Republic of Iraq
Ministry of Higher Education
and Scientific Research
University of Al-Qadisiyah
Faculty of Management
and Economics
Department of Statistics

Bayesian Elastic Net for Censored Normal Regression with Application

A thesis submitted to the Council of college of Administration & Economics\ University of Al-Qadisiyah ln partial fulfillment of the Requirement for the Degree of Master of Science in Statistics

By

Mohammed Rasool Mohsin ALsafi

supervisor

Prof. Dr. Ahmad Naeem Flaih

A.H. 1443 A.D. 2021

بِسْمِ اللهِ الرَّحْمنِ الرَّحِيم

﴿يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ ۚ وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ﴾

صَدَقَ اللهُ العَلِيُّ العَظيمِ

﴿المجادلة 11﴾

Supervisor's recommendation

I certify that the thesis entitled (**Bayesian Elastic Net for Censored Normal Regression with Application**) has been under my supervision for the student (**Mohammed Rasool Mohsin**) in the Department of Statistics / College of Administration & Economisc / University of Al-Qadisiyah as it is part of the requirements for master's degree in statistics sciences.

Signature:

Supervisor: Prof . Dr. Ahmad Naeem Flaih

Date:

Recommendation of the head of the Graduate Studies Committee

Based on the available recommendation , I would like to forward this thesis for discussion.

Signature:

Name: Assist. Prof. Dr. Muhannad F.Al-Saadony

Chairman of the Higher Studies Committee in department of statistics

Approval of the linguistics expert

This is to hereby that the entitled (**Bayesian Elastic Net for Censored Normal Regression with Application**) has been reviewed in terms of stylistics and linguistics (grammar and spelling). Therefore, after the modification of all recommended notes thesis has become free of all linguistics errors and ready to be defended and used as a scientific method to award the degree of master in statistical sciences.

Signature:

Name: Dr. Haider Hadi Mohammed

Date: 12. 10. 2021

(The Decision of the Committee)

University of Al-Qadisiyah

College of Administration and Economics

Statistics department

We are the head and members of the defense committee certify that we have been looked at the thesis entitled (**Bayesian Elastic Net for Censored Normal Regression with Application**) and we have debated the student (**Mohammed Rasool Mohsin**). As a result , the student has defended her thesis and all its content. So that we have found the thesis is worthy to be accepted to award a (excellence) master's degree in statistics science..

N.	Name	Scientific Title	signature	Capacity
١	Muhannad F.Al-Saadony	Assist. Prof. Dr.		The head of the committee
٢	Hana Saad Shebib.	Assist. Prof. Dr.		Member
٣	Hassan Sami Uraibi	Assist. Prof. Dr.		Member
٤	Ahmad Naeem Flaih	Prof. Dr.	A. N. F	Member and supervisor

Approval of the College Committee

The council of the college of administration and economics at the University of Al-Qadisiyah have been approved on the decision of the defense committee.

Prof. Dr. Sawsan Kareem Hodan

The Dean of the college of administration and economics/Al-Qadisiyah university

ACKNOWLEDMENTS

I thank Allah, and may His blessings and peace be upon the Prophet Muhammad, may Allah bless him and grant him peace

I extend my sincere thanks and gratitude to the supervisors Prof. Dr. Ahmed N Flaih for his absolute confidence and cooperation in solving the obstacles that I faced throughout the research period and for his valuable advice.

I would also like to thank all my professors in the Statistics Department at Al-Qadisiyah University for what we have come to, as well as my professors in the discussion committee for their guidance and valuable comments.

I also thank my dear and devoted mother, my dear wife, my children, and all my friends and loved ones.

Mohammed

DEDICATION

I dedicate the fruit of my humble effort to the soul of my martyr father (Rasool Mohsin ALsafi)

Abstract

The bayesian theory has great importance in most science fields. using the bayesian methods and procedures in the statistical tools brings more reliable results since the bayesian method are very flexible and can be computed very easily with the latest developments of computer science. Building a new bayesian regression model depends on its efficiency and how faster the MCMC algorithm implements it. The faster implemented algorithm is the best one. In this thesis, employed the scale mixture of laplace prior distribution .In the tobit regression model new regularization method of the elastic net has been developed. The new hierarchical bayesian model also proposed. Then, the new Gibbs's sample algorithm was implemented. Regression model analysis has the greatest importance in all science fields, especially in statistics theory, where creating a more flexible regression model that provides more interpretable and reliable estimates for the parameters has huge attention for the statistics reaches. Many types of regression model have been developed for asking the best model that fit the data. Also, the proposed regularization method, the elastic net in tobit regression, has been used for the variable selection procedure. Conducted three simulations; scenarios to study the behavior of the posterior distribution through the estimates of the parameters, through the Median of Mean Absolute Deviation (MMAD), and the Standard Deviation (SD) criteria. MMAD and SD results show the comparative feature of my proposed model with some existing regression methods. The Gibbs sample algorithm gives stationary estimates of the parameters for the posterior distribution. Also, employed the proposed model in real data analysis. The variable selection procedure is very clear and set some predicted variable to zero. The proposed model is a very comparative model with the other regression models.

Table of Contents

Content tables

Content figures

Chapter 1

Introduction and Literature Reviews

1.1.Introduction

Statistical regression models are the most widespread tools in many fields of sciences, such as biological, chemical, ecology, physical, social sciences, and economics. These models are very useful to form the function of the dependent (response) variable with one or more independent (explanatory) variable (s). Regression models are used for planning future strategies as a predicted model. So, the regression analysis aims to select the best regression model that interpreted the functional form between the variables and produced the most interpretable model that included the most relevant predictor variables on the response variable, and then uses this model for prediction. Model selection is the key idea behind the regression analysis through applying the variable selection procedure.

The OLS method gives the best unbiased linear estimates when its hypothesis are fulfilled. One of these hypothesis is that the explanatory variables are independent, and when this hypothesis is violated, OLS cannot be used, but alternative methods are used. The variables selection works to reduce the number of explanatory variables to the least possible so that the model is interpretable and has the ability to predictor. When there is a large number of explanatory variables, this problem may occur overfit the problem of the many predictor variables and the problem of multicollinearity motivated the researcher to find the solution for the variable selection methods. In the case of many predictor variables. When ($n > P$) the OLS method gives an optimal solution but when ($p > n$) the OLS will be multiple solutions, so other alternative methods are used but when ($p > n$) there will be a problem multicollinearity, the OLS estimates are meaningful because of the variability of the estimates that lead to biased and high variances estimators. To overcome these problems, the ridge method has been developed to deal with this circumventer. The ridge method can not sparse solution, therefore the lasso method has been developed to produce a sparse solution with biased small variance. Because of some drawbacks on the lasso, the elastic net method that combined ridge and lasso has been developed to produce a sparse solution that copes with the effect of pairwise correlation between predictor variables in the group base of different variables.

1.2. Thesis problem

After study and analyze the literature review about the regularization different methods, the work of Li and Lin (2010) motivate the new idea in this thesis to proposed the new method for estimating the censored normal regression model parameters by employing the prior distribution that proposed by Li and Lin (2010). In this thesis we investigate a special form of regression that is called the bayesian elastic net censored normal regression in presence of the simultaneous procedure (variable selection and shrinkage) of the elastic net regularization method which is can select groups that have correlated variable.

1.3. Thesis objectives

In this thesis, there is one idea and two comparative studies which are the following .

1- To proposed new regularization method for estimating bayesian elastic net censored normal regression by developed new hierarchal model.

2- To perform the comparative study between the new proposed method that proposed is the first objective of this thesis.

3- To perform study between the bayesian elastic net variable selection procedure of censored normal regression model that proposed is the first objective of this thesis and some exists models.

Chapter 2

Some Basic concepts

2.1 Introduction

Regression in general is a formula or method for analyzing the relationship between two or more variables and this relationship can be expressed in an equation that contains one variable known as the dependent (response) variable with one or more explanatory (predictor) variables. This equation can be used for the purpose of estimating and selecting the best model in terms of variable selection and predictions. The linear regression model is defined as follow:

$$Y = X\beta + e, \qquad (2.1)$$

Where Y is an $n\ x\ 1$ vector of dependent variable,

X is an $n\ x\ p$ matrix of explanatory variables,

β is an $p\ x\ 1$ vector of parameter of regression coefficients,

e is an $n\ x\ 1$ vector of random errors, $e_i \sim N(0, \sigma^2)$

p is the number of predictors,

n is the number of observations .

The popular least squares method in general gives the best linear unbiased estimate BLUE with the least variation of the regression model parameters. But there are problems that appear when using this method. For example, when one of the assumptions of the analysis is violated, including the lack of a complete or partial linear correlation between two or more explanatory variables that may lead to the problem called ‘multicollinearity’, that causes inaccurate estimates of parameters that are given and with large variations (Hoerl & Kennard,1970).

Ridge regression is a specialized technique for analyzing multiple regression data that suffers from multicollinearity problem. This method has shown that the activity overcomes the problem of linear regression. The least squares method can be written in the following form, where ridge method does not give sparse solution, i.e., it does not set any parameter estimate to zero, and then it cannot do the variable selection procedure (Kannard & Baldwin,1975).

$$\hat{\beta}_{ridge} = RSS\ (\beta)\ +\ \lambda \sum_{j=1}^{P} \beta_{j}^{2} \qquad (2.2)$$

Tibshirani,(1996) introduced the lasso method, where he invented this method and provided many details about the mechanism of its operation and performance. The term lasso represents the first letter of the concept of Least Absolute Shrinkage and Selection Operation. A penalty function for the linear regression model is a method for estimating the parameters of the regression model as well as for selecting and the organization of the variables included in the model to increase the explanatory accuracy of the regression models. These models are used in the analysis of the phenomenon under studying the convenience of the model to choose a subset of the common variables in the final model instead of using them all.

The sum of squares of random errors is minimized with the sum of the absolute values of the regression model coefficients. Lasso was originally designed for the Least Squares Models, where lasso reveals a large amount of estimated behavior by the lasso coefficient, or the so-called soft thresholding. This includes the relationship of the lasso estimator with the estimator of the letter regression. The sum squares of the residues according to a constraint represents the absolute sum of the coefficients the lasso method is given by:

$$\hat{\beta}_{lasso} = RSS\ (\beta) + \lambda \sum_{J=1}^{P} |\ \beta_j|, \qquad (2.3)$$

The elastic net regression is a systematic regression method that linearly combines the penalties L_1 , L_2 for the ridge method. The lasso was introduced by Zou and Hastie,(2005). The elastic net overcomes the Lasso drawbacks because it uses the two penalty functions. We can work with the elastic net when there are many correlated predictor variables (Li and Lin,2010).

The elastic net estimator is defined by:

$$\hat{\beta} = \text{argmin}_{\beta}||Y - X\beta||^2 + \lambda_2||\beta||_2^2 + \lambda_1||\beta|| \qquad (2.4)$$

Where λ_1,$\lambda_2 \geq 0$ are the shrinkage parameters.

2.2 Regularization combined penalty functions

Kirkland,(2014) states that the ridge regression model gives more reliable estimates than the Lasso regression model does when there are groups of some types of predictors with high correlation. Also, the Lasso usually selects randomly from the predictors in a group and ignore the rest. From the prediction point of view, this procedure is inconsistent with the purpose of building the regression analysis. Because of the drawback (Zou and Hastie ,2005) studied the effect of groups of predictors and proposed to combine the Lasso with the ridge penalty function to propose a model that can be named the Elastic Net. The elastic net regression model supports the variable selection and takes the minimum of (n, p), but the elastic net model takes the predictor variables (p).

