Republic of Iraq

Ministry of Higher Education & scientific research

Department of statistics

College of Administration and Economics

University of AL-Qadisiyah

# New Bayesian Lasso Method In Tobit Quantile Regression

A thesis submitted to the council of the college of Administration & Economics\ University of Al-Qadisiyah as partial fulfillment of the requirements for the degree of Master of Statistics Sciences

By

Esraa Hatem Abed Jabr Al-Shummare

Supervised by

Asst. Prof. Dr. Fadel Hamid Hade Al-Husseini

1444 A.H 2023 A.D

# بسم الله الرحمن الرحيم

﴿يا أيها الذين آمنوا إذا قيل لكم تفسحوا في المجالس فافسحوا يفسح
الله لكم وإذا قيل لكم انشزوا فانشزوا يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ
وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ﴾

صدق الله العلي العظيم

[المجادلة: 11]

# DEDICATION

**-To the pure soul of my mother and father, who missed the warmth of our applause on the day of my success and graduation.**

**-To my sisters and brother, they are a piece of my heart, may God protect them**

**-To my husband and children (thulfiqar and Laith) their presence in my life is a support after God**

**-To all of these and those I dedicate this humble work, and we ask God to make it a beacon for every student of knowledge**

**Amen, Lord of the worlds**

# Thanks and appreciation

بسمه تعالى((ومن يشكر الناس فإنما يشكر لنفسه )) صدق الله العظيم {لقمان12} ,

وقال رسول الله (ص) ((من لم يشكر الناس ,**لم يشكر الله عز وجل))**

**Praise be to God, the Exalted, the Majestic. Thank you very much, a favor and a blessing that fills the heavens and the earth for what you have honored me with in completing this study, which I hope will be satisfied with it... My supervisor, Professor Dr. (Fadel Hamid Hadi) for the guidance and correction that he gave me.. and what he taught me of the abundance of his humanity, his high morals, and a high level.. Thanks to those who lit the lanterns of knowledge. Thank you to those who made efforts to build the country. Teach us the sweetness of morals before the sweetness of knowledge. Whoever taught me a message, I am his slave. With these words, we describe the candles that burn every day to light our paths with knowledge and knowledge of those who deserve wisdom. How much do all our days begin, and how much we put a little for you in our hearts every day, to all my professors at Al-Qadisiyah University / Department of Statistics.**

# Table of Contents



## Chapter One

## Introduction and Literature Review



## Chapter Two

## Some Basic Concepts

## Chapter Three



## Chapter Four

## Simulation and Real Data



## Chapter Five

## Conclusion and Recommendation

# List of Tables

# List of Figures

# Contents Table

| symoble | scientific term |
|---|---|
| AL | Asymmetric Laplace |
| ALD | Asymmetric Laplace Distribution. |
| BAN | Bayesian Adaptive Elastice Net. |
| BLTQR | Bayesian Lasso Tobit Quantile Regression. |
| BNLTQR | Bayesian New Lasso Tobit Quantile Regression. |
| Cdf | Cumulative distribution function. |
| Pdf | Probability distribution function. |
| ESR | Erythrocyte Sedimentation Rate |
| GIG | Generalized Inverse Gaussian. |
| LD | Laplace distribution. |
| LQR | Lasso Quantile Regression. |
| LTQR | Lasso Tobit Quantile Regression. |
| LASSO | Least absolute shrinkage and selection operator |
| MCMC | Markov Chain Monte Carlo. |
| MAE | Mean Absolate Error |
| MSE | Mean square error |
| New BlTQR | New Bayesian lasso Tobit Quantile Regression. |
| NBlTQR | Non-Bayesian lasso Tobit Quantile regression |
| OLS | Ordinary Least Squares Method. |
| QR | Quantile Regression. |
| SMN | Scale Mixture Of Normal. |
| SME | Scale Mixture Of Stander Exponential. |
| SMU | Scale Mixture Of Uniform. |
| TQR | Tobit Quantile Regression. |
| TR | Tobit Regression. |
| VS | Variable Selection. |
| Crq | Censared Regression quantile |

# Abstract

Thesis studies the lasso Tobit quantile regression model. Quantile regression model analysis for left censored data (Tobit ) is very important in many fields of sciences since. It allows the researcher to explore a range of conditional quantities (Quantile functions) here with of conditional forms of unobserved and outliers individuals .Our main objective is to describe conditional quantile forms in Tobit regression by estimating the interested parameter from the lasso Bayesian theorem aspect . We introduced a new Bayesian hierarchical priors model to implement the Gibbs sampler algorithm for the Tobit quantile regression model .Also, we explain the new scale mixture of uniform mixing with standard exponential by derivative new formula for the double exponential prior distribution .We study the Bayesian estimation of quantile regression in the Tobit model using the MCMC and Gibbs sampler algorithm with two simulation scenarios from the results shows that the proposed lasso penalized model is a comparable model under different quantile level and different sample sizes totally , we illustrate the performance of the proposed lasso penalized method by analyzing real data that represent a sample of size (200) observations of Erythrocyte sedimentation rate a response variable with twenty predictor variables . The results show that the proposed model provides variable selection precedent compared to some examined regression models under different quantile regression ,the mean absolute error and the mean square error were the lowest in the proposed model.

# Chapter one

## (1-1)Introduction

The topic of regression is one of the important statistical topics used in many scientific studies and has wide applications in many fields. classical regression, or what is sometimes called mean regression, is one of the important statistical methods that are used to study the relationship between explanatory variables **(X)** and response variable **(Y),** by estimating the conditional mean of the distribution of the response variable $E(Y/X)$. As is known, the classical regression analysis is based on the assumptions of the analysis, the most prominent of which is that the random errors are distributed independently, naturally, with a mean equal to zero, and a constant variance is $\sigma^2 \mathbf{I}$. Which $\varepsilon_i \sim N(\mathbf{0}, \sigma^2 I)$ .It is possible to find an estimate for the response variable by classical regression, if the values of the explanatory variables (X) are known, this helps us to find the values of the response variable (Y) to get the exact results. The model we choose should be consistent with the data available to the researcher in the best possible way. More recently, he demonstrated the quantitative regression technique, proposed by koenker and Bassett (1978), This method included a more comprehensive study of the relationship between the response variable **(Y)** and the explanatory variables **(X)**, by estimating the conditional denominator $Q_\theta(y/x)$ the differences in the distribution of the response variable under the $0 < \theta < 1$ .level rather than estimating the conditional expectation only $E\ (Y\ /\ X)$ In classical regression, the quantitative regression technique is used when estimating the relationship in different parts of the conditional distribution of the response variable. also,( QR) technology is suitable for a lot of data, because this technique does not require normality or symmetry in the data distribution. One of the most important advantages of QR technology is that it is not affected by the problem of heterogeneity, anomalous values, and skewness in the distribution. Also, this technique is important in detecting the problem of heterogeneity in simple regression by drawing. QR estimators are important in determining variables and estimating coefficients, and QR estimators are more powerful and efficient than OLS estimators. An important feature that adds immunity to the model is that the random error is not based on certain assumptions despite these important advantages of this technique. The problem of polylines remains one of .The problem of polylines remains an important proplem that clearly impact the estimators efficiency in of the estimators in the (QR) model, and this problem becomes more difficult when the response variable is of a finite type, as in the case of the Tobit quantile regression model. Since the seminal work of (James Tobin ( 1958) ) Tobit regression model (TR Model) has become very important in censored data at zero point, it is used in many sciences such as psychology science medicine , finance and social science, Etc. T.R. Model is focused on evaluating the relationship between the censored dependent variable and a set of independent variables. The use of Tobit's quantile regression model TQR. model. It is a good statistical model for assessing the relationship between the left-control dependent variable and a set of independent variables at specific quantile levels. Proposed by Powell (1986)) There are many independent variables within some of the TQR models. The left censored dependent variable has a different relationship with these independent variables. The independent variables may have little effect in the Tobit QReg model may have little effect. Therefore, some of the independent variables have risks in this model. So, the Tobin model is an extension of the **(QR)** regression model in investigating the relationship between the explanatory variables and the response variable, and when the response variable is of a finite type. Therefore, the Bayesian method for estimating the **(TQR)** model has the possibility of dealing with the issue of the big limitation in the response variable data in an efficient manner, as shown by many research and studies The **(TQR)** model is important in many applied research, such as medical, economic and other. The Bayesian method provides an efficient method for accurate inference, even in the case of small samples. It is also important to overcome some of the difficulties that accompany the process of estimating the parameters of the **(TQR)** model using the classical method. It can be said that the beginning of the use of the Bayesian style in the **(QR)** modality was created by Yu and Moyeed in (2001), this was followed by the use of this method in the **(TQR)** model by Yu and Stander for the first time in(2007), In fact, the use of the Bayesian method in the **(QR)** model is based on the assumption (regardless of the real distribution of the data) the asymmetric Laplace distribution **(ALD)** in formulating the possibility function of the model. Despite the analytical difficulties resulting in this method, Yu and Moyeed in (2001), showed the possible of using the Markov Chain Monte Carlo **(MCMC)** method in conducting the sampling from the subsequent conditional distributions and to facilitate the application of the biometric method in the **(QR)** model and**(TQR)** model. Kozumi and Kobayashi in( 2011), suggested using the mixed representation of the asymmetric Laplace distribution **(ALD)** in the possible function of the model this suggestion is easy to perform the sampling process from the subsequent conditional distributions, and it has become followed in many recent researches of **(QR)** and **(TQR)** models. Despite the importance of this proposal for estimating the parameters of**(QR)**models using the Bayesian method, the resulting subsequent functions were not analytic easy because of the complexities resulting in the possibility function. In (2003) the researcher (Tsionas) presented a study on the use of the Bayesian method in estimating the parameters of the **(QR)** model, in which the mixed representation of the asymmetric Laplace distribution **(ALD)** was used in the random error distribution of the model in order to speed up and increase the efficiency. And better than that used by (2001) Yu and Moyeed,where the **(Gibbs Sampling)** algorithm was used to complete the model estimation, and the results showed the efficiency of the method used compared to the methods used by (1986) (powell) and Alhamzawi.R method (2016). In (1996) he proposed( Tibshirani) Lasso's method for estimating the coefficients of the models, He showed that this characteristic is important because it tends to make these coefficients (which are unimportant in the model) exactly equal to zero, This characteristic is great importance in choosing the variables in the model and reaching an interpretable model that is more accurate in predicting. Excluding the uonimportant independent variables is a hard matter. Variables selection could be used to solve this problem (VS). Where, It has a strong ability for selecting excellent independent variables for regression models and avoid non-significant independent variables from this regression models, it has a high quality for creating regression models .Recently researchers have revealed a novel way for implementing variable selection in regression models. These approaches have good qualities and require little time to achieve (VS) because the process is automated. The lasso (least absolute shrinkage and selection operator) (Tibshirani,(1996) ), SCAD (Fan and Li,( 2001)), and elastic net approach are some of these methods. The Bayesian Lasso was described in the classic regression model by (Park and Casella, 2008). The majority of these techniques were used to TQR Model. (Alhamzawi, (2013)) used Bayesian techniques to propose adaptive lasso in TQR. Model. (Alhamzawi and Yu,( 2014)) suggested a Bayesian method for estimating coefficients in the TQR .Model using a g-prior distribution with ridge parameter. By using a Bayesian framework, (Alhamzawi, (2014)) suggested an elastic net penalty in TQR. Model To achieve Bayesian Lasso in regression models, most approaches in the field of Bayesian penalizing TQR. Model used the scale mixture of normal (SMN) prior distribution. But (Mallick and Yi (2014)) proposed a new formulation for attaining Bayesian lasso in a classical regression model using a scale mixture of uniform (SMU) prior distribution of the Laplace distribution . This idea was expanded by (Fadel Al-Hussaini (2017)) that using scale mixture of uniform (SMU) prior distribution with quantile regression via Bayesian framework. Also, (Fadel Al-Hussaini (2017)) using (SMU) prior distribution in TQR.Model , to obtaining a straightforward and efficient algorithm. (Flaih et al2020)). propose the new Bayesian Lasso , based on a new development of the hierarchical model and a combination of a new Laplace distribution measure. It's a combination of Rayleigh distribution and normal mixing. ( Remah Oday and Fadel Al-Hussaini (2021)) are proposed a new scale mixture of uniforms distribution mixing with standard exponential distribution on their variances in quantile regression model via Bayesian approach. In this paper , we propose Bayesian new lasso in TQR. Model through using a scale mixture of uniforms distribution mixed with standard exponential distribution on their variances to ensure the posterior distribution of parameter estimationis unimodal.

## (1-2)The Problem

Regression model analysis is a statistical method that uses the relationship between the response variable and predictor variables for modeling and making prediction . But many drawbacks associated with applying the traditional regression method ,such as the subset selection method and the ordinary least squares method that imposed pre-conditions are not met. These problems motivate the researchers to propose new method for analyzing the relationships, such as quantile regression which is considered as a comprehensive statistical method. Along with the elegant properties of quantile regression one can employ this type of regression models to study the more complex situation furthermore many filed of science contains limited response variable which required to be very Cleary to deal with that kind of data ,so the Tobit (left censored ) quantile regression model is a more flexible and robust model to be applled . Lasso Tobit quantile regression model can deal with the non-full rank matrix problem and this model can discover the irrelevant and relevant predictor variables and thereby exploring a parsimonious model (less predictor variables with more interpretability.

## (1-3)The Objectives

To propose a new hierarchical priors model based new reparametrized prior distribution that represents the double exponential distribution .

## (1-5)Literature review

Since Tobin (1958) groundbreaking work the, Tobit regression model (TRM) has seen a lot of use in recent literature as well as several practical applications in a variety of sectors such as medicine, biological sciences, finance and econometrics.Koenker and Bassett (1978), were proposed a new regression model called quantile regression (QR) model, this model is close to the classical regression model. At the same time , It is based on conditional quantiles function instead of conditional mean. QR model has a good property compared to other regression models. QR consider more robust against the outlier data. QR has capable to accommodating non-normal random error. Q Reg gives good inference until when violation of supposition is normal. QR model provides us with complete information about relationships between response variable and predictors variables (explanatory variables) Etc. QR has received much attention in many sciences, because it has attractive features compared to the other regression models. Powell (1986) proposed a new model that is mixing between quantile regression and Tobit regression model called tobit quantile regression model (TQR) . Many researchers are interested in TQR such as Hahn (1995) proposed a new method to compute the confidence intervals for TQR model by using bootstrapping percentile method (B.P.M). (B.P.M) is an efficient method for estimating confidence intervals to TQR. And he suggests some of the traditional estimation method for TQR. Buchinsky and Hahn (1998) Buchinsky and Hahn(1998 ) proposed a new estimation method for Tobit **(QR)** model with high efficiency, when sample size is small. This proposed method is attractive even when the data contain a big amount of censored Data. Therefore, this method considers more efficiency compared with a linear programing method. Tobit quantile regression and other modifications were suggested by Bilias et al. (2000) to enhance the bootstrap approach. TQR is a useful approach for handling data that has been left censored. On the other hand, certain TQR models have a lot of independent variables. As a result, there is a different relationship between these independent factors and the response variable with the left censor. Maybe a few of these independent variables occasionally have a tiny influence. Some independent variables have therefore become significant in the. It is challenging to constantly rule out these independent variables. The variables selection (VS) method might be used to tackle this problem. It performs well when estimating the coefficients of regression models, but (VS) performs better when selecting the active independent variables and avoiding the unfavorable independent variables for these regression models.. Tibshirani (1996), introduced a good method for variable selection called, Lasso (least absolute shrinkage and selection operator). In estimating the parameters, he indicated that this constraint has an important property in that it causes some of these parameters to be exactly equal to zero. This property is of great importance in selecting the variables in the model and achieving a model that is more understandable and more accurate in predicting. Additionally, Tibshirani (1996) proposed estimating Bayes by the

Lasso technique given the (Laplace) distribution's prior parameter distribution. A novel technique for measuring how well they match a quantum regression model was put forth by Machado and Koenker (1999). $R^2$ was shown the analog test by them. By examining the asymmetric Laplace distribution (ALD) for the error term, In order to make the process of estimating parameters easier, Gelfand and Smith invented the (Gibbs Sampling) method in 1990. This method is very important since it can be used to solve many different Bayesian inference issues. In (2001), Yu and Moayed created a model with an asymmetric Laplace distribution (ALD) regardless of the real distribution of the data. Actually, the procedure of choosing this function was just a hypothesis to connect the Bayesian estimation approach to the conventional method, which maximizes the likelihood based on the (ALD) Laplace distribution of error random. Additionally, they mentioned the idea of utilizing Markov Chain Monte Carlo to obtain the outcomes that. even in complex integrations, attaining the ensuing conditional distributions. In( 2005), Koenker proposed that quantile regression can be regarded as robust to the model that delivers distinct impacts of the predictor variable on a different level of the non-uniform quantum function. He also provided information on a number of techniques used to estimate parameters in (TQR). By adopting (ALD) as the error plane distribution in (2007), Geraci and Bottai suggested a new linear model for Bayesian quantile regression with random effects. Assuming an asymmetric Laplace distribution (ALD) and using (MCMC) for sampling in conditional post-distributions, Yu and

Stander (2007) examined the issue of estimating parameters of the (TQR) model from a Bayesian perspective. Bayesian Regression By combining the scale of the normal distribution and the exponential distribution, Lasso was first introduced by Park and Casella (2008) utilizing the Laplace distribution of his parameter. To estimate the parameters of the later distribution, a novel Gibbs sampling method and a new hierarchical model were developed in this study. Li and Zhu first developed lasso quantile regression as a technique to estimate and identifying variables in 2008. The formula for Lasso quantile Regression (LQR) .Griffin and Brown (2010) indicated that the variable selection problem provides the value of a tool to identify the pertinent variable that significantly influences the response variable and to obtain a more accurate and understandable model. Gibbs is a brand-new sampling technique for Bayesian quantile regression that Kozumi and Kobayashi introduced in (2011). They expressed the regression and depicted the asymmetric Laplace distribution (ALD) as a scale mixture of mixing the normal distribution with the exponential distribution. When estimating the parameters of the (QR) model using the Bayesian approach, Kozumi and Kobayashi (2011) produced a study that used a mixed representation of the warped asymmetric Laplace distribution (AL). ALD was defined as a mixture measure of combining the normal distribution with the exponential distribution as by improving the (Gibbs sampling) algorithm. By using Bayesian inference for the adaptive Lasso regression quantile regression and a newly created Gibbs sample approach to estimate the distribution parameter, Al-hamzawi et al.

