Al-Qadisiyah Journal for Administrative and Economic Sciences
QJAE, Volume 24, Issue 2(2022)
Manuscript ID:
ISSN$_{\text{Online}}$: 2312-9883
ISSN$_{\text{Print}}$: 1816-9171

Bayesian Elastic Net Regression and Elastic Net Regression with Application

[1]Ali Ahmad Hussein [2]Dr. Bahr Kadhim Mohammed

[1, 2] (Dept. of Statistics, Faculty of Administration and Economics/

University of Al-Qadisiyah, IRAQ,) ***stat.post13@qu.edu.iq*** , ***bahr.mahemmed@qu.edu.iq***

Corresponding Author : Ali Ahmad Hussein *Affiliation : Dept. of Statistics, Faculty of Administration and Economics*
Email: stat.post13@qu.edu.iq

Abstract

Model selection has become the widely used method to include the relevant predictor variables on the response variable and to remove the irrelevant variables. Bayesian penalized methods such as the elastic net methods are used to address the problem of grouping predictor variables effects. In this paper we concentrated on the sparsity procedure in linear regression model by using elastic net regularization method. We developed Bayesian elastic net by employing the scale mixture of normal distribution mixing with Rayleigh distribution as Laplace prior distribution for the regression parameters. The new scale mixture generates normal mixing with truncated gamma distribution, also we proposed new Gibbs sample algorithm. The proposed Bayesian elastic net method examined by applying real data some and the results shows that the proposed model is comparable with the other regularization methods.

Keywords: *Bayesian, Elastic net, Hierarchical prior model, Gibbs sampler, sparsity.*

1-Introduction

Obviously the regression analysis methods are very widely popular tools that investigated the relationship between the response variable and the independent variable(s). This motivated many authors and researchers to develop various regression analysis tools that cope with the practical underlying situation. The ordinary least squared (OLS) method is very common tool to find the regression coefficient estimates. Moreover, violated the assumptions of (OLS) was the key idea behind searching for substitution methods for regression coefficient estimates.

In addition for that the investigation about the more explanation model developed along with the model selection and variable selection procedures. The Ordinary Least Squares provided unbiased and smallest variance parameters estimates through minimizing of the Residual Sum of Squares (RSS),

$$\text{RSS}(\beta) = (y^{true} - f(X;\ \beta))^2$$

Regression analysis attempts to estimate the population average of the response variable by using the information that the predictor variables are provided. The parameter estimates of regression model are reliable estimates if it offers balance between the variance and bias, in addition to the model explain ability.

It is well known that the OLS estimates are biased and inconsistent (inflated variance) when the multicollinearity problem appear in the design matrix X, or when the number of predictor variable p exceed or near the number of observations n. Therefore, in these circumantance the OLS estimates are usually not unique and instable with high variances. The high variance in the OLS estimates motivated the authors to explore the regularization methods that used to overcomes the limitations of least squares estimates quality, James et al. (2013).

The ridge regression method adding a penalty function to residuals sum of squares (RSS) to address the problem multicollinearly, where the penalty function contains the L2-norm. The ridge parameter estimates cannot set to zero, Hoerl and Kennard (1970).

(Tibshirani, 1996) produced Lasso method which is essentially regards as penalized method that provide variable selection procedure. Consequently, many authors developed other shrinkage methods to provide variable selection procedure; such as, relaxed lasso, fused lasso, adaptive lasso, elastic net, etc. Model selection procedure in regression analysis aims to select the best fit estimated regression model through selecting the relevant predictor variables that affects the response variable and remove the irrelevant variables. In thesis I consider the linear regression model where the ordinary least squares (OLS) estimates are no longer achieved by minimizing the residual sum squares (RSS).

Instead of the OLS the elastic net have discussed in this thesis, elastic net is the flexible regularization and variable selection method that combined two of penalties function. Moreover, the Elastic Net (EN) is another penalized method that proposed by Zou and Hastie (2005) to address the limitations of lasso method. EN method combined the ridge and lasso to the RSS term, EN method deal with many relevant predictors that have highly pairwise correlation and EN usually works better than lasso, Osborne et al.(2000a). Obviously the regression analysis methods are very widely popular tools that investigated the relationship between the response variable and the independent variable(s).

This motivated many authors and researchers to develop various regression analysis tools that cope with the practical underlying situation. The ordinary least squared (OLS) method is very common tool to find the regression coefficient estimates. Moreover, violated the assumptions of (OLS) was the key idea behind searching for substitution methods for regression coefficient estimates. In addition for that the investigation about the more explanation model developed along with the model selection and variable selection procedures. The Ordinary Least Squares provided unbiased and smallest variance parameters estimates through minimizing of the Residual Sum of Squares (RSS).