2.3 Censored Normal Model (Tobit Model)

Censoring happens when the dependent (response) variable operations are limited. In this thesis, we will talk about the regression model when the response variable is limited to the left. Therefore, the censored normal model or tobit model assumes the following regression model:

$$y_i^{*}=x_i\beta+e_i \qquad (2.5)$$

where the error is $e_i \sim N(\,0,\, \sigma^2)$, here y^{*} is the observed (latent) variable which is observed for the values higher than C and censored less than C. Then, the values of y (observations) are defined as follows: (Alhusseini ,2020)

$$y_i^{*} = \begin{cases} \mathrm{y}^{*} & \text{if} \quad \mathrm{y}^{*} > \mathrm{c} \\ \mathrm{C_y} & \text{if} \quad \mathrm{y}^{*} \leq \mathrm{c} \end{cases} \qquad (2.6)$$

Now, we can say that the sample of $y_1,\ldots,y_n$ is a censored. The standant tobit regression model is C=0 in the equation (2.6).

$$y_i^{*} = \begin{cases} \mathrm{y}^{*} & \text{if} \quad \mathrm{y}^{*} > 0 \\ 0 & \text{if} \quad \mathrm{y}^{*} \leq 0 \end{cases} \qquad (2.7)$$

2.4 Ridge Regression model

Lately, the problem of ($\rho \geq n$) and the variable selector procedure have great importance in most of science fields. The researchers always look for fitting the estimated model that coincide with ($\rho \geq n$), such as the ridge regression which deals with the problem of singularity of $(X^T X)^{-1}$ problem. The ridge regression puts a constrain on the resided sum of squares (RSS) and minimizes the problem (James et al ,2013).

$$g(\beta, \lambda)=\text{RSS}(\beta) + \lambda \sum_{j=1}^{k} \beta_j^2 \qquad (2.8)$$

Where the regularization parameter $\boldsymbol{\lambda} \geq 0$ that controls the amount of shrinkage composed on β. As the parameter $\boldsymbol{\lambda}$ becomes bigger the parameter β becomes smaller. Suppose that $\boldsymbol{\lambda} = 0$, the problem (2.8) becomes the OLS problem.

It is very important to standardized the predictor variable so that $\frac{1}{n}\sum_{i=1}^{n} x_{ij} = 0$ and $\frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 = 1$, for j =1,2,…,k. Also , it is very convenience to centered the values of response variable such that $\frac{1}{n}\sum_{i=1}^{n} y_i = 0$.So the penalty function in (2.8) is applied for the $\beta_1, \beta_2, \dots, \beta_k$, but not for the intercept B_0.We can rewrite the problem (2.8) in terms of norm notation vector as follows:

$$\hat{\beta}_{ridge} = \underset{\beta}{argmin}\, ||y_i - x^T\beta||^2 + \boldsymbol{\lambda}\, ||\beta||_2^2$$

Consequently, the ridge estimator ($\hat{\beta}_{ridge}$) can be estimated as follow:

$$\hat{\beta}_{ridge} = (x^T x + \boldsymbol{\lambda}\, I_k)^{-1} x^T y \qquad (2.9)$$

For (2.9) it is very clear for $\boldsymbol{\lambda} \to \infty$ then β to get near zero (Hoerl and Kennard, 1970).

2.5.Lasso Regression model

The lasso is an abbreviation for “least absolute shrinkage and selection factor” Tibshirani,(1996). This abbreviation comes from its careers that it does not only contract coefficients to zero, but it provides a selection of the significant covariates as well.

The lasso estimator is defined as:

$$\hat{\beta}_{Lasso} = \text{argmin } ||Y - X\beta||_2^2 \text{ , subject to } ||\beta|| \leq t \quad (2.10)$$

where what t > 0 is a selection tuning parameter. We acquired the constrained minimization problem in status of the ridge regression. The lasso estimator can be rewritten to an unconstrained decreasing problem:

$$\hat{\beta}_{\text{Lasso}} = \underset{||\beta||_1 \leq t}{\text{argmin}} ||\text{Y} - \text{X}\beta||_2^2 + \lambda||\beta|| \quad , \quad (2.11)$$

The regularization parameter $\boldsymbol{\lambda} \geq 0$ plays the roll of controlling the amonnt of shrinkage in (β_j). It $\boldsymbol{\lambda}$ decides if β is equal to Zero or not. When $\boldsymbol{\lambda}$ getting bigger the $L_1 - norm$ gets smaller, which leads to the variable selection procedure. As in the ridge regression, standardized x_i values and centered y_i values are very important.

2.5.1 Properties of Lasso Regression

1- When $\boldsymbol{\lambda} \to \infty$, then bias $(\hat{\boldsymbol{\beta}}_j) \to \infty$.

2- When $\boldsymbol{\lambda} \to \infty$, then v $(\hat{\boldsymbol{\beta}}_j)$ get smalls.

3- The mean square error criterion get smaller, but get bigger as $\boldsymbol{\lambda} \to \infty$ (Hastie et al ,2015) .

2.6 Classical Elastic net

Zou and Hastie ,(2005) defined new regularization method, which is called the elastic net penalty method, as the mixture of lasso and ridge penalty function and formulated the classical elastic net estimator as follows:

$$\hat{\beta}_{en} = (1 + \lambda_2) \text{ argmin } ||y - x\beta||_2^2 + \lambda_1 ||\beta|| + \lambda_2 ||\beta||_2^2 \qquad (2.12)$$

Where $\lambda_1, \lambda_2 > 0$ are the penalty shrinking parameters, $||\beta||_1$ is L_1- norm of the parameter β , and $||\beta||_2^2$ is L_2- norm of the parameter β. when λ_2= 0 the elastic net becomes the lasso method, and when λ_1= 0 the elastic net becomes the ridge method.

They also showed that the elastic net can automatically works as variable selection procedure and shrinkage simultaneously. Also, the classical elastic net can deals with difficulty of high dimensional data (p > n).

They pointed out that the classical elastic net has an outperformance more than the classical lasso. They also explained that the classical elastic net produces regression model with relevant predictor variables to the response variable and removes the irrelevant predictor variable that does not affects the response variable. The penalty functions parameters ($\lambda_1\lambda_2$) control the amount of shrinkage for the regression coefficients. If λ_1 and λ_2 goes to be very small. Then no sparsely will be performed. Also if $\lambda_1$and λ_2 goes to be very high, then all the predictor variables coefficients will be shrunk towards zero. LABS-EN algorithm proposed by Zon and Hastie,(2005) found the solution of the classical elastic net (Efron et al, 2004).

2.7Literature Reviews

Hoerl-Kennard (1970) proposed a ridge regression method that estimates the coefficients of a multiple regression model based on adding a small positive amount to the diameter of the coefficient of (x´x). Studies reached that when this positive amount is added it increases the possibility of making the data perpendicular; hence obtaining better estimates of the regression model coefficients.

Tibshirani (1996) proposed a new method for estimating parameters of linear models called Lasso, which is the least absolute shrinkage and selection operator, that reduces the sum of squares of residuals subject to the sum of the absolute value of the coefficients. It tends to produce some coefficients that are equal to zero. Tibshirani concluded that lasso has better properties than the sequential step method and the ridge regression method.

Al-Sadoun(2005) performed the coefficients of the multiple linear regression model using the regular ridge regression method and the bayes ridge regression method. Then he compared the two estimators using simulation and concluded that the Bayes regression method is better than the regular ridge.

Zou & Hastie (2005) suggested the elastic net, a new regularization and variable selection method. Reality world data and a simulation study show that the elastic net often surpasses the lasso, whilst having a similar sparse representation. Moreover, the elastic net encourages a grouping effect, where strong correlated predictors tend to be in or out of the model together. The elastic net in particular is useful when the number of predictors (p) is much bigger than the number of observations (n). By contrast, the lasso is not a very favorable variable selection method in the $p \gg n$ case. An algorithm called LARS-EN This function estimates the least angle regression path of solution for l_1-penalized (lasso) logistic regression and the Cox proportional hazards model, is a suggestion for computing elastic net regularization paths efficiently; much like what the algorithm LARS does for the lasso.

Zou et al (2007) study the effective degrees of freedom of the lasso in the framework of Stein's unprejudiced risk estimation (SURE). We show that the number of nonzero coefficients is an unprejudiced estimate for the degrees of freedom of the lasso — a conclusion that requires an assumption on the predictors. Moreover, the unprejudiced estimator is shown to be asymptotically consistent.

Hans (2009) proposed bayesian lasso regression and a new Gibbs sampling for Bayesian lasso regression. He imposes directly exponential doubles before the lasso regression coefficients and a gamma before the shrinkage parameter confirmation was placed on point estimation using the posterior mean, which facilitates prediction for future observations via the posterior predictive distribution. The average test errors were account for to measure the predictive performance. A comparison study by Hans (2009) showed that the standard lasso prediction method does not essential agree with model-based bayesian predictions.

Li & Lin (2010) proposed a bayesian method to solve the elastic net model using a Gibbs sampler. The bayesian elastic net has two major advantages. Firstly, as a bayesian method, the distributional results on the estimates are straightforward, Secondly, it chooses the two penalty parameters simultaneously, avoiding the double shrinkage problem in the elastic net method. Real data examples and simulation studies show that the bayesian elastic net performs well compared to other methods.

Hans (2010) showed that the elastic net proceedings are a form of regularized optimization for linear regression that gives a bridge between ridge regression and the lasso. The estimate that it produces can be viewed as a bayesian posterior mode under a prior distribution implied by the form of the elastic net penalty. This article broadens the scope of the bayesian connection by providing a complete characterization of a class of prior distributions that obstetrics the elastic net estimation as the posterior.

Ji et al. (2012) studied the model selection procedure for the binary and tobit quantile regression models using a new hierarchical model. The New Gibbs sampler algorithm has been developed by using new location-scale mixture formula of the skewed laplace distribution .The proposed method is illustrated in both simulation and real data analysis. The results outperform the proposed method.

Alhamzawi (2013) introduced the tobit quantile regression model using the adaptive lasso penalty function. The new hierarchical model and new Gibbs sample algorithm have been developed by employing the location-scale mixture of normal as a formula for the skewed laplace prior distribution. The proposed model performs well comparing with other regularization methods.

Alhamzawi (2014) introduced the bayesian tobit quantile regression model by employing the g-prior density, additionally, to using the ridge parameter. In this paper, adding ridge parameter was to deal with some challenges that come with censored data, like collinearity between the covariates. This work also deals with the variable selection procedure basel on the g-prior. The results of simulation and real data analysis illustrated the outperformance of the proposed model.

Alhamzawi (2016) introduced the bayesian elastic net for the tobit quantile regression model. The new regularization method deals with the variable selection procedure and parameters estimation for the Tobit quantile model by using the elastic net penalty function by employing the gamma priors. In this work, Alhamzawi treated the hyper-parameters of the proposed gamma priors. The results of simulation and real data analysis were comparable with some exists methods.

Fonti (2017) explained the lasso method of the selection feature, which is a feature that selects fewer independent explanatory variables to describe the response variable and made the model easy when interpreting. Also, Fonti applied the lasso method to linear models and generalized linear models when the number of variables is greater than the number of observations.

Odah, Bager, Bahr. K. (2017) found that it is so far often in economic data to find variables describing specific phenomena which are censoring from the right side or left side. When the data has to be censored from the left side at a censorship point equal to zero, the tobit regression model represents the most appropriate model to use. In this paper, they studied bank loans value, which is one of the basic banking services submitted by banks in any country.

Al-Husseini,F.H.H. (2017) introduced a new hierarchal model for the tobit regression by using the lasso penalty function. In this work the scale mixture for uniforms mixing with a special case of gamma distribution as a representation of the laplace prior distribution employed for development, where the New Gibbs samples algorithm, the parameter estimation, and variable selection were performed. Simulation examples and real data analysis have shown that the proposed model performs well comparing with some other methods.

Alhamzawi and Haithem (2018) suggested a new bayesian elastic net (EN) approach for variable selection and coefficient estimation in tobit regression. Mostly, we present a new hierarchical formularization of the bayesian EN by utilizing the scale mixture of truncated normal distribution (with exponential mixing distributions) of the laplace density part. The Proposal method is an alternate method to the bayesian method of the EN problem. It is shown up that the model performs well in comparison with the old elastic net representation.

