Bayesian Elastic Net Quantile Regression using a New Scale Mixture of Normals with an Application

A thesis Submitted for the Degree of

Master of Science in Statistics

By

Muntadher Hashim Mnati Almusaedi

Supervised by

Prof. Dr. Ahmad Naeem Flaih

Statistics Department

College of Administration and Economics

University of Al-Qadisiyah

2021

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

﴿وَمَا أَرْسَلْنَا مِنْ قَبْلِكَ إِلَّا رِجَالًا نُوحِي إِلَيْهِمْ فَاسْأَلُوا أَهْلَ الذِّكْرِ إِنْ كُنْتُمْ لَا تَعْلَمُونَ ﴾

صدَقَ الله العَظِيم

[النحل: 43]

Aknowledgements

I would like to thank my supervisor Prof. Dr. Ahmad N. Flaih for his limitless help and support.

I would also like to thank everyone who taught me or contributed the statistics classes, Prof. Dr. Mohammad AL-Sharoot, Dr. Rahim AL-Hamzawi, Dr. Ali Alkenani, Dr. Tahir Reisan, Dr. Muhannad Alsaadony, Dr. Taha Alshaybawe and Dr. Fadel Alhusseini, Dr. Bahr Kazem, and Dr. Haydar hadi.

Finally, I would to thank my dear parents, my brothers, my sisters, and my friends Rasol, Samir, Omar for the directly or indirectly support.

Muntader

1 / 2 / 2021

DEDICATION

I would like to dedicated this thesis to my martyr brother

Raad Hashem Manati.

Abstract

Bayesian regression analysis has a 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 larger. The trade off between the bias and variance of the penalized Bayesian regression estimator consequently produce more interpretable model with more prediction accuracy. The quantile regression can be considered as the generalized for the classical linear regression model. The Prof Dr. Alhamzawi suggests that the concept "segment" or "percentile" can be the correct translate for "quantile".

In this thesis, 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, and based on statistical measures, (MSE, SD, MMAD), as well applying real data. Our model is comparable and gives better results.

Table of Contents

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter Five

List of Tables

List of Figures

Chapter One

1. Introduction

Regression analysis is trying to investigate the functional relationship between the response variable **Y** and one or more predictor variables **X**. Consequently, regression analysis can be used for creating the regression model that characterized by more prediction accuracy and more interpretability, additionally the functional form of the regression model linked with variable selection problem, where in variable selection process the irrelevant independent variables removed from the predicted model.

In many situations the researchers depends on linear regression model to estimate the mean of response variable (**Y**) by using the information from the predictor variables. The Ordinary Least Squares (**OLS**) estimation method usually offers unbiased and lowest variance estimators (**BLUE**) through solving the following linear regression model by minimizing the Residual Sum of Squares (**RSS**),

$$\boldsymbol{RSS(\beta)} = \sum_{i=1}^{n} (\boldsymbol{y_i} - \boldsymbol{f}(\boldsymbol{X};\ \boldsymbol{\beta}))^2$$

It is well known that the estimation methods of regression coefficients produce reliable estimators with trade off between the variance and bias, (Kirkland Lisa-Ann, 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**, consequently, the **OLS** estimates are not unique and vary with high variances.

The inflation of variance in the **OLS** estimates inspired the authors to study the regularization methods which are used 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 have been 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 $\boldsymbol{L_2}$**-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 $\boldsymbol{L_1}$ **– 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).

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 $\boldsymbol{E(y|x)}$, Such as the median, and other quantiles (Chatterjee and Hadi, 2013).

Recently the quantile regression analysis became more popular procedure that can be classified as a general method for estimating the $\boldsymbol{\gamma\_th}$ conditional quantile function for $\boldsymbol{\gamma} \in (\mathbf{0}, \mathbf{1})$**,** where the quantile regression model suggest a regular strategy for investigating how the predictor variables effect the location, Shape, and Scale of the entire response variable distribution (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.

In this thesis, we have concerned in studying the estimation of the quantile regression coefficient in the view of Bayesian methodology under the elastic net regularization method. We employed the priors and posteriors distributions proposed by (Li and Lin, 2010) of the elastic net regularization method in quantile regression. Where there is no such employing proposed before.

1.2. Literature Review

Suppose that the linear mean regression model is defined as follows:

$$\mathbf{Y} = \boldsymbol{X}'\boldsymbol{\beta} + \boldsymbol{\epsilon} \qquad (\mathbf{1.1})$$

Where $\boldsymbol{Y}_{\boldsymbol{n}\times\mathbf{1}}$ is the vector of response variables, $\boldsymbol{X}_{\boldsymbol{n}\times\mathbf{p}}$ is the matrix of predictor variables, $\boldsymbol{\beta}_{\boldsymbol{p}\times\mathbf{1}}$ is the vector of regression coefficients, and $\boldsymbol{\epsilon} \sim \mathbf{N}\,(\mathbf{0}, \boldsymbol{\sigma}^{\mathbf{2}}\,)$. In ordinary least squares estimation method is through solving the model $(\mathbf{1.1})$ by minimizing the $\boldsymbol{RSS}$ to obtain the estimator $\boldsymbol{\beta}$ as follows,

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{argmin}||\boldsymbol{y} - \boldsymbol{x}'\boldsymbol{\beta}||_2^2 \qquad (\mathbf{1.2})$$

The $\boldsymbol{\gamma\_th}$ conditional quantile regression function is defined as follows (koenker and Bassett, 1978),

$$\boldsymbol{Q}(\boldsymbol{\gamma}|\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}(\boldsymbol{\gamma}),$$

Where $\boldsymbol{Q}(\boldsymbol{\gamma}|\boldsymbol{x}) = \boldsymbol{F}^{-\mathbf{1}}(\boldsymbol{\gamma})$ is the inverted distribution function and $\boldsymbol{\gamma} \in (\mathbf{0}, \mathbf{1})$. Now, the quantile regression model is defined as follows (Koenker, 2005), (Marasinghe, 2014)

$$\boldsymbol{y}_{\boldsymbol{i}} = \boldsymbol{x}_{\boldsymbol{i}}'\boldsymbol{\beta}(\boldsymbol{\gamma}) + \boldsymbol{\epsilon}_{\boldsymbol{i}}(\boldsymbol{\gamma}), \qquad (\mathbf{1.3})$$

The quantile regression model $(\mathbf{1.3})$ does not required any pre assumptions on the error term $\boldsymbol{\epsilon}$. Similar to $(\mathbf{1.2})$, that is

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = \boldsymbol{argmin}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{\rho}_{\boldsymbol{\gamma}}\,(\boldsymbol{y}_{\boldsymbol{i}} - \boldsymbol{x}_{\boldsymbol{i}}'\boldsymbol{\beta}), \qquad (\mathbf{1.4})$$

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

$$\boldsymbol{\rho_\gamma}\ (\boldsymbol{\epsilon}) = \frac{|\boldsymbol{\epsilon}| + (\mathbf{2}\boldsymbol{\gamma} - \mathbf{1})\boldsymbol{\epsilon}}{\mathbf{2}} \qquad (\mathbf{1.5})$$

Where, $\boldsymbol{\epsilon} = \boldsymbol{y_i} - \boldsymbol{x_i'\beta}$

Using the **OLS** method with overfitting and collinearity problems produced instable and high variances estimates, these problems are the key idea behind resorting to use the subset selection methods and shrinkage methods. All possible subsets selection method includes fitting every potential model with the intercept term and any number of covariates, which is means there are $\mathbf{2^p}$ potential models. The forward selection, backward elimination, and stepwise methods have been developed for variable selecting and model specification, see (Lawson and Hanson, 1974), (miller, 2002), (seber and lee, 2003), (James et al., 2013) for more details. That is mean, the subset selection methods are used to overcoming the overfitting problem.

The selection of relevant variable is one of most important purposes of regression analysis (Chatterjee and Hadi, 2013). The shrinkage procedure of relevant variable through the regularization method might be controlled variance of parameter estimates and letting the irrelevant variables estimates near or equal to zero.

Subsequently, many regularization methods has been developed, such as ridge regression that proposed by (Hoerl and Kennard, 1970), where the ridge estimator is defined as follows,

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} = \boldsymbol{argmin}||\boldsymbol{y} - \boldsymbol{x_i'\beta}||_2^2 + \boldsymbol{\lambda}||\boldsymbol{\beta}||^2 \qquad (\mathbf{1.6})$$

Where $\boldsymbol{\lambda}||\boldsymbol{\beta}||^2 = \boldsymbol{\lambda}\sum_{j=1}^{p}\boldsymbol{\beta}_j^2$ is the penalty function with shrinkage parameter $\boldsymbol{\lambda} \geq \mathbf{0}$. The ridge estimator has lower variance than the least squares estimator but with bias. The ridge method used for regression analysis when there are many predictor variables ($\boldsymbol{p} > \boldsymbol{n}$) and/or the collinearity problem present, but we can say that the ridge method is not variable selection method.

Lasso method proposed by (Tibshirani, 1996) to deals with the problem of $\boldsymbol{p} > \boldsymbol{n}$, Lasso considered as variable selection method. The lasso estimator is defines as follows,

$$\widehat{\boldsymbol{\beta}}_L = \boldsymbol{argmin}||\boldsymbol{y} - \boldsymbol{x}_i'\,\boldsymbol{\beta}||_2^2 + \boldsymbol{\lambda}||\boldsymbol{\beta}||_1 \qquad (\mathbf{1.7})$$

Where $\boldsymbol{\lambda}||\boldsymbol{\beta}||_1 = \boldsymbol{\lambda}\sum_{j=1}^{p}|\boldsymbol{\beta}_j|$ is the penalty function with $\boldsymbol{\lambda} \geq \mathbf{0}$.

The combined penalties method, such as, the elastic net considered two penalty functions $\boldsymbol{L_1} - \mathbf{norm}$ and $\boldsymbol{L_2} - \mathbf{norm}$, 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,

$$\widehat{\boldsymbol{\beta}}_{EN} = \boldsymbol{argmin}||\boldsymbol{y} - \boldsymbol{x}_i'\boldsymbol{\beta}||_2^2\,;$$

$$\boldsymbol{subject\ to\ \alpha}||\boldsymbol{\beta}||_2^2 + (\mathbf{1} - \boldsymbol{\alpha})\,||\boldsymbol{\beta}||_1 \leq \boldsymbol{t}\,, \qquad (\mathbf{1.8})$$

From $(\mathbf{1.8})$ the lasso penalty can be obtained if $\boldsymbol{\alpha} = \mathbf{0}$ and the ridge penalty if $\boldsymbol{\alpha} = \mathbf{1}$. Also, the elastic net method provides variable selection property.

(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.

(Feng, 2010) 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.

(Li et al., 2010) studied the regularization regression method, such as, Lasso, elastic net, and group lasso with Bayesian analysis of the quantile regression.

(Alhamzawi, 2013) proposed some extensions on the Bayesian quantile regression through the prior distribution that allows the full conditional conjugate prior rather than using the normal or Laplace prior distribution. In this paper a novel prior have been used depending on the percentage correlation for variable selection. Also, new Gibbs sample developed to facility the computational for the posterior probabilities.

(Alhamzawi, 2014) proposed the Bayesian Tobit quantile regression model under the gamma prior for the regression coefficients with the elastic net penalty function.

(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.

(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.

(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

$$c(\lambda_1,\lambda_2,\sigma^2)\prod_{j=1}^{p}\int_{1}^{\infty}\sqrt{\frac{t}{t-1}}\ exp\left\{-\frac{\beta_j^2}{2}\left(\frac{\lambda_2}{\sigma^2}\frac{t}{t-1}\right)\right\}t^{\frac{-1}{2}}\ exp\left(-\frac{1}{2\sigma^2}\frac{\lambda_1^2}{4\lambda_2}\right)dt \qquad (1.9)$$

In the thesis new hierarchical model has been proposed for the quantile regression based on the prior distribution (1.9), as well as deriving new Gibbs sample for improving the prediction accuracy of the proposed model.

In this thesis there are new idea and a comparative study which are as follows: To propose new hierarchical model to develop the performance of the elastic net quantile regression model in terms of the Bayesian point of view through employing the Gibbs sample algorithm, To combine the variable selection procedure in the elastic net quantile regression model by employing the prior distribution of (β) as scale mixture of normal distribution and the truncated Gamma distribution that proposed by (Li and Lin, 2010), To proceed a Comparative study between the proposed model with some regularization method.

Chapter Two

Elastic Net Quantile Regression

2.1. Introduction

This chapter provides brief summary on the most popular regularization methods which is called the elastic net method that combine the lasso method ($\boldsymbol{L_1}$**-norm**) and ridge method ($\boldsymbol{L_2}$ **- norm**). This method proposed by (Zou and Hastie, 2005). Also, we briefly provides, the conception of quantile regression.