In regression analysis, the set of the independent variable that should be included in regression equation bring the attention of the researcher, because it is the first part of the regression analysis and then examine to see whether regression equation was correct. So, the variable selection problem related with the regression form specification. The residual mean squares (RMS) is a criterion for model selection, the smallest the RMS between two regression equation is preferred.

$$RME = \frac{SSE}{n-k}$$

Where k is the number of independent variables and SSE is the sum of squares error.

Mallows (1973) developed the Mallows C_k criterion to judge the performance of the regression function by using the from following.

$$C_k = \frac{SSE_k}{s^2} + (2k - n),$$

Where s^2 is the estimated variance.

Akaike (1973) introduced the Akaike information criterion (AIC) as model selection criterion that combined the most fit equation and the smaller number of independent variables, the AIC defines as follows,

$$AIC_k = n\ln\left(\frac{SSE_K}{n}\right) + 2k,$$

The smallest AIC value the better model.

Schwarz (1978) proposed a modification of the AIC is called Bayes Information criterion (BIC) which is defines as follows,

$$BIC_k = n\ln\left(\frac{SSE_K}{n}\right) + k(\ln n),$$

The smallest BIC value the better model.

Zou et al. (2007) discussed the using of of BIC criterion to choose the shrinkage parameter in lasso method. Hocking (1976) list the evaluating regression method that is called all possible equations which is gives 2^k equations (k is the number of independent variables), where we can use the (RME, C_k, R^2) to select the best model.

The limitation of all possible equations is the larger number of equations when k getting larger. Efroymson (1960) introduced the stepwise method as variable selection procedure combined the mechanism of both Forward Selection (FS) Procedure and Backward Elimination (BS) procedure. The calculation of the stepwise method depends on the inclusion and deletion of independent variables, it is essentially a modification method for (FS and BE) methods.

The AIC and BIC are used for select the best fitted model in the stepwise method. It is recommended obtaining the variance inflation factors (VIF) test or the eigenvalues of the correlation matrix of the independent variables as a first step to variable selection procedure.

George and McCulloch (1993) proposed another method for utilizing an information criterion for model selection; this method is called stochastic search variable selection. This method can be used in the well know Bayesian algorithm , so it is depends on the probabilistic considerations in selecting of the subsets of independent variables.

Hoerl and Kennard (1970) introduced a theory about ridge regression with penalized function to estimate the parameters of multiple regression model by adding a small positive quantity (λ) to the inverse of (X^tX) matrix to address the problem of linearly dependent (correlation) of the independent variable. The ridge estimator is biased but with the smallest variance. Also, ridge methods can be applied in the case of (n ≥ k) and regards as regularization method. But ridge regression is not a variable selection method. Ridge uses the L2-norm as penalty function.The response variable in ridge regression is centered , Draper and Smith (1998). Tibshirani (1996) proposed the new variable selection method that is called Lasso. Lasso method can be regards as regularization method that adds the L1-norm penalty function to the RSS. Due to the L1-norm, lasso provides variable selection procedure by setting the parameter estimates to zero.

Also, in this paper there is a remarkable note about Bayes estimation for the linear regression model based on assuming that the parameter β is follows the double exponential distribution as prior density. Hans (2009) introduced Bayesian estimation for lasso regression coefficients. New Gibbs sampling algorithm have developed by imposed directly the Laplace prior on the lasso regression parameters and a gamma prior on the tuning parameter. The results emphases that the classical lasso results did not matching the Bayesian results in terms of prediction.

Yuan and Lin (2006) introduced the so called group lasso as new regularization method; the group lasso is a generalization for the lasso method. Group lasso method essential founded to deal with problem of selected grouped independent variables. Lasso selected individual independent variables but group lasso can select a set of small groups of independent variable. Efron et al. (2004) introduced algorithm to compute the lasso estimate this algorithm is called LARS. LARS used for sake of model selection; they proved that it is takes short time for computational implementation in lasso.

Kyung et al. (2010) introduced the Bayesian estimation for the linear regression with proposed hierarchal models. The Gibbs sample algorithm have developed for the lasso, elastic net, group lasso, and fused lasso methods. The results showed that the proposed hierarchal model outperforms the LARs algorithms from the Bayesian perception. Ghosh (2007) and Zou and Zhang (2009) introduced two adaptive elastic net regularization methods. These new regularization methods focused on the limitation of lasso in dealing with presence of grouped independent variable and the inconsistent of estimators. The adaptive lasso overcomes the problem of inconsistent estimator by imposing weights for the different parameters. Also, adaptive elastic net estimators have oracle properties (normality and consistent). We can say that this method is combining of adaptive lasso and elastic net.

