulations times. Consequently, it can be shown that the proposed model (Benqr) outperformed the other regression models.

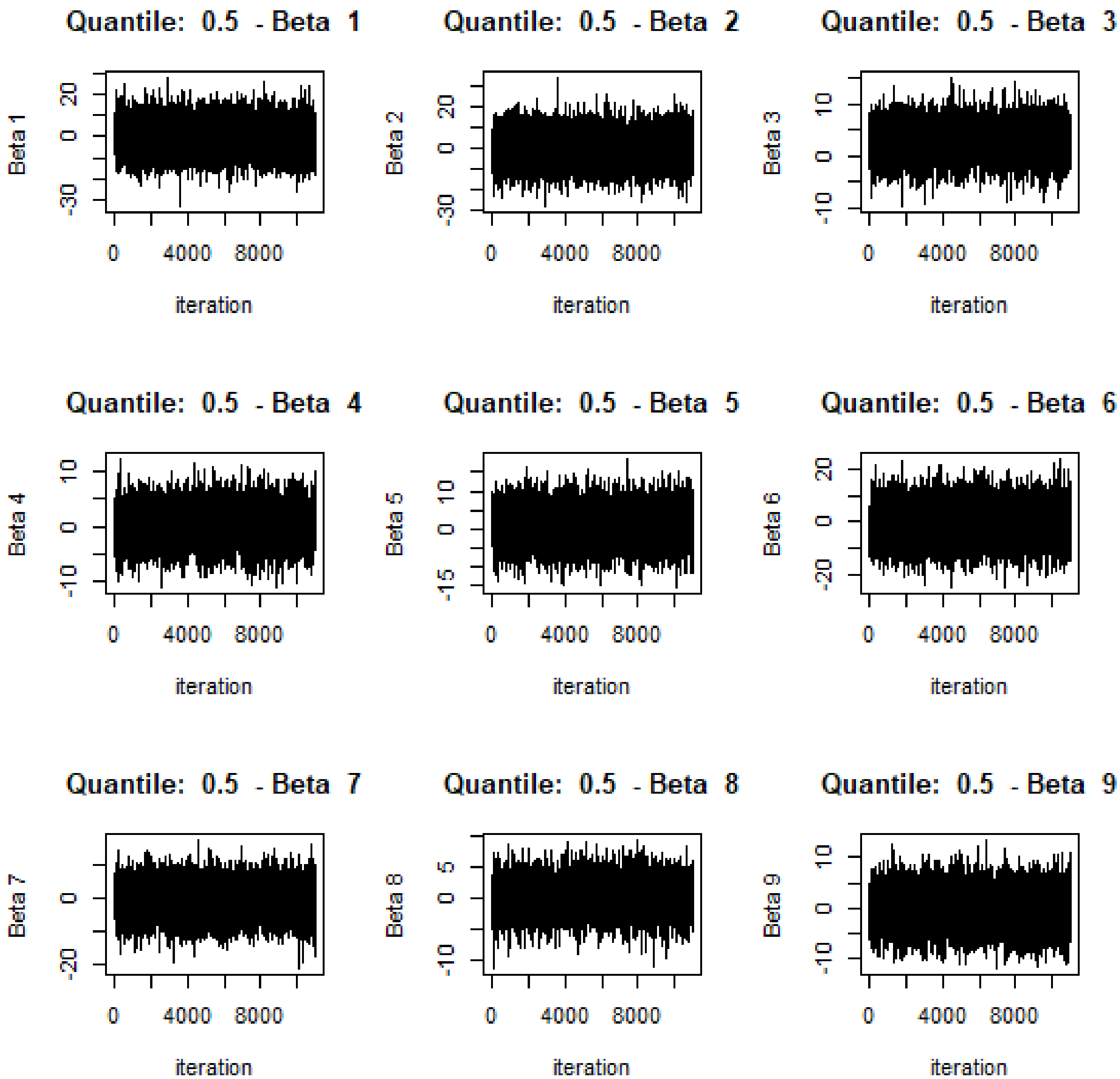

Figure 1. Trace plots of our model with (0.5) quantile

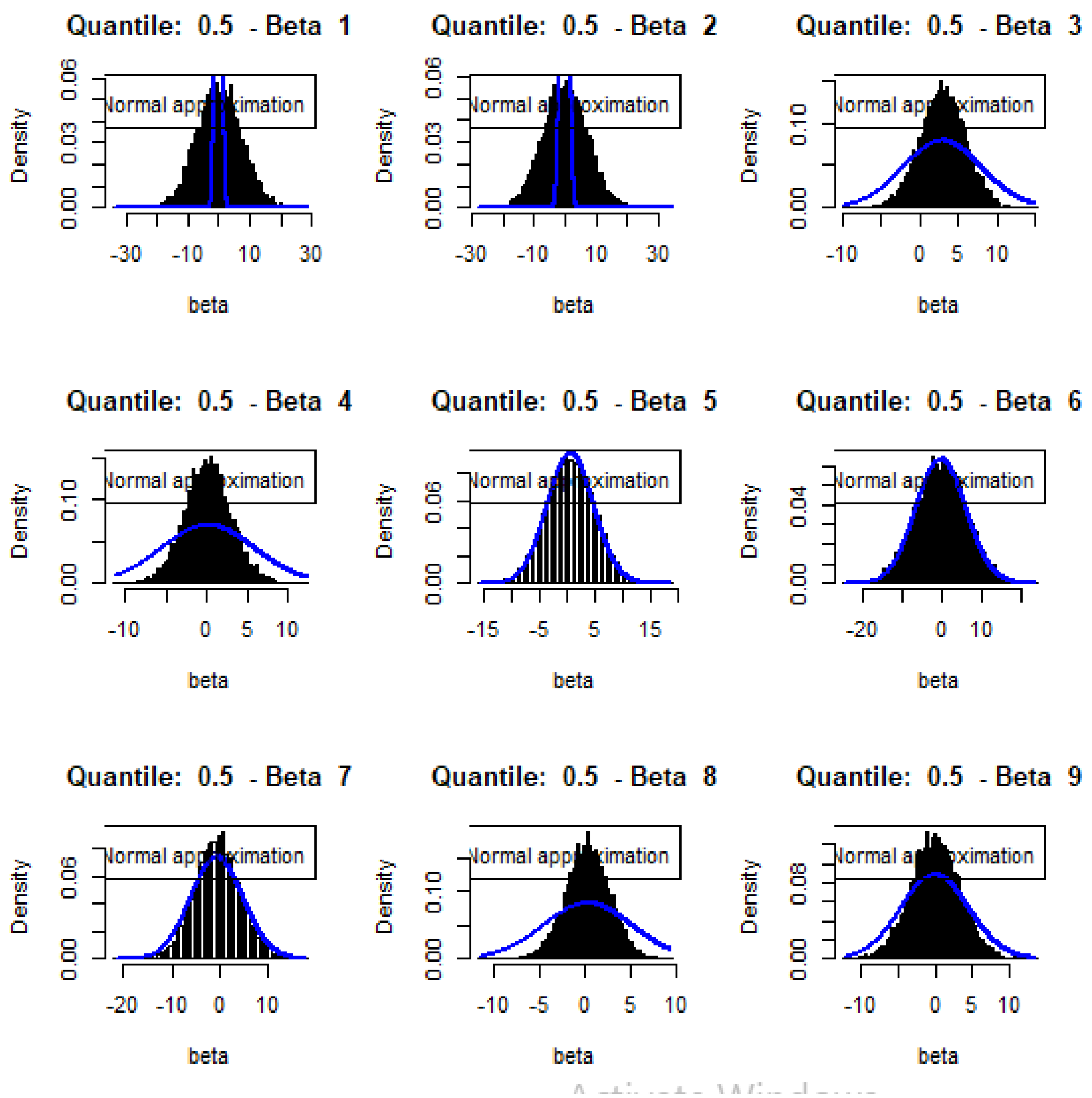

Figure 2. Histograms of our model parameter estimates

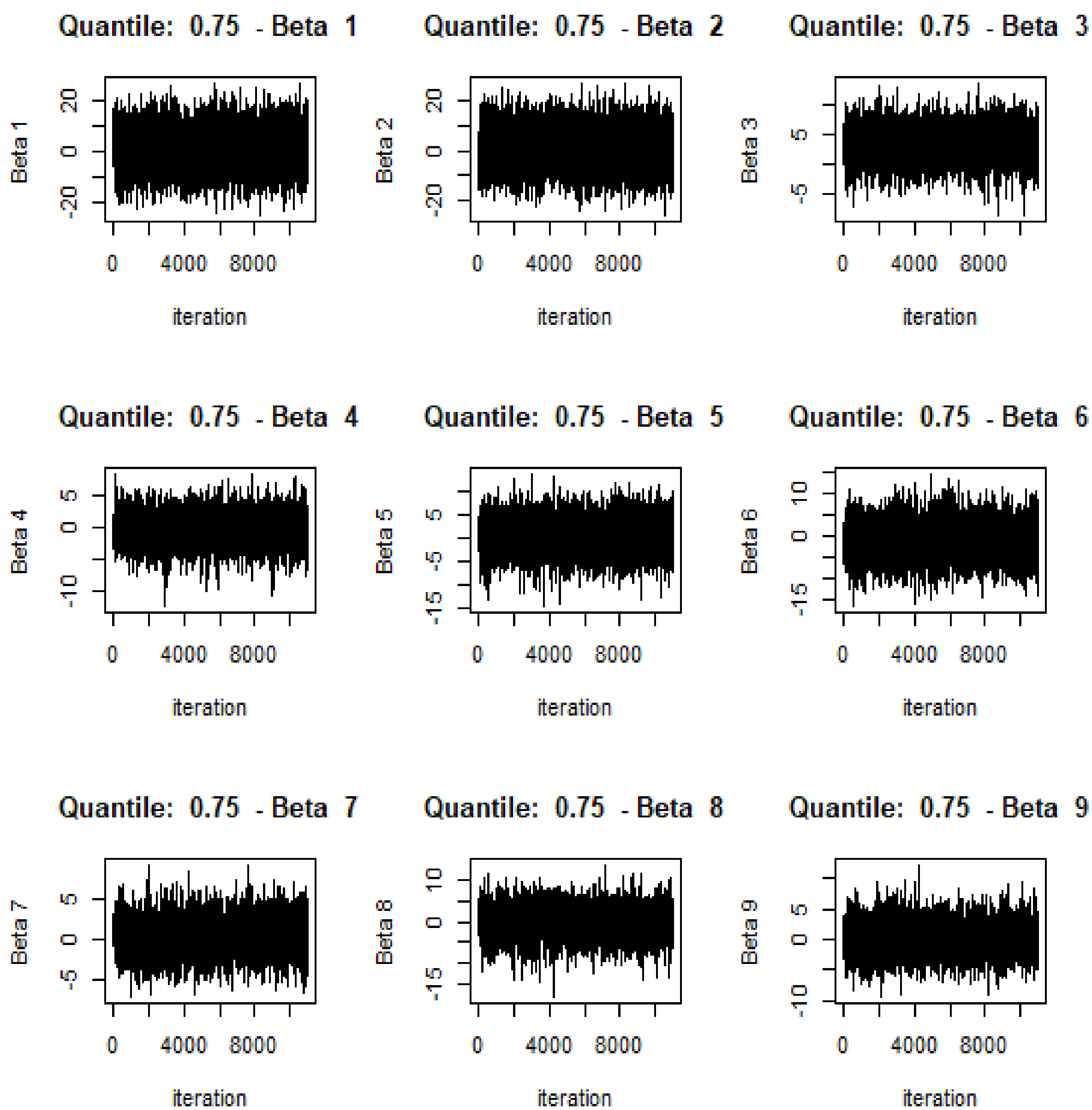

Figure 3. Trace plots of our model with (0.75) quantile

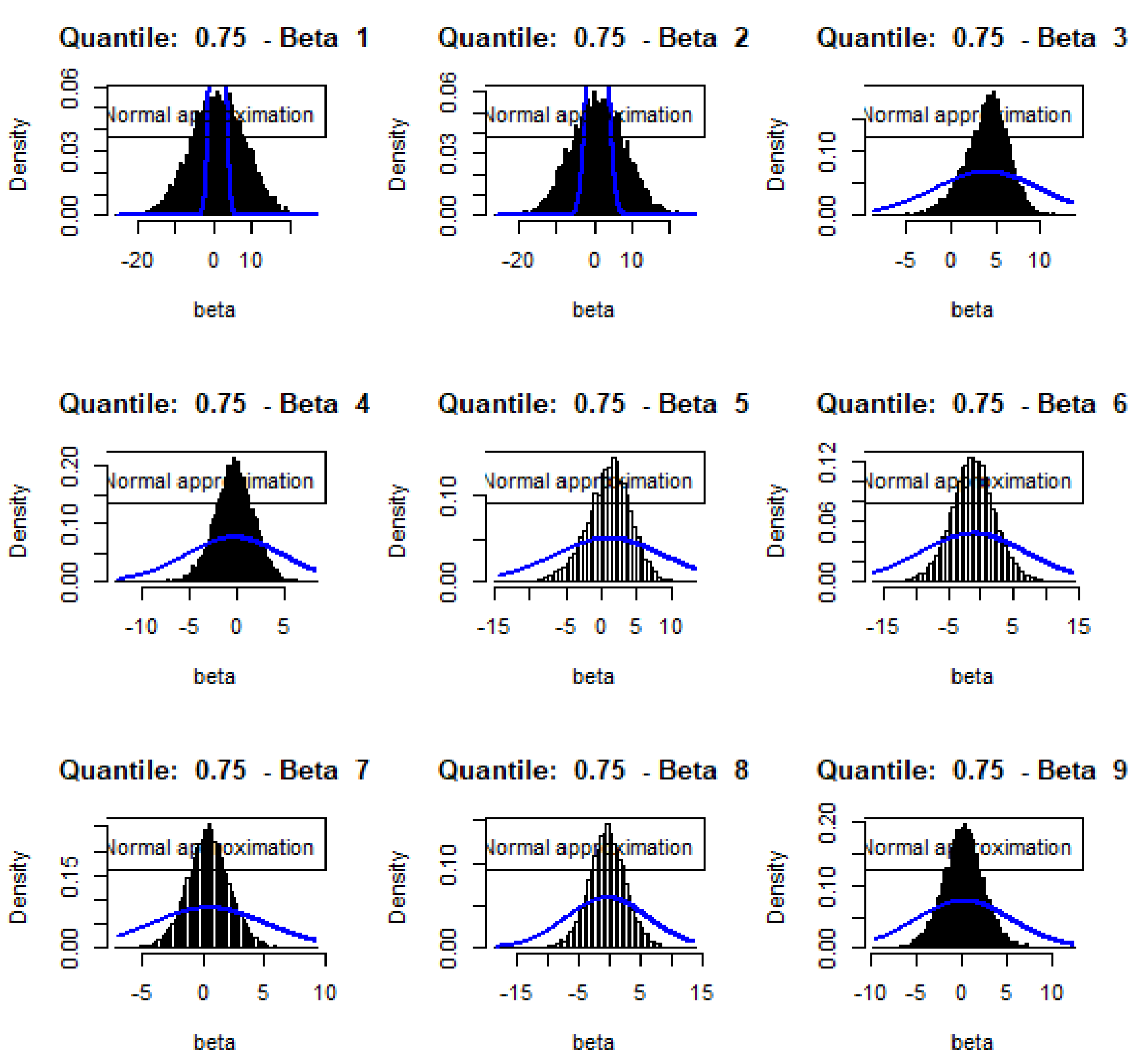

Figure 4. Histograms of our model parameter estimates

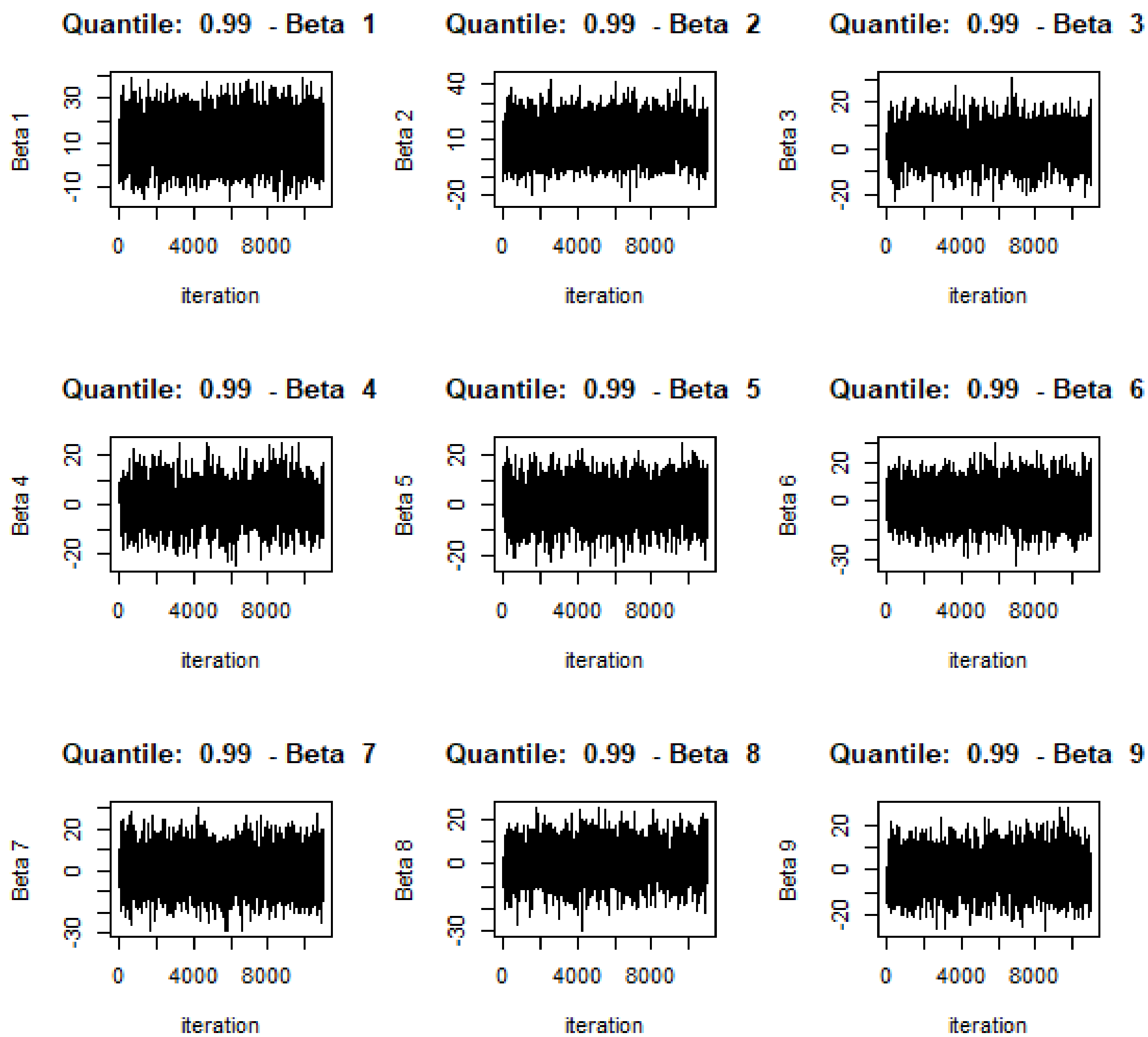

Figure 5. Trace plots of our model with (0.99) quantile

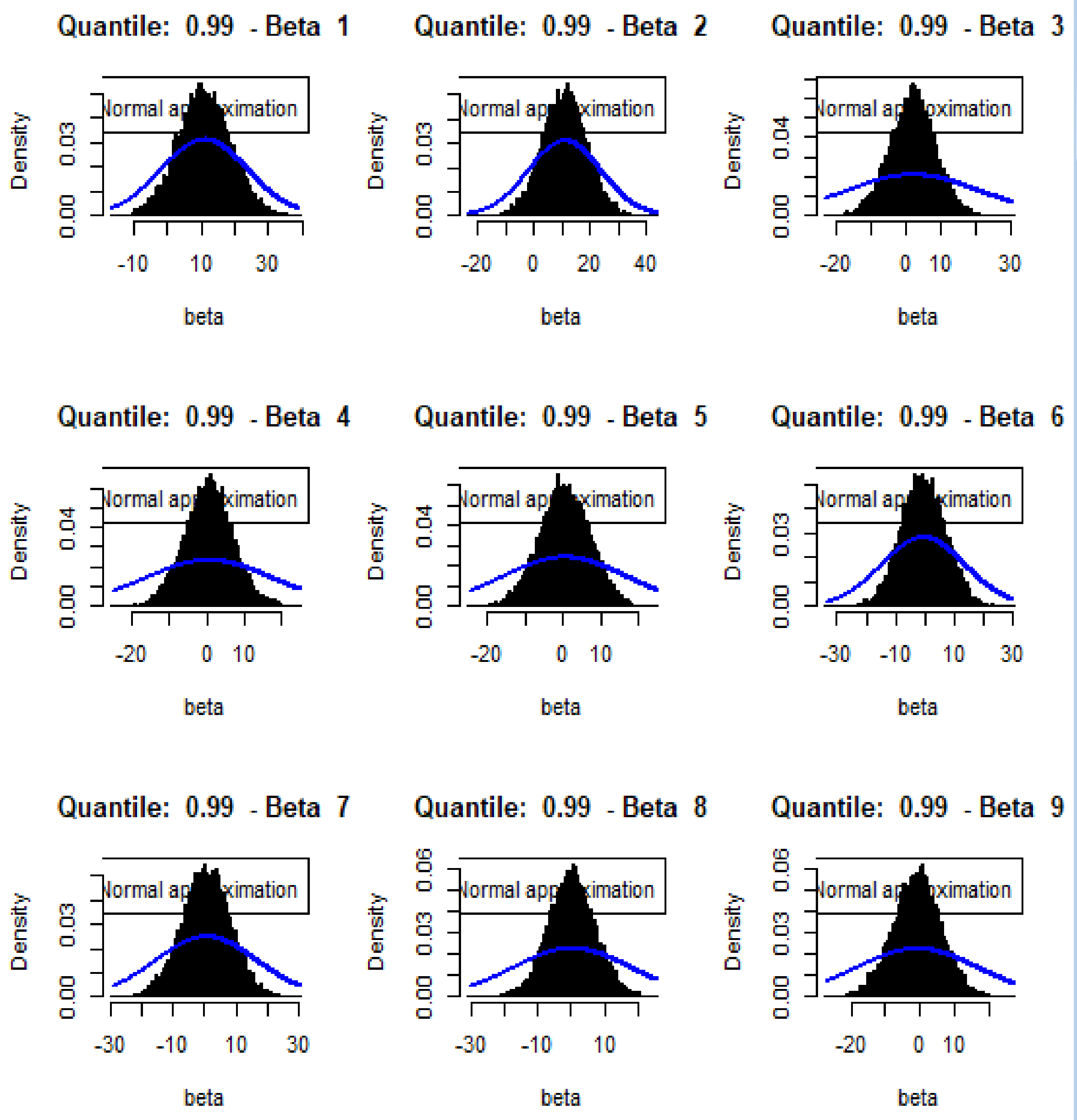

