Employing the Bayesian Elastic Net in Quantile Regression with an Application

Muntadher Hashim Mnati Almusaedi
stat.post32@qu.edu.iq

Dr. Ahmad Naeem Flaih
ahmad.flaih@qu.edu.iq

University of Al-Qadisiyah

Abstract

In this paper we employing the Bayesian elastic net method in quantile regression. The two penalizing (ridge and lasso) function usually combined to produce the elastic net method, in which the variance of the estimators are reduced and the bias approaches the smaller value. The tradeoff between the bias and variance of the estimator produced an interpretable regression model and gives more prediction accuracy. In this paper, we proposed new Bayesian hierarchical model for the quantile regression by utilizing the scale mixture of normal mixing with truncated gamma distribution $(1,\infty)$ which proposed by (Li and Lin, 2010) as Laplace prior distribution for the parameter (β). Moreover, Gibbs sampling algorithms are introduced for the posterior distributions. Real data application for the proposed model has been deducted and a comparison has been made with classical quantile regression model, also with lasso quantile regression model .Our model is comparable and gives better results.

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

1. Introduction

Regression analysis describe the functional form between the response variable Y and one or more predictor variables X. Then, regression analysis can be used for finding the regression model that produce more prediction accuracy and more interpretable model. Also, regression analysis offers the variable selection procedure. The linear regression model is a statistical methodology that estimate the mean of the response variable (y) by using the information from data of the predictor variables. Ordinary Least Squares (OLS) estimators have the smallest and unbiased estimators. OLS offers biased estimators that have bigger variances when the multicollinearity problem appear in the data, and when the number of predictor variable more than the sample size **(k≥n)**. In this case and to overcome the limitations of least squares, the regularization method is the solution, which is a tradeoff between the variance and the bias of estimator. The regularization regression methods works well in the case of many predictors or in the presence of multicollinearity, which are produces biased estimators with the small variance (James et al., 2013).

The ridge method introduced by (Hoerl and Kennard, 1970) through adding the L1-norm constrain to residuals sum of squares (RSS) term to address the collinearly or **k>n** problem, but ridge estimates does not sparse the parameters. (Tibshirani, 1996), proposed the lasso (Least absolute shrinkage and selection operator) which is works by adding L2–norm constrain to RSS. Unlike ridge, lasso method set the estimates near to zero, in other words lasso method can delete the irrelevant predictor variables and then produce more interpretable model. Moreover, the EN is a regularization method proposed by (Zou and Hastie, 2005) which combined the ridge method and lasso method together to the RSS term, EN method deal with the relevant predictor variables that have highly correlated with each other and EN most of the time works better than the lasso (Osborne et al., 2000).

The combined penalties method EN deal with the grouping effect of predictor variables when there are strong correlated between groups of predictor variables, the EN estimator is given as follows,

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

Where λ_1 , λ_2 ≥ 0 are the shrinkage parameters that controls the amount of shrinkage for regression parameters.

In practice, many of the data shows the violation of the linear model assumptions and/or the researchers are concers in modelling quantities rather than the estimed mean response variable $E(y|x)$, like the median, and other quantiles (Chatterjee and Hadi, 2013). The quantile regression needs no any assumption on the distribution error term (Koenker and Bassett, 1978). Quantile regression very common model in many different practical fields such as, econometrics, ecology, biology, survival analysis and many other fields of sciences. The quantile regression model is defined by

$$y_i = x_i'\beta(p) + \epsilon_i(p), \qquad (1)$$

Where β (p) the quantile estimator that minimizing the RSS,

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

where $\rho_p\,(\cdot)$ the check function (Koenker and Bassett, 1978) that defined as follows,

$$\rho_p\,(\epsilon) = \frac{|\epsilon| + (2p - 1)\epsilon}{2} \qquad (3)$$

(Ghosh, 2007) introduced new regularization method for the elastic net that is called adaptive elastic net where the estimators have different weights and then produced the adaptive lasso as well elastic net estiates.

(Alshaybowee et al., 2016) proposed the Bayesian elastic net with single index quantile regression (semi parametric) model to overcome the high dimensionally problem in the data.

(Lee et al., 2016) produced the elastic net shrinkage method to address the dimensionality problem in the data that have strong correlation among the predictor variables in group selections.

(Feng, 2011) developed Bayesian Monte Carlo Markov Chain algorithm for estimating the quantile linear regression coefficients with two Bayesian quantile model methods, the estimators are efficient compared with some existing regression models.