2.2. Ridge and Lasso Penalties Function

Multicollinearity problem appears when there are correlated predictor variables. Visually, the least squared method affected by the problem of multicollinearity or when the number of predictor variables is greater than the sample size or the observations ($\boldsymbol{P > n}$). Multicollinerity produced non full rank matrix **X** and then the $(\boldsymbol{X'X})^{-1}$ is nonsingular matrix, which is leads to inflated variances of the least squares estimators and produced non unique estimates of the parameters. So to address these problems, the ridge regularization method proposed and the ridge estimator can be define as,

$$\boldsymbol{\hat{\beta}_{ridge} = argmin||y_i - x_i'\beta||_2^2 + \lambda \sum_{j=1}^{p} \beta_j^2} \qquad (\mathbf{2.1})$$

Where $\boldsymbol{\lambda} \geq \mathbf{0}$ is the regularization (shrinkage) parameter, and $\boldsymbol{\lambda \sum_{j=1}^{p} \beta_j^2}$ then the ridge regression parameter estimate is

$$\boldsymbol{\hat{\beta}_{ridge} = (x'x + \lambda I)^{-1} x'y}$$

is called the shrinkage penalty function. The ridge method does not provides variable selection procedure since it does not set any parameters **($\boldsymbol{\beta}$)** equal to zero, but it reduce the variance of estimators and show some bias, (jams et al., 2013).

(Tibshirani, 1996) proposed a new regularization and variable selection regression method which is called lasso (least absolute shrinkage and selection operator). Lasso provides parameter estimate equal to zero. The lasso estimators is defined by

$$\widehat{\boldsymbol{\beta}}_{lasso} = \boldsymbol{argmin}||\boldsymbol{y} - \boldsymbol{x}'\boldsymbol{\beta}|| + \boldsymbol{\lambda}\sum_{j=1}^{p}|\boldsymbol{\beta}_j| \qquad (\mathbf{2.2})$$

There are some drawback on using lasso:

1- If $\boldsymbol{p} > \boldsymbol{n}$, lasso select $\boldsymbol{n}$ variables.

2- Ignoring the grouping information of correlated predictor variables, and select one variable of the group.

3- If $\boldsymbol{n} > \boldsymbol{p}$**,** with highly correlated predictor variables, ridge outperforms lasso.

See (Tibshirani, 1996), (zou and Hastie, 2005), (Jiratchayut, 2014) for more details.

2.3. The Elastic Net Method

(Zou and Hastie, 2005) introduced new regularization method that is called elastic net which combine the ridge and lasso penalty function. The elastic net estimator is define as follows

$$\hat{\boldsymbol{\beta}}_{elastic\ net} = \boldsymbol{argmin}\ ||(\boldsymbol{y} - \boldsymbol{x}'\boldsymbol{\beta})||_2^2 + \boldsymbol{\lambda_1}||\boldsymbol{\beta}||_1 + \boldsymbol{\lambda_2}\ ||\boldsymbol{\beta}||^2 \qquad (\mathbf{2.3})$$

Where $\boldsymbol{\lambda_1}$ and $\boldsymbol{\lambda_2} \geq \mathbf{0}$ are the regularization parameters that controls the amount of shrinkage that forced on the regression parameters. The elastic net works well with high correlated predictor variables, as well as when $\boldsymbol{p} \geq \boldsymbol{n}$. The elastic net method performs the shrinkage of the parameters and variable selection.

2.4. Quantile Regression

(Koenker and Bassett, 1978) proposed the quantil regression which provides robust estimators compared to the least squares method, as well it does not required any assumption on the error distribution. The $\boldsymbol{\gamma\_th}$ conditional quantile regression function is defined by

$$\boldsymbol{Q}(\boldsymbol{\gamma}|\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}(\boldsymbol{\gamma}) \quad ; \quad \boldsymbol{\gamma} \in (\mathbf{0}, \mathbf{1}) \qquad (\mathbf{2.4})$$

The quantile regression estimator is defined as follows

$$\hat{\boldsymbol{\beta}}_{(\gamma)} = \boldsymbol{argmin} \sum_{i=1}^{n} \boldsymbol{\rho_\gamma}\ (\boldsymbol{\epsilon})$$

Where $\boldsymbol{\rho_\gamma}(.)$ is the check function defined as follows

$$\boldsymbol{\rho_\gamma}(\boldsymbol{\epsilon}) = \boldsymbol{I}(\boldsymbol{\epsilon} < \mathbf{0})$$

Quantile regression can be seen as an extension to the conditional mean regression. We use the ordinary least squared (**OLS**) method to find the conditional mean of the response variable, but in quantile regression we can compute the conditional $\boldsymbol{\gamma\_th}$ quantile of the response variable which provides full view of the relationships between the predictor variables and response variable.

Quantile regression can be applied regardless of the normality of the response variable, also quantile regression can be used in the cases where the assumptions of linear regression is not met like the normality of error term. It is well know that datasets without outliers does not effects the work of quantile regression. The quantile function $\boldsymbol{\varphi_\gamma(y|x)}$ can be written as the following optimization problem.

$$\boldsymbol{\varphi_\gamma(y|x) = argmin \, \frac{1}{n} \, [\sum_{y_i - x_i\beta > 0} \gamma \, |y_i - x_i'\beta| + \sum_{y_i - x_i\beta < 0} (1-\gamma) \, |y_i - x_i'\beta| \,]}$$

$$\boldsymbol{= argmin \, \in [\rho(y_i - x_i'\beta)]}$$

$$\boldsymbol{= argmin \sum_{y_i - x_i\beta > 0} \rho_\gamma \left(y_i - x_i'\beta_i(\gamma)\right).}$$

Where $\boldsymbol{0 < \gamma < 1}$ is the $\boldsymbol{\gamma\_th}$ quantile of the response variable and $\boldsymbol{\rho_\gamma(y_i - x_i\beta_i(\gamma))}$ is the loss function.

Chapter Three

Bayesian Elastic Net Quantile Regression

3.1. Introduction

Based on the Bayesian interpretation of lasso method that proposed by (Tibshirani, 1996) considered the lasso estimator as the mode of the posterior distribution of the parameter $\boldsymbol{\beta}$ where the prior of $\boldsymbol{\beta}$ is the double exponential distribution, we studied the elastic net regression model that proposed by (Zou and Hastie, 2005). The elastic net estimator $\widehat{\boldsymbol{\beta}}$ in $(\mathbf{1.8})$ can be rewritten as the following penalized regression,

$$\widehat{\boldsymbol{\beta}}_{EN} = \boldsymbol{argmin}\|\boldsymbol{y} - \boldsymbol{x}_i'\boldsymbol{\beta}\|^2 + \boldsymbol{\lambda}_1 \sum_{j=1}^{p} |\,\boldsymbol{\beta}_j\,| + \boldsymbol{\lambda}_2 \sum_{j=1}^{p} \boldsymbol{\beta}_j^2 \;, \qquad (\mathbf{3.1})$$

$||\boldsymbol{\beta}||_1 = \sum_{j=1}^{p} |\,\boldsymbol{\beta}_j\,|$, $\|\boldsymbol{\beta}\|^2 = \sum_{j=1}^{p} \boldsymbol{\beta}_j^2$ $\boldsymbol{\lambda}_1 \geq \mathbf{0}$ and $\boldsymbol{\lambda}_2 \geq \mathbf{0}$ are the shrinkage parameters.

From $(\mathbf{3.1})$ the ridge penalty can be obtained if $\boldsymbol{\lambda}_1 = \mathbf{0}$ and the lasso penalty if $\boldsymbol{\lambda}_2 = \mathbf{0}$. Also, the elastic net method provides variable selection property. In $(\mathbf{3.1})$ $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)'$ is the centered response variable, and $\mathbf{x}_i = (\mathbf{x}_1, \dots, \mathbf{x}_p)$ are the standardized predictor variables.

For the elastic net estimator $(\mathbf{3.1})$ new scale mixture has been derived by (Li and Lin, 2010) as the prior distribution of $\boldsymbol{\beta}$ which is mathematically very tractable and easy to make Bayesian inference by using Gibbs simpler, as well as formula $(\mathbf{3.1})$ provides selects the shrinkage parameters simultaneously.

(Zou and Hastie, 2005) stated that lasso regularization method cannot choose predictors more than the sample size, also lasso cannot deal with grouping nature of predictors in the data and then select one predictor from each group and drop down the other predictor, and as well, lasso estimates are unsatisfactorily when the predictors are highly correlated.

(Park and Casella, 2008) proposed a new scale mixture of prior distribution of $\boldsymbol{\beta}$ as normal mixing with exponential distribution in Bayesian lasso analysis.

(Mallick and Yi, 2014) proposed Bayesian lasso inference under new scale mixture of the prior $\boldsymbol{\beta}$ as uniforms mixing particular gamma $(\mathbf{2}, \boldsymbol{\lambda})$.

(Flaih et al., 2020) proposed new Bayesian lasso under scale mixture of normals mixing Rayleigh as representation of the prior distribution of parameter $\boldsymbol{\beta}$.

Based on the hierarchical model proposed by (Li and Lin, 2010) for the linear regression model $(\mathbf{1.1})$, we developed new hierarchical model for the quantile regression $(\mathbf{1.3})$ with employing the prior in $(\mathbf{1.9})$.

(Alhamzawi, 2014) presented the Bayesian inference for the elastic net Tobit quantile regression with new hierarchical model where the posterior distribution of $\boldsymbol{\beta}$ is

$$\boldsymbol{f}(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}) \propto \boldsymbol{exp}\left\{-\sum_{i=1}^{n} \boldsymbol{\rho_\gamma}(\boldsymbol{y_i} - \boldsymbol{max}\,\{\boldsymbol{y}^*, \boldsymbol{x}_i'\boldsymbol{\beta}\}) - \boldsymbol{\lambda_1}\|\boldsymbol{\beta}\|_1 - \boldsymbol{\lambda_2}\|\boldsymbol{\beta}\|_2^2\right\}. \quad (\mathbf{3.2})$$

Where $\boldsymbol{y}$ following asymmetric Laplace distribution (**ALD**) which is scale mixture of normal mixing with exponential density (Hendricks and Koenker, 1992)

$$\exp\left\{-\sum_{i=1}^{n}\frac{|\epsilon_i|+(2\gamma-1)\epsilon_i}{2}\right\}=\prod_{i=1}^{\infty}\int_{0}^{\infty}\frac{1}{\sqrt{4\pi w_i}}\exp\left\{-\frac{(\epsilon_i-\delta w_i)^2}{4w_i}-\varsigma w_i\right\}dw_i \quad (3.3)$$

Where $\delta = (1-2\gamma)$ *and* $\varsigma = \gamma\,(1-\gamma)$.

The prior distribution of β was a Laplace distribution which written as (Chen et al., 2011)

$$\pi(\beta|\lambda_1,\lambda_2,\upsilon) \propto \prod_{i=1}^{p}\int_{0}^{\infty}\upsilon\ N\left(0,\frac{1}{a_j}+\lambda_2\right)^{-1}\ gamma\left(1,\frac{\lambda_1^2}{2}\right)da_j. \qquad (3.4)$$

3.2. The Prior Distributions for the Bayesian Elastic Net Quantile Regression

(Li and Lin, 2010) propose the following hierarchical model for Bayesian elastic net based on the classical linear regression model,

$$y|\beta,\sigma^2 \sim N\,(x_i'\beta,\sigma^2 I_n)\,,$$

$$\beta|\sigma^2 \sim \exp\left\{\left(-\frac{\lambda_1||\beta||_1}{2\sigma^2}-\frac{\lambda_2||\beta||_2^2}{2\sigma^2}\right)\right\},$$

$$\sigma^2 \sim (\sigma^2)^{-1},$$

Here the marginal posterior density of β is

$$f(\beta|y)=\int_0^{\infty}\frac{C(\lambda_1,\lambda_2,\sigma^2)}{(2\pi\sigma^2)^{\frac{n}{2}}}\ \exp\left\{-\frac{RSS(B)+\lambda_1||\beta||_1+\lambda_2||\beta||_2^2}{2\sigma^2}\right\}\pi\sigma^2\,d\sigma^2 \qquad (3.5)$$

Where $C\,(\lambda_1,\lambda_2,\sigma^2)$ is the normalizing constant.