Al-Husseini,F.H.H. (2018) introduced the composite tobit quantile regression model from the bayesian point of view. In this work, the MCMC algorithm has been developed by employing a scale mixture for the skewed Laplace prior distribution as a formula of normal mixing with exponential distribution. The results of simulation scenarios and real data analysis illustrate that the proposed method that combines the information of the covariates for the different quantiles outperforms the other methods.

Hilali,H.K.A.(2019) introduced a new regularization method by using transformation for the scale mixture of laplace prior distribution that proposed by Ma lick and yi (2014). Also, the new Gibbs sampling algorithm was proposed for the Bayesian adaptive lasso tobit regression. The results of the simulation and the practical side were very promising.

Chapter 3

Bayesian Elastic Net Tobit Regression

3.1. Bayesian Elastic Net Tobit Regression

The elastic net (EN) penalty method combined the ridge and the lasso regularization methods in the minimization problem of the residual sum of squares. EN is another variable selection method that works in the cases where the lasso had some limitations, such as: p>n (grouped predictor variables with high pairwise correlations) and n>p (with high correlation between predictor variables).(Zou and Hastie,2005) introduced the elastic net method to solve the limitations of lasso method that was proposed by Tibshirani (1996). The higher correlations among the predictor variables, the more significant the elastic net method will improve the prediction accuracy of the lasso method. The elastic net method provides good performance, but elastic net method does not have the oracle properties (consistent for sparsely and asymptotic normality for parameters) See:(Jiratchayut,2014) and (Kirkland, 2014), for further explanation. (Park and Casella,2008) proposed the bayesian lasso penalized method which considers that the prior distribution of the linear regression coefficient Q as scale mixture of normal mixing with exponential distribution. (Mallick and Yi,2014) proposed Bayesian lasso method under new scale mixture for the prior Q as uniform mixing with particular gamma (2, λ). (Li and Lin,2010) introduced the bayesian elastic net method with new formulation of the prior distribution of Q as scale mixture of normal mixing with truncated gamma distribution. (Alhamzawi,2014) presented bayesian inference for the elastic net tobit quantile regression and proposed new hierarchical model.

By motivation of (Li and Lin,2010) and (Alhamzawi,2014) works, investigated the bayesian elastic net tobit regression (BENTR) model through employing new hierarchical model for the bayesian elastic net tobit model, and proposed new Gibbs sampling algorithm for BENTR. The classical elastic net estimator is defined as:

$$\hat{\beta}_{EN} = \underbrace{argmin}_{\beta} (y - x_i'\beta)^T (y - x_i'\beta) + g(\beta), \qquad (3.13)$$

Where $g(\beta)$ is the penalty function defined by $g(\beta) = \lambda_1 \|\beta\|_1 + \lambda_2 ||\beta||_2^2$, $\lambda_1 \geq 0$, and $\lambda_2 \geq 0$ is the shrinkage parameters guarantees the strictly convex. The ridge penalty can be obtained if $\lambda_1 = 0$ and the lasso penalty if $\lambda_2 = 0$. The parameters λ_1 and λ_2 controls the amount of shrinkage and selection and the amount of grouping, respectively. $\mathbf{y}$= $(\mathbf{y_1},\ldots, \mathbf{y_n})'$ is the centered response variable such that: $\frac{1}{n}\sum_{i=1}^{n} y_i = 0$, and $\mathbf{x_i}$=$(\mathbf{x_1}\,,\ \ldots\ ,\ \mathbf{x_p})$ are the standardized predictor variables to be with mean =0 and variance =1 ($\frac{1}{n}\sum_{i=1}^{n} x_{ij} = 0$, $\frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 = 1$) for j=1,2,..,p.

We proposed a BENTR analysis in this thesis for the parameters' estimation and the variable selection procedure. We employed the prior distribution of $\boldsymbol{\pi(\beta|\sigma^2, \lambda_1, \lambda_2)}$, which is defined by:

$$\pi(\beta|\sigma^2, \lambda_1, \lambda_2) \propto exp\{-\frac{1}{2\sigma^2}(\lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p} {\beta_j}^2)\} \qquad (3.14)$$

Then, in general the posterior marginal distribution of the parameter $\boldsymbol{\beta}$ of the tobit regression model, conditioning on latent variable $\boldsymbol{y}^*$, is:

$$p(\beta|y^*) \propto exp\{-\frac{1}{2\sigma^2}(y_i - max(y^0, y^*) - \frac{1}{2\sigma^2}(\lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p} {\beta_j}^2)\} \tag{3.15}$$

Where y^0 is a censoring point. We exploits the above formulas (3.14 and 3.15) to setup the bayesian elastic net tobit regression through the following general posterior marginal density of $\boldsymbol{\beta}$:

$$f\,(\beta/y) = \int_0^\infty h\,(2\pi\sigma^2)^{-n/2}\; exp\left\{-\frac{1}{2\sigma^2}\Big((y_i - max(y^0, y^*)) - \frac{1}{2\sigma^2}(\lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p} {\beta_j}^2\Big)\right\} \boldsymbol{\pi}\sigma^2\, d\sigma^2 \tag{3.16}$$

Where **h** is the normalizing constant of $\boldsymbol{\lambda_1}$ **,** $\boldsymbol{\lambda_2}$ **,** and $\boldsymbol{\sigma^2}$.

3.2. Hierarchical Model and Prior Distributions for BENTR

By using the structure equation of tobit regression (2.6) and the prior proposed by (Li and Lin,2010), we get:

$$Y|X,\beta \sim N(X\beta, \sigma^2 \mathrm{I_n})$$

$$\pi(\beta) \propto exp\{- \lambda_1||\beta|| - \lambda_2||\beta||_2^2\}$$

$$f(\beta|\sigma^2, y)\,) \propto exp\{-\frac{1}{2\sigma^2}(y - x_i'\beta)^T \left(y - x_i'\beta\right) - \lambda_1||\beta|| - \lambda_2||\beta||_2^2\},$$

$$= exp\{-\frac{1}{2\sigma^2}(y - x_i'\beta)^T \left(y - x_i'\beta\right) + (2\sigma^2\,\lambda_1)||\beta|| + (2\sigma^2\lambda_2)||\beta||_2^2\}$$

$$exp\{-\frac{1}{2\sigma^2}(\lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p} {\beta_j}^2)\} \propto$$

$$K \prod_{j=1}^{p} \int_{1}^{\infty} \sqrt{c} \exp \{- \frac{\beta_j^2}{2} (\frac{\lambda_2}{\sigma^2} c) \} w^{\frac{-1}{2}} \exp(- \frac{1}{2\sigma^2} \frac{\lambda_1^2}{4\lambda_2} w) \, dw \qquad (3.17)$$

where $\boldsymbol{K}$ is the normalizing constant and $\boldsymbol{c} = \frac{\boldsymbol{w}}{\boldsymbol{w-1}}$. The prior formula (3.17) represent the scale mixture of normal mixing with truncated gamma.

Suppose that $y^0 = 0$, then we list the following proposed hierarchical model for the Bayesian elastic net regression model:

$$y_i = \begin{cases} y_i^* & if \quad y_i^* > 0 \\ 0 & if \quad y_i^* \leq 0 \end{cases},$$

$$y_i^* = x_i^T \beta + \epsilon_i,$$

$$y_i^* | \beta, \sigma^2 \sim \text{N}(x_i^T \beta, \sigma^2 I_n),$$

$$\beta_j | w_j, \sigma^2 \sim \prod_{j=1}^{p} N(\text{mean=0, var} = (\frac{\lambda_2}{\sigma^2} \frac{w_j}{w_j - 1})^{-1}),$$

$$w_j | \sigma^2 \sim \prod_{j=1}^{p} TG\ (mean = \frac{1}{2}, var = \frac{8\lambda_2 \sigma^2}{\lambda_1^2}),$$

$$\sigma^2 \sim \frac{1}{\sigma^2}. \qquad (3.18)$$

Where TG is the truncated gamma supported on $(1, \infty)$.Our contribution is to employ the hierarchy model (3.18) to develop new bayesian computation for the elastic net tobit regression.

3.3. Conditional Posterior Distributions.

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

$$y_i^* / \beta, \sigma^2 \sim \text{N}(x_i' \beta, \sigma^2 I_n),$$

Where i = 1, 2, …, n.

Following (Alhamzawi,2014) and (Li and Lin,2010) and conditioning on $\boldsymbol{y}^*$, $\boldsymbol{\beta}$ the posterior distribution of $\boldsymbol{\beta}$ is:

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

$$\propto \exp\{-\frac{1}{2\sigma^2}(y^* - x'\beta)'(y^* - x'\beta)\} \exp\{-\frac{1}{2\sigma^2}\beta' Q_\gamma \beta)$$

$$\text{Where Q = diag } (\frac{\gamma_1}{\gamma_1 - 1}, \dots, \frac{\gamma_p}{\gamma_p - 1})$$

$$= -\frac{1}{2\sigma^2}[\beta'(x'\mathrm{x})\beta - 2y^*\mathrm{x}\beta + y'^*y^* + \beta' Q_\gamma \beta]$$

$$= -\frac{1}{2\sigma^2}[\beta'(x'\mathrm{x} - Q_\gamma)\beta - 2y^*\mathrm{x}\beta + {y^*}'y^*]$$

$$\text{Let s} = x'\mathrm{x} + \lambda_2 Q_\gamma \text{, then}$$

$$= -\frac{1}{2\sigma^2}[\beta' s\beta - 2y^*\mathrm{x}\beta + {y^*}'y^*]$$

$$= -\frac{1}{2\sigma^2}(\beta - s^{-1}x'y^*)'\mathrm{c}(\beta - s^{-1}x'y^*) \quad (3.19)$$

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

$$\beta/y^*, \sigma^2, \gamma \sim \mathrm{N}(s^{-1}x'y^*, \sigma^2 s^{-1}) \quad (3.20)$$

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

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

$$\propto (\sigma^2)^{\frac{-n}{2} - p - 1}\{\Gamma z(\frac{1}{2}, \frac{\lambda_1^2}{8\sigma^2\lambda_2})\}^{-p} \exp[-\frac{1}{2\sigma^2}\{(y^* - x'\beta)'(y^* - x'\beta) +$$

$$\lambda_2 \sum_{j=1}^{p} \frac{\gamma_j}{\gamma_j - 1}\beta_j^2 + \frac{\lambda_1^2}{4\lambda_2}\sum_{j=1}^{p}\gamma_j\}], \quad (3.21)$$

Where Γz ($\boldsymbol{\alpha}$, x) = $\int_x^\infty \boldsymbol{t^{a-1}}\ \boldsymbol{e^{-t}}\ \boldsymbol{dt}$ is the upper incomplete gamma function. see (Armido and Alfred,1986). for more details, and $\boldsymbol{1_P}$ is the vector of p-dimensional of $1's$.