(2012) introduced the variable selection problem. Following up on Park and Casella's (2008) work, Mallick and Yi (2014) created a novel Bayesian lasso by introducing the (ALD), which combines a uniform distribution with a gamma distribution (2,λ). new hierarchical model proposed as well a new Gibbs sampling algorithm Marasinghe proposed in (2014) that no distribution proposal for the error term is necessary for the (QR) model Alhamzawi (2015) introduced the Bayesian Lasso type models by assigning scale mixtures of normal with mean equal to zero and unknown variances in the model selection problem in quantile regression models. New (MCMC) algorithms were also applied for parameter estimation, and a new model selection criterion was proposed for the quantile regression. Al-Hamzawy researched longitudinal data with ordinal replies in 2016. employing a Bayesian ordinal quantile regression model with random effects has certain advantages. Site-derived effective new Gibbs sampling Asymmetric double exponential distribution scale of mixture representation. In( 2017), Fadel Al-Hussaini proposed a new and changing appreciation of Bayesian Check in lasso quantile regression by looking at the scale mixture that Suggested by Malik Wei (2014). The new simple and effective Gibbs sample Introduce the algorithm. Simulation was performed and real data analyzed It shows the performance of the proposed model . It has also been proposed (Fadhel Al-Hussaini (2017)) using a scale-a combination of a standardized pre-distribution (SMU) with quantitative regression via a Bayesian framework. Fadel Al-Husseini (2017) (SMU) also used the prior distribution in the TQR model to obtain a clear and efficient algorithm. In (2017), Kobayashi studied the $\theta^{th}$.Tobit quantitative regression from Bayesian perspective with endogenous variables over exogenous variables and that In definite quantities and this depends on the asymmetric exponential power To distribute as in distributing the error term. In 2018 , Al-husseini proposed using the Mallick and Yi (2014) scale mixture in the Tobit quantile regression from the Bayesian point of view, in this work variable selection problem has studied throug simulation study and real data analysis. In 2020, Flaih et al. propose the new Bayesian Lasso through a new development of the hierarchical model using a combination of a new measure of the Laplace distribution It is a mixture of normal mixing with Rayleigh distribution .In 2020, Flaih et at. A new mixture of distribution Uniforms mingle with distribution Standard exponential following a mathematical relationship and work Some algebraic steps to reach the mixture of the Laplace densitometer. (Flaih et al2020)). propose the new Bayesian Lasso , based on a new development of the hierarchical model and a combination of a new Laplace distribution measure. It's a combination of Rayleigh distribution and normal mixing. ( Remah Oday and Fadel Al-Hussaini (2021)) are proposed a new scale mixture of uniforms distribution mixing with standard exponential distribution on their variances in quantile regression model via Bayesian approach. In this study . In this thesis the our contribution proposed Bayesian new lasso in TQR. Model through using a scale mixture of uniforms distribution mixed with standard exponential distribution on their variances to ensure the posterior distribution of parameter estimation is unimodal . The our contribution this thesis studies the lasso Tobit quantile regression model. Quantile regression model analysis for left censored data (Tobit ) is very important in many fields of sciences since. It allows the researcher to explore a range of conditional quantities (Quantile functions) here with of conditional forms of unobserved and outliers individuals .Our main objective is to describe conditional quantile forms in Tobit regression by estimating the interested parameter from the lasso Bayesian theorem aspect . We introduced a new Bayesian hierarchical priors model to implement the Gibbs sampler algorithm for the Tobit quantile regression model .Also, we explain the new scale mixture of uniform mixing with standard exponential by derivative new formula for the double exponential prior distribution .We study the Bayesian estimation of quantile regression in the Tobit model using the MCMC and Gibbs sampler algorithm with two simulation scenarios from the results shows that the proposed lasso penalized model is a comparable model under different quantile level and different sample sizes totally , we illustrate the performance of the proposed lasso penalized method by analyzing real data that represent a sample of size (200) observations of Erythrocyte sedimentation rate a response variable with twenty predictor variables . The results show that the proposed model provides variable selection precedent compared with some examined regression models under different quantile regression ,the mean absolute error and the mean square error were the lowest in the proposed model.

# Chapter two

## (2-1) sammcery

In this chapter, we explain the Ordinary Least Squares Method, Linear Regression Model, Quantile Regression, Tobit Quantile Regression (TQReg), and Lasso Method. We also talk about a method Variable selection procedure and , the lasso (least absolute shrinkage and operator)

## (2-2) Ordinary Least Squares Method(OLS)

The (OLS) method is a function that attempts to study the relationship between the response variable and the explanatory variables, by estimating the conditional mean of the response variable, and this function is most commonly used by estimating the parameters of the estimated regression problem (OLS) (β) to reduce the objective cost as following :

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{argminRSS}(\boldsymbol{\beta_\theta})$$

$$= \boldsymbol{\varepsilon^t \varepsilon}$$

$$= (\boldsymbol{y} - \boldsymbol{x^t \beta_\theta})^{\boldsymbol{t}}(\boldsymbol{y} - \mathbf{x}^{\boldsymbol{t}} \boldsymbol{\beta_\theta})$$

$$= \|\boldsymbol{y} - \boldsymbol{x^t \beta_\theta}\|_2^2 \quad \textbf{............(2-1)}$$

Where **RSS** stands for Residual Sum of Squares, the **(OLS)**

estimator ($\widehat{\boldsymbol{\beta}}$) in **(2-4 )** is given by.

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{OLS}} = (\boldsymbol{X^t X})^{-1} \boldsymbol{X^t Y}$$

The **OLS** estimator is unbiased and have the smallest variance.

When $\mathbf{k} > \mathbf{n}$ and if the **)** $.\boldsymbol{X^t X}$ The multicollinearity problem arise in many types of data in real world phenomenon, because of this problem The **( OLS )** estimates show high variances for the parameter estimate .

## (2-3 ) Linear Regression Model

The topic of regression is one of the important statistical topics used in many scientific studies and it has wide applications and many fields. Ordinary linear regression is one of the important statistical methods used in studying the relationship between explanatory variables **X** and response variable **Y**. As we know that the analysis of normal regression is based on the analysis hypotheses that The most important of them is that the random error is normally distributed with a mean equal to zero and a $(\sigma^2 I)$ constant variance is $\varepsilon_{i} \sim N(0, \sigma^2 I)$

The regression model can be defined as follows:

$$Y = B_0 + B_1X_1 + B_2X_2 + \cdots + B_kX_k + \varepsilon_i \ldots\ldots\ldots\ldots\ldots (2-2)$$

Whereas :

$\varepsilon_i$: random error term which independent of **X** , and $\beta$.

***K*:** is the number of coefficient (parameter) of the regression mode predictor variables.

thus the multiple linear regression model is We can rewrite model **(2-2)** in matrix form as follows **:**

$$y_i = x_i^t \beta_\theta + \varepsilon_i \ldots\ldots\ldots (2-3)$$

**With:**

$$E(\varepsilon_i) = 0$$

$$var(\varepsilon_i) = \sigma^2 I$$

Regression analysis is important because it explains the link between the explanatory variables and the response variable and presents the issue of variable selection..

## (2-4)Quantile Regression

Demonstrated a quantile regression technique in studying the relationship between response variable and explanatory variables, by estimating conditional sections $Q_\theta(\mathrm{Y}|\mathrm{X})$, value the quantile

$0 < \theta < 1$.The quantitative regression technique in applied fields is suitable for a lot of data, as this technique does not require the normality or symmetry in the distribution of the data and is used when the relationship is to be estimated in different sections of the conditional distribution of the response variable. this technique is important in revealing the problem of heterogeneity variance in simple regression by drawing the quantitative function $\theta^{th}$ in terms of the inverse cumulative distribution function **(CDF).**

$$\mathbf{F (Y) = P (Y \leq y)},$$

let $Q_\theta$be the quantitative function defined as:

$$Q_\theta(\mathrm{Y}|\mathrm{X}) = \mathrm{F}^{-1}(\theta)\text{..........}\mathbf{(2\text{-}4)}$$

The conditional $\theta^{th}$ function **(2-4)** is for the $\theta^{th}$ quantile of random variable **Y** condition **X** We can rewrite the function **(2-4)** as follows :

$$\mathrm{Q}_\theta(\mathrm{Y}|\mathrm{X}) = \mathrm{argmin}\frac{1}{n}\left[\sum_{y_i \geq x_i\beta_\theta} \theta|y_i - x_i^t\beta_\theta| + \sum_{y_i < x_i\beta_\theta} (1-\theta)|y_i - x_i^t\beta_\theta|\right]$$

$$= \underbrace{arg\ min}_{\beta} \mathrm{E}[\rho(y_i - x_i^t\beta_\theta)]$$

$$= argmin\frac{1}{n}\sum_{y_i \geq x_i\beta_\theta} \rho_\theta(\varepsilon_i)\ \text{..................}\mathbf{(2\text{-}5)}$$

Where $\varepsilon_i$ is the error for observation $i$ and quantile $\theta$ , and $\rho_\theta(\varepsilon_i)$ is called the loss function or check function . the median regression is a special case of quantile regression if $\theta = 0.5$ ,

$Q_{0.5}(y|x)$ , Quantile regression provides an estimate of the relationship between predictive variables and the quantitative specification of the response variable suppose we have the following linear regression model.

$$y_i = x_i^t \beta_\theta + \varepsilon_i$$

Then based on $\theta^{th}$ quantile of $y_i$ then **(2-5)** can be rewritten as follows :

$$\widehat{\beta}_\theta = \mathbf{argmin} \sum_{i=1}^{n} \rho_\theta (y_i - x_i^t \beta_\theta)$$

Where as:

$$Q_\theta(y|x) = x_i^t \beta_\theta$$

$\beta_\theta$: Represents vector(k×1) of the parameters of quantile.

$x_i$: A vector (k) represents an explanatory variable.

That is, we aim to estimate the conditional quantile in the distribution of the response variable $Q_\theta(y/x) = x_i^t \beta_\theta$ instead of the conditional expectation in classical regression.The important feature of **(QR)**models, which gave the character of immunity to this model, is that the random error distribution in this model is not based on certain assumptions. Despite the importance of this technique and its unique advantages, the problem of polylinearity remains one of the important problems that have a clear impact on the efficiency of estimators in modeling. As it is known that the problem of multicollinearity is one of the important problems that a lot of research directs, which leads the researcher to the wrong conclusions about deleting some variables in the relevant model or viceversa.

## (2-5 )Tobit Quantile Regression (TQReg)

This model was presented for the first time by (1986) (Powell) to analyze the relationship between the response variable (limited variable) and the vector of the explanatory variables and in the entire conditional distribution of the response variable, so the TQR model represents an extension of the QR regression model in the investigation of the relationship between the explanatory variables and the response variable and when the response variable is of the finite type Therefore, the TQR model can be expressed in the following form:

$$\boldsymbol{y_i = \max(0, y^*) \quad y^* = x_i^t \beta_\theta + \varepsilon_i} \dots\dots\dots\dots\dots.. [\mathbf{2-6}]$$

where :
$\boldsymbol{y^*}$: is called  the latent variable is take unobserved observation
$\boldsymbol{\beta_\theta}$: are vector unknown parameters $(\beta_{\theta 1}, \beta_{\theta 2}, \dots..\beta_{\theta k})$,

$\mathbf{x_i}$: A vector (k) represents an explanatory variable.

$\boldsymbol{\theta}$ : is quantile level that recognized interval (0,1) $y_i$ is left-censored dependent variable at zero.The usual method for estimating the parameters of the TQR model can be achieved from the following equation:

$$\widehat{\boldsymbol{\beta}} = \underbrace{\boldsymbol{arg\,min}}_{\boldsymbol{\beta}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\rho_\theta}\,(\mathbf{y^* - \max\{0, y^*\}}) \dots\dots\dots [\mathbf{2-7}]$$

The  linear programming strategy can be used to solve the problem, this strategy proposed by (Koenker and D'Orey, (1987)). In the T.Q.R.Model, there are many methods for its parameter estimation. However, most of these techniques have flaws and are useless. (Koenker and Machado (1999)) and (Yu and Moyeed (2001)) proposed Bayesian approach in estimation of T.Q.R.Model parameters. The error term $e_i$ in T.Q.R.Model is very closed from asymmetric Laplace distribution (ALD). That it has probability density function (p.d.f) as follow;

$$f(\varepsilon_i|\mu,\sigma,\theta) = \frac{\theta(1-\theta)}{\sigma} \exp{-\rho_\theta}\left\{\left(\frac{\varepsilon_i-\mu}{\sigma}\right)\right\} \text{ ……………….} [2-8]$$

When: $\mu = 0$ and $\sigma = 1$ then, pdf to $\varepsilon_i$ is:

$$f(\varepsilon_i|\sigma,\theta) = \theta(1-\theta)\exp(-\rho_\theta\{(\varepsilon_i)\}) \text{ ………………………… } [2-9]$$

The random variable $(\varepsilon_i)$ is have :

$$E(\varepsilon_i) = \frac{1-2\theta}{\theta(1-\theta)}$$

and

$$\text{var}(\varepsilon_i) = \frac{1-2\theta+2\theta^2}{\theta^2(1-\theta)^2}$$

The joint distribution of

$$y = (y_1,\dots,y_n)' \text{ given } X = (x_1,\dots,x_n)'$$

is:

$$(y|X,\beta,\sigma,\theta) = \theta^n(1-\theta)^n\exp\left\{-\sum_{i=1}^{n}\rho\theta(y^* - \max\{0,x_i^t\beta_\theta + \varepsilon_i\}\right\} \text{ ……..} [2-10]$$

Via (Kozumi and Kobayashi, (2011)) Transformation ,the ALD takes scale mixture normal (SMN) distribution. Therefore ,the dependent variable in TQR. Model takes following equation

$$f(y^*|x_i^t,\theta,\beta_\theta,v_i) = \left[\frac{1}{\sqrt{4\pi v_i}}\right]^n e^{-\sum_1^n\frac{(y^*-x_i^t\beta_\theta-(1-2\theta)v_i)^2}{4v_i}} \text{ ……… } [2-11]$$

The equation (2-11) is very important in Gibbs samplers of posterior distributions for coefficient estimates of TQR Model

## (2-6) Lasso Method

The lasso kernel was proposed in 1996 by Tibshirani . The lasso method (least absolute shrinkage and selection operator) is a penal method that imposes a penalty function on the remaining sum of squares. The lasso estimator formula is as follows :

$$\hat{B}_{Lasso} = argmin(y - x\beta_{\theta})^{t}(y - x\beta_{\theta}) + \lambda \sum_{j=1}^{k} |\beta_j|$$

Where**:**

$\lambda \geq 0$ is the shrinkage parameters $\lambda \sum_{j=1}^{k} |\beta_j|$ is the penalty function

Also, the Lasso method works on selecting variables and estimating parameters. The Lasso method makes the non-important variable that does not affect the model, making it very close to zero. That is, it works on selecting variables to reach a regression model that is more explicable and more accurate in predicting . that is $\hat{B}_{Lasso}$ is sparse or not sparse estimator .