We propose employing the hierarchical model $(\mathbf{3.5})$ to develop new hierarchical model for the elastic net quantile regression with the prior distribution $(\mathbf{1.9})$ of $\boldsymbol{\beta}$, $\boldsymbol{\pi}$ $(\boldsymbol{\beta}|\boldsymbol{\sigma}^2)$ that proposed by (Li and Lin, 2010),

The formula of prior $(\mathbf{1.9})$ represent new scale mixture of normal $\boldsymbol{N}$ $(\boldsymbol{\beta_j}$; $\mathbf{0}, \boldsymbol{\sigma}^2(\boldsymbol{t}-\mathbf{1})|\boldsymbol{\lambda}_2\boldsymbol{t}$) mixing the variance $\boldsymbol{\sigma}^2(\boldsymbol{t}-\mathbf{1})|\boldsymbol{\lambda}_2\boldsymbol{t}$ with truncated gamma ($\frac{1}{2}$, $\mathbf{8}\boldsymbol{\lambda}_2\boldsymbol{\sigma}^2|\boldsymbol{\lambda}_1^2$) , where $\boldsymbol{t} \in (\mathbf{1}, \infty)$.

3.3. Bayesian Model Hierarchy

Based on the quantile regression model $(\mathbf{1.3})$ and the prior density $(\mathbf{1.9})$, we proposed the following Bayesian elastic net quantile model hierarchy representation

$$
\left.\begin{array}{c}
Q_i(\gamma|X_i) = y_i = X_i'\beta_\gamma, \qquad i = 1, \dots, n. \\
y_i|\beta, w_i \sim N\left(y_i; x_i'\beta_\gamma + \delta w_i, 2w_i\right), \\
w_i \sim Exp\left(w_i; \frac{1}{\gamma}(1-\gamma)\right), \\
\beta_j|\tau, \sigma^2 \sim \prod_{j=1}^{p} N\,(0, (\frac{\lambda_2}{\sigma^2}\ \frac{\tau_j}{\tau_j - 1})^{-1}) \\
\tau|\sigma^2 \sim \prod_{j=1}^{p} Truncated\ Gamma\left(\frac{1}{2}, \frac{8\lambda_2\sigma^2}{\lambda_1^2}\right), \tau \in (1, \infty) \\
\sigma^2 \sim (\sigma^2)^{-1}
\end{array}\right\} \qquad (3.6)
$$

3.4. 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.

$$y_i|w_i, \beta \sim N(x_i'\beta_\gamma + \delta w_i, 2w_i)$$

Where $i = 1, \dots, n$ the posterior distribution of β is as follows and directly by following (Kozumi and Kobayashi, 2011) and (Alhamzawi, 2014), I supposed that

$$\beta_\gamma|y, w \sim N(\hat{\beta}_\gamma, \hat{C}_\gamma) \qquad (3.7)$$

where $C_\gamma^{-1} = \sum_{i=1}^{n} \frac{x_i x_i'}{\sigma^2 w_i} + [\, Var(\beta_{prior})\,]^{-1}$

and $\hat{\beta}_\gamma = \hat{C}_\gamma \left[\sum_{i=1}^{n} \frac{x_i(y_i - \delta w_i)}{\sigma^2 w_i} + \, Var(\beta_{prior}) * mean(\beta_{prior}) \right]$

From the hierarchal mode (3.6) the prior distribution of $\beta_j \sim N(\beta_j; O, (\frac{\lambda_2}{\sigma^2} \frac{\tau_j}{\tau_j - 1})^{-1})$ then I have the following multivariate normal posterior distribution for β with mean $(\hat{\beta}_\gamma)$ and variance $\hat{C}_\gamma$,

$$C_\gamma^{-1} = \sum_{i=1}^{n} \frac{x_i x_i'}{\sigma^2 w_i} + [\, [\frac{\lambda_2}{\sigma^2}\left(\frac{\tau}{\tau - 1}\right)]^{-1}\,]^{-1}$$

$$C_\gamma^{-1} = \sum_{i=1}^{n} \frac{x_i x_i'}{\sigma^2 w_i} + \frac{\lambda_2 \tau}{\sigma^2(\tau - 1)}$$

$$C_\gamma^{-1} = [\sum_{i=1}^{n} \frac{x_i x_i'}{\sigma^2 w_i}\,]^{-1} + \frac{\sigma^2(\tau - 1)}{\lambda_2 \tau}$$

This the variance of $\boldsymbol{\beta_\gamma}$. And the mean $\boldsymbol{\beta_\gamma}$ is defined as follows,

$$\widehat{\boldsymbol{\beta}}_\gamma = \widehat{\boldsymbol{C}}_\gamma \left[\sum_{i=1}^{n} \frac{x_i(y_i - \delta w_i)}{\sigma^2 w_i} + Var(\beta_{prior}) * mean(\beta_{prior}) \right]$$

from $(\mathbf{3.6})$, we can see that the mean of $\boldsymbol{\beta_{prior}}$ equal to zero, then the $\widehat{\boldsymbol{\beta}}_\gamma$ is

$$\widehat{\boldsymbol{\beta}}_\gamma = \left[\left(\sum_{i=1}^{n} \frac{x_i x_i'}{\sigma^2 w_i} \right)^{-1} + \frac{\lambda_2 \tau}{\sigma^2(\tau - 1)} \right] \left[\sum_{i=1}^{n} \frac{x_i(y_i - \delta w_i)}{\sigma^2 w_i} + zero \right]$$

$$\widehat{\boldsymbol{\beta}}_\gamma = \left[\left(\sum_{i=1}^{n} \frac{x_i x_i'}{\sigma^2 w_i} \right)^{-1} + \frac{\lambda_2 \tau}{\sigma^2(\tau - 1)} \right] \left[\sum_{i=1}^{n} \frac{x_i(y_i - \delta w_i)}{\sigma^2 w_i} \right]$$

the $\boldsymbol{\beta_\gamma}$ distribution is the multivariate normal with mean $\widehat{\boldsymbol{\beta}}_\gamma$ and variance $\widehat{\boldsymbol{C}}_\gamma$;

$$\boldsymbol{\beta_\gamma | y, w \sim multiariate\ Normal\left[\widehat{\beta}_\gamma, \widehat{C}_\gamma\right]} \qquad (\mathbf{3.8})$$

The second variable is $\boldsymbol{\sigma^2}$, where the terms that involves $\boldsymbol{\sigma^2}$ are

$$\pi(\sigma^2 | y, \beta, \tau) \propto \pi(y|\beta, \sigma^2, \tau)\pi(\beta|\sigma^2)\pi(\sigma^2)d\sigma^2$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+p+1} \left\{\Gamma_u\left(\frac{1}{2}, \frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\right\}^{-p} \times exp\left[-\frac{1}{2\sigma^2}\left\{ \lambda_2 \sum_{j=1}^{p} \frac{\tau_j}{\tau_j - 1}\beta_j^2 + \frac{\lambda^2}{4\lambda_2}\sum_{j=1}^{p}\tau_j \right\} \right] \qquad (\mathbf{3.9})$$

Where $\boldsymbol{\Gamma u\,(\alpha, x) = \int_x^\infty t^{a-1}\ e^{-t}\,dt}$ is the upper incomplete gamma function, see (Armido and Alfred, 1986) for more details.

The third variable $(\boldsymbol{\tau} - \mathbf{1}_{\boldsymbol{P}})$, where the full conditional distribution is

$$(\boldsymbol{\tau} - \mathbf{1}_{\boldsymbol{P}})|\boldsymbol{y}, \boldsymbol{\sigma}^2, \boldsymbol{\beta} \sim \prod_{j=1}^{p} GIG\left(\lambda = \frac{1}{2}, \varphi = \frac{\lambda_1}{4\lambda_2\sigma^2}, \chi = \frac{\lambda_2\beta_j^2}{\sigma^2}\right), \qquad (3.10)$$

Where **GIG (.)** is the generalized inverse Gaussian disruption, see (Jorgensen, 1982) for more details, **i.e.** we can say that $\boldsymbol{x} \sim \boldsymbol{GIG}\ (\boldsymbol{\lambda}, \boldsymbol{\varphi}, \boldsymbol{\chi})$ if its pdf as follows,

$$\mathrm{f}\ (x|\lambda, \varphi, \chi) = \frac{(\varphi|\chi)^{\lambda/2}}{2k_\lambda\,(\sqrt{\varphi\chi})}\, x^{\lambda-1}\, exp\left\{-\frac{1}{2}\,(\chi x^{-1} + \varphi\,\chi)\right\}, \qquad (3.11)$$

Where $\boldsymbol{x} > \mathbf{0}, \boldsymbol{k}_{\boldsymbol{\lambda}}\,(.)$ is the Bassel function of the third Kind with order $\boldsymbol{\lambda}$.

So, we can easily say that

$$(\boldsymbol{\tau}_j - \mathbf{1})^{-1}|\boldsymbol{y}, \boldsymbol{\sigma}^2, \boldsymbol{\beta} \sim \boldsymbol{IG}\left(\boldsymbol{\mu} = \frac{\sqrt{\boldsymbol{\lambda}}}{(2\lambda_2\,|\beta_j|)}, \lambda = \frac{\lambda_1}{4\lambda_2\sigma^2}\right)$$

With the following pdf,

$$\mathrm{f}\ (x|\mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}}\ exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}.$$

See (Chhikara and Folks, 1988) for more details.

3.5. Choosing the Elastic Net Shrinkage Parameters $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$.

(Park and Casella, 2008) suggested the empirical Bayes estimates for the shrinkage parameters $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ by using the marginal Maximum likelihood of the data and use the Monte Carlo Expectation- maximization (**MCEM**) algorithm. Following (Li and Lin, 2010), we treated $\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\sigma}^2$ as missing data and $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ as fixed parameters, the likelihood is

$$\lambda_1^p(\sigma^2)^{-\frac{n}{2}-p-1}\{\Gamma_u\left(\frac{1}{2},\frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\}^{-p}\prod_{j=1}^{p}(\tau_j-1)^{-\frac{1}{2}}$$

$$exp\left[-\frac{1}{2\sigma^2}\left\{RSS+\lambda_2\sum_{j=1}^{p}\frac{\tau_j}{\tau_j-1}\beta_j^2+\frac{\lambda_1^2}{4\lambda_2}\sum_{j=1}^{p}\tau_j\right\}\right] \qquad (3.12)$$

and then we can take the log for the function (3.12) and maximization problem is solving as follow see (Li and Lin, 2010) for more details

$$\frac{\partial R}{\partial\lambda_1}=\frac{p}{\lambda_1}+\frac{p\lambda_1}{4\lambda_2}E\left[\left\{\Gamma_U\left(\frac{1}{2},\frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\right\}^{-1}\varphi\left(\frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\frac{1}{\sigma^2}\middle|\lambda^{(k-1)},Y\right]-\frac{\lambda_1}{4\lambda_2}\sum_{j=1}^{p}E\left[\frac{\tau_j}{\sigma^2}\middle|\lambda^{(k-1)},Y\right],$$

$$\frac{\partial R}{\partial\lambda_2}=-\frac{p\lambda_1^2}{8\lambda_2^2}E\left[\left\{\Gamma_U\left(\frac{1}{2},\frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\right\}^{-1}\varphi\left(\frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\frac{1}{\sigma^2}\middle|\lambda^{(k-1)},Y\right]-\frac{1}{2}\sum_{j=1}^{p}E\left[\frac{\tau_j}{\tau_j-1}\ \frac{\beta_j^2}{\sigma^2}\middle|\lambda^{(k-1)},Y\right]$$

$$+\frac{\lambda_1^2}{8\lambda_2^2}\sum_{j=1}^{p}E\left[\frac{\tau_j}{\sigma^2}\middle|\lambda^{(k-1)},Y\right], \qquad (3.13)$$

Where $\varphi(t)=t^{-\frac{1}{2}}e^{-t}$, And **R** is log function of (3.13).