The third variable is $(\boldsymbol{\gamma} - \mathbf{1}_P)$, where the full conditional distribution is:

$$(\gamma - 1_P)/y^*, \sigma^2, \beta \sim \prod_{j=1}^{p} GIG\left(\lambda = \frac{1}{2}, \varphi = \frac{\lambda_1}{4\lambda_2\sigma^2}, \chi = \frac{\lambda_2\beta_j^2}{\sigma^2}\right), \qquad (3.22)$$

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

$$f(x/\lambda, \varphi, \chi) = \frac{(\varphi/\chi)^{\lambda/2}}{2k_\lambda(\sqrt{\varphi\chi})} x^{\lambda-1} \exp\{-\frac{1}{2}(\chi x^{-1} + \varphi \boldsymbol{x})\}, \qquad (3.23)$$

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

So, we can easily say that

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

With the following pdf,

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

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

3.4. Using the empirical Bayes for choosing $\boldsymbol{\lambda_1}$ and $\boldsymbol{\lambda_2}$.

(Park and Casella,2008) and (Casella,2001), suggested that 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), they treated $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\sigma^2}$ as missing data and $(\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2})$ as fixed parameters. Hence, the likelihood is:

$$\lambda_1^P(\frac{1}{\sigma^2})^{\frac{n}{2}+p+1}\{\Gamma_u\left(\frac{1}{2},\frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\}^{-p}\prod_{j=1}^{p}(\frac{1}{\gamma_j-1})^{1/2}\,exp[-\frac{1}{2\sigma^2}\{(y^*-x'\beta)(y^*-x'\beta)+\lambda_2\sum_{j=1}^{p}\frac{\gamma_j}{\gamma_j-1}\beta_j^2+\frac{\lambda_1^2}{4\lambda_2}\sum_{j=1}^{p}\gamma_j\}], \tag{3.24}$$

We can take the log for the function (3.24) and the solve the maximization problem by MCEM algorithm. See (Li and Lin,2010) for more details.

3.5. Bayesian sampling for variable selection and estimation

In this section we use the special MCMC algorithm that named Gibbs sampler to implement the Bayesian hierarchical model (3.18). Gibbs sampler algorithm generates random variables indirectly from the full conditional distribution of the interested parameter and fixed all the other parameters. Also, we will generate the conditional posterior distribution of each parameter for the elastic net quantile regression model thorough the following steps:

1- By updating $\boldsymbol{y_i^*}$ from the following full conditional posterior distribution:

$$y_i^*/X,\beta,\sigma^2 \sim \begin{cases} N_n\,(\mathrm{X}\beta,\sigma^2 I_n) & if\ y^*>0,\\ \varsigma(y_i) & otherwise\end{cases}$$

where $\varsigma(y_i)$ has a degenerate density which get all of its mass on y_i, where **i = 1, 2, …, n.**

2- By updating $\boldsymbol{\beta/y,\ \sigma^2,\ \gamma}$ from the full conditional posterior distribution that following the multivariate normal distribution denoted in (**3.24**) with mean $\boldsymbol{s^{-1}X'Y^*}$ and variance $\boldsymbol{\sigma^2 s^{-1}}$, where

$$s = x'x + \lambda_2\ (Q_\gamma),$$

$$Q = diag\ (\ \frac{\gamma_1}{\gamma_1 - 1}, \ \dots\ , \frac{\gamma_p}{\gamma_p - 1}) \qquad (3.25)$$

3- By updating $(\boldsymbol{\gamma_j - 1})^{-1}/\boldsymbol{y}, \boldsymbol{\sigma^2}, \boldsymbol{\beta}$ from the full conditional inverse Gaussian distribution (3.23) (Chhikarn and Folks,1988)

$$f(x/\lambda', \mu') = \sqrt{\frac{\lambda'}{2\pi x^3}}\, exp\left\{\frac{-\lambda'(x-\mu')^2}{2(\mu')^2 x}\right\}; x > 0 \qquad (3.26)$$

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

(Li and Lin,2010) stated that sampling from which is the inverse Gaussian distribution, is faster than the Hyperbolic function proposed by (Scott,2008) .

4-By updating $\boldsymbol{\sigma^2}$ through using the acceptance-rejection algorithm that relies on the incomplete gamma function,

$$f(\sigma^2) = \frac{\Gamma(a)\Gamma(\frac{1}{2})^{-p}}{b^a} h(\sigma^2) \qquad (3.30)$$

Where $a = \frac{n}{2} + p$, $b = \frac{1}{2} \left[\|y^* - x'\beta\| + \lambda_2 \sum_{j=1}^{p} \frac{\gamma_j}{\gamma_j - 1} \beta_j^2 + \frac{\lambda_1^2}{4\lambda_2} \sum_{j=1}^{p} \gamma_j \right]$.

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

5- By updating $\boldsymbol{\lambda_1}$ and $\boldsymbol{\lambda_2}$ through finding the estimates of $\boldsymbol{\lambda_1}$ and $\boldsymbol{\lambda_2}$ that maximizing the log function of **(**3.20**)**.

Chapter 4
Simulation Studies and Application

4.1. Simulation study

In this section regression models patterns are estimated under different simulation scenarios to express the for each of the following methods; the proposed Bayesian elastic net Tobit (Bentr) using R package, the Tobit model by using the (cr) R package, Bayesian elastic net (Banet) by implementing the R programming, and the lasso quantile (crq) by implementing the (crq) R package conducted the following simulation studies to support the theoretical side in which the above methods works well. Furthermore, the comparison that used to assess the performance and the estimation accuracy of the different methods was in terms of parameters estimates and through statistic of Median Mean Squared Error (MMAD) and the Standard Deviation (S.D.). The MMAD is as follows:

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

The overall efficiency of each estimation method can be compared by the total MMAD. MCMC (Gibbs sampling) algorithm has been used with **20000** iterations to reach the stationary for the posterior distributions of the interested parameters burned in the first 1000 iterations. Moreover, have generated the observations of predictor variables $\boldsymbol{x_1}, \dots, \boldsymbol{x_{20}}$ from normal distribution, $\boldsymbol{N_{n=20}}(\boldsymbol{0}, \Sigma)$, where the variance covariance matrix $\Sigma_{\boldsymbol{ij}} = \boldsymbol{\rho}^{|\boldsymbol{i-j}|}$ under four different distributions of the i.i.d errors. For each simulation study, we run 300 simulations.

4.1.1. Scenario I

In this simulation scenario, assumed the true vector of coefficients $\beta = (0,3,0,0,0,0,0,0,0 \ldots, 0)_{20}$ which is the case of very sparse vector with error terms as followed $e_i \sim N(0,1)$, $e_i \sim N(0,5)$, $e_i \sim N(0,1) + N(0,1)$ .As well, generated the observations of $x_1, \ldots, x_{20}$ predictor variables through $N_{n=20}(0, \Sigma)$, where Σ is the variance covariance matrix defined as: $\Sigma_{ij} = 0.9^{|i-j|}$. Consequently, have simulated the following regression model, under different samples sizes (n = 25,50,100,150,200,250) and different estimation methods (Tobit, BAnet, Crq, the method adopted by this study). The censored point was equal to zero (C = 0) to figure out the behavior of the estimation methods.

$$y_i = 3x_2 + e_i$$

Table (1). The value of criterions MMAD and S.D. for simulation scenario one

	Methods		$e_i \sim N(0,1)$	$e_i \sim N(0,5)$	$e_i \sim N(0,1) + N(0,1)$
Sim1	n=25	Tobit	3.30982 (0.75746)	4.7487 (1.25884)	2.37132 (2.10976)
		BAnet	2.07033 (1.13683)	6.72492 (3.06104)	4.5450638 (3.11594)
		Crq	3.01247 (0.58850)	4.94020 (1.44609)	3.07189 (2.06448)
		BENTM	0.37287 (0.22559)	1.08482 (0.69666)	0.33702 (0.70987)
	n=50	Tobit	2.48780 (0.14205)	3.1066 (0.64449)	1.84961 (0.54231)
		BAnet	4.32501 (0.46857)	5.49438 (0.76207)	4.16627 (0.57398)
		Crq	3.26266 (0.11311)	4.12225 (0.77484)	2.73711 (0.50161)
		BENTM	0.14582 (0.04867)	0.63749 (0.19990)	0.20283 (0.06488)
	n=100	Tobit	1.95808 (0.22672)	1.76343 (0.46198)	1.74778 (0.29782)
		BAnet	3.82422 (0.16980)	4.64568 (0.15125)	3.87999 (0.418029)
		Crq	2.80118	2.66914	2.61293

			(0.22277)	(0.60335)	(0.317344)
		BENTM	0.12408 (0.05791)	0.58881 (0.14659)	0.18863 (0.061547)
	n=150	Tobit	1.94378 (0.22304)	1.83543 (0.38865)	1.78748 (0.44371)
		BAnet	3.70282 (0.35578)	5.07402 (0.49252)	3.93621 (0.17507)
		Crq	2.86402 (0.21994)	2.79554 (0.26790)	2.64137 (0.27084)
		BENTM	0.12257 (0.01294)	0.43897 (0.13995)	0.12061 (0.05273)
	n=200	Tobit	1.57864 (0.45740)	1.70312 (0.33581)	1.82346 (0.27705)
		BAnet	3.76246 (0.29654)	4.31168 (0.746763)	3.56677 (0.15152)
		Crq	2.62427 (0.39261)	2.50376 (0.66136)	2.70301 (0.17888)
		BENTM	0.10162 (0.01697)	0.42092 (0.07651)	0.1364 (0.032570)
	N=250	Tobit	1.69367 (0.25736)	1.8003 (0.44779)	1.68130 (0.096407)
		BAnet	3.85296 (0.24572)	4.95694 (0.50716)	3.64999 (0.43214)
		Crq	2.72156 (0.17765)	2.4298 (0.64758)	2.55005 (0.239980)
		BENTM	0.08116 (0.01489)	0.40163 (0.13975)	0.12660 (0.011466)

Table (1) displayed the values of the criteria MMMAD and SD that measured the quality of the estimation process under four different types of errors, different sample sizes, and different regression models observed the values of MMAD of the proposed model are smaller compared with the others model, also this is very clear as the sample size getting larger. For example, when (n=25) with different error distributions the values of MMAD and its SD for the proposed model are (0.37287, 0.22559), and when (n=250) with different error distributions, the values of MMAD and its SD for the proposed model are (0.08116, 0.01489).

4.1.2. Scenario II

In this simulation scenario, assumed the true vector of coefficients $\beta = (0,3,0,0,0,1,0,0,0\ \dots,0)_{20}$ which is the case of sparse vector with error terms, followed $e_i \sim N\ (0,1)$, $e_i \sim N, (0,5)$, $e_i \sim N(0,1) + N(0,1)$. As well, generated the observations of $x_1, \dots, x_{20}$ predictor variables through $N_{n=20}(0\ , \Sigma)$, where Σ is the variance covariance matrix defined as: $\Sigma_{ij} = 0.9^{|i-j|}$. Consequently, have simulated the following regression model under different samples sizes (n= 25,50,100,150,200,250) and different estimation methods (Tobit, BAnet, Crq, the method adopted here). The censored point was equal to zero (C = 0) to figure out the behavior of the estimation methods,

$$\mathbf{y_i = 3x_{2i} + x_{6i}} + \boldsymbol{e_i}$$

Table (2). The value of criterions MMAD and S.D. for simulation scenario two

	Methods		$e_i \sim N(0,1))$	$e_i \sim N(0,5)$	$e_i \sim N(0,1) + N(0,1)$
Sim2	n=25	Tobit	4.9820 (1.01681)	7.84054 (2.27518)	4.46189 (1.58553)
		Banet	5.3221 (0.52401)	7.6939 (1.07881)	5.12101 (0.77253)
		Crq	5.22560 (1.05874)	7.71490 (1.84375)	4.73255 (1.38319)
		BENTM	0.50118 (0.08693)	1.63418 (0.33883)	0.65546 (0.16668)
	n=50	Tobit	3.12874 (0.62296)	3.90474 (0.90037)	2.90393 (0.60466)
		Banet	4.52996 (0.21200)	6.06506 (0.67733)	4.85503 (0.41606)
		Crq	3.77596 (0.50462)	4.52070 (0.75565)	3.77654 (0.45691)
		BENTM	0.42141 (0.07036)	1.15171 (0.15558)	0.50533 (0.08443)
		Tobit	2.51882 (0.18451)	2.77265 (0.72729)	2.42383 (0.460615)

	n=100	Banet	4.29508 (0.19913)	5.46653 (0.93122)	4.12029 (0.38019)
		Crq	3.12670 (0.22563)	3.54866 (0.58411)	3.09834 (0.37150)
		BENTM	0.21551 (0.01403)	0.78428 (0.14250)	0.32320 (0.08093)
	n=150	Tobit	2.01061 (0.16882)	2.87802 (0.54578)	2.40299 (0.38323)
		Banet	4.14242 (0.19149)	5.51047 (0.58329)	4.37114 (0.47051)
		Crq	2.98740 (0.12283)	3.46618 (0.46774)	3.24750 (0.21941)
		BENTM	0.19261 (0.02928)	0.74683 (0.13671)	0.30538 (0.05840)
	n=200	Tobit	2.19052 (0.25246)	2.25748 (0.16924)	1.86645 (0.11085)
		Banet	3.74736 (0.23581)	5.076705 (0.99285)	3.94052 (0.24737)
		Crq	2.91260 (0.17236)	2.89879 (0.29618)	2.70515 (0.17736)
		BENTM	0.16439 (0.03634)	0.58875 (0.10006)	0.22428 (0.05414)
	N=250	Tobit	2.11974 (0.27044)	2.06982 (0.58897)	1.95753 (0.26718)
		Banet	3.92183 (0.16112)	4.96555 (0.62139)	3.93794 (0.17582)
		Crq	2.90824 (0.18913)	2.91075 (0.45084)	2.86795 (0.26016)
		BENTM	0.14347 (0.01214)	0.55799 (0.07386)	0.17744 (0.03823)