## (2-7) Variable selection procedure and the lasso

some independent variables have risks in this model. The exclusion of non-significant independent variables is difficult. Variable selection can be used to solve this problem (VS). As it has strong ability to select excellent independent variables for regression models and avoid bad independent variables from these regression models, it has high quality for creating regression models. The researchers gently revealed a new way to implement variable selection in regression models. These methods have good qualities and require little time to achieve VS because the process is automated. Among these methods, the lasso Tibshirani, (1996), (Fan and Li, (2001)), and the elastic network approach. The Bayesian Lasso was described in the classical regression model by (Park and Casella, 2008). The majority of these techniques were used in T.Q. R. Model (Al-Hamzawy, (2013)) Use

Bayesian techniques to suggest adaptive lasso in T.Q model. R. (Al-Hamzawy and Yu, (2014)) Bayesian method for estimating the coefficients in the Tobit model. Using the Bayesian framework, (Al-Hamzawi, (2014)) suggested an elastic net penalty in the T model . But (Mallick and Yi (2014)) proposed a new formulation to achieve Lasso Bayesian in the classical regression model using the scale-mix of the Uniform Distribution (SMU) of the Laplace distribution. This idea was extended by (Fadel Al-Hussaini (2017)) using a scale combination of a standardized (SMU) distribution with quantitative regression via a Bayesian framework. Fadel Al-Husseini (2017) also used (SMU) the pre-distribution in the T.Q.R model, to obtain a clear and effective algorithm. (Flei et al. 2020)). The new Bayesian Lasso proposal, based on a new development of the hierarchical model and a combination of the new Laplace distribution scale. It is a combination of Rayleigh distribution and normal mixing. (Ramah Uday and Fadel Al-Hussaini (2021)) propose a new combination of the combination of the uniform distribution with the standard exponential distribution on their variances in a quantitative regression model via Bayesian approach. In this paper, we propose a new Bayesian lasso in the T.Q.R model by using a scale mixture of the uniform uniform distribution mixed with the standard exponential distribution over their variances to ensure that the post-distribution of the parameter estimate is one mode.

# Chapter three

## (3-1) sammcery

In this chapter, some basic concepts related to research are presented. The importance of the Laplace distribution in the Bayesian method applied in the research was a simple summary of this distribution in its twisted and symmetric form, and the identification of some important shapes in the mixed representation of this distribution. The study also included the pre-hierarchical distribution method and the conclusions .

## ( 3-2 ) Laplace distribution

The Laplace distribution is one of the important continuous probability distributions named after the French mathematician Pierre-Simon Laplace. The Laplace distribution in it is various forms (twisted and symmetric) and its different formula. It has wide applications in various researches, medical and economic studies, engineering and other important research and studies. The Laplace distribution is of great importance in research and fortified studies as an alternative to the normal distribution. Features a location parameter $p \in (-\infty, \infty)$ and the scale parameter $(\theta > 0)$ This distribution is denoted by the symbol $L(p, \theta)$**.** The probability density function of the traditional Laplace distribution can be expressed by the following formula:

$$f(x/p, \theta) = \frac{1}{2\theta} exp\left(-\frac{|x-p|}{\theta}\right) \quad \text{........(3-1)}$$

**Whare :** $-\infty < x < \infty$

Also, the cumulative distribution function can be expressed in the following formula:

$$f(x/p, \theta) = \begin{cases} \frac{1}{2} exp\left(-\frac{|x-p|}{\theta}\right) & , x \leq 0 \\ 1 - \frac{1}{2} exp\left(-\frac{|x-p|}{\theta}\right) & x < 0 \end{cases} \quad \text{.......(3-2)}$$

The mean and variance of the Laplace distribution are:

$$\boldsymbol{E}(\boldsymbol{x}) = \boldsymbol{\theta} \quad \textbf{.....(3-3)}$$

$$\boldsymbol{V}(\boldsymbol{x}) = \boldsymbol{2\theta^2} \quad \textbf{......(3-4)}$$

And when the parameter of the site is equal to zero $\boldsymbol{p} = \boldsymbol{0}$ for the equation**(3-1)** It can be expressed in the following form:

$$\boldsymbol{f}(\boldsymbol{x}/\boldsymbol{\lambda}) = \frac{\boldsymbol{\lambda}}{\boldsymbol{2}} \boldsymbol{exp}(-\boldsymbol{\lambda}|\boldsymbol{x}|) \quad \textbf{.......(3-5)}$$

$\boldsymbol{\lambda} = \boldsymbol{1}/\boldsymbol{\theta} \quad , \mathbf{p} = \mathbf{0}$

It is worth noting that the probability density function of the Laplace distribution can be expressed in different forms of mixed representation, and this representation is of great importance in research and studies, especially those that use the Bayesian hierarchical method because of this representation of its importance in facilitating the estimation process and increasing its efficiency. As it is known that many phenomena that need statistical analyzes face asymmetry in the data and the skewed or asymmetric Laplace distribution is one of the important distributions in research and asymmetric statistical studies. An important feature of the skewed Laplace distribution is that it is considered to be a tick exponential family, which has the important characteristic that the potential estimations in this family distributions are always consistent. There are many forms of the probability density function**(pdf)**in the**(AL)**distribution that are important in scientific applications:

$$\boldsymbol{f_\theta}(\boldsymbol{x}) = \boldsymbol{\theta}(\boldsymbol{1} - \boldsymbol{\theta})\boldsymbol{exp}\{-\boldsymbol{\rho_\theta}(\boldsymbol{x})\} \quad \textbf{......(3-6)}$$

$\mathbf{0 < \theta < 1}$ , is skew parameter where it $\boldsymbol{\rho_\theta}$ can be expressed in the followingform:

$$\rho_\theta(\mathrm{x}) = \frac{|x|+(2\theta-1)x}{2} = x(\theta - I(x < 0)) \quad \text{......(3-7)}$$

The arithmetic mean and variance of the skewed Laplace distribution**(AL)**can be expressed in the probability function**(3-6)** the following formula:

$$E(x) = (1-2\theta)/\theta(1-\theta) \quad \text{.................(3-8)}$$

$$V(x) = (1-2\theta+2\theta^2)/\theta^2(1-\theta)^2 \quad \text{......(3-9)}$$

(kotz,et2001) Show that the Laplace distribution can be represented by different forms of mixed distributions and summarize these distributions in a table showing these distributions. In 2011, Kozumi and Kobayashi used mixed representation in the Laplace distribution with the formula:

Let **x** is a standard normal variable with:

$$F(z) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2}\right) \qquad -\infty < \mathrm{x} < \infty \quad \text{........(3-10)}$$

Let **e** is a standard exponential variable :

$$F(z) = exp(-z) \quad \text{........(3-11)} \;, \quad e > 0$$

Then $$\varepsilon = \varepsilon_1 e + \varepsilon_2\sqrt{ez} \quad \text{......(3-12)}$$

Has (ALD) where :

$$\varepsilon_1 = (1-2\theta)/\theta(1-\theta), \qquad \varepsilon_2^2 = 2/\theta(1-\theta)$$

## (3-3)likelihood function of Tobit QReg :

Offer (Koenker, 2005) algorithms for TQR parameter estimation. When the left-hand censored response variable has a significant number or amount of data at the end, some of these techniques are useless Stander and Yu (2007). The convergence between the function) and the Laplace distribution Alfer symmetric (if the error term) and the distribution of the (k) probability density function was suggested by Konker and Machodom (1999), Yu and Moyeed (2001), and they take the form:

$$f(\varepsilon_i/\mu,\lambda,\theta) = \frac{\theta(1-\theta)}{\lambda} exp - \rho_\theta\left\{\left(\frac{\varepsilon_i-\mu}{\lambda}\right)\right\} \dots\dots\dots\dots (3-13)$$

if $\mu = 0$ and $\lambda = 1$ then,the probability density function (pdf) to $\varepsilon_i$ is:

$$f(\varepsilon_i/\lambda,\theta) = \theta(1-\theta)\exp((-\rho_\theta\{(\varepsilon_i)\}) \dots\dots\dots\dots\dots (3-14)$$

With:

$$\mathrm{E}(\varepsilon_i) = \frac{1-2\theta}{\theta(1-\theta)}$$

$and$

$$var(\varepsilon_i) = \frac{1-2\theta+2\theta^2}{\theta^2(1-\theta)^2}$$

$\rho_\theta(.)$ Is the check (loss) function . the jont distribution of

$y = (y_1, \dots, y_n)'$ given $x = (x_1, \dots, x_n)'$ **is:**

$$f(y^*\Big/x,\beta_\theta,\lambda,\theta) = \theta^n(1-\theta)^n exp\left\{-\sum_{i=1}^{n}\rho_\theta(y^* - max\{0, x_i^t\beta_\theta + \varepsilon_i\}\right\} \dots\dots (3-20)$$

The use of (ALD)directly leads to very difficult integrations, but Kozomi and Kobayashi (2011) this is facilitated by a suggestion that can be formulated (ALD) as a distribution (SMN) (scale mixture of standard normal).

$$y_i = max(0, y^*) \qquad , i=1,\ldots,n$$
$$y^*/\beta_\theta, v_i \sim N(x_i^t\beta_\theta + (1-2\theta)v_i, 2v_i) \ldots\ldots\ldots (3-15)$$

Given the suggestion Kozumi and Kobayshi (2011) the distribution of error $\varepsilon_i$ has a distribution with (SMN) a mean $(1-2\theta)$and a variance $(2\theta^{-1}v_i)$where the distribution of eror $\varepsilon_i$ takes the following formula:

$$\varepsilon_i = Zv_i + \theta\sqrt{v_i}\epsilon_i$$

Where $v_i$ it is the exponential distribution with it is parameter $\theta(1-\theta)$ and $\epsilon_i$ it is the standard normal distribution with mean (SMN) with mean (0) and variance (1):

$$E(\varepsilon_i) = ZE(v_i) + \theta\sqrt{E(v_i)}\,E(\epsilon_i)$$

$$E(v_i) = \frac{1}{\theta(1-\theta)}\text{ , } E(\epsilon_i) = 0$$

$$E(\varepsilon_i) = Z\frac{1}{\theta(1-\theta)} + \theta\sqrt{\theta(1-\theta)}\ 0$$

$$\left[\frac{(1-2\theta)}{\theta(1-\theta)} = Z\frac{1}{\theta(1-\theta)}\right] \div \left[\frac{1}{\theta(1-\theta)}\right]$$

$$Z = \frac{\frac{(1-2\theta)}{\theta(1-\theta)}}{\frac{1}{\theta(1-\theta)}}$$

$$Z = 1-2\theta$$

$$var(\varepsilon_i) = Z^2 var(v_i) + \theta^2\sqrt{var(v_i)}\,var(\epsilon_i)$$

Where as : $Z^2 = (1-2\theta)^2 \text{ ,} var(v_i) = \frac{1}{\theta^2(1-\theta)^2}$

$$\sqrt{var(v_i)} = \frac{1}{\theta(1-\theta)}\ and\ var(\epsilon_i) = 1$$

$$\frac{(1-2\theta+2\theta^2)}{\theta^2(1-\theta)^2} = (1-2\theta)^2\frac{1}{\theta^2(1-\theta)^2} + \theta^2\frac{1}{\theta(1-\theta)}1$$

$$\frac{1-2\theta+2\theta^2}{\theta^2(1-\theta)^2} - \frac{(1-2\theta)^2}{\theta^2(1-\theta)^2} = \theta^2\frac{1}{\theta(1-\theta)}$$

$$\frac{(1-2\theta+2\theta^2)-(1-4\theta+4\theta^2)}{\theta^2(1-\theta)^2} = \theta^2\frac{1}{\theta(1-\theta)}$$

$$\frac{(1-2\theta+2\theta^2)-1+4\theta-4\theta^2}{\theta^2(1-\theta)^2}=\theta^2\frac{1}{\theta(1-\theta)}$$

$$\frac{2\theta-2\theta^2}{\theta^2(1-\theta)^2}=\theta^2\frac{1}{\theta(1-\theta)}$$

$$\theta^2=\frac{\frac{2}{\theta(1-\theta)}}{\frac{1}{\theta(1-\theta)}}$$

$$\theta^2=2$$

Here $Z=(1-2\theta)$

And $\theta^2=2$

Then $\varepsilon_i=(1-2\theta)v_i+\sqrt{2v_i}\epsilon_i$ , $\varepsilon_i \sim N[(1-2\theta)v_i\,;2v_i]$

At the suggestion of Kozumi and Kobayashi (2011) we can rewrite (ALD) the random error of (SMN) and insert the exponential mixture of standard the following hierarchy is produced:

$$y_i=max\{0,y^*\}$$
$$y^*=x_i^t\beta_\theta+\varepsilon_i$$
$$y^*=x_i^t\beta_\theta+(1-2\theta)v_i+\sqrt{2\theta^{-1}v_i}\epsilon_i$$
$$y^*/\beta_\theta,v_i \sim N(x_i^t\beta_\theta+(1-2\theta)v_i\,,2v_i)$$

The probability density function of latent variable $y^*$ is :

$$f(y^*/x_i^t,\theta,\beta_\theta,v_i)=\frac{1}{\sqrt{4\pi v_i}}exp-\frac{[y^*-(x_i^t\beta_\theta+(1-2\theta)v_i)]^2}{4v_i}$$

The likelihood function of the probability density function $(f(y^*/\beta_\theta,v_i))$ is :

$$f(y^*/x_i^t\,,\theta,\beta_\theta,v_i)=\left[\frac{1}{\sqrt{2\pi v_i}}\right]^n exp^{-\sum_{i=1}^{n}\frac{[y^*-(x_i^t\beta_\theta-(1-2\theta)v_i)]^2}{4v_i}}$$

Where $v_i$ an exponential distribution is distributed with an average $\theta(1-\theta)$ and the equation **(3-21)** is an important equation for building Gibbs samples for distributions to estimate the coefficients in (TQR) the later model.