Celeux et al. (2012) showed that lasso has not the ability to detect the effects of grouped variables. Also, they stated that the variable selection with Bayesian perception outperforms the variable selection with lasso and elastic net methods based on the efficiency criterion.

Hans (2011) introduced new Gibbs sampler algorithm to find the solution for the Bayesian estimates using the elastic net method. In this paper the values of the shrinkage parameters (λ_1 and λ_2) are used based on the 10-fold cross validation method. Also, the scale mixture of normal has used to make the computational of the Gibbs sampler algorithm easier. The proposed Gibbs sample algorithm considered as an alternative to SSVS method.

Zou and Hastie (2005) introduced the so called elastic net, which is regards as regularization method that combined the ridge and lasso methods. It can be considered as variable selection method that works simultaneously as variable selection and shrinkage method. Furthermore, the elastic net dealing well with a grouping effect of correlated independent variables as contrast of lasso.

Li and Lin (2010) introduced the parameter estimation of the elastic net model from the Bayesian perception. By using the Gibbs sampler algorithm based on considering that the prior density is a scale mixture of normal mixing with truncated Gamma. The linear regression model studied for variable selection and prediction accuracy, the proposed model outperform in variable selection procedure and is a comparable model in the terms of prediction accuracy.

Park and Casella (2008) developed Gibbs sample algorithm based on new Bayesian hierarchal prior model. The scale mixture of normal mixing with exponential density have used as representation form for the double exponential prior distribution through the lasso linear regression. The results are very similar for the classic lasso results.

Rahim and Haithem (2018) introduced new Bayesian elastic net regularization method for variable selection and parameter estimation in linear regression. New hierarchical form prior model have developed based on the location-scale mixture of normal mixing with gamma density. The simulation results and real data analysis results showed the outperforms of the proposed model.

Mallick and Yi (2014) introduced new Bayesian lasso method that depends on new representation of the double exponential prior density as scale mixture of uniform mixing with special case of gamma distribution. Variable selection procedure has performed and parameter estimation explained based on the new lasso method.

Alhamzawi (2016) proposed the Tobit quantile Bayesian elastic net regression model. The variable selection procedure and coefficients estimation have developed through new Bayesian hierarchal prior model. The gamma priors have used in Gibbs sample algorithm. The results showed that from the simulation examples and real data analysis that the proposed model outperforms other methods.

Flaih et al. (2020) introduced new scale mixture of normal mixing Rayleigh density to represent the double exponential prior density. New hierarchal prior model have developed and therefore new Gibbs sample algorithm have implement to calculate the mode of the posterior density of lasso regression model parameter. The proposed model is comparable in terms of variable selection and estimation accuracy. Fadel et al. (2020) developed an extension for lasso Tobit and adaptive lasso Tobit regression models based on the proposed scale mixture in Flaih et al. (2020). In this paper there are two ideas and one comparative study which are as follows:

1. To propose new Bayesian hierarchical model that consider the Laplace prior distribution as Scale mixture of normal mixing with Rayleigh distribution.

2. To combine the Bayesian model selection problem with the penalized elastic net linear regression model under the prior distribution mention in idea One.

3. To perform a comparative study between the Bayesian penalized elastic net that proposed in idea one and the classical elastic net.

2-Bayesian Hierarchical Prior models

Elastic net penalized method is very common used in regression model as a regularization method which combine the ridge and lasso penalty functions. Zou and Hastie (2005) introduced the elastic net method as sparsity procedure that can deal with the effect of correlated variables in of covariates , the elastic net estimator is defined as follows :

$$\hat{\beta} = \operatorname{argmin} \|y - X\beta\|^2{}_2 + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|^2 \ldots \quad (1)$$

Where the elastic net penalized function is

$$h(\beta) = \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|^2$$

here $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ the penalties parameters.