Figure 6. Histograms of our model parameter estimates

Figures 1-6 some of the displayed the histograms tables fit the distributions of the parameter estimates and it is very clear that the distribution of the parameter estimates distributed according to the normal distribution under the different quantile levels, and the rest of figures displayed the trace plot which are regards as convergence diagnose tool that indicates the MCMC samples of the posterior distribution of regression parameter estimates convergence to stationary distribution (true parameter values), which is mean the Gibbs sampling algorithm is easy to implement and it is efficient.

7- Conclusions

This paper presented new contribution for the Bayesian elastic net quantile regression models through employing the Laplace density of parameter (β) as scale mixture of normals mixing with truncated gamma distribution that proposed by li and lin (2010) into the quantile regression. New hierarchical model has developed for the proposed model, as well as I provided Gibbs sampler algorithm for the proposed posterior distribution. I displayed the advantages of the proposed model in the simulation analysis. The results explained that the proposed model is comparable model in terms of the parameter estimation and in terms of the quality of the estimates through the values of MSE criterion.

Bibliography

Al-hamzawi, R., & Yu, K. (2013). Conjugate priors and variable selection for Bayesian quantile regression. Computational Statistics & Data Analysis, 64, 209-219.

Al-hamzawi, R. (2016). Bayesian elastic net Tobit quantile regression. Communications in Statistics-Simulation and Computation, 45(7), 2409-2427.

Alshaybawee, T. Midi, H. & Alhamzawi, R. (2016). Bayesian elastic net single index quantile regression.

Chatterjee, S. & Hadi, A.S (2013). Regression Analysis by example. Wiley series in probability and statistics, Fifth edition.