## (3-4)Bayesian Lasso quantile Regression For Mutations:

Suppose that we have the following quantile regression model with response variable $\boldsymbol{y_i}$ and vecter of predictors $\boldsymbol{x_i}\ of\ \boldsymbol{k \times 1}$ .Also suppose that the quantile function of $\boldsymbol{\theta^{th}}$is :

$$\boldsymbol{Q_\theta(y/x) = x_i^t\beta_\theta} \qquad \text{, i=1,…,n …………(3-16)}$$

**Where :**

$$\boldsymbol{\varepsilon_i \sim ALD(\lambda, \theta)} \text{ that is:}$$

$$\boldsymbol{f(\varepsilon_i/\lambda) = \lambda\theta(1-\theta)exp\{-\lambda\rho_\theta(\varepsilon_i)\}} \text{……….(3-17)}$$

And suppose that likelihood of $\boldsymbol{y_i}$, **given** $\boldsymbol{x_i}$ **i=1,….,n is:**

$$\boldsymbol{f(y^*/x_i, \beta_\theta, \lambda) = \lambda\theta(1-\theta)exp\{-\lambda \textstyle\sum_{i=1}^{n} \rho_\theta(\varepsilon_i)\}}$$

$$\boldsymbol{= \lambda\theta(1-\theta)exp\left\{-\lambda\sum_{i=1}^{n}\rho_\theta(y^* - x_i^t\beta_\theta)\right\}}$$

$$\boldsymbol{= \lambda\theta(1-\theta)exp\left\{-\lambda\sum_{i=1}^{n}\rho_\theta(y^* - Max\{0, x_i^t\beta_\theta + \varepsilon_i\})\right\}}$$

By following liet (2010) , Kozumi and Kobayashi (2011) and Benoit et al (2013).the response variable $\boldsymbol{y_i}$ can be view as :

$$\boldsymbol{\varepsilon_i = \xi_1\lambda^{-1}t_i + \xi_2\lambda^{-1}\sqrt{t_i}\epsilon_i} \text{………(3-18)}$$

Where :

$$\boldsymbol{t_i \sim standard\ exponential} \text{ and } \boldsymbol{\epsilon_i \sim standard\ normal}$$

$$\boldsymbol{\xi_1 = \frac{1-2\theta}{\theta(1-\theta)}} \text{ …………(3-19)}$$

And

$$\boldsymbol{\xi_2 = \sqrt{\frac{2}{\theta(1-\theta)}}} \text{ ……….(3-20)}$$

By letting $(\boldsymbol{t_i = \lambda^{-1} v_i})$ then we can say that $\boldsymbol{t_i}$ as exponential distribution with parameter $\left(\frac{1}{\theta}\right)$ .then the formula **(3-18)** can be rewritten as follow :

$$\boldsymbol{y_i = x_i^t \beta_\theta + \xi_1 v_i + \lambda^{-\frac{1}{2}} \xi_2 \sqrt{v_i} \epsilon_i}$$

Then the hierarchical structure model with $\boldsymbol{v_i , \epsilon_i}$ $_{\boldsymbol{,i=1,\dots,n}}$ is :

$$\boldsymbol{y_i = x_i^t \beta_\theta + \xi_1 v_i + \lambda^{-\frac{1}{2}} \xi_2 \sqrt{v_i} \epsilon_i}$$

$$\boldsymbol{v_i / \theta \sim \theta^n exp\left\{-\theta \sum_{i=1}^{n} v_i\right\}}$$

$$\boldsymbol{\epsilon_i \sim \left(\frac{1}{\sqrt{2\pi}}\right)^n exp\left\{-\frac{1}{2}\sum_{i=1}^{n} \epsilon_i^2\right\}}$$

## (3-5)The Bayesian Hierarchical model:

By following Li and Zhu (2008) the lasso quantile penalized regression solution is :

$$\boldsymbol{\hat{\beta} = argmin \sum_{i=1}^{n} \rho_\theta (y^* - x_i^t \beta_\theta) + \lambda \sum_{j=1}^{k} |\beta_j| \dots \dots (3-21)}$$

The Bayesian lasso quantile regression based on **(3-21)** required to impose prior distribution for $\beta$ following Tibshirani (1996) the prior distribution of $\boldsymbol{\beta}$ is a Laplace density in lasso quantile regression model is :

$$\boldsymbol{\pi(\beta / \theta , \lambda) = \frac{\theta\lambda}{2} exp(-\theta\lambda|\beta|) \dots \dots (3-22)}$$

Now we can write the prior **(3-22)** as scale mixture of uniform distribution mixing with standard exponential distribution as in the following propostion see Mallick and Yi (2014) for more details .

$$\boldsymbol{\exp(-\theta\lambda|x|) = \int_{\sigma > \theta|x|} \lambda \exp(-\lambda\sigma). d\sigma}$$

$$\boldsymbol{\frac{\theta\lambda}{2} exp(-\theta\lambda|x|) = \int_{\sigma > \theta|x|} \frac{\vartheta\lambda}{2} \lambda exp(-\lambda\sigma). d\sigma}$$

$$\lambda\sigma = \omega \Rightarrow \ \sigma = \frac{\omega}{\lambda}$$

$$d\sigma = \frac{1}{\lambda}.d\omega$$

$$= \int_{\omega > \vartheta\lambda|x|} \frac{\theta\lambda}{2} \lambda exp\left(-\lambda\frac{\omega}{\lambda}\right)\frac{1}{\lambda}.d\omega$$

$$\frac{\theta\lambda}{2} exp(-\theta\lambda|x|) = \int_{\omega > \vartheta\lambda|x|} \frac{\theta\lambda}{2} exp(-\omega).d\omega \ldots\ldots\ldots(3-23)$$

Now by letting $x = \beta$ in (3-27) we get

$$\frac{\theta\lambda}{2} exp(-\theta\lambda|\beta|) = \int_{\omega > \theta\lambda|\beta|} \frac{\theta\lambda}{2} exp(-\omega).d\omega \ldots\ldots.(3-28)$$

Hence ,the Bayesian hierarchical model is :

$$\left\{\begin{array}{c} y^{*} = x_i^t\beta_\theta + \xi_1 v_i + \lambda^{-\frac{1}{2}}\xi_2\sqrt{v_i}\epsilon_i \\ v/\theta \sim \theta^n exp\left(-\theta\sum_{i=1}^{n} v_i\right) \\ \epsilon_i \sim \left(\frac{1}{\sqrt{2n}}\right)^n exp\left(-\frac{1}{2}\sum_{i=1}^{n}\epsilon_i^2\right) \\ \beta/\theta,\lambda \sim unifor\left(-\frac{1}{\theta\lambda},\frac{1}{\theta\lambda}\right) \\ \omega \sim standard\ exponentioal \\ \theta \sim \theta^{a-1}exp(-b\theta) \\ \lambda \sim \lambda^{c-1}(-d\lambda) \end{array}\right\} \ldots\ldots\ldots\ldots\ldots\ldots(3-24)$$

## (3-6) The Gibbs Sampling For The Lasso Tobit Quantile Regression Model

1-The full conditional posterior distribution $\mathbf{f(y^*/x, v, \beta, \theta, \lambda, \omega)}$ in lasso regularization quantile regression is

$$f(y^*/x, v, \beta_\theta, \theta, \lambda, \omega) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\lambda^{-1}\xi_2^2 v_i}} exp\left[-\frac{(y^* - x_i^t\beta_\theta - \xi_1 v_i)^2}{2\lambda^{-1}\xi_2^2 v_i}\right] \ldots\ldots\ldots.(3-25)$$

Which is normal distribution with mean $(\mathbf{x}_i^t\boldsymbol{\beta}_\theta + \xi_1 \mathbf{v}_i)$and variance $(\lambda^{-1}\xi_2^2\mathbf{v}_i)$.

2-The full conditional posterior distribution $\mathbf{f}(\mathbf{v}_i/\mathbf{x},\mathbf{y},\boldsymbol{\beta},\boldsymbol{\omega},\boldsymbol{\theta},\boldsymbol{\lambda})$ is :

$$f(v_i/x,y,\beta,\omega,\theta,\lambda)\alpha f(y/x,v,\beta,\omega,\theta,\lambda)\pi(v_i/\theta)$$

$$\alpha \frac{1}{\sqrt{2\pi\lambda^{-1}\xi_2^2 v_i}} exp\left[-\frac{(y^* - x_i^t\beta_\theta - \xi_1 v_i)^2}{2\lambda^{-1}\xi_2^2 v_i}\right] exp(-\theta v_i)$$

$$\alpha \frac{1}{\sqrt{v_i}}\left[-\frac{\left((y^* - x_i^t\beta_\theta) - \xi_1 v_i\right)^2}{2\lambda^{-1}\xi_2^2 v_i}\right] exp(-\theta v_i)$$

$$\alpha \frac{1}{\sqrt{v_i}}\left[exp\left(-\frac{(y^* - x_i^t\beta_\theta)^2}{2\lambda^{-1}\xi_2^2 v_i}\right)\right] exp\left(\frac{2(y^* - x_i^t\beta_\theta)\xi_1 v_i}{2\lambda^{-1}\xi_2^2 v_i}\right)$$

$$.\,exp\left(-\frac{\xi_1^2 v_i^2}{2\lambda^{-1}\xi_2^2 v_i}\right).\,exp(-\theta v_i)$$

$$\alpha \frac{1}{\sqrt{v_i}} exp\left(-\frac{\lambda(y^* - x_i^t\beta_\theta)^2 v_i^{-1}}{2\xi_2^2}\right) exp\left(-\frac{\lambda\xi_1^2 v_i}{2\xi_2^2}\right) exp(-\theta v_i)$$

$$\alpha \frac{1}{\sqrt{v_i}} exp\left[-\frac{1}{2}\left[\left(\frac{\lambda\xi_1^2}{\xi_2^2} + 2\lambda\right) v_i + \frac{\lambda\left(y^* - x_i^t\beta_\theta\right)^2}{2\xi_2^2} v_i^{-1}\right]\right] \ldots\ldots\ldots (\mathbf{3\text{-}26})$$

the full conditional distribution of $(\mathbf{v}_i)$is **(GIG) (Generalized Inverse Gaussian)**

3-The full conditional posterior distribution of $\mathbf{f}(\boldsymbol{\omega}/\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\theta})$ is:

$$\mathbf{f}(\boldsymbol{\omega}/\boldsymbol{\beta},\boldsymbol{\lambda})\boldsymbol{\alpha}\,\boldsymbol{\pi}(\boldsymbol{\beta}/\boldsymbol{\omega},\boldsymbol{\lambda},\boldsymbol{\theta}).\boldsymbol{\pi}(\boldsymbol{\omega})$$

$$\alpha\, \pi_{j=1}^{n} \exp(-\omega_j) I\{\omega_j > \theta\lambda|\beta_j|\} \ldots\ldots\ldots (3-27)$$

The full conditional posterior distribution of $\boldsymbol{\omega}_j$ is left truncated standard exponential

4-The full conditional posterior distribution of $\boldsymbol{\beta}$ is

the posterior distribution of $\boldsymbol{\beta}$ is directly can be found by following Kozumi and Koboyashe (2011) .

$$\boldsymbol{\beta}_q/\mathbf{y},\boldsymbol{\omega}\sim\mathbf{N}(\widehat{\boldsymbol{\beta}}_q,\widehat{\mathbf{C}_q})$$

Where $C_q^{n-1} = \sum_{i=1}^{n} \frac{x_i x_i^t}{(\lambda^{-1}\xi_2^2 v_i)} + [var(\beta_{prior})]^{-1}$

**And** $\widehat{\beta_q} = \hat{C}_q \left[\sum_{i=1}^{n} \frac{x_i(x_i^t\beta_\theta + \xi_1 v_i)}{(\lambda^{-1}\xi_2^2 v_i)} + var(\beta_{prior}) * mean(\beta_{prior})\right]$

Form the hierarchal modal **(3-29)** the prior distribution of **.**

$$\beta_j \sim uniform\left(-\frac{1}{\theta\lambda}, \frac{1}{\theta\lambda}\right)$$

Then ,we have the following multivariate normal posterior distribution for $\boldsymbol{\beta}$ with mean $\widehat{\boldsymbol{\beta}_q}$ and variance $\widehat{\boldsymbol{C}_q}$ :

$$C_q^{\wedge -1} = \sum_{i=1}^{n} \frac{x_i x_i^t}{\lambda^{-1}\xi_2^2 v_i} + \left(\frac{1}{3\theta^2\lambda^2}\right)$$

$$\widehat{C_q} = \left[\sum_{i=1}^{n} \frac{x_i x_i^t}{\lambda^{-1}\xi_2^2 v_i}\right]^{-1} + 3\theta^2\lambda^2$$

This is the variance of $\boldsymbol{\beta}_q$ and the mean of $\boldsymbol{\beta}_q$ is defined as follows :

$$\widehat{\beta_q} = \widehat{C_q}\left[\sum_{i=1}^{n} \frac{x_i(x_i^t\beta_\theta + \xi_1 v_i)}{\lambda^{-1}\xi_2^2 v_i} + var(\beta_{prior}) * mean(\beta_{prior})\right]$$

The $\boldsymbol{\beta}_q$distribution is the multivariate normal with mean $\widehat{\beta_q}$ and variance $\hat{C}_q$ **.**

$$\boldsymbol{\beta}_q/\boldsymbol{y}, \boldsymbol{\omega} \sim \textbf{\textit{multivariate normal}}[\widehat{\boldsymbol{\beta}}_q, \widehat{\boldsymbol{C}}_q] \text{.........(3-28)}$$

5. The full conditional posterior distribution of $\boldsymbol{\theta}$ is :

$$f(\boldsymbol{\theta}/\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\beta}, \boldsymbol{\lambda})\,\boldsymbol{\alpha}\, \boldsymbol{f}(\boldsymbol{y}/\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\lambda})\boldsymbol{f}(\boldsymbol{v}/\boldsymbol{\theta})\boldsymbol{\pi}(\boldsymbol{\theta})$$

$$\alpha \left(\frac{1}{\sqrt{2\pi\lambda^{-1}\xi_2^2 v_i}}\right)^n \exp\left[-\frac{1}{2}\sum_{i=1}^{n} \frac{\left((y^* - x_i^t\beta_\theta)\xi_1 v_i\right)^2}{\lambda^{-1}\xi_2^2 v_i}\right] * \theta^n \exp\left(-\theta\sum_{i=1}^{n} v_i\right) * \theta^{a-1}\exp(-b\theta)$$

$$\alpha\, \theta^{\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n} \frac{\left((y^* - x_i^t\beta_\theta)\xi_1 v_i\right)^2}{\lambda^{-1}\xi_2^2 v_i}\right] * \theta^n \exp\left(-\theta\sum_{i=1}^{n} v_i\right) * \theta^{a-1}\exp(-b\theta)$$

$$\alpha\ \theta^{\frac{n}{2}}\theta^{n}\theta^{a-1}exp\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{\left((y^{*}-x_{I}^{t}\beta_{\theta})\xi_{1}v_{i}\right)^{2}}{\lambda^{-1}\xi_{2}^{2}v_{i}}\right]exp\left(-\theta\sum_{i=1}^{n}v_{i}\right)exp(-b\theta)$$

$$\alpha\ \theta^{\frac{n}{2}+n+a-1}exp\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{\lambda\left((y^{*}-x_{i}^{t}\beta_{\theta})\xi_{1}v_{i}\right)^{2}}{\xi_{2}^{2}v_{i}}\right]exp\left(-\theta\sum_{i=1}^{n}v_{i}\right)exp(-b\theta)$$

$$\alpha\ \theta^{a+\left(\frac{3n}{2}\right)-1}exp-\left[\theta\left[\sum_{i=1}^{n}\left(\frac{\left(y^{*}-x_{i}^{t}\beta_{\theta}-\xi_{1}v_{i}\right)^{2}}{2\xi_{2}^{2}v_{i}}+v_{i}\right)+b\right]\right] \text{........(3-29)}$$

The full conditional distribution of $\theta$ is a gamma distribution.

6-The full conditional distribution of $\lambda$ is :

$$f(\lambda/\beta)\alpha\ \pi(\beta/\theta,\lambda)*\pi(\lambda)$$
$$\alpha\ \pi_{j=1}^{k}(\lambda_{j})^{c-1}\exp(-d\lambda_{j})$$
$$\alpha\ (\lambda_{j})^{k+c-1}\exp\left(-d\sum_{i=1}^{k}\lambda_{j}\right)\pi_{j=1}^{k}I\left\{\lambda_{j}<\frac{\omega_{J}}{\theta_{j}|\beta_{j}|}\right\}\ldots\ldots(3-30)$$

The full conditional distribution of $\lambda_j$ is truncated gamma distribution.

## (3-7) The Bayesian lasso Tobit quantile model computation algorithm

We are sampling the following parameters and variables, Gibbs sampling algorithm with giving some initial values for $y, \beta, v_i, \omega, \theta$ and $\lambda$ is difined by the following steps.

1-Sampling the response variable $y$ :In this step we generate $y$ from normal distribution with mean $(x_i^t\beta_\theta + \xi_1 v_i)$and variance $(\lambda^{-1}\xi_2^2 v_i)$.

2-Sampling $\beta$: In this step we generate $\beta$ orom multivariate normal $(\widehat{\beta_q},\ \widehat{C}_q)$ where.

$$\widehat{\beta_q}=\left(\left[\sum_{i=1}^{n}\frac{x_i x_i^t}{\lambda^{-1}\xi_2^2 v_i}\right]^{-1}+3\theta^2\lambda^2\right)\left[\sum_{i=1}^{n}\frac{x_i(x_i^t\beta_\theta+\xi_1 v_i)}{\lambda^{-1}\xi_2^2 v_i}\right] \textbf{ and}$$

$$\widehat{C_q} = \left[\sum_{i=1}^{n} \frac{x_i x_i^t}{\lambda^{-1}\xi_2^2 v_i}\right]^{-1} + 3\theta^2\lambda^2$$

3-Sampling $\theta$:In this step we generate $\theta$ fromgamma distribution with shape parameter $\left(a + \frac{3n}{2}\right)$ and rate parameter $\left[\frac{1}{2} \cdot \frac{\left(y^* - x_i^t\beta_\theta - \xi_1 v_i\right)^2}{\xi_2^2 v_i} + v_i\right] + b$

4-Sampling: In this step we generate $\lambda$ from truncated gamma distribution with shape parameter $(k + c)$ and rate parameter $d$ .