Table (2) displayed the values of the criterions MMMAD and SD that measured the quality of the estimation process under four different types of errors, different sample sizes, and different regression models observed the values of MMAD of the proposed model are smaller compared with the others model. Also, this is very clear as the sample size getting larger. For example, when (n=25) with different error distributions the values of MMAD and its SD for the proposed model are (0.50118, 0.08693), and when (n=250) with different error distributions the values of MMAD and its SD for the proposed model are (0.14347, 0.01214).

4.1.3. Scenario III

In this simulation scenario, assumed the true vector of coefficients $\beta = (0{,}85{,}0.85{,}0.85{,}0.85{,}0.85{,}0.85{,}0.85{,}0.85{,}0.85, \ldots, 0.85)_{20}$ which is the case of density vector with error terms follows $e_i \sim N(0,1), e_i \sim N(0,5), e_i \sim N(0,1) + N(0,1)$. As well, generated the observations of $x_1, \ldots, x_{20}$ predictor variables through $N_{n=20}(0\ ,\Sigma)$, where Σ is the variance covariance matrix defined as $\Sigma_{ij} = 0.9^{|i-j|}$. Consequently, have simulated the following regression model under different sample sizes (n= 25,50,100,150,200,250) and different estimation methods (Tobit, BAnet, Crq, the method adopted here). The censored point was equal to zero (C = 0) to figure out the behavior of the estimation methods:

$$\boldsymbol{y_i = \sum_{i=1}^{8} 0.85X_i + e_i}$$

Table (3). The value of criterions MMAD and S.D. for simulation scenario three

	Methods		$e_i \sim N(0,1))$	$e_i \sim N(0,5)$	$e_i \sim N(0,1) + N(0,1)$
Sim3	n=25	Tobit	4.99173 (0.64713)	5.15854 (0.91697)	5.84224 (1.00871)
		BAnet	5.90819 (0.79945)	7.74524 (0.30492)	5.98727 (1.41810)
		Crq	4.96285 (0.53977)	5.78119 (0.68264)	6.61679 (2.38345)
		BENTM	0.67795 (0.16489)	1.21295 (0.28481)	1.02232 (0.16457)
	n=50	Tobit	3.46772 (0.42519)	3.14952 (0.78548)	4.40413 (0.65500)
		BAnet	4.83583 (0.18644)	5.86021 (0.94120)	4.84948 (0.56408)
		Crq	4.04724 (0.22848)	4.22451 (0.75211)	4.62629 (0.62147)
		BENTM	0.38497 (0.08232)	0.74404 (0.11085)	0.75890 (0.21315)
	n=100	Tobit	2.77326 (0.37594)	2.78412 (0.28275)	2.34919 (0.37167)
		BAnet	4.44287 (0.51859)	5.67594 (1.11220)	4.35189 (0.48969)
		Crq	3.28826	3.55983	3.11690

			(0.50175)	(0.48211)	(0.42883)
		BENTM	0.24549 (0.06827)	0.73532 (0.18677)	0.44187 (0.06088)
	n=150	Tobit	2.44676 (0.25534)	2.58381 (0.52537)	2.48412 (0.25806)
		BAnet	4.01119 (0.44898)	5.09955 (0.80961)	4.38368 (0.61308)
		Crq	3.26777 (0.32659)	3.06652 (0.28603)	3.22498 (0.45663)
		BENTM	0.17205 (0.03548)	0.62611 (0.09641)	0.40948 (0.06592)
	n=200	Tobit	2.36452 (0.18308)	2.14097 (0.44484)	2.30627 (0.39299)
		BAnet	3.93020 (0.20159)	4.95660 (0.31110)	4.01492 (0.33161)
		Crq	3.00732 (0.09246)	3.08527 (0.43955)	2.87015 (0.31929)
		BENTM	0.13473 (0.02470)	0.56940 (0.13748)	0.31753 (0.06180)
	N=250	Tobit	2.38328 (0.20155)	2.19319 (0.83867)	1.97403 (0.27295)
		BAnet	3.92374 (0.16314)	5.42051 (0.45303)	3.95325 (0.38264)
		Crq	3.03154 (0.12195)	3.10431 (0.83456)	2.68799 (0.27220)
		BENTM	0.15859 (0.04718)	0.56458 (0.10610)	0.31516 (0.05041)

Table (3) displayed the values of the criteria MMMAD and SD that measured the quality of the estimation process under four different types of errors, different sample sizes, and different regression models observed the values of MMAD of the proposed model are smaller compared with the others model. Also, this is very clear as the sample size getting larger. For example, when (n=25) with different error distributions the values of MMAD and its SD for the proposed model are (0.67795, 0.16489), and when (n=250) with different error distributions the values of MMAD and its SD for the proposed model are (0.15859, 0.04718). For the simulation scenario one and under the error term that is distributed according to the normal distribution, $e_i \sim N(0,1)$ we draw six figures dedicated to each sample size to compare the true values of parameter vector and the estimates values of the parameters based on different estimation methods.

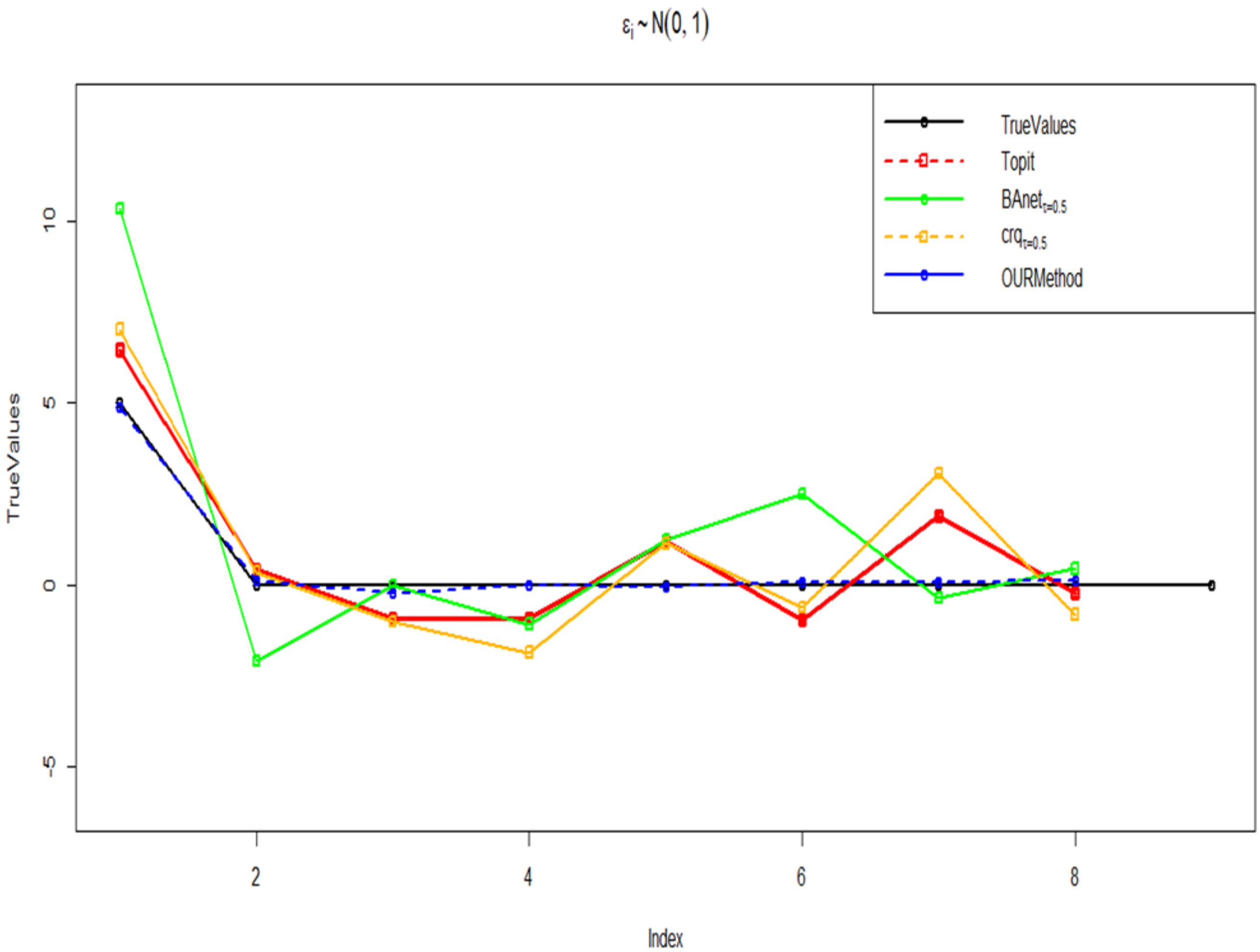

Figure (1). Lines plot for the different estimation methods with $e_i \sim N(0,1)$ and (n=25).

Figure (1) Contains the results of simulation scenario one, where the error is $e_i \sim N(0,1)$. The figure contains the sparse line (black) in the middle .The vertical line represents the true vector. Furthermore, the blue line represents the parameters estimates based on the proposed model using sample size (n=25). The red line is the tobit model results, the orange line is the (Crq=0.5) results , and the green line is the (BAnet) results. From figure (1) it is very clearly to observe that the blue line is the closed line to the standard line (sparse) and matching some points . But the tobit model parameters estimates come next.