3.6. 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 (3.13). 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** from the following full conditional distribution

$$y_i|w_i, \beta \sim N\,(x_i'\beta_\gamma + \delta w_i, 2w_i)$$

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

2- Updating $\beta|y, \sigma^2, \tau$ from the full conditional posterior density which following the multivariate normal distribution (3.8) with mean and variance as follows,

$$\widehat{\beta}_\gamma = \left[\sum_{i=1}^{n} \frac{x_i x_i'}{\sigma^2 w_i} + \frac{\lambda_2 \tau}{\sigma^2(\tau - 1)}\right]\left[\sum_{i=1}^{n} \frac{x_i(y_i - \delta w_i)}{\sigma^2 w_i}\right]$$

$$and$$

$$C_\gamma^{-1} = \sum_{i=1}^{n} \frac{x_i x_i'}{\sigma^2 w_i} + \frac{\lambda_2 \tau}{\sigma^2(\tau - 1)} \qquad (3.14)$$

3- Updating w_i^{-1} ; $i = 1, 2, \dots, n$ from the full conditional posterior distribution of $\mathbf{w_i^{-1}}$ which is follows Inverse Gaussian (μ', λ') see (Alhamzawi, 2014), where

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

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

$$\mathrm{f}\,(x|\lambda', \mu') = \sqrt{\frac{\lambda'}{2\pi x^3}}\, exp\left\{\frac{-\lambda'(x - \mu')^2}{2(\mu')^2 x}\right\}; x > 0 \qquad (3.15)$$

4- Updating $(\tau_j - 1)^{-1}|y, \sigma^2, \beta$ from the full conditional inverse Gaussian distribution (3.14), with

$$\mu = \frac{\sqrt{\lambda_1}}{(2\lambda_2|\beta_j|)} \text{ and } \lambda = \frac{\lambda_1}{4\lambda_2\sigma^2}\ ; \qquad j = 1, 2, \dots, p \qquad (3.16)$$

(Li and Lin, 2010) stated that sampling from (3.14) the inverse Gaussian distribution is much faster than the **Hyperbolic Dist. rgig ( ) function** (Scott, 2008).

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

$$f(\sigma^2) \le \frac{\Gamma_a \, \Gamma_{\frac{1}{2}}^{-p}}{b^a} \, h(\sigma^2); \qquad (3.17)$$

Where $a = \frac{n}{2} + p$,

$$b = \frac{1}{2}\left[\|y - x'\beta_\gamma\| + \lambda_2 \sum_{j=1}^{p} \frac{\tau_j}{\tau_j - 1}\, \beta_j^2 + \frac{\lambda_1^2}{4\lambda_2} \sum_{j=1}^{p} \tau_j\right].$$

and h **(.)** is the inverse gamma (a, b).

6- Updating λ_1 and λ_2, we can find the estimates of λ_1 and λ_2 that maximizing the log function of (3.13) after m-steps for implement of the algorithm.

Chapter Four

Simulation and Real Data Analysis

4.1. Introduction

In this section, simulation study will be conducted to show the behavior of our proposed model, Bayesian elastic net quantile regression (**Benqr**) using **R** and compared with different exists models; the classic quantile regression model (**cqr**) by implementing the (rq) **R** package quantreg, and the lasso quantile regression model (**lqr**) by implementing the **R** package. Our comparison is based on the parameters estimates of the different models under different quantiles ($\boldsymbol{\gamma = 0.25}$, $\boldsymbol{\gamma = 0.50}$, $\boldsymbol{\gamma = 0.75}$ and $\boldsymbol{\gamma = 0.99}$). Also, we used the median mean absolute deviation (**MMAD**) criterion,

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

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 $\boldsymbol{x_1, \dots, x_9}$ predictor variables from $\boldsymbol{N_{n=9}(0}, \Sigma)$, where the matrix $\Sigma_{\boldsymbol{ij}} = \boldsymbol{\rho}^{|\boldsymbol{i-j}|}$ with three distributions of the Independent and identically distributed random variables (**i.i.d**) errors. For each simulation study, we run **400** simulations.

4.1.1. Simulation Example One

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

Table 1. Parameter estimates of example 1 with $\boldsymbol{\epsilon_i}\sim\boldsymbol{N}(\mathbf{0},\mathbf{1})$.

True parameter	Quantile level	0	3	0	0	0	0	0	0	0
Benqr	**0.25**	**-0.53433**	**3.05755**	**-0.04878**	**0.15044**	**0.05801**	**0.02518**	**0.05821**	**0.09411**	**-0.11741**
lqr	0.25	-0.48510	3.11706	-0.10233	0.22113	0.03017	0.06392	-0.00889	0.05554	-0.13766
cqr	0.25	-0.31952	3.26606	-0.22838	0.30980	0.02538	0.36293	-0.07736	-0.12678	0.01154
Benqr	**0.50**	**0.00348**	**2.61704**	**0.02319**	**0.05258**	**0.19002**	**0.13365**	**-0.03415**	**-0.13002**	**-0.12188**
lqr	0.50	0.01652	2.71332	-0.00693	0.05948	0.16909	0.13757	-0.01339	-0.14076	-0.12048
cqr	0.50	-0.06342	2.80605	-0.42396	0.40156	0.04339	0.65369	-0.16830	-0.24411	0.17263
Benqr	**0.75**	**0.59039**	**2.64516**	**0.33294**	**0.08248**	**-0.0428**	**-0.06223**	**-0.07704**	**0.00512**	**-0.19677**
lqr	0.75	0.54035	2.88732	0.25855	0.01500	-0.0327	-0.18471	-0.08628	0.10899	-0.18795
cqr	0.75	0.60376	3.15254	-0.15802	-0.3066	0.31255	0.15051	0.08138	0.64905	-0.18132
Benqr	**0.99**	**2.16136**	**2.82956**	**0.06402**	**-0.02756**	**-0.0494**	**-0.05186**	**-0.00629**	**-0.09735**	**0.01993**
lqr	0.99	1.47320	3.27608	0.06881	-0.04868	0.02646	0.09077	-0.17528	-0.22168	-0.01531
cqr	0.99	0.80679	3.50108	-0.05175	-0.36564	0.16453	0.31269	-0.13776	-0.18034	-0.02319

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

True parameter	Quantile level	0	3	0	0	0	0	0	0	0
Benqr	**0.25**	**-0.94167**	**3.08580**	**0.18950**	**0.12903**	**-0.0174**	**0.23762**	**-0.09033**	**0.24558**	**-0.11196**
lqr	0.25	-0.91259	2.87556	0.11570	0.15341	0.05266	0.18572	-0.13328	0.53398	-0.28134
cqr	0.25	-0.91319	2.52562	0.08169	0.28278	0.16490	-0.35038	-1.25653	1.04882	-0.64224
Benqr	**0.50**	**-0.08201**	**2.35569**	**-0.03954**	**0.24500**	**0.05604**	**-0.11054**	**0.01792**	**-0.14199**	**0.14266**
lqr	0.50	-0.10540	2.58759	-0.09529	0.26693	0.05386	-0.06312	0.05668	-0.10855	0.13835
cqr	0.50	-0.11045	2.72309	-0.17261	0.15492	0.10169	-0.03924	-0.13531	0.00210	0.01763
Benqr	**0.75**	**0.77623**	**2.68634**	**0.06926**	**0.02502**	**-0.0233**	**-0.17077**	**-0.16754**	**-0.07847**	**0.15773**
lqr	0.75	0.77059	3.00004	0.16754	-0.02458	-0.0834	-0.09377	-0.22678	-0.08948	0.21058
cqr	0.75	0.70621	3.37876	0.28050	-0.04396	-0.2131	0.55459	-0.44564	0.07573	0.30717
Benqr	**0.99**	**1.19079**	**2.80197**	**-0.16346**	**0.10390**	**0.53281**	**0.15548**	**0.30575**	**-0.14864**	**0.32483**
lqr	0.99	1.94265	2.60253	0.01807	-0.05214	0.29374	0.18550	0.02513	-0.12150	0.05225
cqr	0.99	2.78453	1.96509	0.20717	0.11471	0.42179	0.24586	0.23482	-0.08939	-0.07131

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

		Errors distribution		
The methods	**Quantile level**	$\epsilon_i \sim N(O,1)$	ϵ_i~Normal mixture	$\epsilon_i \sim \chi^2_3$
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)

4.1.2. Simulation Example Two

In this simulation, we supposed that the true vector of parameter $\boldsymbol{\beta}$ **= (0, 3, 0, 0, 0, 1, 0, 0, 0)**^{t} with error terms followed $\boldsymbol{\epsilon_i}$**~ N (μ = 0,** $\boldsymbol{\sigma^2}$ **= 1),** $\boldsymbol{\epsilon_i}$ **~ *normal mixture*** $(\mathbf{1},\mathbf{1}) + \boldsymbol{N}(\mathbf{2},\mathbf{2})$. Also, we generated the observation of $\boldsymbol{x_1},\dots,\boldsymbol{x_9}$ predictor variables based on $\boldsymbol{N_{n=9}}(\mathbf{0}\ ,\Sigma)$, where Σ is the **var-cov** matrix defined as $\Sigma_{\boldsymbol{ij}} = \mathbf{0.5}^{|\boldsymbol{i}-\boldsymbol{j}|}$**.** As well as, we simulated $\boldsymbol{y_i}$ as $\boldsymbol{y_i} = \mathbf{3}\boldsymbol{x_{2i}} + \boldsymbol{x_{6i}} + \boldsymbol{e_i}$

Table 4. Parameter estimates of example 2 with $\boldsymbol{\epsilon_i}\sim\boldsymbol{N}(\mathbf{0},\mathbf{1})$**.**

True parameter	Quantile level	0	3	0	0	0	1	0	0	0
Benqr	**0.25**	**-0.02136**	**2.81152**	**0.71724**	**0.11596**	**0.21045**	**1.03062**	**-0.13299**	**-0.02843**	**-0.10299**
lqr	0.25	-0.06120	2.95282	0.77753	0.05321	0.17684	1.12803	-0.15366	-0.02589	-0.06785
cqr	0.25	0.02508	2.81003	0.98391	0.14589	-0.0352	1.29055	-0.22535	-0.07966	-0.15095
Benqr	**0.50**	**0.63827**	**2.63622**	**0.85958**	**0.13296**	**0.07848**	**0.73636**	**0.11969**	**-0.04909**	**0.17233**
lqr	0.50	0.61856	2.71381	0.84442	0.03640	0.07007	0.81281	0.12284	-0.10753	0.10566
cqr	0.50	0.55286	2.47040	0.90523	0.25266	-0.2281	0.72520	0.14699	-0.11374	0.03152
Benqr	**0.75**	**0.63827**	**2.63622**	**0.85958**	**0.13296**	**0.07848**	**0.73636**	**0.11969**	**-0.04909**	**0.17233**
lqr	0.75	0.61856	2.71381	0.84442	0.03640	0.07007	0.81281	0.12284	-0.10753	0.10566
cqr	0.75	0.55286	2.47040	0.90523	0.25266	-0.2281	0.72520	0.14699	-0.11374	0.03152
Benqr	**0.99**	**2.44915**	**2.74335**	**0.71445**	**-0.01726**	**0.04707**	**0.91117**	**0.03247**	**0.12144**	**-0.03079**
lqr	0.99	1.76998	2.73674	0.84050	-0.03548	0.05125	1.00818	-0.02740	0.07772	0.03496
cqr	0.99	0.94421	2.81680	0.83707	0.12081	-0.1931	1.02810	0.07911	-0.38471	0.23358

Table 5. Parameter estimates of example 2 with $\epsilon_i \sim Normal\ mixture$

True parameter	Quantile level	0	3	0	0	0	1	0	0	0
Benqr	**0.25**	**-0.76867**	**2.83206**	**0.37196**	**0.09776**	**0.24434**	**0.68113**	**-0.02235**	**-0.04047**	**0.00646**
lqr	0.25	-0.74785	3.01065	0.31901	0.07608	0.24827	0.84207	-0.05281	-0.01318	-0.45281
cqr	0.25	-0.74747	2.89980	0.66302	0.65816	0.08295	0.84713	-0.07826	0.15698	-0.02873
Benqr	**0.50**	**-0.02207**	**2.69680**	**0.79179**	**0.01176**	**0.00036**	**0.64628**	**0.05262**	**-0.01460**	**0.08242**
lqr	0.50	-0.02128	2.85358	0.82594	-0.03001	0.01753	0.81606	0.07063	-0.02575	0.05487
cqr	0.50	-0.01279	3.21665	1.00531	-0.34249	0.06680	1.09793	0.10051	-0.10712	0.30524
Benqr	**0.75**	**0.64677**	**2.51760**	**0.86592**	**-0.01184**	**0.10844**	**0.90032**	**0.07448**	**-0.09930**	**0.04268**
lqr	0.75	0.51576	2.83907	0.98206	-0.11208	0.12414	1.05086	0.06801	-0.07546	0.05907
cqr	0.75	0.63668	2.68257	1.04202	-0.08128	0.34395	0.84192	0.19372	-0.64362	0.65851
Benqr	**0.99**	**2.58847**	**2.73112**	**0.48717**	**0.21655**	**0.32023**	**0.23671**	**-0.07212**	**0.21496**	**-0.10703**
lqr	0.99	1.76996	2.94541	0.69654	0.31156	0.15903	0.11042	0.02868	0.45162	-0.15418
cqr	0.99	1.05066	2.90792	0.78714	0.29758	1.16333	-0.72234	0.48761	1.81204	-1.01052