5-Sampling : In this step we generate $v_i$ from (**GIG)** (Generalized Inverse Gaussian) with parameters

$$\left(\frac{\theta\xi_1^2}{\xi_2^2} + 2\theta\right) \text{ and } \frac{\theta\left(y^* - x_i^t\beta_\theta\right)^2}{\xi_2^2} .$$

6-Sampling $\omega$: In this step we generate $\omega_j$ from left truncated standard exponential

$$\omega_j = \omega_j^* + \theta\lambda\left|\beta_j\right|$$

And $\omega_j^*$ is generated from standard exponential distribution

# Chapter Four

# Simulation and Real Data

## (4 -1)Simulation study

In this part of our study, we evaluate the performance of our proposed method using simulated scenarios. The our proposed method (**NBLTQR**) compared with Bayesian and non-Bayesian lasso Tobit quantile regression (T.Q regression). The non-Bayesian T Q Regression is introduced by (Powell, 1986). Via employing package quantreg that is introduced by (Koenkers, 2011) through function (crq). The new Bayesian lasso in (T Q Regression) referred to (New B Tobit Q Reg) that proposed by (allhuseini, 2017) . To preferring between used methods two criteria has been employed ,firstly mean absolute error referred (MAE) and standard deviation referred (S.D) . We consider four choices of quantile levels $\theta = 0.25, \theta = 0.5, \theta = 0.75 \text{ and } \theta = 0.90$ .For each simulation scenario . We generated the random error $(\varepsilon_i, i = 1, 2, \dots, n)$ from three different distributions ,firstly: standard normal distribution $(\varepsilon_i \sim N(0,1)$,secondly : Normal distribution with mean (2) and variance (9) $(\varepsilon_i \sim N(2,9))$ **, thirdly (t)** distribution with five degrees of freedom. We generated our data simulation as the following model

$$y_i = \max(0, y^*)$$

$$y^* = X_i^t \beta_\theta + \varepsilon_i$$

where **:**

The algorithm of our proposed method is run for 11,000 iterations and the first one a thousand were excluded as burn in. To evaluate our proposed method compared with other methods in the same field two simulations scenario have been used. In our study ,we will use four sample size (n=25,n=50,n=100 and 200). The independent variables generated from multivariate normal with mean (0) and var-cov $\left(\boldsymbol{R}_{I},\boldsymbol{CJ}\right) = (0.5)^{I+J}$

## (4-2)First Simulations Scenario

In our first simulation , we show the effectiveness of our proposed method with sparse case. Therefore the true model defined as follow

$$\boldsymbol{y_i} = \mathbf{max}(\mathbf{0}, \boldsymbol{y}^*)$$

where $\boldsymbol{y}^* = \boldsymbol{x_{1i}} + \mathbf{3}\boldsymbol{x_{4i}} + \mathbf{1}\boldsymbol{x_{7i}} + \boldsymbol{\varepsilon_i} \qquad \mathbf{i} = \mathbf{1}, \mathbf{2}, \dots . \mathbf{n}$

The true parameters of above model as follow :

$\boldsymbol{\beta} = (\mathbf{1}, \mathbf{0}, \mathbf{0}, \mathbf{3}, \mathbf{0}, \mathbf{0}, \mathbf{1}, \mathbf{0})^{\boldsymbol{t}}$.

**Table 1** shown a summary of the mean absolute error and standard deviation for our proposed method and other two methods for the comparison.The mean absolute error (MAE) calculated by our proposed method much smaller than The mean absolute error (MAE) calculated by other two methods (crq, New B Tobit Q Reg), via all error distributions and quantile levels and all sample size under consideration. Also the standard deviation (S.D) calculated by our proposed method much smaller than standard deviation (S.D) calculated by other two methods (crq, New B L Tobit Q Reg), via all error distributions and quantile levels under consideration. Therefore ,the our proposed method more accurately compared with other two methods (crq, New B LTobit Q Reg)

**Table -1- show the mean absolute error (MAE) and standard deviation (S.D) for first simulation scenario**

| Sample size | Methods | $\theta$ | $\varepsilon_i \sim N(0,1)$ | $\varepsilon_i \sim N(2,9)$ | $\varepsilon_i \sim t_{(5)}$ |
|---|---|---|---|---|---|
| n=25 | TQR | 0.25 | 1.019 (0.751) | 0.923 (0.562) | 1.212 (0.792) |
| | New B Tobit Q Reg | 0.25 | 0.828 (0.927) | 0.891 (0.376) | 1.007 (0.871) |
| | **NB LTQR** | **0.25** | **0.447 (0.265)** | **0.402 (0.100)** | **0.671 (0.269)** |
| | TQR | 0.5 | 1.851 (0.929) | 0.817 (0.581) | 1.013 (0.828) |
| | New B Tobit Q Reg | 0. 5 | 1.193 (0.821) | 0.782 (0.378) | 0.921 (0.562) |
| | **NB LTQR** | **0.5** | **0.742 (0.273)** | **0.462 (0.142)** | **0.513 (0.105)** |
| | TQR | 0.75 | 1.143 (0.845) | 0.943 (0.905) | 1.106 (0.856) |
| | New B Tobit Q Reg | 0.75 | 0.871 (0.415) | 0.651 (0.361) | 1.001 (0.815) |
| | **NB LTQR** | **0.75** | **0.651 (0.451)** | **0.414 (0.132)** | **0.781 (0.461)** |
| | TQR | 0.90 | 1.271 (0.971) | 0.821 (0.431) | 0.709 (0.361) |
| | New B Tobit Q Reg | 0.90 | 0.921 (0.561) | 0.861 (0.351) | 0.816 (0.672) |
| | **NB LTQR** | **0.90** | **0.714 (0.481)** | **0.741 (0.291)** | **0.681 (0.256)** |
| N=50 | TQR | 0.25 | 1.141 (0.854) | 1.045 (0.871) | 1.341 (0.851) |
| | New B Tobit Q Reg | 0.25 | 0.981 (0.681) | 0.871 (0.473) | 0.917 (0.534) |
| | **NB LTQR** | **0.25** | **0.616 (0.361)** | **0.554 (0.146)** | **0.351 (0.191)** |
| | TQR | 0.5 | 1.251 (0.863) | 1.272 (0.989) | 1.108 (0.829) |
| | New B Tobit Q Reg | 0. 5 | 0.956 (0.688) | 0.852 (0.409) | 0.956 (0.361) |
| | **NB LTQR** | **0.5** | **0.721 (0.365)** | **0.615 (0.252)** | **0.356 (0.089)** |
| | TQR | 0.75 | 1.137 (0.879) | 1.262 (0.953) | 1.122 (0.845) |
| | New B Tobit Q | 0.75 | 0.942 | 0.844 | 0.782 |

| | | | | | |
|---|---|---|---|---|---|
| | Reg | | (0.507) | (0.264) | (0.386) |
| | **NB LTQR** | **0.75** | **0.727 (0.256)** | **0.572 (0.183)** | **0.381 (0.117)** |
| | TQR | 0.90 | 1.231 (0.925) | 1.102 (0.781) | 1.022 (0.791) |
| | New B Tobit Q Reg | 0.90 | 0.826 (0.461) | 0.782 (0.351) | 0.891 (0.461) |
| | **NB LTQR** | **0.90** | **0.681 (0.176)** | **0.573 (0.102)** | **0.517 (0.281)** |
| | TQR | 0.25 | 1.453 (0.952) | 1.264 (0.934) | 1.344 (0.862) |
| | New B Tobit Q Reg | 0.25 | 0.936 (0.572) | 0.838 (0.573) | 0.854 (0.475) |
| | **NB LTQR** | **0.25** | **0.682 (0.254)** | **0.482 (0.184)** | **0.391 (0.162)** |
| | TQR | 0.5 | 1.251 (0.895) | 1.361 (0.781) | 1.172 (0.837) |
| **N=100** | New B Tobit Q Reg | 0. 5 | 0.834 (0.465) | 0.794 (0.461) | 0.684 (0.178) |
| | **NBLTQR** | **0.5** | **0.581 (0.093)** | **0.672 (0.184)** | **0.468 (0.104)** |
| | TQR | 0.75 | 1.561 (0.986) | 1.352 (0.841) | 1.281 (0.795) |
| | New B Tobit Q Reg | 0.75 | 0.945 (0.582) | 0.845 (0.358) | 0.861 (0.582) |
| | **NB LTQR** | **0.75** | **0.735 (0.566)** | **0.684 (0.472)** | **0.578 (0.273)** |
| | TQR | 0.90 | 1.464 (0.954) | 1.246 (0.943) | 1.172 (0.834) |
| | New B Tobit Q Reg | 0.90 | 0.916 (0.747) | 0.857 (0.463) | 0.682 (0.217) |
| | **NBLTQR** | **0.90** | **0.736 (0.264)** | **0.682 (0.201)** | **0.461 (0.095)** |
| | TQR | 0.25 | 1.115 (0.835) | 1.064 (0.734) | 1.145 (0.699) |
| | New B Tobit Q Reg | 0.25 | 0.845 (0.451) | 0.745 (0.363) | 0.671 (0.451) |
| | **NBLTQR** | 0.25 | **0.785 (0.217)** | **0.638 (0.375)** | **0.583 (0.125)** |
| **N=200** | TQR | 0.5 | 1.361 (0.838) | 1.173 (0.792) | 1.092 (0.881) |
| | New B Tobit Q Reg | 0. 5 | 0.892 (0.411) | 0.751 (0.394) | 0.864 (0.382) |
| | **NBLTQR** | **0.5** | **0.473 (0.106)** | **0.381 (0.092)** | **0.358 (0.077)** |
| | TQR | 0.75 | 1.274 | 1.107 | 1.074 |

| | | | | | |
|---|---|---|---|---|---|
| | | | (0.892) | (0.738) | (0.693) |
| | New B Tobit Q Reg | 0.75 | 0.681<br>(0.186) | 0.764<br>(0.176) | 0.727<br>(0.375) |
| | **NBLTQR** | **0.75** | **0.439**<br>**(0.101)** | **0.381**<br>**(0.071)** | **0.263**<br>**(0.028)** |
| | TQR | 0.90 | 1.064<br>(0.761) | 1.096<br>(0.734) | 1.124<br>(0.892) |
| | New B Tobit Q Reg | 0.90 | 0.739<br>(0.268) | 0.679<br>(0.361) | 0.617<br>(0.316) |
| | **NBLTQR** | **0.90** | **0.563**<br>**(0.112)** | **0.428**<br>**(0.103)** | **0.406**<br>**(0.174)** |

The values in parentheses is standard deviation (S.D)

## (4-3) Second Simulations Scenario

In our second simulation , we show the effectiveness of our proposed method with dense case. Therefore the true model defined as follow

$$\boldsymbol{y_i} = \mathbf{max}(\mathbf{0}, \boldsymbol{y}^*)$$

where $y^* = 0.85x_{1i} + 0.85x_{2i} + 0.85x_{3i} + 0.85x_{4i} + 0.85x_{5i} + 0.85x_{6i} + 0.85x_{7i} + 0.85x_{8i} + \boldsymbol{\varepsilon_i}$ $\quad$ $\mathrm{i} = 1,2,\ldots.\mathrm{n}$

The true parameters of above model as follow :

$\boldsymbol{\beta} = (\mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85})^{\boldsymbol{t}}$**.**

Table 2 shown a summary of the mean absolute error and standard deviation for our proposed method and other two methods for the comparison.

The mean absolute error (MAE) calculated by our proposed method much smaller than The mean absolute error (MAE) calculated by other two methods (crq, New B Tobit Q Reg), via all error distributions and quantile levels and all sample size under consideration. Also ,the standard deviation (S.D) calculated by our proposed method much smaller than standard deviation

(S.D) calculated by other two methods (crq, New B Tobit Q Reg), via all error distributions and quantile levels under and all sample size consideration. Therefore ,the our proposed method more accurately compared with other two methods (crq, New B Tobit Q Reg).

**Table -2- show the mean absolute error (MAE) and standard deviation (S.D) for second simulation scenario**

| Sample size | Methods | $\theta$ | $\varepsilon_i \sim N(0,1)$ | $\varepsilon_i \sim N(2,9)$ | $\varepsilon_i \sim t_{(5)}$ |
|---|---|---|---|---|---|
| n=25 | TQR | 0.25 | 1.747 (0.936) | 1.591 (0.984) | 1.602 (0.957) |
| | New B Tobit Q Reg | 0.25 | 0.984 (0.581) | 0.918 (0.603) | 0.823 (0.593) |
| | NBLTQR | **0.25** | **0.619 (0.318)** | **0.584 (0.285)** | **0.692 (0.205)** |
| | TQR | 0.5 | 1.471 (0.945) | 1.256 (0.879) | 1.175 (0.938) |
| | New B Tobit Q Reg | 0. 5 | 0.805 (0.491) | 0.937 (0.404) | 0.857 (0.412) |
| | NBLTQR | **0.5** | **0.657 (0.317)** | **0.725 (0.306)** | **0.527 (0.203)** |
| | TQR | 0.75 | 1.372 (0.813) | 1.461 (0.972) | 1.171 (0.764) |
| | New B Tobit Q Reg | 0.75 | 0.794 (0.487) | 0.871 (0.251) | 0.816 (0.316) |
| | NBLTQR | **0.75** | **0.518 (0.268)** | **0.472 (0.276)** | **0.361 (0.192)** |
| | TQR | 0.90 | 1.515 (0.832) | 1.456 (0.962) | 1.634 (0.957) |
| | New B Tobit Q Reg | 0.90 | 0.861 (0.462) | 0.945 (0.539) | 0.863 (0.572) |
| | NBLTQR | **0.90** | **0.521 (0.264)** | **0.485 (0.353)** | **0.535 (0.412)** |
| N=50 | TQR | 0.25 | 1.362 (0.822) | 1.256 (0.832) | 1.127 (0.892) |
| | New B Tobit Q Reg | 0.25 | 0.904 (0.583) | 0.756 (0.362) | 0.748 (0.436) |
| | NBLTQR | **0.25** | **0.465 (0.273)** | **0.372 (0.193)** | **0.283 (0.093)** |
| | TQR | 0.5 | 1.257 | 1.362 | 1.204 |

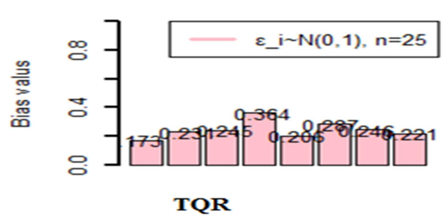
| | | | | | |
|---|---|---|---|---|---|
| | | | (0.792) | (0.782) | (0.722) |
| | New B Tobit Q Reg | 0. 5 | 0.605 (0.372) | 0.577 (0.296) | 0.523 (0.149) |
| | NBLTQR | **0.5** | **0.463 (0.362)** | **0.356 (0.672)** | **0.347 (0.133)** |
| | TQR | 0.75 | 1.472 (0.847) | 1.573 (0.937) | 0.945 (0.694) |
| | New B Tobit Q Reg | 0.75 | 0.834 (0.356) | 0.674 (0.564) | 0.585 (0.436) |
| | NBLTQR | **0.75** | **0.386 (0.175)** | **0.393 (0.174)** | **0.311 (0.118)** |
| | TQR | 0.90 | 1.374 (0.804) | 1.685 (0.867) | 0.943 (0.558) |
| | New B Tobit Q Reg | 0.90 | 0.782 (0.282) | 0.681 (0.361) | 0.861 (0.372) |
| | NBLTQR | **0.90** | **0.293 (0.113)** | **0.267 (0.103)** | **0.246 (0.096)** |
| **N=100** | TQR | 0.25 | 1.272 (0.782) | 1.371 (0.799) | 0.892 (0.372) |
| | New B Tobit Q Reg | 0.25 | 0.587 (0.292) | 0.498 (0.189) | 0.396 (0.094) |
| | NBLTQR | **0.25** | **0.237 (0.102)** | **0.375 (0.098)** | **0.293 (0.084)** |
| | TQR | 0.5 | 1.351 (0.862) | 1.241 (0.691) | 0.947 (0.481) |
| | New B Tobit Q Reg | 0. 5 | 0.648 (0.378) | 0.582 (0.380) | 0.486 (0.194) |
| | NBLTQR | **0.5** | **0.356 (0.138)** | **0.471 (0.117)** | **0.365 (0.096)** |
| | TQR | 0.75 | 1.256 (0.783) | 1.184 (0.661) | 0.893 (0.472) |
| | New B Tobit Q Reg | 0.75 | 0.768 (0.474) | 0.792 (0.295) | 0.610 (0.318) |
| | NBLTQR | **0.75** | **0.405 (0.261)** | **0.327 (0.328)** | **0.293 (0.219)** |
| | TQR | 0.90 | 1.465 (0.846) | 1.257 (0.783) | 0.975 (0.494) |
| | New B Tobit Q Reg | 0.90 | 0.829 (0.396) | 0.817 (0.496) | 0.739 (0.393) |
| | NBLTQR | **0.90** | **0.394 (0.143)** | **0.283 (0.107)** | **0.188 (0.096)** |
| | TQR | 0.25 | 1.526 (0.822) | 1.289 (0.982) | 0.942 (0.582) |
| | New B Tobit Q Reg | 0.25 | 0.821 (0.372) | 0.519 (0.257) | 0.482 (0.178) |
| | NBLTQR | **0.25** | **0.388** | **0.292** | **0.257** |