(Al-hamzawi, 2013) introduced extensions to the Bayesian quantile regression by using the prior distribution which works with the full conditional conjugate prior.

(Al-hamzawi, 2016) introduced the Tobit quantile regression model from the Bayesian point of view with gamma prior for the regression parameters in the elastic net method.

(Li et al,. 2010) proposed the Lasso, elastic net, and group lasso regression models with Bayesian analysis of the quantile regression.

(Li and Lin, 2010) introduced new formulation for prior distribution of the elastic net with Bayesian analysis linear regression to overcome the double shrinkage problem in the elastic net penalty function, the prior formulation of $\pi(\beta|\sigma^2)$ is defined as follows,

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

In this paper we introduced new hierarchical model and new Gibbs sampler algorithm for the quantile regression to improve the prediction accuracy of the regression model.

2. The Hierarchical model and prior distributions

Along with the quantile regression model (1) and the prior distribution (4), we have the following hierarchical model for Bayesian elastic net quantile regression,

$$y_i^* = \acute{x}_\iota\, \beta_p,$$

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

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

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

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

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

3. The Full conditional Posterior Distributions

Assuming that all priors for the different parameters are independent, we can write the full conditional posterior distributions as follows,

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

Where $\mathrm{i} = 1, 2, \ldots, \mathrm{n}$

Depending on (Alhamzawi, 2016) and (Li and Lin, 2010) and conditioning on y^*, V_i, β the posterior distribution of parameter β is

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

$$= -\frac{1}{2\sigma^2}\, (\beta - R^{-1} x' y^*)' R\, (\beta - R^{-1} x' y^*)$$

Then distribution of β is multivariable normal with mean $R^{-1}x'y^*$ and variance σ^2R^{-1} ;

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

The second variable σ^2posterior distribution is defines as follows

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

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

The third variable $(\tau-1_k)$ posterior distributed is defined as follows,

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

4. The Gibbs Sampler and the Full Conditional Distribution

In this paper we will use the is Gibbs sampler to implement the model hierarchy (5). The Gibbs sampler generates the random variables from the full conditional distributions directly for the interested parameter and fixed all the other parameters (Evans, 2012).

The conditional posterior distributions for each parameter for the elastic net quantile regression will be generated by using the following algorithms steps:

1- We update 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- We update β from the full conditional posterior distribution that follows the multivariate normal distribution (6) with mean $R^{-1}X'Y^*$ and variance σ^2R^{-1}, where

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

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

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

(Chhikarn and Folks, 1988) shows the inverse Gaussian distribution that is:

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

4- We update $(\tau_j - 1)^{-1}$ from the inverse Gaussian distribution (Chhikarn and Folks, 1988)

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

5- We update σ^2 by using the acceptance-rejection algorithm that works with the incomplete gamma functions,

$$(\sigma^2) \le \frac{\Gamma_a\, \Gamma_{\frac{1}{2}}^{-k}}{b^a}\, g(\sigma^2);$$

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

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

Where g (.) is the inverse gamma (a, b).

6- Real Data Analysis

An real data analysis performed on the obesity in children. 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,

$$mse(\hat{\beta}_j) = var(\hat{\beta}_j) + Bias(\hat{\beta}_j)^2.$$

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 a 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 solving 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. Child age (x_1), Child's gender (x_2), Mother employed (x_3) ,The mother's working hours (x_4) , Is the father alive (x_5) , The number of the child's meals per day (x_6) , The number of non-main meals for the child per day (x_7) , The number of hours sitting in front of TV and smart phones (x_8) , Number of sleeping hours per day (x_9) , Does the child have a thyroid disorder (x_{10}), The order of the birth of a child among his siblings (x_{11}) , Monthly family income (x_{12}) , The number of sports hours for the child (x_{13}), Child housing (x_{14}) ,The marital status of the mother (x_{15}).

Table (1) shows the parameter estimate under (0.75) quantile for the proposed model Bayesian elastic net quantile regression (Benqr), Lasso quantile regression (Lqr), and Classic quantile regression (Cqr).

Table (1) parameters estimates under 0.75 quantile

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	0.0000	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	-0.00806463	0.000	-0.272605262
β_{10}	0.18295002	0.229228758	0.450100296
β_{11}	0.000	0.002517603	0.143453
β_{12}	-0.07641313	0.000	-0.172447169
β_{13}	0.000	0.000	0.4345666
β_{14}	-0.00614971	0.000	-0.191640565
β_{15}	0.36936485	0.376423894	0.771559606

From table (1) its shown that the proposed model is a comparable model with lasso quantile regression model (lqr) by giving some sparse solution. The following table (2) gives the values of the MSE measure to test the quality of the regression model under different quantile levels.