Table 6. MMAD and S.D. for simulation example 2

Errors distribution				
The methods	**Quantile level**	$\epsilon_i \sim N(O,1)$	$\epsilon_i \sim$**Normal mixture**	$\epsilon_i \sim \chi_3^2$
Benqr	**0.25**	**0.2067 (0.17256)**	**0.5948 (0.63190)**	**0.3350(0.32750)**
lqr	0.25	0.2751 (0.31484)	0.6127 (0.58770)	0.3834(0.39792)
cqr	0.25	0.4492 (0.48524)	0.6857(0.76076)	0.5937(0.58904)
Benqr	**0.50**	**0.4221 (0.39620)**	**0.6214(0.58966)**	**0.3235(0.30728)**
lqr	0.50	0.4944 (0.41982)	0.4994(0.52796)	0.2689(0.28212)
cqr	0.50	0.5965 (0.58524)	0.7197(0.67194)	0.4106(0.41314)
Benqr	**0.75**	**0.4221(0.39620)**	**0.5985(0.61694)**	**0.3994(0.44440)**
lqr	0.75	0.3944 (0.41982)	0.5576(0.59562)	0.4264(0.42234)
cqr	0.75	0.5965 (0.58524)	0.8245(0.90666)	0.5126(0.54932)
Benqr	**0.99**	**0.5739 (0.59290)**	**0.6885 (0.67262)**	**0.6968(0.69734)**
lqr	0.99	0.7895 (0.86814)	0.8739(0.94150)	1.3263(1.31346)
cqr	0.99	1.4657 (1.49032)	1.7255(1.63402)	1.8361(1.82406)

4.1.3. Simulation Example Three

In this simulation, we supposed that the true vector of parameter $\boldsymbol{\beta}$ **= (0, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85)**^{t} with error terms followed $\boldsymbol{\epsilon_i}$ **~ N (μ = 0,** $\boldsymbol{\sigma^2}$**= 1),** $\boldsymbol{\epsilon_i \sim normal\ mixture(1,1) + N(2,2)}$. Also, we generated the observation of $\boldsymbol{x_1, \dots, x_9}$ predictor variables based on $\boldsymbol{N_{n=9}(0}\ ,\Sigma)$, where Σ is the **var-cov** matrix defined as $\Sigma_{ij} = \mathbf{0.5}^{|i-j|}$ **.** As well as, we simulated $\boldsymbol{y_i}$ as $\boldsymbol{y_i} = \sum_{i=2}^{9} \mathbf{0.85}\boldsymbol{xi} + \boldsymbol{e_i}$

Table 7. Parameter estimates of example 3 with $\boldsymbol{\epsilon_i \sim N(0,1)}$**.**

True parameter	Quantile level	0	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
Benqr	**0.25**	**-0.64979**	**0.64823**	**0.74447**	**0.73520**	**0.97061**	**0.62568**	**0.60114**	**1.02711**	**0.69645**
Lqr	0.25	-0.53170	0.88079	0.80848	0.80029	0.98900	0.92916	0.60317	1.09904	0.73951
Cqr	0.25	-0.46926	1.00515	0.81415	0.84072	1.10119	0.70071	0.99467	0.96591	0.93354
Benqr	**0.50**	**0.00560**	**0.73128**	**0.50646**	**0.76899**	**1.41066**	**0.52072**	**0.60691**	**1.02969**	**0.85337**
Lqr	0.50	0.03795	0.77051	0.58120	0.79847	1.48683	0.59922	0.57299	1.18765	0.82352
Cqr	0.50	0.06532	0.69808	0.38372	0.80232	1.27890	0.31426	0.63536	1.38375	0.61412
Benqr	**0.75**	**0.58129**	**0.40298**	**0.65137**	**0.60307**	**0.38781**	**0.97435**	**0.85312**	**0.67398**	**0.40399**
Lqr	0.75	0.50730	0.46501	0.77824	0.72253	0.51722	1.15583	0.77779	0.63649	0.54176
Cqr	0.75	0.36418	0.80755	0.91819	0.62220	0.73756	1.22816	1.36461	0.67616	0.53008
Benqr	**0.99**	**2.36802**	**0.46516**	**0.83815**	**0.62850**	**0.84328**	**0.85638**	**0.74499**	**0.71654**	**0.42551**
Lqr	0.99	1.56529	0.58486	0.83660	0.87044	0.75907	0.76754	1.03286	0.66912	0.68675
Cqr	0.99	0.82397	0.33002	1.19283	1.16331	0.95764	0.56843	1.34262	0.07024	1.23907

Table 8. Parameter estimates of example 3 with $\epsilon_i \sim Normal\ mixture$

True parameter	Quantile level	0	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
Benqr	**0.25**	**-0.77908**	**0.81223**	**0.44166**	**0.92073**	**0.51251**	**0.77586**	**0.47105**	**0.80101**	**0.50189**
lqr	0.25	-0.64181	1.16662	0.50998	1.07550	0.48753	0.95927	0.52969	0.98686	0.43907
cqr	0.25	-0.76095	1.38259	0.68323	0.99703	0.74657	0.81767	0.22455	1.11988	0.74775
Benqr	**0.50**	**-0.01943**	**0.91355**	**0.97288**	**0.62345**	**0.69155**	**0.81773**	**0.49024**	**0.97807**	**0.76968**
lqr	0.50	0.02013	0.91125	1.07509	0.65413	0.63218	0.88747	0.57730	0.90728	0.84898
cqr	0.50	-0.15101	0.88657	1.43474	0.43765	0.99507	1.06066	0.96357	0.45006	0.80190
Benqr	**0.75**	**0.75847**	**0.57105**	**0.69253**	**0.55011**	**0.79049**	**0.73938**	**0.74893**	**0.85903**	**0.65568**
lqr	0.75	0.70365	0.79020	0.72262	0.63562	0.91916	0.71245	0.78972	0.98888	0.77737
cqr	0.75	0.50857	0.69065	0.85752	0.65947	1.08956	0.89166	0.90471	1.40010	0.66218
Benqr	**0.99**	**0.90636**	**0.73899**	**0.64219**	**0.99836**	**0.79263**	**0.81081**	**0.63894**	**1.08887**	**1.21735**
lqr	0.99	2.55313	0.76596	0.63943	0.78859	0.65507	0.72681	1.12224	0.63894	1.21720
cqr	0.99	1.62379	0.71111	0.59562	0.99347	0.79446	0.81081	1.54224	1.08887	1.21735

Table 9. MMAD and S.D. for simulation example 3

Errors distribution				
The methods	Quantile level	$\epsilon_i \sim N(O,1)$	$\epsilon_i \sim$Normal mixture	$\epsilon_i \sim \chi_3^2$
Benqr	**0.25**	**0.5047 (0.54550)**	**0.7716 (0.77690)**	**0.4398(0.45960)**
lqr	0.25	0.5992 (0.57440)	0.8805(0.80862)	0.4476(0.40184)
cqr	0.25	0.5604 (0.57274)	0.7731(0.80496)	0.5761(0.58494)
Benqr	**0.50**	**0.4128 (0.39252)**	**0.4346(0.44240)**	**0.3486(0.36548)**
lqr	0.50	0.4807 (0.44508)	0.5014(0.44444)	0.3784(0.37400)
cqr	0.50	0.7277 (0.72344)	0.6256(0.67666)	0.3619(0.40254)
Benqr	**0.75**	**0.5612 (0.70150)**	**0.4949(0.58180)**	**0.3439(0.38476)**
lqr	0.75	0.7125 (0.69258)	0.5499(0.60050)	0.4057 (0.43084)
cqr	0.75	0.6467 (0.65348)	0.7074(0.76426)	0.5191(0.53866)
Benqr	**0.99**	**0.5040 (0.57410)**	**0.6194(0.61758)**	**0.5916(0.62902)**
lqr	0.99	0.9063 (0.91910)	0.7998(0.80136)	1.3660(1.33054)
cqr	0.99	1.5618 (1.59886)	1.5307(1.52642)	1.9134(1.87878)

From table **1**- table **9** for the previous simulation examples (**1 , 2 , and 3**), 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.