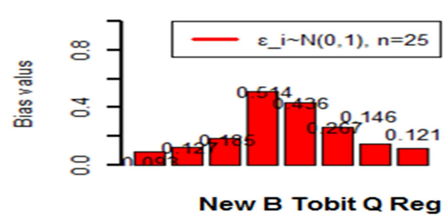
| | | | | | |
|---|---|---|---|---|---|
| **N=200** | | | **(0.142)** | **(0.167)** | **(0.074)** |
| | TQR | 0.5 | 1.472 (0.743) | 1.382 (0.892) | 0.882 (0.403) |
| | New B Tobit Q Reg | 0. 5 | 0.764 (0.378) | 0.694 (0.284) | 0.582 (0.204) |
| | NBLTQR | **0.5** | **0.283 (0.095)** | **0.256 (0.105)** | **0.220 (0.097)** |
| | TQR | 0.75 | 1.132 (0.741) | 1.261 (0.728) | 0.772 (0.345) |
| | New B Tobit Q Reg | 0.75 | 0.654 (0.284) | 0.494 (0.165) | 0.647 (0.151) |
| | NBLTQR | **0.75** | **0.285 (0.092)** | **0.185 (0.076)** | **0.198 (0.078)** |
| | TQR | 0.90 | 1.189 (0.735) | 1.106 (0.672) | 0.835 (0.285) |
| | New B Tobit Q Reg | 0.90 | 0.573 (0.282) | 0.511 (0.204) | 0.493 (0.286) |
| | NBLTQR | **0.90** | **0.235 (0.083)** | **0.124 (0.056)** | **0.104 (0.056)** |

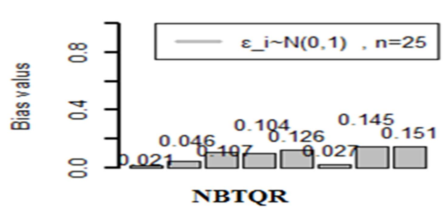
The values in parentheses is standard deviation (S.D)

## (4-4) Bias amount

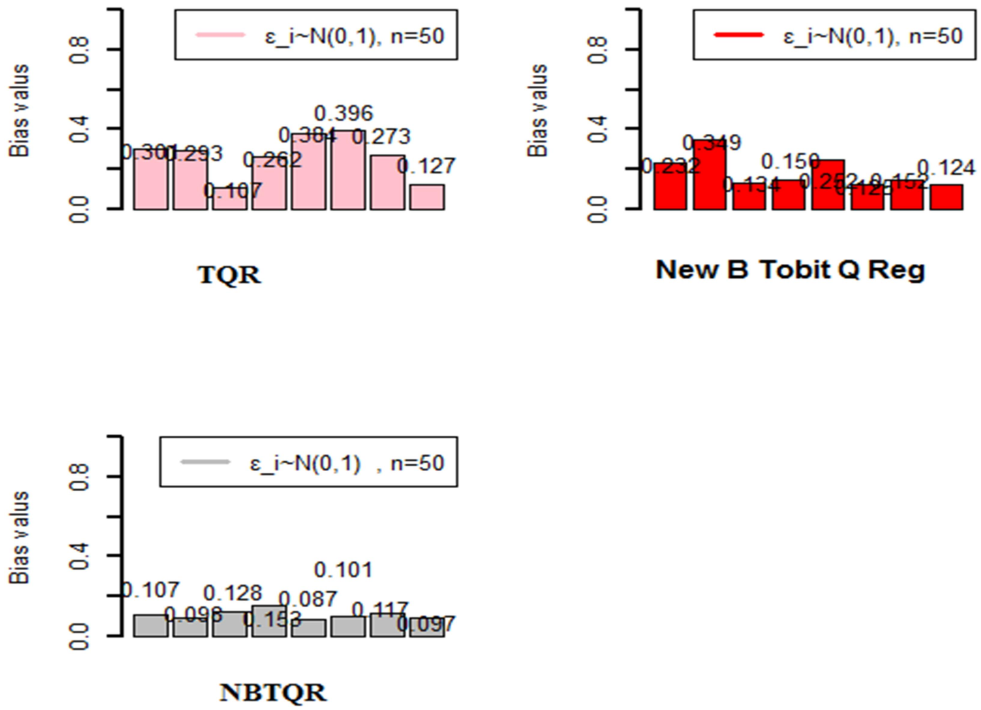
In this our study, we will used another criteria, to evaluation of performance our proposed method with previous two methods . Via using Bias amount criteria as shown in below figures



## Figure-1- Bias amount for three methods under first Simulations Scenario at $\theta = 0.25$




## Figure-2- Bias amount for three methods under first Simulations Scenario at $\theta = 0.50$

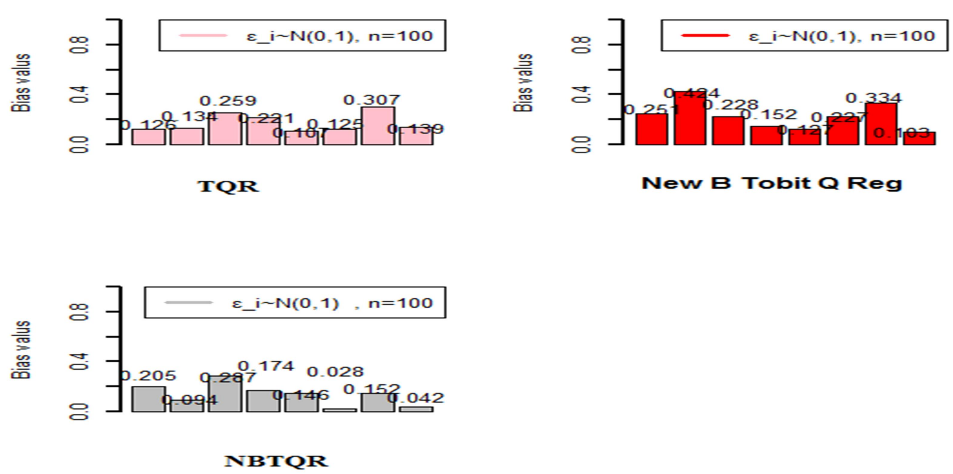


**Figure-3- Bias amount for three methods under first Simulations Scenario at $\theta = 0.75$**

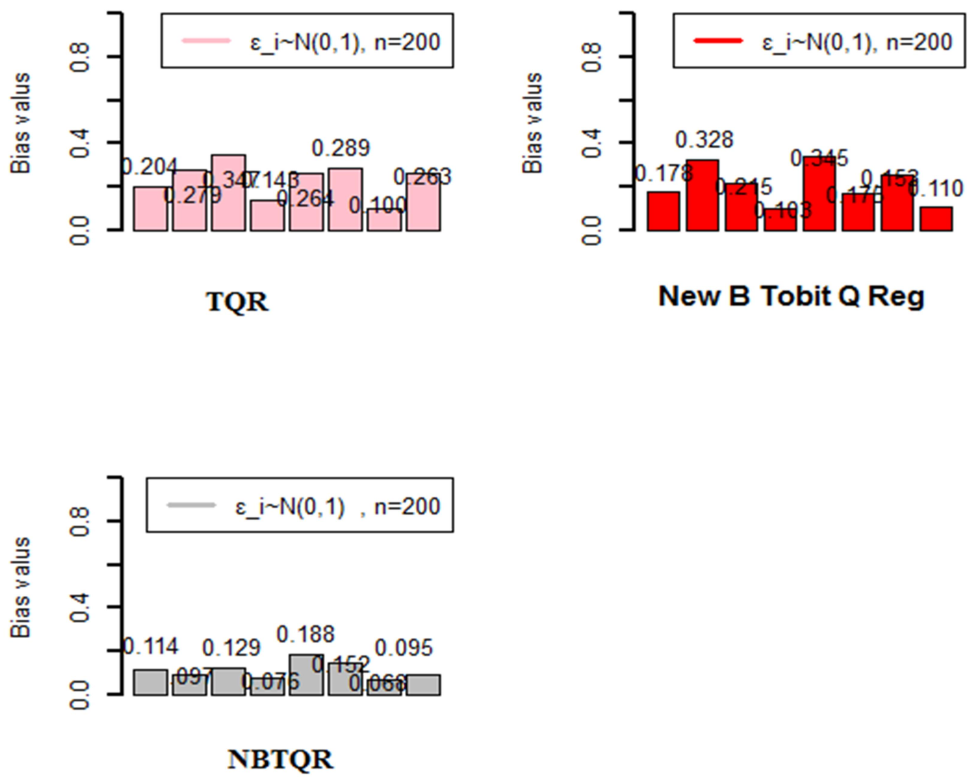



**Figure-4- Bias amount for three methods under first Simulations Scenario at $\theta = 0.90$**

Figure 1,2,3,4 displayed the graphs of Bias amount to three the methods, we can see Bias amount generated by our proposed method (NBLTQR) much smaller than other method (T,Q,R and New B Tobit Q Reg). Therefore , therefore ,the our proposed method (NBLTQR) has a useful performance compared other two methods. Where the parameter estimation via our proposed method (NBLTQR) were very closed to true parameters.

## (4-5) Real Data

In this part of our study, we use Tobit quantile regression model for analyze the medical phenomenon. The our data collected from Al-Rafidain Laboratory in the city of Samawah and these data consists 200 observations and 20 independent variables and one dependent variable. The depend variable is Erythrocyte Sedimentation Rate referred to as (ESR) ,and the twenty independent variables are:

$x_1$: **Covid − 19,**

$x_2$: **Rheumatic Disease ,**

$x_3$: **Cancer disease,**

$x_4$: **random blood sugar referred to as (R.B.Sugar),**

$x_5$: **blood Urea referred to as (B.Urea),**

$x_6$: **size Creatinine referred to as (S. Creatinine),**

$x_7$: **Low-density lipoprotein referred to as (L. D.L),**

$x_8$: **High-density lipoprotein referred to as (H. D.L),**

$x_9$: **calcium referred to as (Ca),**

$x_{10}$: **Hematocrit test referred to as (HCT),**

$x_{11}$: **Haemoglobin blood referred to as (HB),**

$x_{12}$: **Packed cell volume referred to as (PCV),**

$x_{13}$: **White Blood Cell referred to as (WBC),**

$x_{14}$: **Size. cholestrol referred to as (S.cholestrol) ,**

$x_{15}$: **Blood Group ,**

$x_{16}$: **Platelet Count Test referred to as (PCT),**

$x_{17}$: **Procalcitonin Test referred to as (PT),**

$x_{18}$: **Mean Platelet Volume referred to as (MPV) ,**

$x_{19}$: **Weight**

$x_{20}$: **Age.**

Similar to our simulation Scenario, we compare our proposed method with two other methods (NBTQR, New B Tobit Q Reg and TQ R). The our proposed method and other two methods are evaluated via mean absolute error (MAE) and mean square error (MSR). The results of MAE and MSE inserted in table 3 . From result shown in table 3 , the MAE and MSE generated by our proposed method much smaller than the MAE and MSE generated by other two methods (crq, New B Tobit QReg), via all quantile levels under consideration. Therefore, the our proposed method is consider a good method compared with other method in same filed .

**Table -3- the mean absolute error (MAE) and mean square error (MSE) for real data**

| Level quantile | criterias | TQE | New B Tobit Q Reg | NBLTQR |
|---|---|---|---|---|
| $\theta = 0.25$ | MSE | 1.342 | 0.942 | 0.571 |
| | MAE | 1.076 | 0.971 | 0.756 |
| $\theta = 0.5$ | MSE | 1.261 | 0.873 | 0.672 |
| | MAE | 1.123 | 0.934 | 0.819 |
| $\theta = 0.75$ | MSE | 1.002 | 0.902 | 0.641 |
| | MAE | 1.001 | 0.949 | 0.801 |
| $\theta = 0.90$ | MSE | 1.174 | 0.838 | 0.518 |
| | MAE | 1.084 | 0.915 | 0.719 |

From simulation approaches and real dataset ,we conclude the our method is the effective method in variables selection and coefficient estimation in Tobit quantile regression . In tables below parameters estimation of TQR via all quantile levels ($\theta = 0.25, \theta = 0.5, \theta = 0.75$ and $\theta = 0.90$)

**Table -4- parameters estimation of Tobit quantile regression at $\theta = 0.25$**

| Variables | Name of variables | TQR | New B Tobit Q Reg | NBLTQR |
|---|---|---|---|---|
| $X_1$ | Covid-19 | 0.676 | 0.741 | 1.687 |
| $X_2$ | Rheumatic disease | 0.176 | 0.260 | 0.341 |
| $X_3$ | Cancer disease | 0.521 | 0.159 | 0.426 |
| $X_4$ | R. B. sugar | 0.128 | 0.741 | 1.741 |
| $X_5$ | B. urea | 0.358 | 0.397 | 0.471 |
| $X_6$ | S. Creatinine | 0.342 | 0.191 | 0.000 |
| $X_7$ | LDL | 0.752 | 0.237 | 0.523 |
| $X_8$ | HDL | 1.052 | 1.281 | 1.852 |
| $X_9$ | Ca++ | 0.515 | 0.715 | 0.732 |
| $X_{10}$ | Hct | 0.232 | 0.001 | 0.439 |
| $X_{11}$ | Hb | 0.651 | 0.957 | -0.765 |
| $X_{12}$ | Pcv | -0.571 | -0.261 | 0.547 |
| $X_{13}$ | WBC | 0.262 | 0.000 | 0.000 |
| $X_{14}$ | S.cholestrol | 1.064 | 0.351 | 0.483 |
| $X_{15}$ | Blood Group | 0.000 | 0.000 | 0.000 |
| $X_{16}$ | PCT | 0.297 | -0.121 | -0.132 |
| $X_{17}$ | PT | 0.424 | 0.124 | 0.655 |
| $X_{18}$ | MPV | 0.301 | 0.170 | 0.363 |
| $X_{19}$ | WEIGHT | 0.000 | 0.000 | 0.000 |
| $X_{20}$ | AGE | 0.281 | 1.502 | 0.246 |

From results inserted in above table, there are negative and positive effect on the response variable (Erythrocyte Sedimentation Rate) via the methods (TQR, New B Tobit Q Reg,NBLTQR). In TQR method there are two independent variables($X_{15}$:Blood Group and $X_{19}$: weight) are ineffective in response variable (Erythrocyte Sedimentation Rate) ,but the rest independent variables are effective in response variable (Erythrocyte Sedimentation Rate). But In New B Tobit Q Reg method there are three independent variables**($X_{13}$:White cell blood ,$X_{15}$:Blood Group and $X_{19}$: weight)** are ineffective in response variable (Erythrocyte Sedimentation Rate) ,but the rest independent variables are effective in response variable (Erythrocyte Sedimentation Rate). In our proposed method (NBTQR), there are four independent variables **($X_6$: S. Creatinine,$X_{13}$:White cell blood ,$X_{15}$:Blood Group and $X_{19}$: weight)** are ineffective in response variable (Erythrocyte Sedimentation Rate) ,but the rest independent variables are effective in response variable (Erythrocyte Sedimentation Rate).