Flaih et al. (2020) introduced the Bayesian lasso regression model based on scale mixture representation (3.3). In this thesis I assumed the above formula (3.3) by considering the linear regression model:

$$E\,(\,y\,/\,X, \beta) = X\beta$$

Suppose that the scale mixture of Laplace distribution that mixing normal with Rayleigh distribution defined as follows,

If $x/y \sim N\,(\mu\,, y^2)$ with y~Ray (b), then x~Laplace (μ, b), that is:

$$\frac{1}{2b} e^{-\frac{|x-\mu|}{b}} = \int_0^\infty \frac{1}{\sqrt{2\pi y^2}} e^{-\frac{(x-\mu)^2}{2y^2}} \frac{y}{b} e^{-\frac{y^2}{2b}} dy \quad \ldots \quad (2)$$

by letting μ=0, X =β, and $b = \frac{\sigma^2}{\lambda_1}$, then (3.3) become as follows :

$$\frac{\lambda_1}{2\sigma^2} e^{-\frac{\lambda_1|\beta|}{2\sigma^2}} = \int_0^\infty \frac{1}{\sqrt{2\pi y^2}} e^{-\frac{\beta_j^2}{2y^2}} \frac{\lambda y}{\sigma^2} e^{-\frac{\lambda_1 y^2}{2\sigma^2}} dy \quad \ldots \quad (3)$$

Zou and Hastie (2005) introduced the prior distribution of elastic net method $\pi(\beta)$ as:

$\pi(\beta) \propto e^{-\lambda_1\|\beta\|^1 - \lambda_2\|\beta\|_2^2}$, ... (4)

Then by multiplying both sides of (3.5) with $e^{-\lambda_2\|\beta\|_2^2}$, we get the scale mixture that cope with the prior (3.6),

$$\frac{\lambda_1}{2\sigma^2}e^{-\frac{\lambda_1|\beta j|}{2\sigma^2}-\frac{\lambda_2\beta_j^2}{2\sigma^2}} = \int_0^\infty \frac{1}{\sqrt{2\pi y^2}} e^{-\frac{\beta_j^2}{2\sigma^2}} e^{-\frac{\lambda_2\beta_j^2}{2\sigma^2}} \frac{\lambda_1}{\sigma^2} e^{-\frac{\lambda_1 y^2}{2\sigma^2}}\, dy$$

$$= \int_0^\infty \frac{1}{\sqrt{2\pi y^2}} e^{-\frac{\beta_j^2}{2}\left(\frac{1}{y^2}+\frac{\lambda_2}{\sigma^2}\right)} \frac{\lambda_1 y}{\sigma^2} e^{-\frac{\lambda_1 y^2}{2\sigma^2}}\, dy$$

$$\text{Let } \frac{1}{\sqrt{y^2}} = \frac{\sqrt{\frac{1}{y^2}+\frac{\lambda_2}{\sigma^2}}}{\sqrt{1+\frac{\lambda_2 y^2}{\sigma^2}}} \text{ , then}$$

$$\int_0^\infty \sqrt{\frac{1}{y^2}+\frac{\lambda_2}{\sigma^2}}\; e^{-\frac{\beta_j^2}{2}\left(\frac{1}{y^2}+\frac{\lambda_2}{\sigma^2}\right)} . \frac{1}{\sqrt{1+\frac{\lambda_2 y^2}{\sigma^2}}} \frac{\lambda_1 y}{\sigma^2}\; e^{-\frac{\lambda_1 y^2}{2\sigma^2}}\; dy$$

$$\text{Let } t = 1 + \frac{\lambda_2 y^2}{\sigma^2} \quad \rightarrow \frac{1}{t-1} = \frac{\sigma^2}{\lambda_2 y^2}$$

and

$$\frac{1}{y^2} + \frac{\lambda_2}{\sigma^2} = \frac{\lambda_2}{\sigma^2}\left(1 + \frac{\sigma^2}{\lambda_2 y^2}\right) = \frac{\lambda_2}{\sigma^2}\left(\frac{t}{t-1}\right)$$

From $t = 1 + \frac{\lambda_2 y^2}{\sigma^2}$ if y=0 , and $y = \infty$, we get $t \in (1, \infty)$

$$\frac{\lambda_1}{2\sigma^2} e^{-\frac{1}{2\sigma^2}\left(\lambda_1|\beta_j| + \lambda_2\beta_j^2\right)} \propto \int_1^\infty \sqrt{\frac{\lambda_2}{\sigma^2}\left(\frac{t}{t-1}\right)}\, e^{-\frac{\beta_j^2}{2}\left(\frac{\lambda_2}{\sigma^2}\left(\frac{t}{t-1}\right)\right)} t^{-\frac{1}{2}} \frac{\lambda_1 y}{\sigma^2} e^{-\frac{\lambda_1}{2\sigma^2}\frac{t\sigma^2}{\lambda_2}} \frac{\sigma^2}{\lambda_2 2y} dy$$

$$\propto \int_1^\infty \sqrt{\frac{t}{t-1}} \sqrt{\frac{\lambda_2}{\sigma^2}}\; e^{-\frac{\beta_j^2}{2}\left(\frac{\lambda_2}{\sigma^2}\left(\frac{t}{t-1}\right)\right)} t^{-\frac{1}{2}} e^{-\frac{\lambda_1}{2\lambda_2}} t\, dt$$

$$\propto \int_1^\infty \sqrt{\frac{\lambda_2}{\sigma^2}\frac{t}{t-1}}\, e^{-\frac{\beta_j^2}{2}\left(\frac{\lambda_2}{\sigma^2}\frac{t}{t-1}\right)} t^{-\frac{1}{2}} e^{-\frac{\lambda_1 t}{2\lambda_2}} dt \qquad \text{... (3.7)}$$