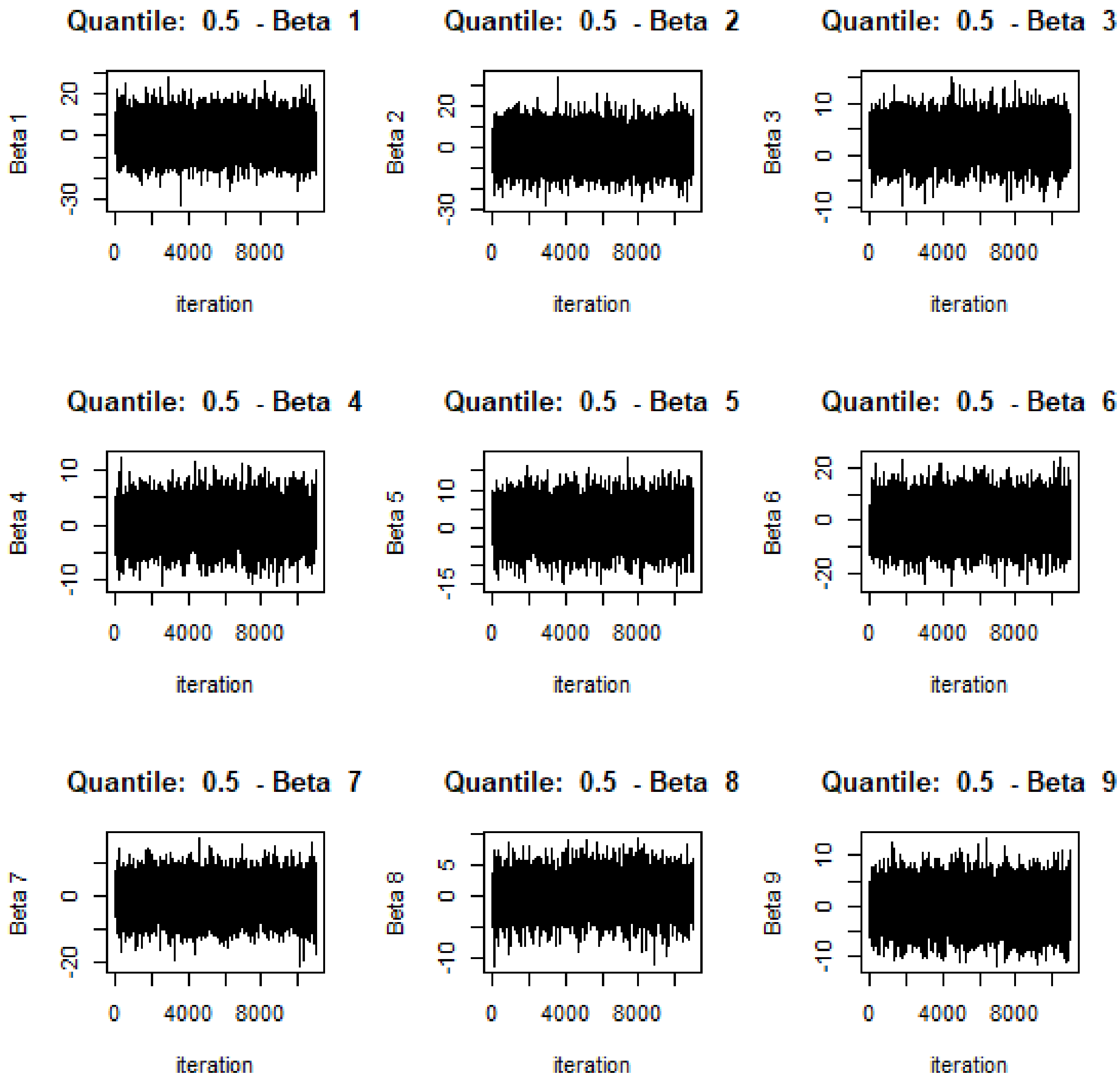

Figure 1. Trace plots of Benqr with (0.5) quantile

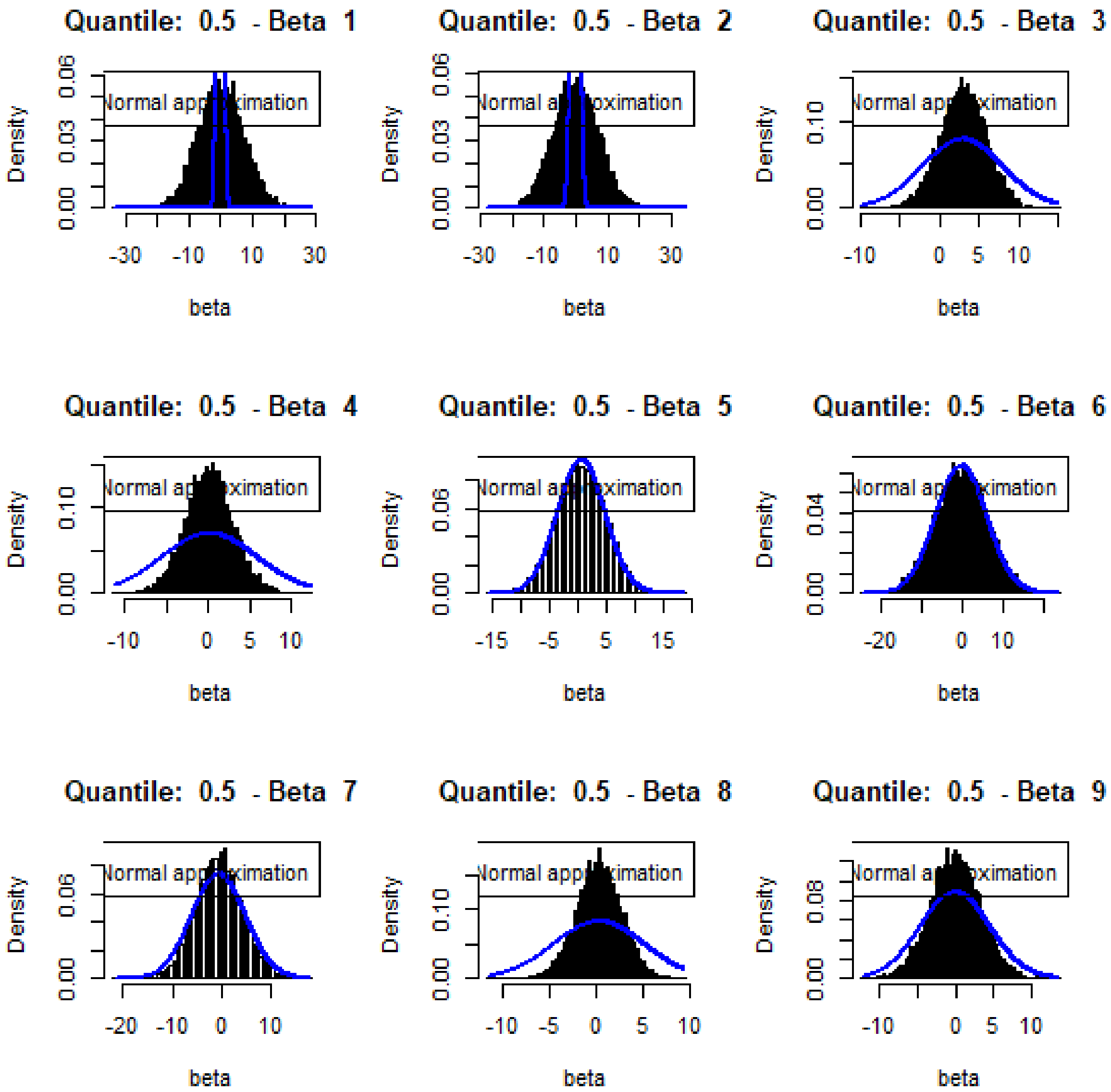

Figure 2. Histograms of Benqr parameter estimates

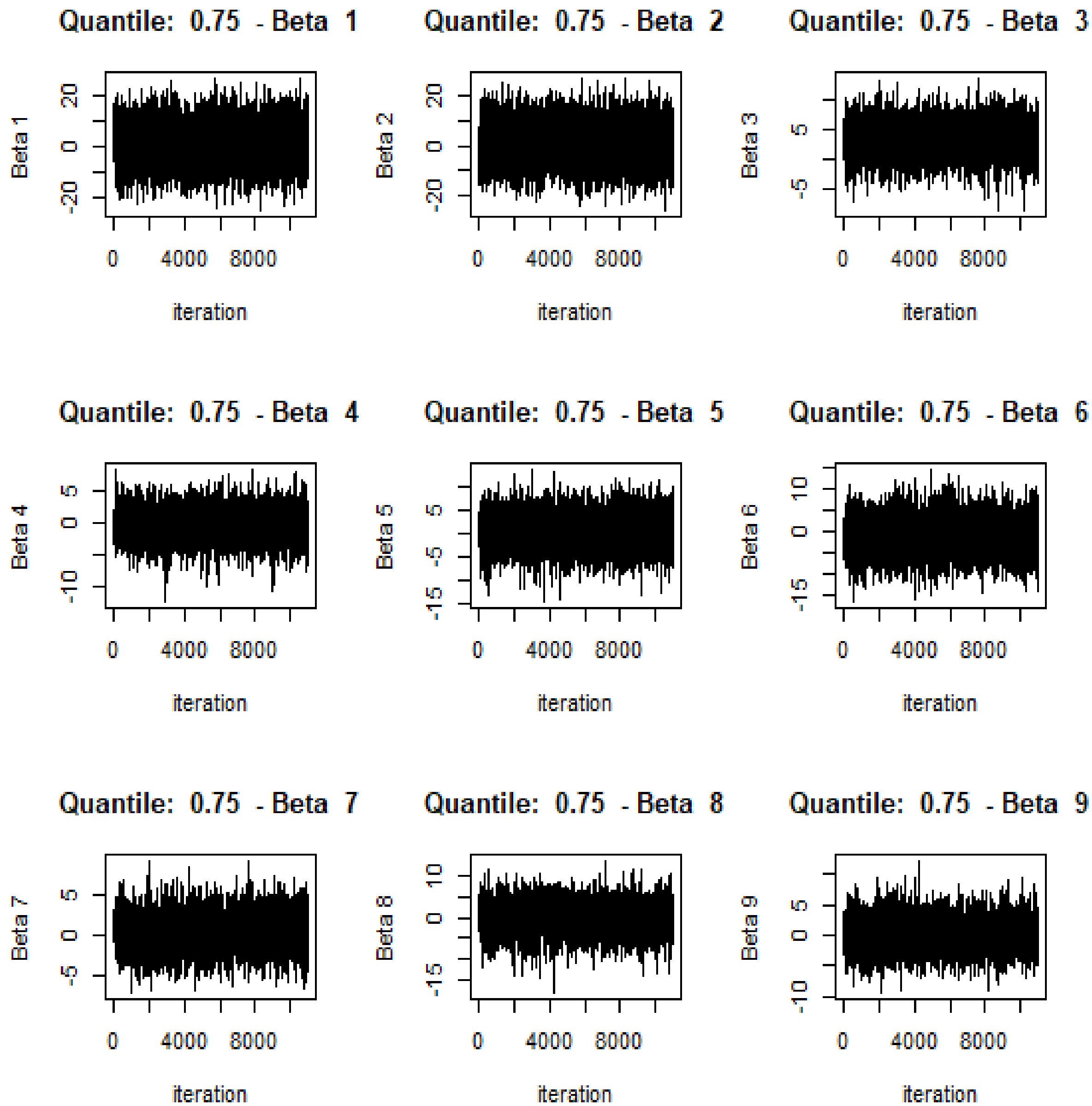

Figure 3. Trace plots of Benqr with (0.75) quantile

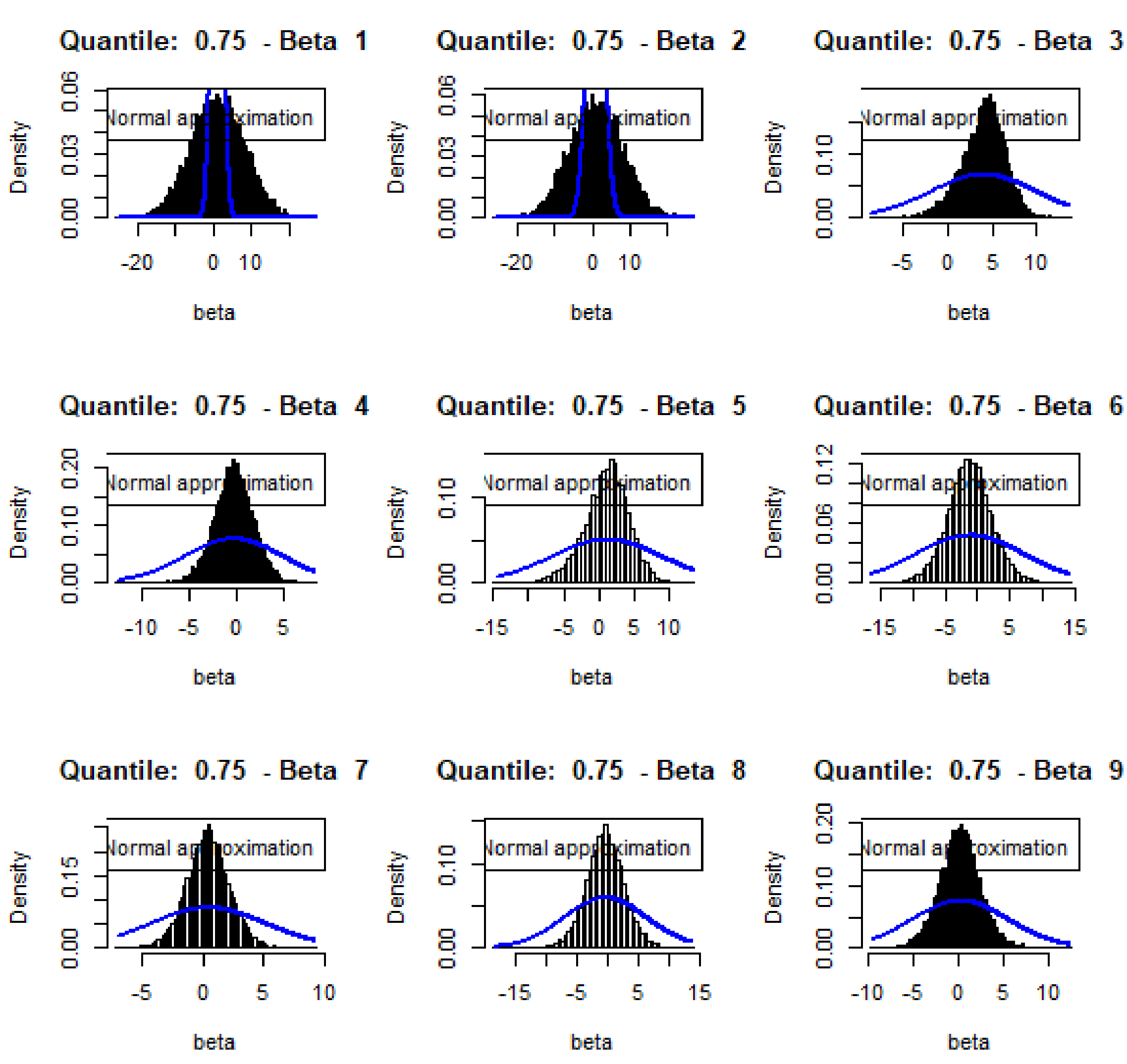

Figure 4. Histograms of Benqr parameter estimates

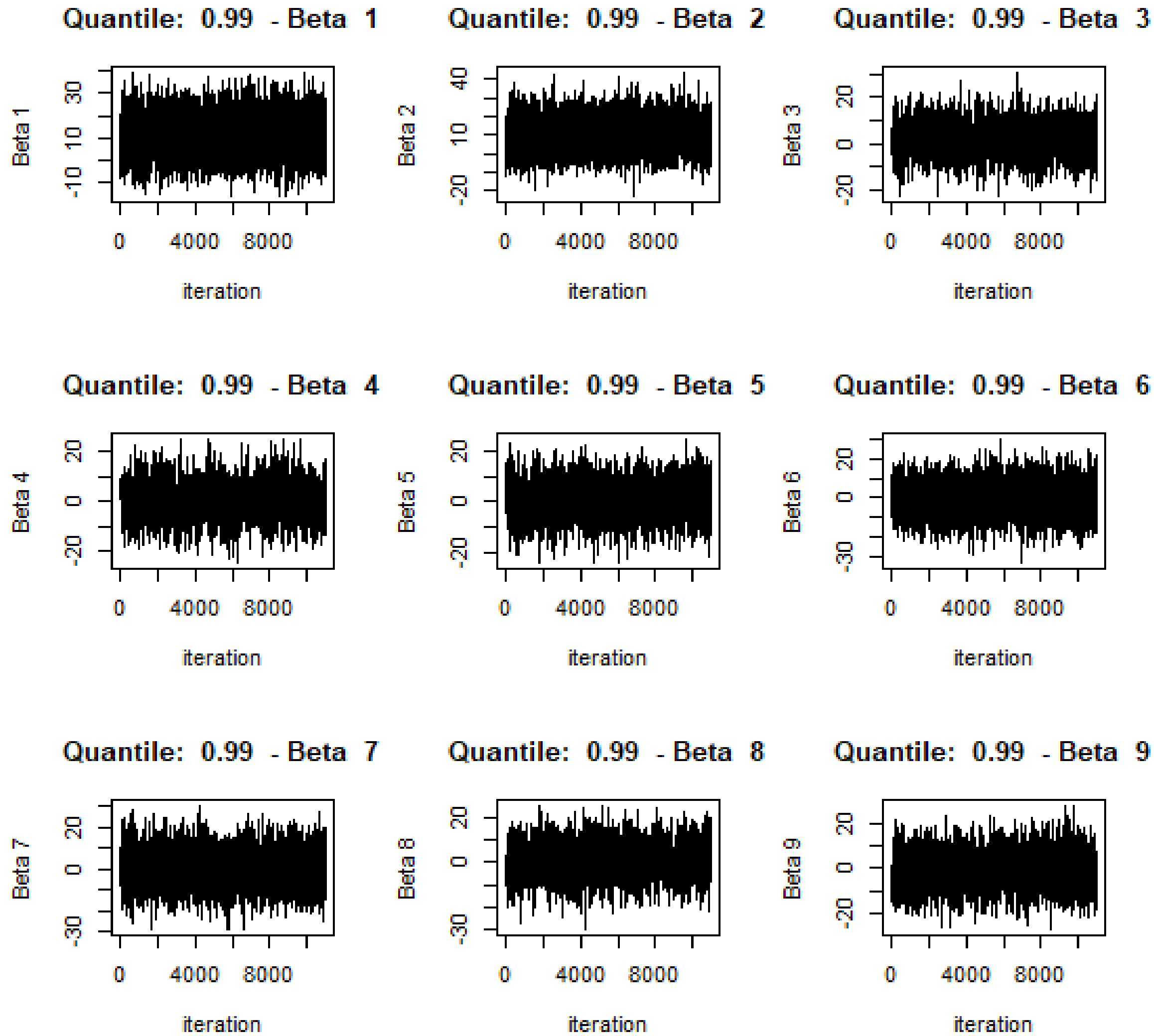

Figure 5. Trace plots of Benqr with (0.99) quantile

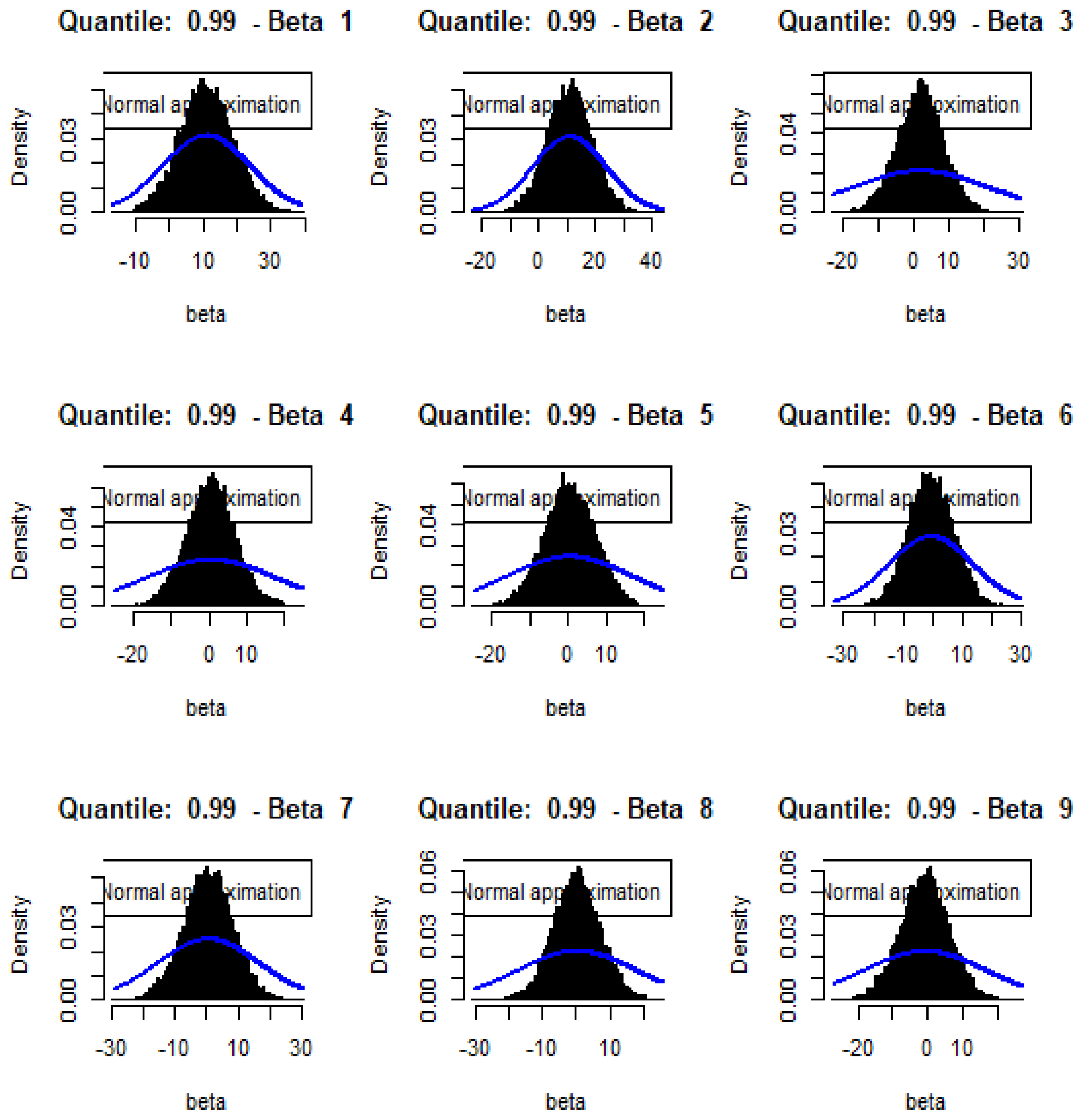

Figure 6. Histograms of Benqr parameter estimates

Figures **1 - 6** displayed the histograms graphs that 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.

4.2 Real Data Analysis

By visting the hospital of children in mesan fequently and from the records of the data department in the hospital I gather the data about the phenomena of fatness in new born children Which number is (**100**) same. The mean squared error (**MSE**) criterion has been employed to measure the performance of the proposed Bayesian elastic net quantile regression model comparing with the classical quantile regression model and the lasso quantile regression model,

$$\mathbb{E}\,[\,\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\|_2^2\,] = \sum_{j=1}^{k}[\,(\mathbf{Bias}(\widehat{\boldsymbol{\beta}}_j))^2 + \mathbf{var}(\widehat{\boldsymbol{\beta}}_j)].$$

Childhood obesity is a very serious medical condition that affects children and adolescents. Obese children are those who are overweight for their age and height. Childhood obesity is especially worrisome because extra pounds often put children on the path to health problems that were previously considered adult problems - diabetes, high blood pressure and high cholesterol. Many obese children become obese adults, especially if one or both parents are obese. Childhood obesity can also lead to low self-esteem and frustration. One of the best strategies for reducing childhood obesity is to improve the eating and exercise habits of your entire family. Treating and preventing childhood obesity helps protect your child's health, now and in the future.

The World Health Organization recently recorded a remarkable and worrying increase in the weight of children under the age of five, as these increased numbers are harbinger of danger to the public health of children now and in the future. In our current study, we tried to focus on this important and dangerous phenomenon at the same time, and dedicate our competence to contribute to solve this problem that threatens human societies in all countries of the world, as our current study includes an approved variable **(y)** representing the weight of children under the age of five (where it is considered this variable is a quantitative variable), and a group of independent variables with direct and indirect effects on obesity in children under five years old. Below is a brief description of the independent variables that were used in our current study:

1- Child age **(X_1),** there is a strong correlation between the child's age and weight. If the child's age increases and the daily behaviors of the child and his family are not good, this contributes to weight gain.

2- Child's gender **(X_2)** recent medical studies have shown that gender has an effect on increasing the child's weight due to factors related to genetics.

3- Mother employed **(X_3)** the type of work of the mother may be included as a catalyst for increasing, decreasing or moderating the child's weight.

4- The mother's working hours **(X_4)** can enter the mother's working hours as a direct factor in increasing the child's weight, as the mother's preoccupation with work, especially work outside the home, drives the family to depend on fast food. And prepared foods, which causes a factor to gain weight in children.

5- Is the father alive **(X_5)** this variable is considered an indirect factor, because that will be included in the family's income and thus the quality of the family's food consumption.