**Table -5- parameters estimation of Tobit quantile regression at $\theta = 0.5$**

| Variables | Name of variables | TQR | New B Tobit Q Reg | NBLTQR |
|---|---|---|---|---|
| $X_1$ | Covid-19 | 0.068 | 0.176 | 0.240 |
| $X_2$ | Rheumatic disease | 0.680 | 0.401 | 0.807 |

| | | | | |
|---|---|---|---|---|
| $X_3$ | Cancer disease | 0.096 | 0.733 | 0.160 |
| $X_4$ | R. B. sugar | -0.030 | 0.317 | 0.000 |
| $X_5$ | B. urea | 0.137 | -0.372 | 0.330 |
| $X_6$ | S. Creatinine | 0.876 | 0.001 | 0.031 |
| $X_7$ | LDL | 0.086 | 0.117 | 0.307 |
| $X_8$ | HDL | -0.210 | 0.091 | 0.034 |
| $X_9$ | Ca++ | 0.572 | -0.241 | 0.030 |
| $X_{10}$ | Hct | 0.472 | 0.591 | 0.570 |
| $X_{11}$ | Hb | 1.573 | 0.572 | 0.482 |
| $X_{12}$ | Pcv | 0.699 | 0.472 | 0.588 |
| $X_{13}$ | WBC | 0.285 | 0.000 | 0.000 |
| $X_{14}$ | S.cholestrol | 0.381 | 0.699 | 0.470 |
| $X_{15}$ | Blood Group | 0.000 | 0.000 | 0.000 |
| $X_{16}$ | PCT | 0.741 | 0.381 | 0.609 |
| $X_{17}$ | PT | 0.269 | 0.783 | 1.306 |
| $X_{18}$ | MPV | 0.267 | 0.741 | 0.566 |
| $X_{19}$ | WEIGHT | 0.000 | 0.000 | 0.000 |
| $X_{20}$ | AGE | 1.271 | 0.925 | 0.468 |

From results inserted in above table, there are negative and positive effect on the response variable (Erythrocyte Sedimentation Rate) via the methods (TQR, New B Tobit Q Reg,NBTQR). In TQR method there are two independent variables **($X_{15}$:Blood Group and $X_{19}$: weight)** are ineffective in response variable (Erythrocyte Sedimentation Rate) ,but the rest independent variables are effective in response variable (Erythrocyte Sedimentation Rate). But In New B Tobit Q Reg method there are three independent variables **($X_{13}$:White cell blood ,$X_{15}$:Blood Group and $X_{19}$: weight)** are ineffective in response variable (Erythrocyte Sedimentation Rate) ,but the rest independent variables are effective in response variable (Erythrocyte Sedimentation Rate). In our proposed method

(NBLTQR), there are four independent variables (**$X_6$: S. Creatinine,$X_{13}$:White cell blood ,$X_{15}$:Blood Group and $X_{19}$: weight)** are ineffective in response variable (Erythrocyte Sedimentation Rate) ,but the rest independent variables are effective in response variable (Erythrocyte Sedimentation Rate).

**Table-6-parameters estimation of Tobit quantile regression at $\theta = 0.75$**

| Variables | Name of variables | TQR | New B Tobit Q Reg | NBTQR |
|---|---|---|---|---|
| $X_1$ | Covid-19 | 0.068 | 0.176 | 0.240 |
| $X_2$ | Rheumatic disease | 0.680 | 0.401 | 0.807 |
| $X_3$ | Cancer disease | 0.096 | 0.733 | 0.160 |
| $X_4$ | R. B. sugar | -0.030 | 0.317 | 0.000 |
| $X_5$ | B. urea | 0.137 | -0.372 | 0.330 |
| $X_6$ | S. Creatinine | 0.000 | 0.000 | 0.000 |
| $X_7$ | LDL | 0.086 | 0.117 | 0.307 |
| $X_8$ | HDL | -0.210 | 0.091 | 0.000 |
| $X_9$ | Ca++ | -0.572 | -0.241 | -0.030 |
| $X_{10}$ | Hct | 0.472 | 0.591 | 0.570 |
| $X_{11}$ | Hb | 1.573 | 0.572 | 0.482 |
| $X_{12}$ | Pcv | 0.699 | 0.472 | 0.588 |
| $X_{13}$ | WBC | 0.000 | 0.000 | 0.000 |
| $X_{14}$ | S.cholestrol | 0.381 | 0.699 | 0.470 |
| $X_{15}$ | Blood Group | 0.000 | 0.000 | 0.000 |
| $X_{16}$ | PCT | 0.741 | 0.381 | 0.609 |
| $X_{17}$ | PT | 0.269 | 0.783 | 1.306 |
| $X_{18}$ | MPV | 0.267 | 0.741 | 0.566 |
| $X_{19}$ | WEIGHT | 0.000 | 0.000 | 0.000 |
| $X_{20}$ | AGE | 1.271 | 0.925 | 0.468 |

From results inserted in above table, there are negative and positive effect on the response variable (Erythrocyte Sedimentation Rate) via the methods (TQR, New B Tobit Q Reg,NBTQR). In TQR method there are four independent variables **($X_6$: S. Creatinine,$X_{13}$:White cell blood,$X_{15}$:Blood Group and $X_{19}$: weight)** are ineffective in response variable (Erythrocyte Sedimentation Rate) ,but the rest independent variables are effective in response variable (Erythrocyte Sedimentation Rate). Also in New B Tobit Q Reg method there are four independent variables **($X_6$: S. Creatinine,$X_{13}$:White cell blood,$X_{15}$:Blood Group and $X_{19}$: weight)** are ineffective in response variable (Erythrocyte Sedimentation Rate) ,but the rest independent variables are effective in response variable (Erythrocyte Sedimentation Rate). In our proposed method (NBTQR), there are six independent variables **($X_4$: random blood sugar ,$X_6$: S. Creatinine, $x_8$: High-density lipoprotein referred to as (H. D.L),$X_{13}$:White cell blood ,$X_{15}$:Blood Group and $X_{19}$: weight)** are ineffective in response variable (Erythrocyte Sedimentation Rate) ,but the rest independent variables are effective in response variable (Erythrocyte Sedimentation Rate).

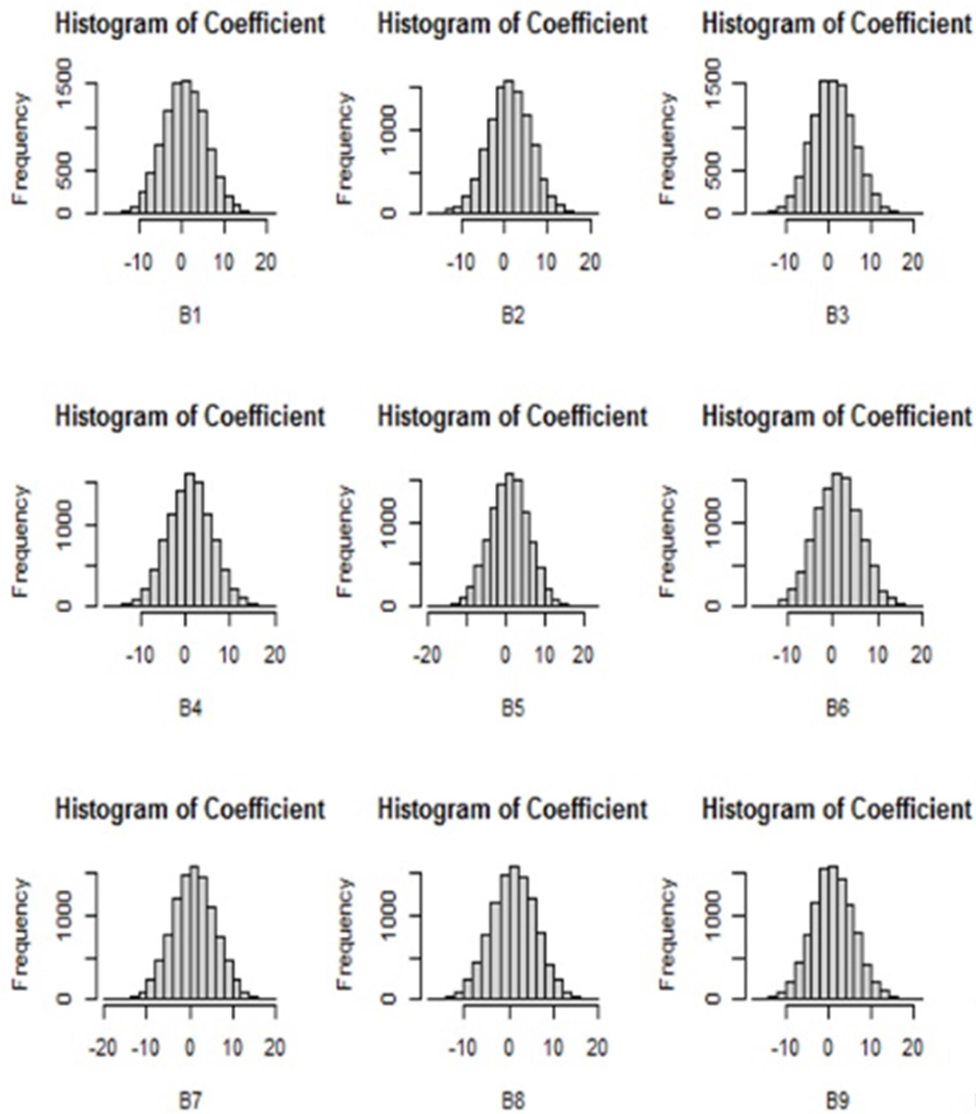
**Table -7- parameters estimation of Tobit quantile regression at $\theta = 0.90$**

| Variables | Name of variables | TQR | New B Tobit Q Reg | NBLTQR |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| $X_1$ | Covid-19 | 0.825 | 0.367 | 0.482 |
| $X_2$ | Rheumatic disease | 0.233 | 0.180 | 0.238 |
| $X_3$ | Cancer disease | 0.261 | 0.002 | 0.529 |
| $X_4$ | R. B. sugar | 0.281 | -0.078 | 0.000 |
| $X_5$ | B. urea | -0.371 | 0.156 | 0.652 |
| $X_6$ | S. Creatinine | 0.000 | 0.000 | 0.000 |
| $X_7$ | LDL | 0.401 | 0.247 | 0.306 |
| $X_8$ | HDL | 0.733 | -0.342 | 0.000 |
| $X_9$ | Ca++ | 0.317 | -0.109 | 0.085 |
| $X_{10}$ | HCT | -0.372 | 0.189 | 0.268 |
| $X_{11}$ | HB | 1.471 | 0.156 | 0.280 |
| $X_{12}$ | PCV | 0.241 | 0.075 | 0.176 |
| $X_{13}$ | WBC | 0.000 | 0.000 | 0.000 |
| $X_{14}$ | S.cholestrol | 0.472 | 0.136 | 0.733 |
| $X_{15}$ | Blood Group | 0.000 | 0.000 | 0.000 |
| $X_{16}$ | PCT | 0.699 | 0.350 | -0.372 |
| $X_{17}$ | PT | 0.285 | 0.000 | 0.001 |
| $X_{18}$ | MPV | 0.381 | 0.000 | 0.117 |
| $X_{19}$ | WEIGHT | 0.000 | 0.000 | 0.000 |
| $X_{20}$ | AGE | 0.741 | 0.000 | -0.241 |

From results inserted in above table, there are negative and positive effect on the response variable (Erythrocyte Sedimentation Rate) via the methods (TQR, New B Tobit Q Reg,NBTQR). In TQR method there are four independent variables **($X_6$: S. Creatinine,$X_{13}$:White cell blood,$X_{15}$:Blood Group and $X_{19}$: weight)** are ineffective in response variable (Erythrocyte Sedimentation Rate) ,but the rest independent variables are effective in response variable (Erythrocyte Sedimentation Rate). Also in New B Tobit Q Reg method there are four independent variables **($X_6$: S. Creatinine,$X_{13}$:White cell blood,$X_{15}$:Blood Group and $X_{19}$: weight)** are ineffective in response variable (Erythrocyte Sedimentation Rate) ,but the rest independent variables are effective in response variable (Erythrocyte Sedimentation Rate). In our proposed method (NBTQR), there are six independent variables **($X_4$: random blood sugar ,$X_6$: S. Creatinine, $x_8$: High-density lipoprotein referred to as (H. D.L),$X_{13}$:White cell blood ,$X_{15}$:Blood Group and $X_{19}$: weight)** are ineffective in response variable (Erythrocyte Sedimentation Rate) ,but the rest independent variables are effective in response variable (Erythrocyte Sedimentation Rate).

The below graph is displayed histogram for 20 parameters estimation of independent variables.


Histogram of Coefficient
Frequency
B1
Histogram of Coefficient
Frequency
B2
Histogram of Coefficient
Frequency
B3
Histogram of Coefficient
Frequency
B4
Histogram of Coefficient
Frequency
B5
Histogram of Coefficient
Frequency
B6
Histogram of Coefficient
Frequency
B7
Histogram of Coefficient
Frequency
B8
Histogram of Coefficient
Frequency
B9

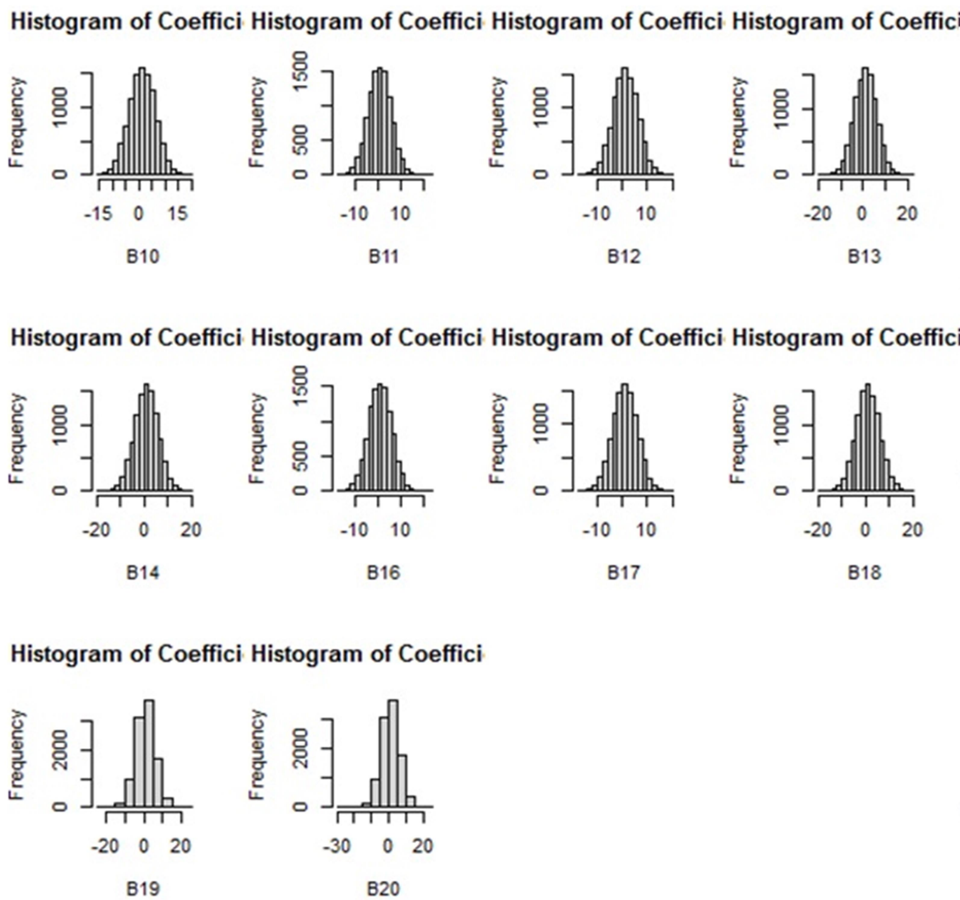
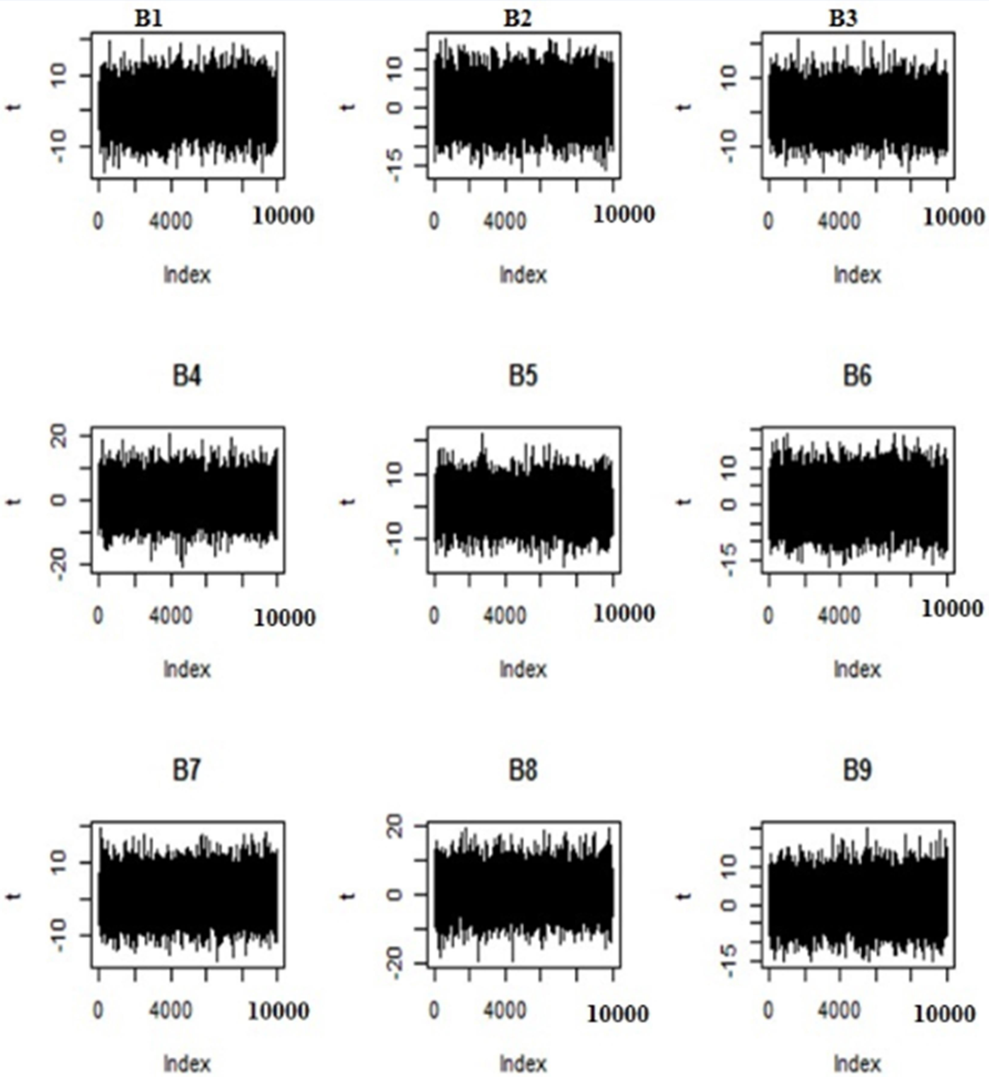