Chhikara, R. and Folks, L. (1988). The Inverse Gaussian distribution Theory: Method-ology, and Applications. Marcel Dekker, Inc, 1st edition. 156

Evans, S. (2012). Bayesian Regression Analysis. Master thesis. University of Louisville.

Feng, Y (2011), Bayesian Quantile Linear Regression. University of IIIinois at Urbana-Champaign.

Ghosh, S. (2007). Adaptive Elastic Net: An Improvement of Elastic Net to achieve Oracle Properties. Technical Reports, Indiana University Purdue University, Indianapolis, USA. PR no. 07-01.

Hoerl, A.E & Kennard, R.W (1970). "Ridge regression: Biased estimation for Non-Orthogonal problems", Technometrics, 12(1): 55-67.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). An introduction to statistical learning: with applications in R (Springer Texts in Statistics: 103). New York, NY, USA: Springer.

Jiratchayut, K. and Bumrungsup, C. (2015). A Study of Adaptive Elastic Net Estimators with Different Adaptive Weights. University Physical and Technology 20(3), P-ISSN 2686-9000.

Kirkland, Lisa-Ann. (2014). Lasso-Simultaneous shrinkage and election via ρ_1 norm. Msc. thesis. University of Pretoria.

Koenker, R., and G Bassett (1978): Regression Quantiles, Econometrica, Vol. 46, No. 1, pp.33-50.

Lee[1], Y. Nguyen[2], V. and Wang[3], D. (2016). On Variable and Grouped Selections of the Elastic Net. University of Wisconsin-Madison. Dec 18.

Li, Q., and Lin, N. (2010). The Bayesian Elastic Net. Bayesian Analysis, 5(1), 151-170.

Li, Q., Xi, R., and Lin, N. (2010). Bayesian regularized quantile regression. Bayesian Analysis, 5(3), 533-556.

Osborne, M.R., Presnell, B. and Turlach, B.A. (2000b). On the LASSO and its dual. Journal of Computational and Graphical Statistics, 9(2):319-337.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 58(1):267-288.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2):301-320.