6- The number of the child's meals per day **(X_6)** the child's nutritional behaviors are random and irregular, and sometimes the child's meals reach very large numbers, and with very high calories.

7- The number of non-main meals for the child per day **(X_7)** children in their diet depend on non-main meals, all chocolate, gypsum and other prepared foods. These meals may reach large quantities.

8- The number of hours sitting in front of TV and smart phones **(X_8)** this variable is considered one of the main factors in increasing children's weight due to the lack of sports activity when staying for long periods on television or smart phones.

9- Number of sleeping hours per day **(X_9)** recent medical studies have proven that less sleep is one of the causes that lead to weight gain in children and adults, in order to stimulate some hormones responsible for weight gain in the human body.

10 - Does the child have a thyroid disorder **(X_{10})** If there is an imbalance in the secretion of the thyroid gland, this will contribute to weight gain in children, even if their diet is healthy.

11- The order of the birth of a child among his siblings **(X_{11})** the order of the birth of a child among his brothers has a role in increasing the weight of the child himself due to hereditary and non-genetic factors.

12- Monthly family income **(X_{12})** food behaviors vary from one family to another depending on the family's monthly income. If the family's income is high, the children of those families will consume high-calorie food quantities that may be very high.

13- The number of sports hours for the child **(X_{13})** this variable means the number of hours of stressful games that the child plays, such as games and science, my effort such as running, jumping and so on.

14- Child housing **(X_{14})** child housing is one of the important variables, where housing in cities makes food options for the child due to the proximity of markets and in abundance. Conversely, in local areas, food options are limited.

15- The marital status of the mother **(X_{15})** it is known that the care of the health of the child is entirely entrusted to the mother, and therefore the marital status of pain has a role in the health of the child in general and not only on an increase or decrease in its weight. From **table 10**, it can be observed that the proposed model (**Benqr**) give the less values of **MSE** criterion among the lasso quantile regression (**lqr**) and classical quantile regression (**cqr**) models under different quantile levels (**25%,50%,75%,** and **99%).**

Methods	MSE at 0.25	MSE at 0.50	MSE at 0.75	MSE at 0.99
Benqr	33.16415	21.22685	29.16255	96.81203
Lqr	35.73716	23.53479	31.26604	111.3631
Cqr	33.52035	33.52035	32.48061	393.5459

Table 10. MSE valued for (0.25, 0.50, 0.75, and 0.99) quantiles

β	$Benqr$	Lqr	Cqr
β_1	2.5014599	2.426672371	2.475293040
β_2	1.31755930	0.849223021	0.838277266
β_3	-1.90213819	0.201054180	-2.041567993
β_4	0.000	0.182675021	-0.119409340
β_5	0.57212346	0.224982314	0.659439987
β_6	-0.14141920	0.649638361	-0.621611460
β_7	0.12299057	0.179366047	0.305794655
β_8	-0.37494003	0.125041129	-0.342463811
β_9	-0.33806463	0.000	-0.272605262
β_{10}	0.18295002	0.229228758	0.450100296
β_{11}	0.000	0.162517603	-0.042885332
β_{12}	-0.17641313	0.000	-0.172447169
β_{13}	0.000	0.000	0.001586925
β_{14}	-0.42614971	0.000	-0.191640565
β_{15}	0.36936485	0.376423894	0.771559606

Table 11. parameter estimates

Table 11, displayed the estimates of coefficients of the predictor variables under our method (**Benqr**), **Lqr**, and **Cqr**. We observed that our method is compareable, also the proposed model provided variable selection procedure, for example ($\boldsymbol{\beta_4 = 0.0}$), ($\boldsymbol{\beta_{11} = 0.00}$), and ($\boldsymbol{\beta_{13} = 0.000}$). That is mean, the variables (The mother's working hours, the order of the birth of a child among his siblings, and the number of sports hours for the child) are unimportant variable and do not effect the response variable (weight of children under the age of five) and for that we removed from the estimated regression model.