From (4), we can deal with β_j/σ^2 as Scale mixture of normal distributions N $(0,\frac{\sigma^2(t-1)}{\lambda_2 t})$ mixing truncated gamma with shape parameter (1/2) and Scale parameter ($\frac{2\lambda_2}{\lambda_1}$), see Almusaedi and Flaih (2021a, 2021b), Alsafi and Flaih (2021) for more information. By formula (4), we have the following elastic net linear regression (ENLR) hierarchical model,

$$\left.\begin{array}{c} y = X\beta + e\,, \\ y|X\,,\beta\,,\sigma^2 \sim N\,(X\beta\,,\sigma^2\, In\,), \\ \beta\,\Big|\,\lambda_2,\sigma^2,t \sim \prod_{j=1}^{p} N\left(0,\left(\frac{\lambda_2}{\sigma^2}\frac{t_j}{t_j-1}\right)^{-1}\right) \\ t|\lambda_1,\lambda_2 \prod_{j=1}^{p} \text{truncated gamma}\left(\frac{1}{2},\frac{2\lambda_2}{\lambda_1}\right); t \in (1,\infty), \\ \sigma^2 \sim \text{ Inverse Gamma} \end{array}\right\} \ldots \quad (5)$$

3- Full Conditional Posterior Distributions of ENLR

By using the hierarchical model (5), the full joint distribution is well defined as follows:

$$\pi\left(\beta | y, X, \sigma^2, t\right) \propto \pi(y/X, \beta, \sigma^2)$$

$$f\left(y | \beta, \sigma^2\right) \pi\left(\sigma^2\right) \prod_{j=1}^{p} \pi\left(\beta_j | \ t_j, \sigma^2\right) \pi(t_j) \ =$$

$$\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)} \cdot \frac{\tau^\alpha}{\sqrt{\alpha}} \left(\sigma^2\right)^{-\alpha-1} e^{-\frac{\tau}{\sigma^2}}$$

$$\prod_{j=1}^{p} \sqrt{\frac{\lambda_2 t}{\sigma^2 (t-1)}} \quad e^{-\frac{\beta_j^2}{2}\left(\frac{\lambda_2}{\sigma^2} \cdot \frac{t}{(t_j-1)}\right)} \quad t^{-\frac{1}{2}} \ e^{\frac{-\lambda_1}{2\lambda_2} t} \ \ldots \quad (6)$$

Remark that, y variable is centered and x is standardized. Now the full conditional posterior distributions are as follows:

The parts that includes β, $\pi(\beta)$ in the joint distribution (6) is

$$e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)-\frac{1}{2\sigma^2}\lambda_2\beta' A\beta} \qquad \text{, where } A = \left(\frac{t}{t-1}\right)$$

$$= \exp\left[-\frac{1}{2\sigma^2}\left\{\left(\beta' (X'X)\beta - 2yx\beta + y'y\right) + \lambda_2 \beta' A \beta\right\}\right]$$

$$= \exp\left[-\frac{1}{2\sigma^2}\left\{\beta'\left(X'X + \lambda_2 A\right)\beta - 2yX\beta + y'y\right\}\right]$$

$$= \exp\left[-\frac{1}{2\sigma^2}\left\{\left(\beta' C\beta - 2yx\beta + y'y\right)\right\}\right]$$

Where $C = X'X + \lambda_2 A$

$$\exp\left\{-\frac{1}{2\sigma^2}\left(\beta' C\beta - 2yX\beta + y'y\right)\right\} \ldots \quad (7)$$

Let $\left(\beta - C^{-1}X'y\right)' C\left(\beta - C^{-1}X'y\right) = \beta C'\beta - 2yX\beta + y'(XC^{-1}X)'y$

then (7) Can rewrite as follows:

$$\exp\left[-\frac{1}{2\sigma^2}\left\{\left(\beta - C^{-1}X'y\right)' C\left(\beta - C^{-1}X'y\right) + y'(In - XC^{-1}X')y\right] \ \ldots \quad (8)$$

The second part of (8) does not involve β, so we can reduce (8) as follows

$$\exp\left[-\frac{1}{2\sigma^2}\left\{\left(\beta - C^{-1}X'y\right)' A\left(\beta - C^{-1}X'y\right)\right] \ldots \quad (9)$$

We can say that (9) is the multivariable normal distribution with mean $C^{-1}X'y$ and variance $\sigma^2 C^{-1}$.