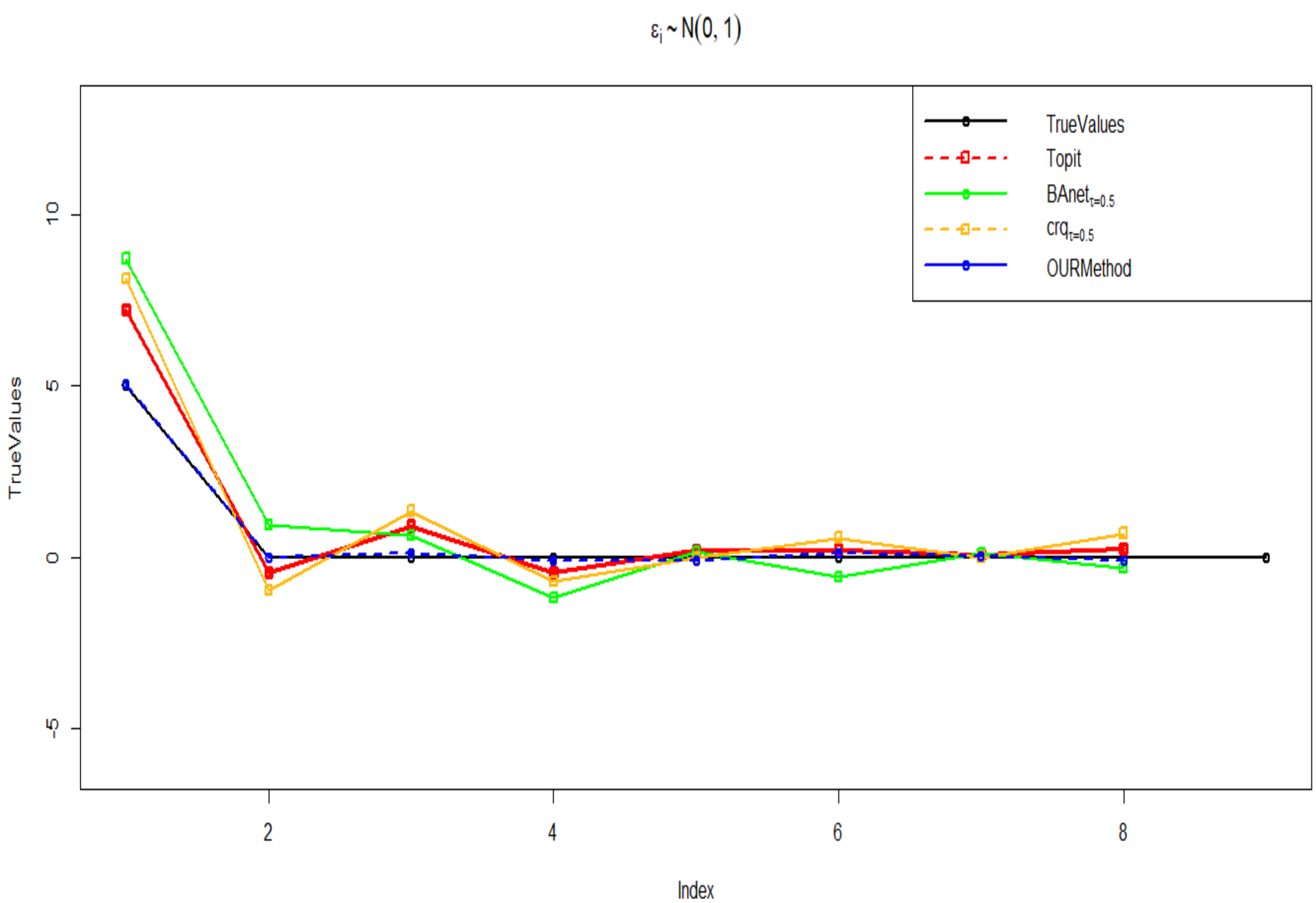

Figure (2). Lines plot for the different estimation methods with $e_i \sim N(0,1)$ and (n=50).

In figure (2) , we draw the results of simulation scenario tow with $e_i \sim N(0,1)$ and (n=50) . The result represent the parameters estimates for the different models. We observed that the parameter estimates of the proposed mode (blue line) are very close and matching in some points the standard line (sparse) .Also, the other model results are closed to each other and matching the sparse line in some points.

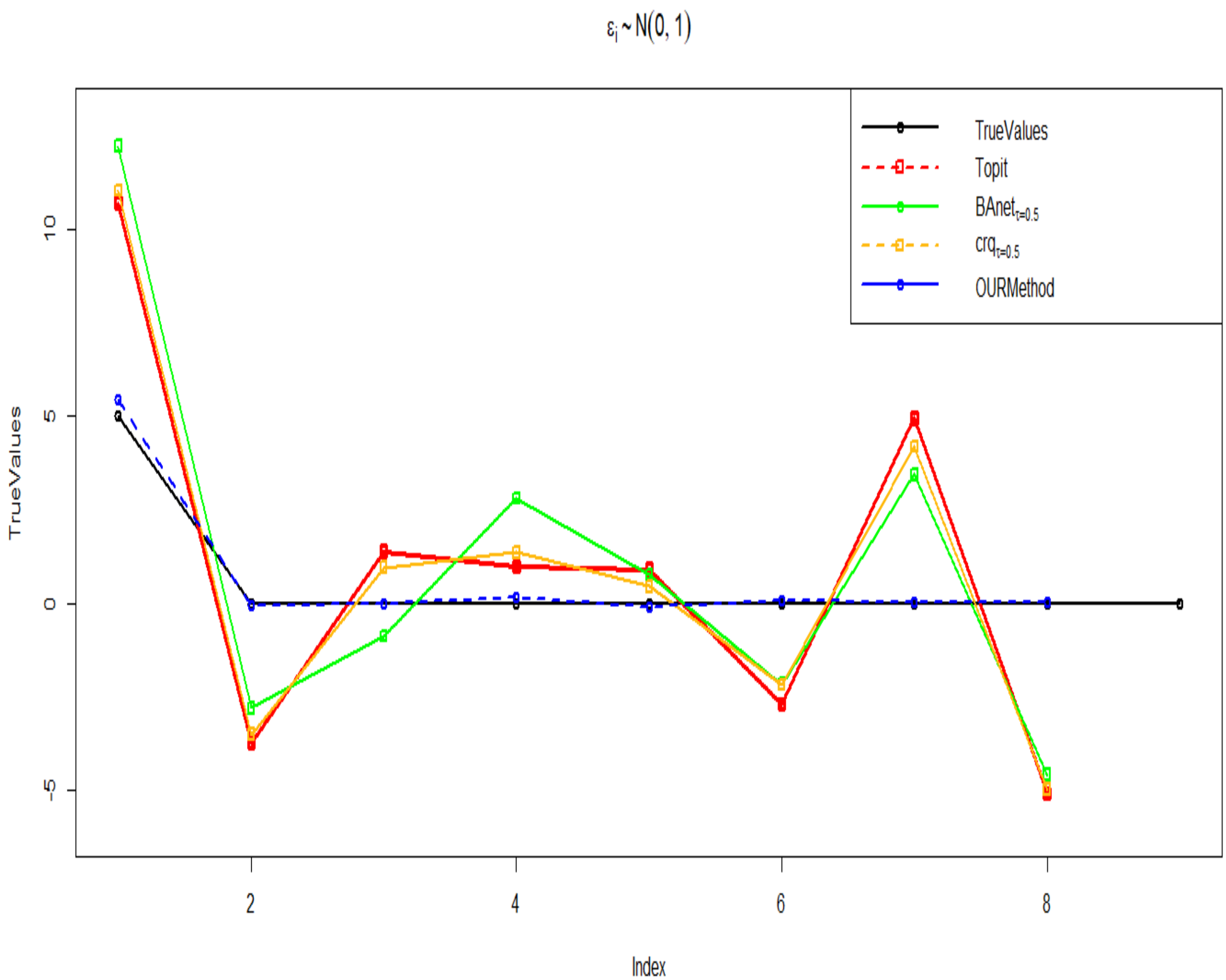

Figure (3). Lines plot for the different estimation methods with $e_i \sim N(0,1)$ and (n=100).

Figure (3) shows that the blue line is the close line to the sparse line under $e_i \sim N(0,1)$ and with sample size (n=100). Also, we observed the matching of blue line points (parameters estimates) with the black line . For the other models, clearly all the lines (red, orange , green) are away from the sparse, but they match each other in some points.

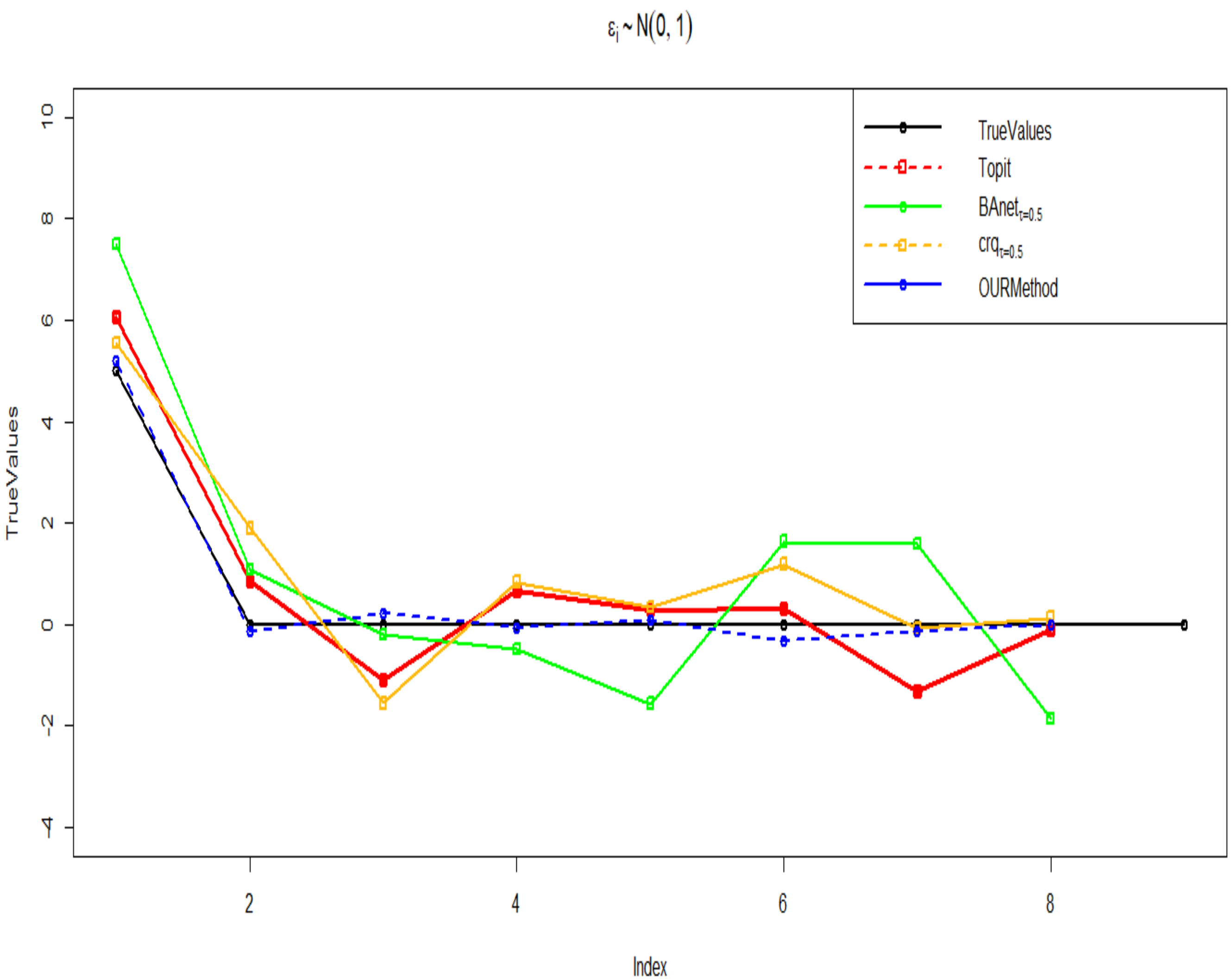

Figure (4). Lines plot for the different estimation methods with $e_i \sim N(0,1)$ and (n=150).

In figure (4) displayed the results of parameters estimates for the simulation scenario one under $e_i \sim N(0,1)$ and sample size (n=150). Very clearly, the blue line is the closed line to the sparse vector of true parameters estimates as comparing with the other lines.