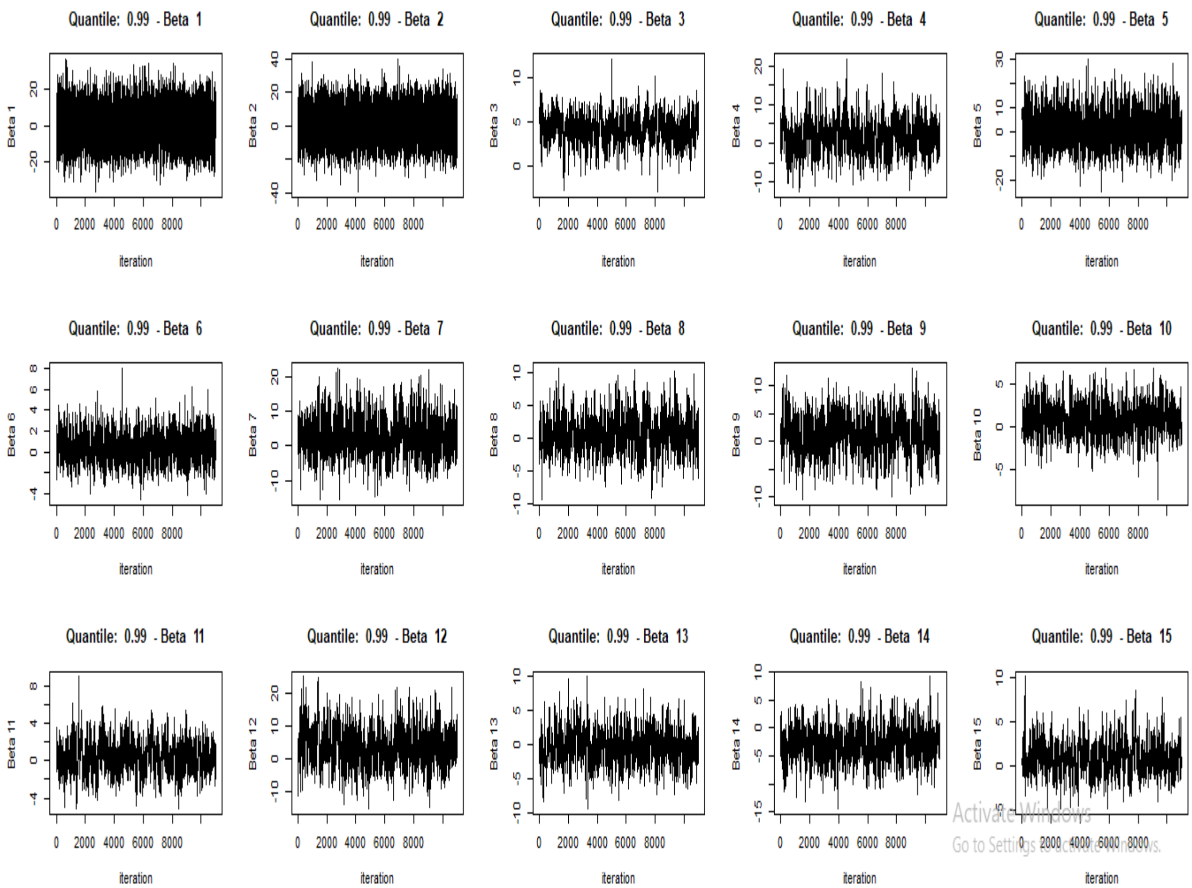

Figure 7. Trace plots of Benqr with (0.99) quantile

Figure 7, shows the trace plots of the parameter estimates which are indicates that the posterior distribution of the interested parameters is stationary.

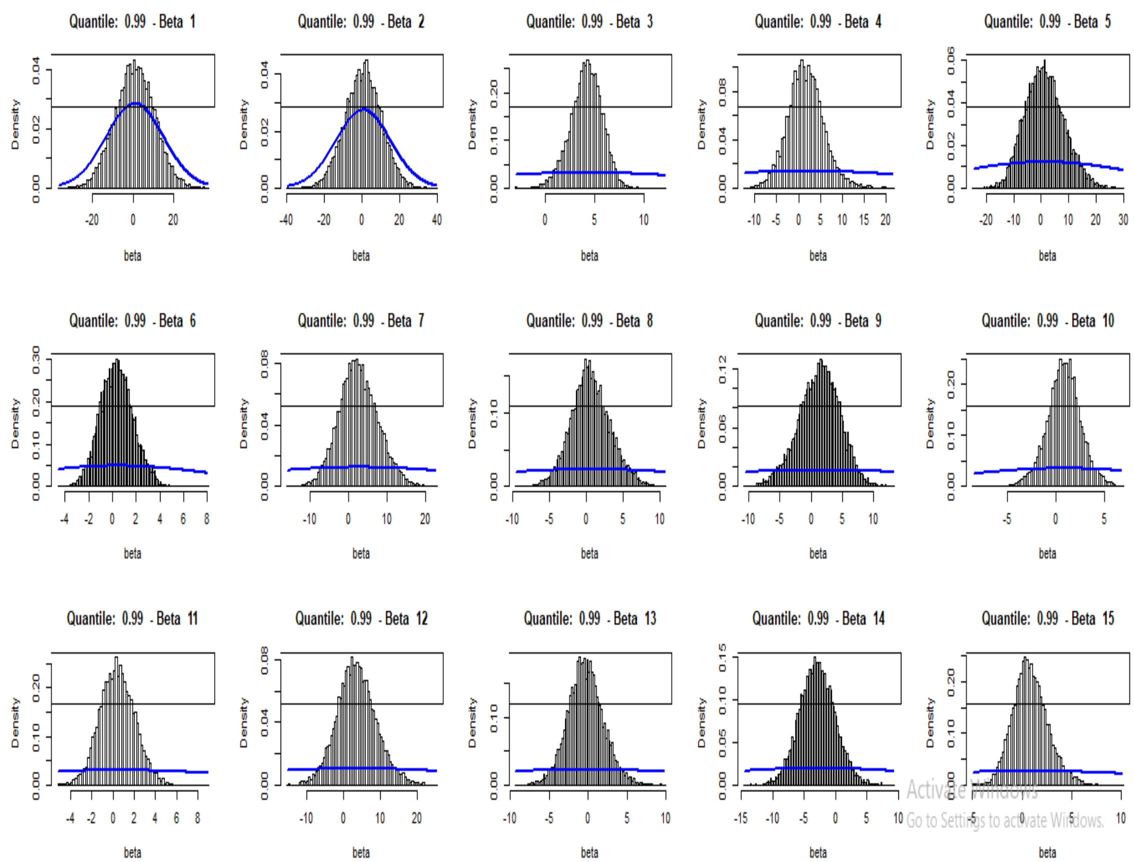

Figure 8. Histograms of Benqr parameter estimates

Figure 8, displayed the distributions of the parameters estimates which are indicates that all the parameters follows the normal distribution.

Chapter Five

5. Conclusion and Recommendation

5.1. Conclusions

This thesis presents a new contribution for the Bayesian elastic net quantile regression models through employing the Laplace density of parameter ($\boldsymbol{\beta}$) 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 and in the real data 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.

5.2. Recommendation

The proposed model, Bayesian elastic net quantile will motivate the researchers to develop other penalized Bayesian regression model, such as the develop of Bayesian elastic net Tobit regression, Bayesian elastic net binary regression, and many other penalized Bayesian regression models.

Bibliography

Al-hamzawi, R. (2014). Bayesian elastic net Tobit quantile regression. Communications in Statistics-Simulation and Computation, 45(7), 2409-2427.

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

Alshaybawee, T. Midi, H. & Alhamzawi, R. (2016). Bayesian elastic net single index quantile regression. Journal of Applied Statistics, Vol 44, 2017-Issue 5.

Armido, D. and Alfred, M. (1986). "Computation of the incomplete gamma function ratios and their inverse." ACM Trans. Math. Soft. 12: 377-393. 156

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

Chen, M., D. Carlson, A. Zaas, C. Woods, G.S. Ginsburg, J. Lucas, and L. Carin (2011). De-tection of viruses via statistical gene expression analysis. IEEE Transactions on Biomedical Engineering. 58, 468-479.

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 Illinois at Urbana-Champaign.

Flaih, A.N. Alshaybawee, T. Alhusseini F.H.H. (2020). Sparsity via new Bayesian Lasso. Periodicals of Engineering and Natural Sciences. 8(1).

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. 01-07.

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). NY, USA: Springer.

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

Jiratchayut, K. M. (2014). Bayesian Elastic Net Regression. A dissertation doctor in statistics. Thammasat University. Faculty of science and Technology.

Jorgensen (1982). "Statistical properties of the generalized inverse Gaussian distribution." Lecture Notes in Statistics, 9. 156

Kirkland, Lisa-Ann. (2014). Lasso-Simultaneous shrinkage and election via $\boldsymbol{\rho}$, norm. Msc. thesis. University of Pretoria.

Koenker, R. (2005): Quantile Regression, University of Illinois. Cambridge University Press, Cambridge.

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

Koenker, R. and Hallock, K.F. (2001). Quantile regression. J. Econ. Perspect. 15, 143 156, [Cross Ref]

Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. Journal of Statistical Computation and Simulation.

Lawson, C.L. and Hanson, R.J. (1974). Solving least square problems. (Prentice-Hall Series in Automatic Computing). Englewood Cliffs, NJ, USA: Prentice-Hall.

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.

Mallick, H., & Yi, N. (2014). A new Bayesian lasso. Statistics and its interface, 7(4), 571.

Marasinghe, D. (2014). Quantile Regression for climate data. Master thesis. Clemson University.

Miller, A. (2002). Subset selection in regression. 2nd ed. (Monographs on Statistics and Applied Probability: 95). Boca Raton, FL, USA: Chapman & Hall/CRC.

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.

Park, T. & Casella, G. (2008). The Bayesian lasso. Journal of the American Statistical Association, 103(482), 681-686.

Scott, D. (2008). HyperbolicDist: The hyperbolic distribution. R Package version 0.5-2. URL http://www.r-project.org 156

Seber, G.A.F. and lee, A.J. (2003). Linear regression analysis. 2nd ed. (Wiley Series in Probability and Statistics). Hoboken, NJ, USA: John Wiley & Sons.

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), Stanford University, USA 67(2):301-320.

الشبكة المرنة البيزية في الانحدار القسيمي باستخدام توزيع مختلط جديد من التوزيع الطبيعي مع التطبيق

رسالة مقدمة لنيل درجة الماجستير في علوم الإحصاء

بواسطة

منتظر هاشم مناتي المساعدي

يشرف عليها

أ.د. احمد نعيم فليح

قسم الإحصاء

كلية الإدارة والاقتصاد

جامعة القادسية

2021

الخلاصة

كان لتحليل الانحدار البايزي أهمية كبيرة في السنوات الأخيرة ، وخاصة ما يعرف بالطريقة الاعتيادية ، مثل طرق ريج ، لاسو ، اللاسو التكيفي ، طرق الشبكة المرنة ، حيث يعتبر اختيار التوزيع المسبق للمعلمة المهتمين بتقديرها هو الفكرة الرئيسية في تحليل الانحدار البيزي. ومن خلال فرض شرط جزائي على نموذج الانحدار البيزي ، يتم تقليل تباين المقدرات بشكل ملحوظ ويزداد التحيز. ينتج عن المفاضلة بين التحيز والتباين في مقدر الانحدار البيزي الجزائي نموذجًا أكثر قابلية للتفسير مع دقة تنبؤ أكبر. ويعتبر الانحدار القسيمي من المواضيع المهمه والتي تعتبر تعميم الى الانحدار الاعتيادي، ويأتي مصطلح القسيم (المئين) كترجمة مطابقة الى مصطلح ال (**Quantile**) حسب رأي الأستاذ الدكتور (رحيم الحمزاوي).

في هذه الرسالة ، اقترحنا نموذجًا هرميًا جديدًا للانحدار القسيمي البيزي من خلال استخدام توزيع مختلط لمعلمة المقياس من التوزيع الطبيعي مع توزيع كاما المبتور الذي ذكرها الباحثان (لي ولين 2010) كتوزيع لابلاس مسبق. وبناءا على ذلك، تم اقتراح خوارزميات أخذ العينات جيبس الجديدة.

تم إجراء مقارنة مع نموذج الانحدار القسيمي الكلاسيكي ونموذج الانحدار القسيمي لاسو من خلال إجراء دراسات المحاكاة ، وبناءً على المقاييس الإحصائية (**MSE، SD، MMAD**). وكذلك تطبيق البيانات الحقيقية. نموذجنا كان قابل للمقارنة مع النماذج الاخرى ويعطي نتائج أفضل.