The second Conditional posterior distribution is for σ^2, $\pi(\sigma^2)$. The terms that involve σ^2 in the full joint distribution (6) are as follows

$$\left(\sigma^2\right)^{-\frac{n}{2}} \left(\sigma^2\right)^{-\alpha-1} \left(\sigma^2\right)^{-\frac{p}{2}} - \frac{1}{e^{2\sigma^2}}(y - X\beta)'(y - X\beta) - \frac{\tau}{\sigma^2} - \frac{\beta'\lambda_2 A\beta}{2\sigma^2}$$

Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 24, Issue 2(2022) ISSN$_{Print}$: 1816-9171
Manuscript ID:

$$= (\sigma^2)^{-\frac{n}{2}-\frac{p}{2}-\alpha-1} \; \frac{-1}{e^{2\sigma^2}}\{(y-X\beta)'(y-X\beta)+\tau+\beta'\lambda 2A\beta\} \quad \ldots \quad (10)$$

The formula (10) is the inverse gamma distribution with shape parameter

$(\frac{n}{2}+\frac{p}{2}+\alpha)$ and Scale parameter $\frac{(y-X\beta)'(y-X\beta)}{2}+\frac{\beta'\lambda 2A\beta}{2}+\tau$.

The third part in the conditional posterior distribution of (t_j). The parts of (6) involve (t_j) are

$$\sqrt{\frac{\lambda_2}{\sigma^2}\frac{tj}{t_j-1}}\; e^{-\frac{\beta_j^2}{2}(\frac{\lambda_2}{\sigma^2}\frac{t_j}{t_j-1})tj^{-\frac{1}{2}}}\; e^{-\frac{\lambda 1}{2\lambda 2}t_j}$$

Then based on the (Chhikara and Folks 1988) works, the distribution of $(t-1)$ is the generalized inverse Gaussian distribution and defined as follows,

$$(t-1)\sim GIG(\lambda=\frac{1}{2}, a=\frac{\lambda 1}{4\lambda_2\sigma^2}, \chi=\frac{\lambda_2\beta_j^2}{\sigma^2}), \ldots \quad (11)$$

Then, $(t-1)^{-1}$ variable follows the full conditional inverse Gaussian distribution with $\mu=\frac{\sqrt{\lambda_1}}{(2\lambda_2|\beta_j|)}$ and $\lambda=\frac{\lambda_1}{4\lambda_2\sigma^2}$. See (Chhikara and Folks 1988) for more details.

The choosing of the Shrinkage parameters λ_1 andλ_2 conducted by Following Li and Lin (2010) and Park and Casella (2008), they used the empirical Bayes procedure. we can take the log for the following functions and the maximization problem is solving as follows:

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

$$-\frac{\lambda_1}{4\lambda_2}\sum_{j=1}^{p}E\left[\frac{\tau_j}{\sigma^2}\middle|\lambda^{(k-1)},y\right]$$

$$\frac{\partial R}{\partial\lambda_2}=-\frac{\rho\lambda_1^2}{8\lambda_2^2}E\left[\left\{\Gamma_U\left(\frac{1}{2},\frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\right\}^{-1}\varphi\left(\frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\frac{1}{\sigma^2}\middle|\lambda^{(k-1)},y\right]$$

$$-\frac{1}{2}\sum_{j=1}^{p}E\left[\frac{\tau_j}{\tau_j-1}\frac{\beta_j^2}{\sigma^2}\middle|\lambda^{(k-1)},y\right]+\frac{\lambda_1^2}{8\lambda_2^2}\sum_{j=1}^{p}E\left[\frac{\tau_j}{\sigma^2}\middle|\lambda^{(k-1)},y\right], \quad \ldots(3.15)$$

Where $\varphi(t)=t^{-\frac{1}{2}}e^{-t}$.