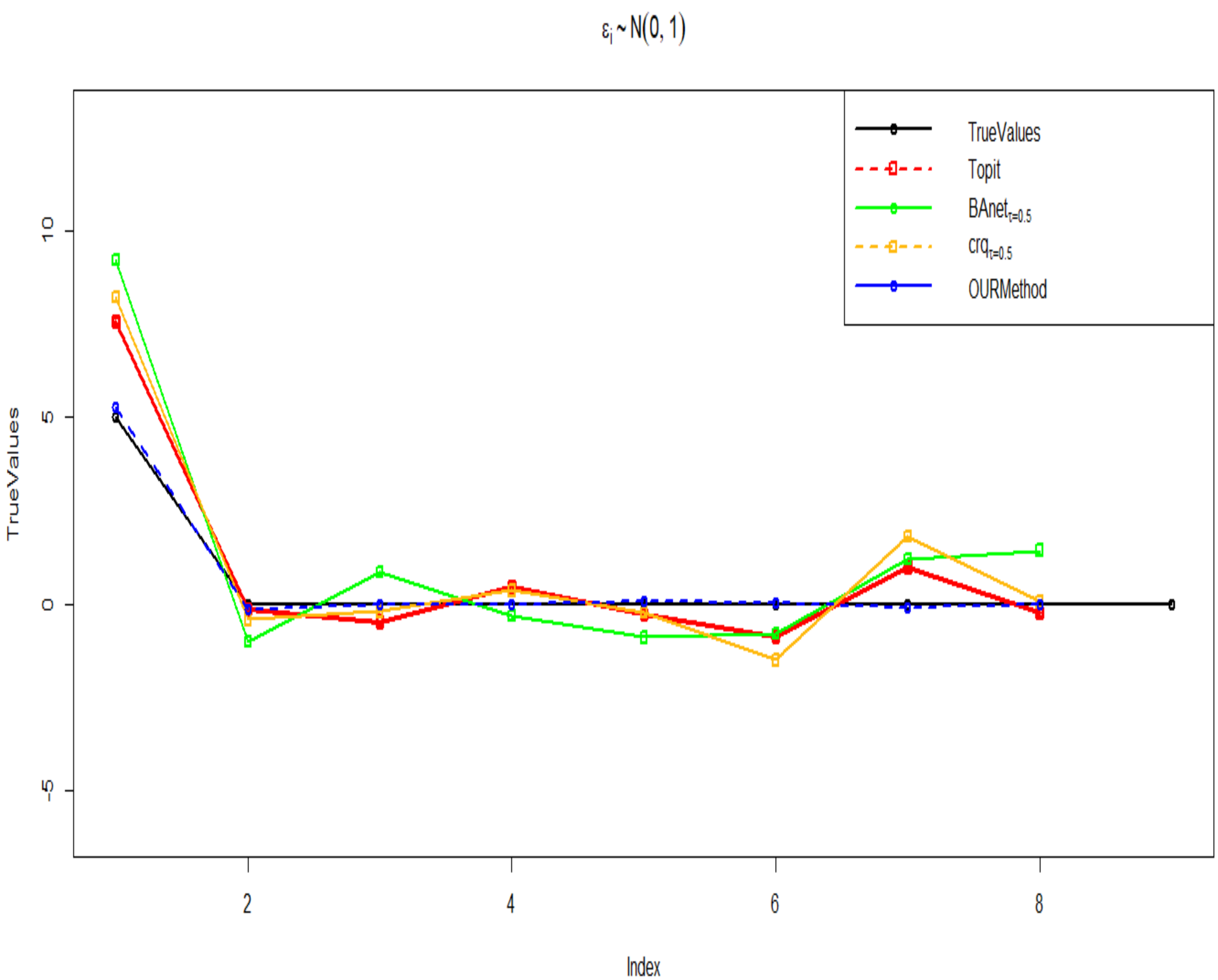

Figure (5). Lines plot for the different estimation methods with $e_i \sim N(0,1)$ and (n=200).

In figure (5) the lines are drawn for the simulation scenario one with $e_i \sim N(0,1)$ and (n=200).Also, it is very clear that the parameters estimates that are computed from the proposed posterior distribution for $B^{,s}$ are very close to the sparse line (Black line) and match in some points. The other lines are very close to each other and close to sparse line.

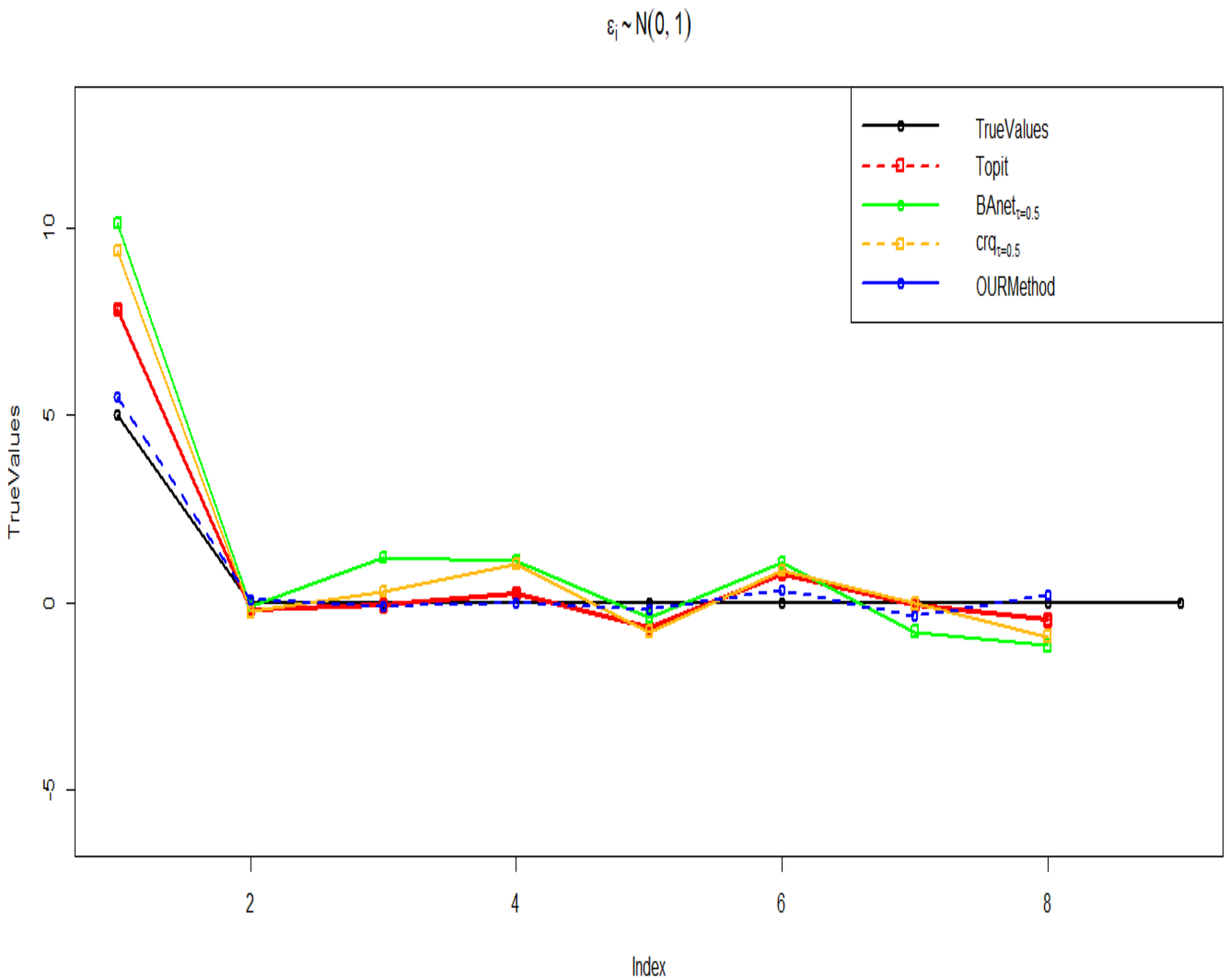

Figure (6). Lines plot for the different estimation methods with $\boldsymbol{e_i \sim N(0,1)}$ and (n=250).

Figure (6) Shows the results of simulation scenario one with $e_i \sim N(0,1)$ and sample size (n=250). The blue line matches the sparse line , i.e., it is the closed line.

4.2 Real Data Description and Analysis

The following data have information that records for mother visits to the Salam Health Center in Waist Health Department. Furthermore, I used (50) personal forms of mother that are available in the above center, which represents the ‘mean’. I took simple random sample. Women was drawn to study the factors affecting the number of children born (response variable) Y , while the independent variables were as follows:

$\boldsymbol{X_1}$: Age of the mother

$\boldsymbol{X_2}$: Mother's age at marriage

$\boldsymbol{X_3}$: Academic achievement of mother

$\boldsymbol{X_4}$: Academic level of the husband

$\boldsymbol{X_5}$: Weight of mother

$\boldsymbol{X_6}$: The length of the mother

$\boldsymbol{X_7}$: Mother smoking status

$\boldsymbol{X_8}$: Age of the husband

X_9: The occupation of the husband

X_{10}: Number of dead children

X_{11}: Use status of contraceptives

X_{12}: Mother with thyroid disease

X_{13}: The number of hours a mother sleeps a day

X_{14}: Breastfeeding duration

X_{15}: Mother's occupation

X_{16}: Viruses status

X_{17}: Mother's food system

X_{18}: Matching blood status

X_{19}: Gestational diabetes status

X_{20}: Psychological status

Table (4). parameters estimates of $\boldsymbol{\beta_1}$,…..., $\boldsymbol{\beta_{20}}$ under four different models.

Parameters	Tobit	BAnet	Crq	BENTM
β_1	0.1117	0.0000	0.0167	**-9.8778**
β_2	-0.1188	0.0000	-0.0358	**0.0000**
β_3	0.1030	0.2408	-0.1288	**-0.3445**
β_4	0.0015	-0.2216	0.1258	**0.0000**
β_5	0.0453	0.0279	0.0686	**1.1694**
β_6	-0.1802	-0.1775	-0.2526	**0.9188**
β_7	0.6665	1.7751	0.9264	**-18.4727**
β_8	0.0572	0.0437	0.0633	**1.1728**
β_9	0.7540	-0.1622	0.4789	**-0.2375**
β_{10}	0.0320	0.8403	-0.0198	**0.8641**
β_{11}	-0.6230	-0.2304	-1.2832	**0.1860**
β_{12}	-1.8328	-0.6509	-2.5601	**0.2194**
β_{13}	0.5692	-0.0236	0.6496	**0.2061**
β_{14}	-0.1765	0.0515	-0.1420	**1.5064**
β_{15}	3.0677	1.1744	4.3379	**-7.1955**
β_{16}	-0.6657	0.1197	-0.0061	**0.4179**
β_{17}	-0.3281	1.1367	-0.4272	**-0.3860**
β_{18}	-2.2937	-0.5384	-2.7590	**0.6754**
β_{19}	-3.1945	0.2881	-3.1232	**5.8932**
β_{20}	0.6224	-1.3341	0.6829	**8.1628**

Table (4) summarized the parameter estimates that captured from the posterior distributions for the proposed model and the other three existing models. Gibbs sample algorithm estimates the mean of (B) for the posterior distribution estimates. We observed the variable selection procedure in the proposed model in the second and fourth variables (Mother age at marriage and academic level of the husband), where the parameters estimates were (β_2= 0, β_4= 0). The results of the proposed model were very meaningful estimates. The proposed mode results are comparable to the other existing models. Furthermore, β_1= -9.8778, which means that the age of the mother is very important variable and effect the response variable (weight of newborn child). Also, the variables (smoking status of mother, gestational diabetes status, and psychological status of mothers) are very important variables which are very effective on the response variable.