Table 2. MSE valued for (0.25, 0.50, 0.75, and 0.99) quantiles

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

From table 2, 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%).

The following figures (1) shows the trace plot of the posterior distributions for the different parameters and we can conclude the stability of the Gibbs sampler for the different posterior distribution for the parameters. Figure (2) displayed the distributions of the parameters estimates which are indicates that all the parameters follows the normal distribution.

Figure 1. Trace plots of our model with (0.99) quantile

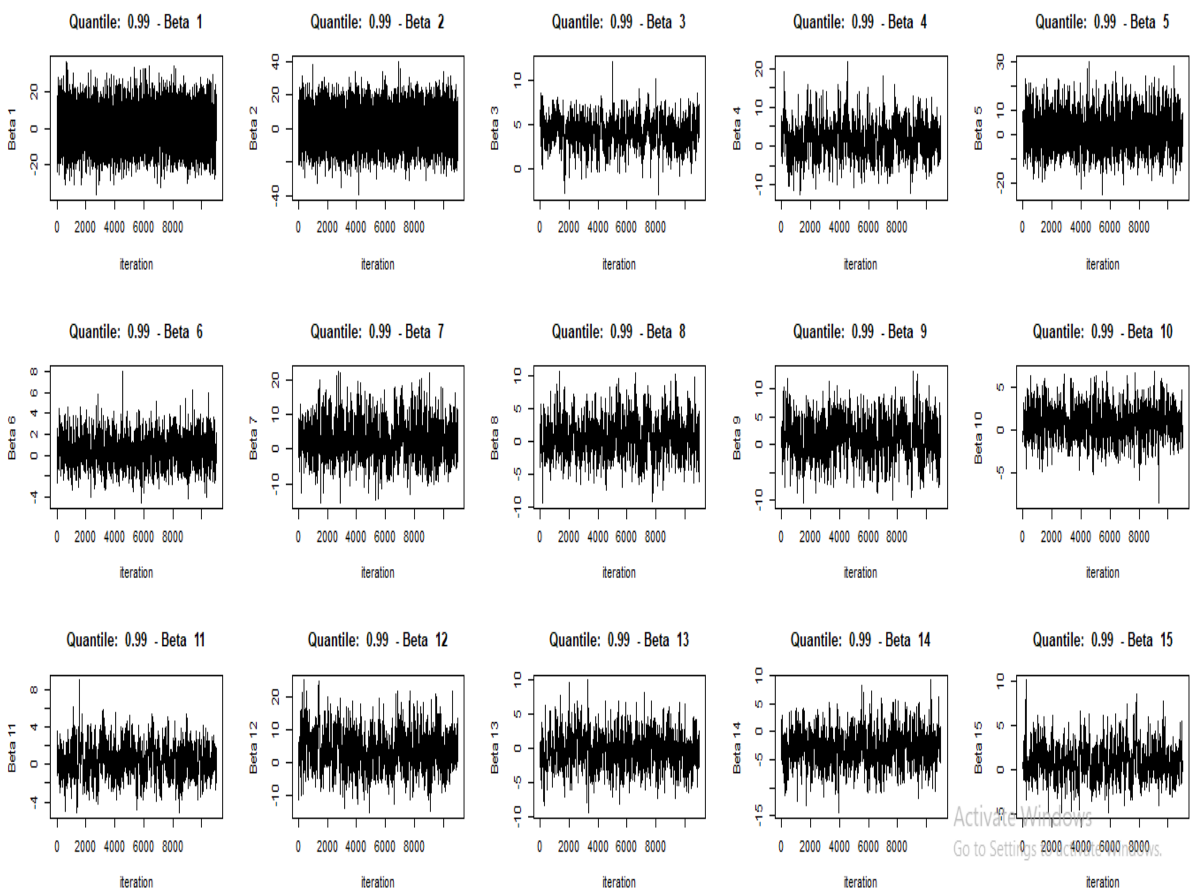
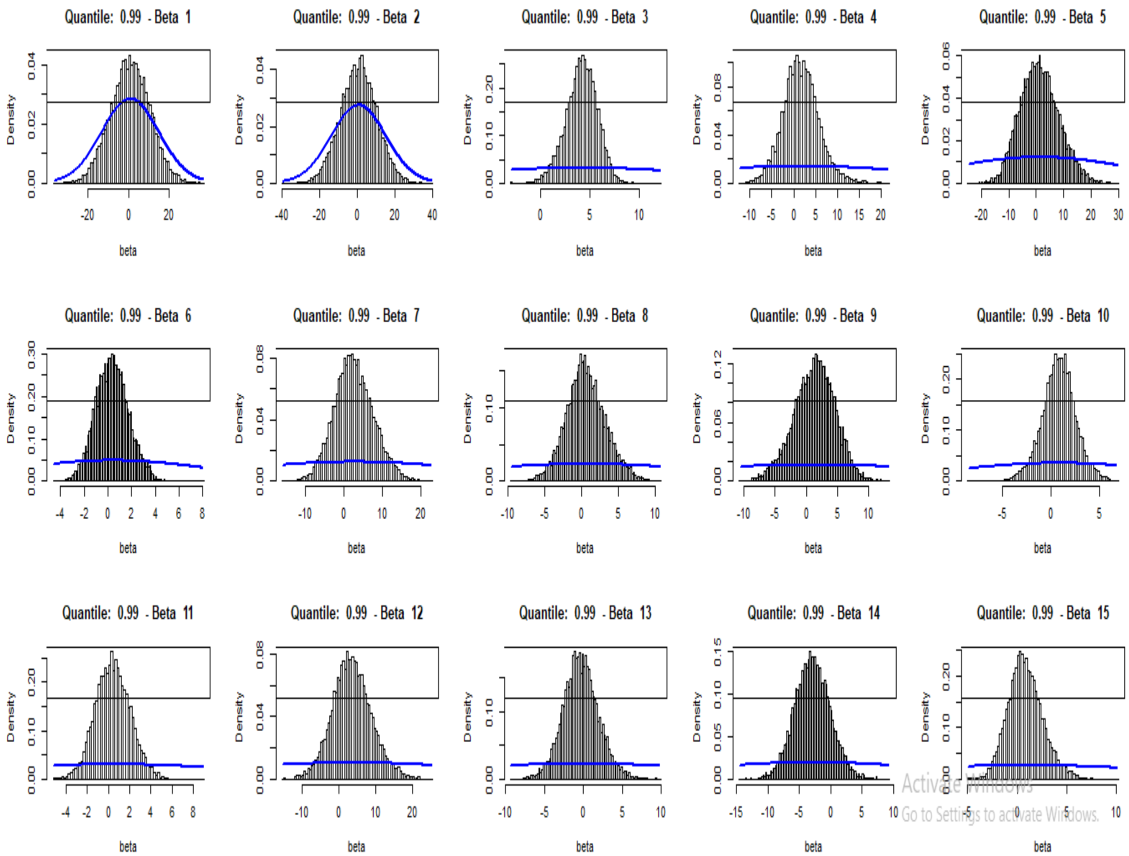