4-Real Data Analysis

We will examine the proposed model and compare it with other models. Real-life case have studied based on the blood viscosity syndrome disease data by considering the blood viscosity syndrome as response variable (**y**), and the explanatory variables (**X**) The data collected from pathological analyzes of patients visiting the ASC disease in the province of Babylon Centre . In addition to a set of questions posed by the researcher to affected Persons, this work was conducted on a sample that included (n=97) Person. The following data contains information that records visits of blood viscosity patients to Marjan Teaching Hospital in Babil Governorate Moreover , I used (97) models of different people , that is I took a simple the random sample , patients were drawn to study the factors affecting patients' blood viscosity (response variable), while the predictor variables are as follows :

y_i	Blood viscosity sy ndrome
X_1	Person gender
X_2	Age of person
X_3	Environment / elevated / flat
X_4	Occupation
X_5	Anemia
X_6	Temperature
X_7	Genetics factor
X_8	Person weight
X_9	Blood pressure
X_{10}	Mental state
X_{11}	Kidney disease
X_{12}	Drink water and fluids
X_{13}	Congenital heart defects
X_{14}	Decreased plasma levels in the blood
X_{15}	Lung disease
X_{16}	Dietary pattern \ fats
X_{17}	Drinking alcoholic beverages
X_{18}	Playing sports

$ISSN_{Online}$: 2312-9883
$ISSN_{Print}$: 1816-9171

X_{19}	Smoking
X_{20}	Medicines and drugs
X_{21}	Increasing the amount of proteins in the blood

The following table shows the values of the mean square error and the mean absolute criterions.

Table (1) Value of mean square error (MSE) and mean absolute error (MAE)

Methods	MSE	MAE
BANETR	19.077	9.236
ANETR	23.112	13.243

From table (1) the value of the quality model criterion (MSE) of the proposed method gives the less value comparing (MSE=19.077) with the MSE of the ENRM (MSE=23112). Also, the other quality criteria MAE of the proposed method gives the less value (MAE=9.236). Therefore, obviously t the proposed model works better than the other method.

Table (2) Parameter estimates of the predictor variables

Descriptive variables	**Variables**	$\hat{\beta}$ **BANETR**	$\hat{\beta}$**ANETR**
Person gender	x_1	1.466	2.875
Age of person	x_2	0.0044	0.000
Environment / elevated / flat	x_3	0.000	0.122
Occupation	x_4	0.287	0.000
Anemia	x_5	0.000	0.000
Temperature	x_6	0.370	0.000
Genetics factor	x_7	0.716	0.000
Person weight	x_8	0.000	0.000

ISSN$_{Online}$: 2312-9883
ISSN$_{Print}$: 1816-9171

Blood pressure	x_9	0.107	0.000
Mental state	x_{10}	0.000	-0.0506
Kidney disease	x_{11}	0.081	0.000
Drink water and fluids	x_{12}	0.189	0.000
Congenital heart defects	x_{13}	-0.020	0.000
Decreased plasma levels in the blood	x_{14}	-0.271	-0.224
Lung disease	x_{15}	0.364	0.000
Dietary pattern \ fats	x_{16}	0.000	-0.243
Drinking alcoholic beverages	x_{17}	0.650	0.000
Playing sports	x_{18}	0.000	0.000
Smoking	x_{19}	-0.182	0.000
Medicines and drugs	x_{20}	-0.112	0.000
Increasing the amount of proteins in the blood	x_{21}	0.000	0.000

From table (2) the proposed model removed the irrelevant predictor variable that does not influence the response variable, So we can say that the proposed model provide variable selection procedure. For example, the parameter estimate of the (age of person) variable takes zero value, and so on for the other variables (Age of person , Occupation , Anemia , Temperature , Genetics , factor , Person weight , Blood pressure , Kidney disease , Drink water and fluids , Congenital heart defects , Lung disease , Drinking alcoholic beverages , Playing sports , Smoking , Medicines and drugs , Increasing the amount of proteins in the blood). Eventually, the relevant predictor variables that effects the response variable (Blood viscosity) are (Person gender , Environment / elevated / flat , Mental state , Decreased plasma levels in the blood , Dietary pattern \ fats).

5- Conclusions

The proposed model introduces new scale mixture of normal distribution mixing with Rayleigh distribution as Laplace prior distribution. New Bayesian hierarchical model and new Gibbs sampler algorithm for elastic net linear regression have developed. The real data analysis examined by the proposed model and compared the results with some exist regularization methods. The proposed method clearly shows that the penalized proposed method outperformed the other method from the variable selection procedure point of view.

References

[1] James, G.,Witten, D., Hastie, T., and Tibshirani, R. (2013). An Introduction to Statistical Learning: With Applications in R. Springer Publishing Company, Incorporated, New York.

[2] Hoerl, A. E., and Kennard, R. W. (1970b). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55-67.

[3] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.