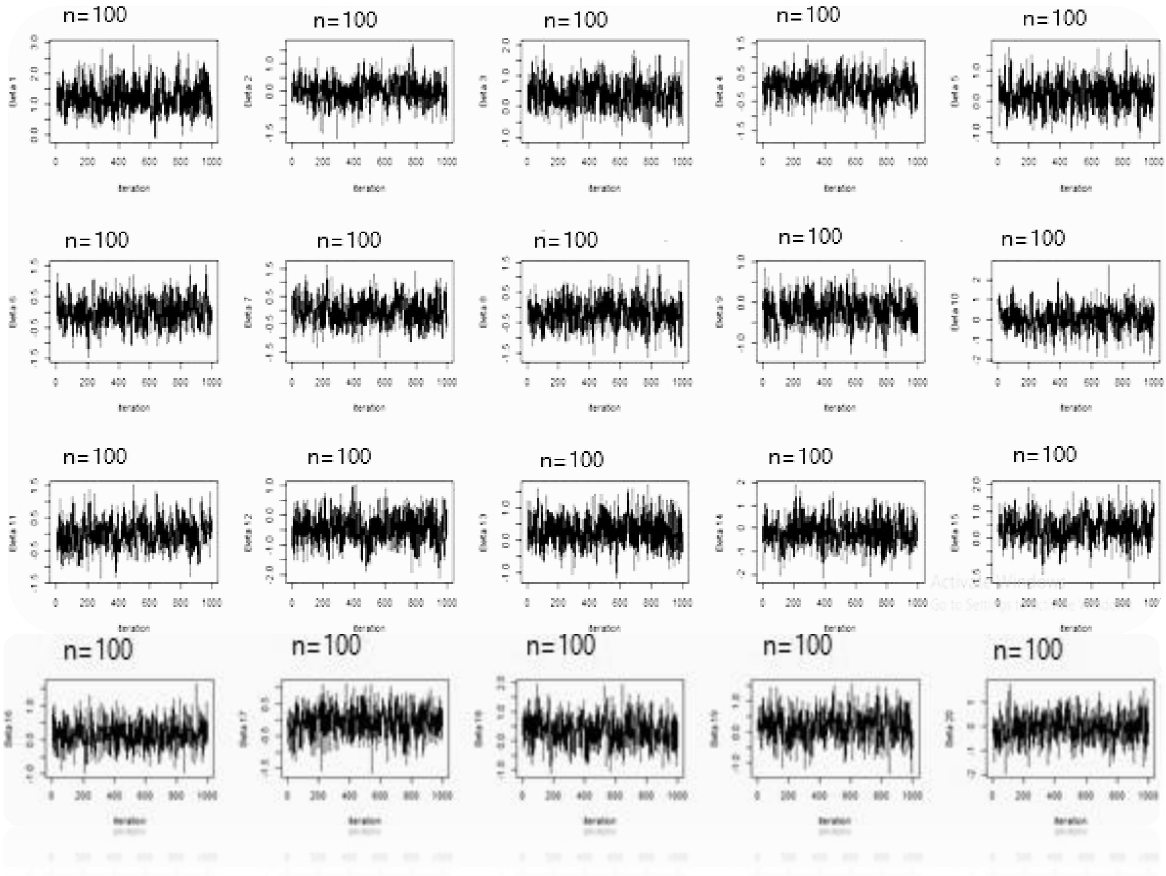

Figure (7). Trace plots for $\beta_1, \dots, \beta_{20}$ parameters.

The following figure illustrates the trace plot of the posterior densities for different (20) parameters. Trace plots are displayed the stability of the Gibbs sampler algorithm, which is mean that the appropriate prior distribution that formulated the posterior distribution.

Chapter 5
Conclusions and Recommendations

5.1. Conclusions

This thesis is employed the scale mixture of laplace prior distribution as mixing the normal distribution with truncated gamma that is introduced by Li and Lin (2010) in the elastic net tobit regression .The new proposed regularization method works as variable selection procedure in the elastic net tobit regression model. Consequently, the model that proposed creates new bayesian hierarchical model which leads to faster Gibbs sample computation and gives more meaningful parameter estimates. conducted three simulation scenarios. Also, I applied the proposed model in the real data. Moreover, we have had comparable model with some exists models and the results were competitive with those of other models.

5.2. Recommendation

The proposed model (Bayesian elastic net Tobit model) that have produced in this thesis is a promising model in terms of variable selection procedure and from the point of view of the interpretability of the model. So, we produced new regularization tobit regression model that uses y_i as censored variable, i.e., my work is an extension for the regularization tobit regression. I recommend researcher who are interested in bayesian model to try it with the right censored and the interval censored models.

References

Bibliography

1. Alhamzawi, R. (2013). Tobit quantile regression with the adaptive lasso penalty. In The Fourth International Arab Conference of Statistics, 450 ISSN (1681 6870).

2. Alhamzawi, R., & Yu, K. (2014). Bayesian Lasso-mixed quantile regression. Journal of Statistical Computation and Simulation, 84(4), 868-880.

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

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

5. Alhusseini, F. H. H. (2017). New Bayesian Lasso in Tobit Quantile Regression. *Romanian Statistical Review Supplement*, *65*(6), 213-229.

6. Alhusseini, F. H. H., & Georgescu, V. (2018). Bayesian composite Tobit quantile regression. Journal of Applied Statistics, 45(4), 727-739.

7. Alhusseini, F. H. H., Flaih, A. N., & Alshaybawee, T. (2020). Bayesian extensions on Lasso and adaptive Lasso Tobit

regressions. *Periodicals of Engineering and Natural Sciences*, *8*(2), 1131-1140.

8. Al-Sadoun , Muhannad Faiz , (2005) . " Empirical Bayes and Bayes for Ridge regression" , Quarterly Specialized Refereed Journal, vol. (7) no. (1).

9. Armido, D. and Alfred, M. (1986). “Computation of the incomplete gamma function ratios and their inverse.” ACM Trans. Math. Soft., 12: 377–393.

10. Arthur , E. Hoerl , Robert W. Kennard (1970) ,"Ridge Regression Biased Estimation for Nonorthogonal Problems" ; Technometrics Vol.12 , No.1 , pp.55-60.

11. Casella, G. (2001). “Empirical Bayes Gibbs sampling.” Biostatistics, 2: 485–500.

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

13. Chhikara, R. (1988). *The inverse Gaussian distribution: theory: methodology, and applications* (Vol. 95). CRC Press.

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

15. Fonti, V., & Belitser, E. (2017). Feature selection using lasso. *VU Amsterdam Research Paper in Business Analytics*, *30*, 1-25.

16. Hans ,Christ , (2009) , " Bayesian lasso regression" , Biometrika, 96, (4), Biometrika Trustpp. 835–845.

17. Hans, C. (2010). Model uncertainty and variable selection in Bayesian lasso regression. *Statistics and Computing*, *20*(2), 221-229.

18. Hastie, T., Tibshirani, R., and Wainwright, M. (2015). Statistical Learning with Sparsity: The Lasso and Generalizations. Taylor Francis Group, Forida, USA.

19. Hilali, H. K. A. (2019). Bayesian adaptive Lasso Tobit regression with a practical application.

20. 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.

21. Jørgensen, B. B. (1982). Mineralization of organic matter in the sea bed—the role of sulphate reduction. *Nature*, *296*(5858), 643-645.

22. Ji, Y., Lin, N., & Zhang, B. (2012). Model selection in binary and tobit quantile regression using the Gibbs sampler. *Computational Statistics & Data Analysis*, *56*(4), 827-839.

23.JIRATCHAYUT,M.K.(2014). Bayesian elastic net (Doctoral dissertation, THAMMASAT UNIVERSITY).

24.Kannard , Arthur E. Boart W.; F. Baldwin, Kent, (1975), " Ridge Regression , Some solutions " , Communications in statistics , 4(2), 105-123.

25.Kirkland, Lisa-Ann. (2014) Lasso-simultaneous shrinkage and election via p, norm Msc. Thesis. University of Pretoria.

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

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

28.Odah, M. H., Bager, A. S. M., & Mohammed, B. K. (2017). Tobit regression analysis applied on Iraqi bank loans. American Journal of Mathematics and Statistics, 7(4), 179-182.

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

30.Rogaeva, E., Meng, Y., Lee, J. H., Gu, Y., Kawarai, T., Zou, F., ... & St George-Hyslop, P. (2007). The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nature genetics, 39(2), 168-177.

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

32. Tibshirani , Robert, (1996) ," Regression Shrinkage and Selection via the Lasso" , J. R. Statist. Soc. B 58, No. 1, pp. 267-288.

33. van Wieringen, W. N. (2015). Lecture notes on ridge regression. *arXiv preprint arXiv:1509.09169*.

34. Zou, H., & 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.

35. Zou, H., Hastie, T., & Tibshirani, R. (2007). On the “degrees of freedom” of the lasso. *The Annals of Statistics*, *35*(5), 2173-2192.

جمهورية العراق

وزارة التعليم العالي والبحث العلمي

جامعة القادسية

كلية الإدارة والاقتصاد

قسم الإحصاء

الشبكة المرنة البيزية لانحدار الرقابة الطبيعي مع تطبيق عملي

رسالة ماجستير مقدمة الى مجلس كلية الادارة والاقتصاد/جامعة القادسية

وهي جزء من متطلبات نيل درجة الماجستير في قسم االحصاء

من طالب الماجستير

محمد رسول محسن الصافي

المشرف

أ. د. احمد نعيم فليح

1443 هـ

2021 م

المستخلص

يحظى تحليل نماذج الانحدار باهتمام كبير في كل مجالات العلوم . وخصوصا مجال النظرية الاحصائية حيث ان بناء نموذج الانحدار المرن الذي يوفر قدرة تفسيرية عالية ويوفر مقدرات يمكن الاعتماد عليها بالتحليل لمعالم النموذج قد جذب انتباه الكثير من الباحثين في مجال الاحصاء . حيث تم خلال العقود المنصرمة تطوير العديد من نماذج الانحدار بهدف البحث عن النموذج الذي يلائم البيانات عند الدراسة وتعتبر نظرية بيز من النظريات التي لاقت رواجا في الكثير من المجالات العالمية ، حيث ان استخدام طرائق بيز وإجراءاتها في الاساليب الاحصائية تعطي نتائج معول عليها طالما ان اساليب نظرية بيز تتمتع بمرونة عالية ويمكن اجراء عملياتها الحسابية بشكل سهل خصوصا مع التطوير الحاصل في مجال الحاسوب . ان بناء نموذج انحدار بيزي تعتمد في كفاءته على سرعة تنفيذ خوارزمية مونت كارلو لسلسلة ماركوف MCMC في هذه الرسالة وظفت خليط لمعلمة القياس لتوزيع لا بلاس للمعلمة المسبقة لكن هنا وظيفتها في الانحدار توبت . تم طرح طريقة تنظيم جديدة وبالتالي تطوير نموذج انحدار توبت للشبكة المرنة وكذلك تم اقتراح نموذج بيزي هرمي وبالتالي عمل خوارزمية Gibbs للمعاينة وتنفيذها. ان طريقة التنظيم المقترحة في هذه الرسالة وبالتالي نموذج انحدار توبت للشبكة المرنة تم استخدامها كأسلوب او اجراء لاختيار المتغيرات المهمة وقد اجريت ثلاث سيناريوهات لدراسة محاكاة بهدف دراسة سلوك التوزيع اللاحق للمعلمة من خلال تقديرات المعالم مستخدما معيار وسيط المعدل للانحرافات المطلقة ومعيار الانحراف القياسي للحكم على اداء الطريقة المقترحة ومقارنتها مع نتائج بعض الطرق الموجودة ، حيث اظهرت نتائج المعايير الصفة التنافسية لطريقتنا المقترحة مع الطرق الاخرى . إضافة الى ذلك اظهرت خوارزمية عينة Gibbs استقراريه التقدير لمعلمات التوزيع في التوزيع اللاحق كذلك وظفنا الطريقة المقترحة في تحليل بيانات واقعية ، حيث وفرت الطريقة المقترحة خاصية اختيار المتغيرات بشكل واضح حيث جعلت بعض معالم النموذج مساوية للصفر . واظهرت الطريقة المقترحة كذلك قدرتها على التنافس والمقارنة مع النماذج الاخرى المستخدمة .