Figure 2. Histograms of our model parameter estimates



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 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 and the variable selection procedure.

8- Recommendation

The proposed model, Elastic Net Bayesian Quantile, will inspire researchers to develop a similar Bayesian regression model, for example, Tobit's regression for the Bayesian net, as well as the binary regression of the Elastic Net.

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.

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.

الملخص

في هذا البحث نستخدم طريقة الشبكة المرنة البايزية في الانحدار الكمي. عادة ما يتم الجمع بين دالتي الجزاء (ريدج ولاسو) لإنتاج طريقة الشبكة المرنة ، والتي يتم فيها تقليل تباين المقدرات ويقترب التحيز من القيمة الأصغر. أنتجت المفاضلة بين التحيز والتباين للمقدر نموذج انحدار قابل للتفسير وتعطي مزيدًا من دقة التنبؤ. في هذا البحث ، اقترحنا نموذجًا هرميًا بايزيًا جديدًا للانحدار الكمي من خلال استخدام توزيع مختلط لمعلمة المقياس من التوزيع الطبيعي مع توزيع كاما المبتور (1،∞) الذي اقترحه (Li and Lin, 2010) كتوزيع لابلاس المسبق للمعامل ($\boldsymbol{\beta}$). وبناءا على ذلك تم اقتراح خوارزميات أخذ العينات جيبس الجديدة. تم خصم تطبيق البيانات الحقيقية للنموذج المقترح تم إجراء مقارنة مع نموذج الانحدار الكمي الكلاسيكي ونموذج الانحدار الكمي لاسو ، نموذجنا قابل للمقارنة ويعطي نتائج أفضل.

الكلمات المفتاحية: التحليل البايزي ، الانحدار الكمي ، الشبكة المرنة ، عينة جيبس.