[4] Zou, H., & Hastie, T. (2005). Regression shrinkage and selection via the elastic net, with applications to microarrays. JR Stat Soc Ser B, 67.

[5] Osborne, M.R., Presnell,B. and Turlach,B.A.(2000a). On the LASSO and Its Dual. Journal of Computational and Graphical Statistics Vol. 9, No. 2 (Jun., 2000), pp. 319-337. Published By: Taylor & Francis, Ltd.

[6] Osborne, M.R., Presnell,B. and Turlach,B.A.(2000b).A new approach to variable selection in least squares problems. IMA Journal of Numerical Analysis, Volume 20, Issue 3, July 2000, Pages 389–403.

[7] Mallows, C.L. (1973). Some Comments on C_p. Technometrics, Volume 15, 1973 - Issue 4.

[8] Akaike, H. (1973), "Information Theory and an Extension of Maximum Likelihood Principle," in *Second International Symposium on Information Theory* (B. N. Petrov and F. Caski, Eds.) Akademia Kiado, Budapest, 267-28 1.

[9] Schwarz, G. (1978), "Estimating the Dimensions of a Model," Annals of Statistics, 121,461-464.

[10] Zou, H., Hastie, T., & Tibshirani, R. (2007). On the degrees of freedom of the lasso. The Annals of Statistics, 35, 2173-2192.

[11] Hocking, R.R. (1976) The Analysis and Selection of Variables in Linear Regression. Biometrics, 32, 1-50.

[12] Efroymson, M.A. (1960). "Multiple Regression Analysis," In: A. Ralston and H. S. Wilf, Eds., Mathematical Methods for Digital Computers, John Wiley, New York.

[13] George, E.I. and McCulloch, R.E..(1993).Variable Selection Via Gibbs Sampling. Journal of the American Statistical Association, Vol. 88, No. 423. (Sep., 1993), pp. 881-889.

[14] Hoerl, A. E., and Kennard, R. W. (1970a). Ridge regression: Applications to nonorthogonal problems. *Technometrics, 12*(1), 69-82.

[15] Draper, N. R., & Smith, H. (1998). Applied regression analysis (3rd ed.). New York: John Wiley & Sons.

[16] Hans, C. (2009). Bayesian lasso regression. Biometrika, 96, 835-845.

[17] Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 49-67.

[18] Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. The Annals of statistics, 32(2), 407-499.

[19] Kyung, G., Ghosh, M., Gill, J., &, Casella M. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian analysis*, *5*(2), 369-411.

[20] Ghosh, S. (2007). Adaptive elastic net: An improvement of elastic net to achieve oracle properties. Preprint, 1.

[21] Zou, H., & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. Annals of statistics, 37(4), 1733.

[22] Celeux, G., Anbari, M. E., Marin, J.-M., & Robert, C. P. (2012). Regularization in regression: Comparing Bayesian and frequentist methods in a poorly informative situation. Bayesian Analysis, 7(2), 477-502.

[17] Hans, C. (2011). Elastic net regression modeling with the orthant normal prior. Journal of the American Statistical Association, 106(496), 1383-1393.

[18] Li, Q., & Lin, N. (2010). The Bayesian elastic net. Bayesian analysis, 5(1), 151-170.

[19] Park, T., & Casella, G. (2008). The bayesian lasso. Journal of the American Statistical Association, 103(482), 681-686. Choi, W. W., Weisenburger, D. D. Greiner, T. C., Piris, M.

[20] Alhamzawi, R., & Ali, H. T. M. (2018). The Bayesian elastic net regression. Communications in Statistics-Simulation and Computation, 47(4), 1168-1178.

[21] Mallick, H., & Yi, N. (2014). A new Bayesian lasso. Statistics and its interface, 7(4), 571-582.

[22]Alhamzawi, R. (2016). Bayesian elastic net Tobit quantile regression. Communications in Statistics-Simulation and Computation, 45(7), 2409-2427.

[23] Flaih, A.N, Alshaybawee, and Al husseini F.H, (2020). sparsity via new Bayesian Lasso . Periodicals of Engineering and Natural Sciences , vol.8 , is suc 1, 345-359.

[24] Flaih, A. N., Al-Saadony, M. F., & Elsalloukh, H. (2020). New Scale Mixture for Bayesian Adaptive Lasso Tobit Regression. Al-Rafidain University College For Sciences, (46).

[25] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B, 67, 301 320.

[26] Chhikara, R. and Folks, L. (1988). The Inverse Gaussian Distribution Theory: Methodology, and Applications. Marcel Dekker, Inc, 1st edition.