**Figure 5** show histogram of NBLTQR coefficients estimation

From the above figure is readily observed that the coefficients estimation is very closed to normal distribution .

But The below graph is displayed trace plot for 20 parameters estimation of independent variables.


B1
B2
B3
B4
B5
B6
B7
B8
B9
t
Index
0
4000
10000
10
0
-10
-15
20
-20

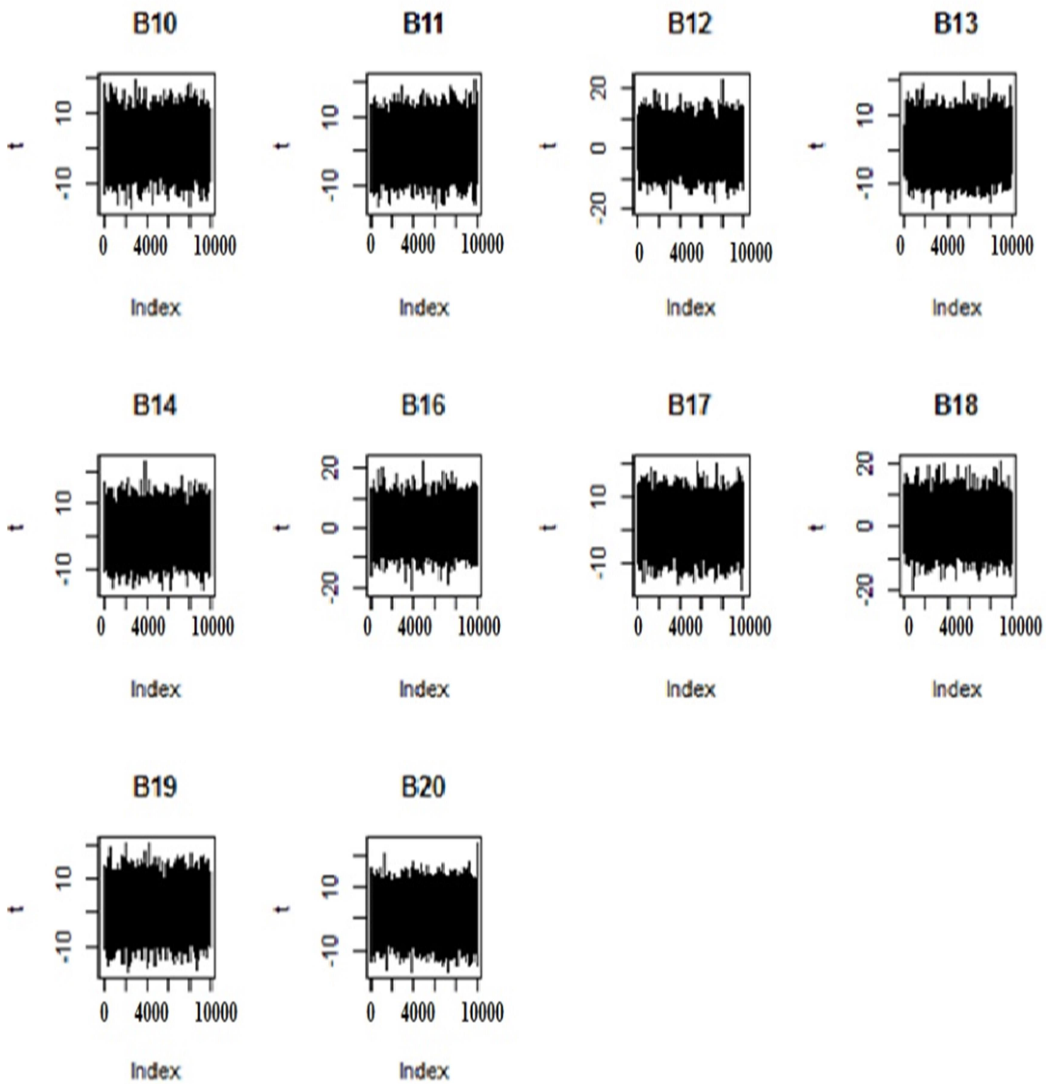



**Figure 6**: show trace plot of NBLTQR coefficients estimation

From the above figure is readily observed that the our algorithm is very convergence via all iteration .

# Chapter Five

## (5-1)Conclusions

This thesis focuses on employing the scale mixture of uniforms representation for the Laplace distribution as prior density for the interested parameter of the lasso Tobit quantile regression .So, We focus on variable selection procedure by using the combination of the lasso method and to bit regression from the Bayesian aspect we give an overview for the multiple linear regression the OLS method ,as well as the Tobit quantile regression and variable selection procedure. Moreover, we developed a new hierarchical prior model for the suggested regression model –a new Gibbs sampler algorithm has developed based on the proposed hierarchical prior model . we demonstrated the proposed performance of the proposed method in simulation analysis and analysis real data .The results obtained outperformed the proposed method on both simulation and real data in comparisons to those from other existing method in terms of some quantile criterion such as MAE and MSE .Also, we show how powerful the penalized lasso method is in obtaining parsimonious model the real data analysis .

## (5-2)Recommendation

We recommend the use of the Bayesian theory in the estimation procedure in regression analysis when dealing with sample size and when there problem such as the multicollinearity problem the proposed method can be developed by parameterization for the distribution of the regression parameter for different types of regression method , such as the count data regression , longitudinal count data , Bayesian group lasso in Tobit quantile regression , as well as, Bayesian elastic net Tobit quantile regression, as well as, Bayesian Bridge Tobit quantile regression . Finally ,we think that there are more improvement needed in Bayesian estimation algorithms for the variable selection procedure in order to successfully analyze the medical data sets where there are lavage numbered of variables that effects the studied phenomenon to help the decision maker to select the right decision that effect the patients life.

## (5 -3) References


1) Abbas, Haider Kadhim. "Bayesian Lasso Tobit regression." Journal of Al-Qadisiyah for computer science and mathematics 11.2 (2019): 1-13.
2) **Alhamzawi, R., & Yu, K. (2011). Power prior elicitation in Bayesian quantile regression. Journal of Probability and Statistics, 2011.**
3) Alhamzawi, Rahim, and Haithem Taha Mohammad Ali. "A new Gibbs sampler for Bayesian lasso." Communications in Statistics-Simulation and Computation 49.7 (2020): 1855-1871.

4) **Flaih, A. N., Alshaybawee, T., & Alhusseini, F. H. H. (2020). Sparsity via new Bayesian Lasso. *Periodicals of Engineering and Natural Sciences*, *8*(1), 345-359**
5) Alhamzawi, Rahim, and Haithem Taha Mohammad Ali. "Bayesian tobit quantile regression with penalty." Communications in Statistics-Simulation and Computation 47.6 (2018): 1739-1750.
6) Alhamzawi, Rahim, and Keming Yu. "Bayesian Lasso-mixed quantile regression." Journal of Statistical Computation and Simulation 84.4 (2014): 868-880.
7) **Alhamzawi, Rahim. "Model selection in quantile regression models." Journal of Applied Statistics 42.2 (2015): 445-458.**
8) **Alhusseini, Fadel Hamid Hadi. "Bayesian Quantile Regression with Scale Mixture of Uniform Prior Distributions." International Journal of Pure and Applied Mathematics 115.1(2017):77-91**

9 ) Alhusseini, Fadel Hamid Hadi. "New Bayesian Lasso Composite Quantile Regression." American Academic Scientific Research Journal for Engineering, Technology, and Sciences 29.1 (2017): 138-152.

**11) Alhusseini, Fadel Hamid Hadi. "New Bayesian Lasso in Tobit Quantile Regression." Romanian Statistical Review Supplement 65.6 (2017): 213-229.**

**12) Alhusseini,F. & Vasile Georgescu (2018)" Bayesian Composite Tobit quantile Regression" Journal of Applied Statistics, Pages. Volume 45, Issue 4,pp.238-222 .**

**13) Alhusseini,F. (2016) "TOBIT QUANTILE REGRESSION AND IRAQI BANKS' PROFIT" Young Economists Journal / Revista Tinerilor Economisti . Vol. 13 Issue 27, pp.141-152.**

**14) Alhusseini,F. (2017) "Using the Bayesian Technique to the Tobit Quantile Regression with Scale Mixture Uniform for Selecting Important Variables Affecting Iraqi Investment Banking" Journal of Quantitative Methods .Volume 12, Issue – 1.**

**15 ) Amemiya, Takeshi. "Tobit models: A survey." Journal of econometrics 24.1-2 (1984): 3-61.**

**16) Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, Donald B. Rubin(2014)" Bayesian Data Analysis" Third Edition,A CHAPMAN & HALL BOOK.**

17) Brooks, Stephen. "Markov chain Monte Carlo method and its application." Journal of the royal statistical society: series D (the Statistician) 47.1 (1998): 69-100

18) Carlin, Bradley P., and Siddhartha Chib. "Bayesian model choice via Markov chain Monte Carlo methods." Journal of the Royal Statistical Society: Series B (Methodological) 57.3 (1995): 473-484.

19) Caron, François. "Regression shrinkage and selection via the Lasso. Robert Tibshirani, 1996." (2014).

20) Chernozhukov, V. and Hong, H. (2003). “An MCMC approach to classical estimation.” Journal of Econometrics, 115: 293–346. MR1984779. doi: http://dx.doi.org/ 10.1016/S0304-4076(03)00100-3. 164.

21 ) Chib, S. (2001). “Markov chain Monte Carlo methods: computation and inference.” In Heckman, J. J. and Leamer, E. (eds.), Handbook of Econometrics, volume 5, 3569– 3649. Amsterdam: North Holland. 173

22) Chib, Siddhartha. "Markov chain Monte Carlo methods: computation and inference." Handbook of econometrics 5 (2001): 3569-3649.

23) Chris Hans. Bayesian lasso regression. Biometrika, 96(4):835–845, 2009.

24) D.F. Andrews and C. L. Mallows, " Scale mixtures of normal distributions, ”Journal of the Royal Statistical Society. Series B (Methodological), 99-102, 1974.

25) Daniel L McFadden. Econometric analysis of qualitative response models. Handbook of econometrics, 2:1395– 1457, 1984.

26) Davino, Cristina, Marilena Furno, and Domenico Vistocco. Quantile regression: theory and applications. Vol. 988. John Wiley & Sons, 2013

27) Escobar, M. D. and West, M. (1995). “Bayesian density estimation and inference using mixtures.” Journal of the American Statistical Association, 90: 577–588. MR1340510. 170

**28) Gamerman, D. (1997), Markov Chain Monte Carlo, Chapman & Hall, London**

**29) Gelfand, A. (1996), Empirical Bayes methods for combining likelihood: comment, Journal of the American Statistical Association, Vol. 91, No. 433, pp. 551-552**

**30) Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), Bayesian Data Analysis: Second Edition , Chapman & Hall/CRC, London, pp. 182184**

**31) Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). Bayesian Data Analysis. Boca Raton: CRC Press, 3rd edition. MR3235677. 180**

**32) Gelman, Andrew, and Donald B. Rubin. "Markov chain Monte Carlo methods in biostatistics." Statistical methods in medical research 5.4 (1996): 339-355.**

**33) Hans, Chris. "Bayesian lasso regression." Biometrika 96.4 (2009): 835-845.**

**34) Hashem, H., Vinciotti, V., Alhamzawi, R., & Yu, K. (2016). Quantile regression with group lasso for classification. Advances in Data Analysis and Classification, 10(3), 375-390.**

**35) Hassan, R. O. Alhussein F.H.H (2021).Bayesian lasso in quantile regression with a new prior.**

**36) Himel Mallick and Nengjun Yi. A new bayesian lasso. Statistics and its interface, 7(4):571–582, 2014 doi: http://dx.doi.org/ 10.1080/07474938.2013.807146. 162, 165, 169, 178**

37) James L Powell. Symmetrically trimmed least squares estimation for tobit models. Econometrica: journal of the Econometric Society, pages 1435–1460, 1986.

38) 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.

39) Kobayashi, G. (2017). Bayesian endogenous Tobit quantile regression. Bayesian Analysis, 12(1), 161-191.

40) Kobayashi, Genya, and Hideo Kozumi. "Bayesian analysis of quantile regression for censored dynamic panel data." Computational Statistics 27.2 (2012): 359-380.

41) Koenker, R. and Machado, J. A. (1999). "Goodness of fit and related inference processes for quantile regression." Journal of the american statistical association, 94(448):1296–1310. 3.

42) Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. Econometrica: journal of the Econometric Society, 33-50.

43) Koenker, Roger, and Gilbert Bassett Jr. "Regression quantiles." Econometrica: journal of the Econometric Society (1978): 33-50.

44) Koenker, Roger, and Kevin F. Hallock. "Quantile regression." Journal of economic perspectives 15.4 (2001): 143-156.

45) Koenker, Roger, et al., eds. "Handbook of quantile regression." (2017).

**46) Koenker, Roger. "Quantile regression: 40 years on." Annual Review of Economics 9 (2017): 155-176.**

47) Koenker,R. (2005) .Quantile Regression .New York : Cambridge University Press .

**48) Kozumi, H. and Kobayashi, G. (2011). "Gibbs sampling methods for Bayesian quantile regression." Journal of Statistical Simulation and Computation, 81: 1565–1578. MR2851270. doi: http://dx.doi.org/10.1080/00949655.2010.496117. 162, 164, 165, 168, 169**

**49) Kozumi, Hideo, and Genya Kobayashi. "Gibbs sampling methods for Bayesian quantile regression." Journal of statistical computation and simulation 81.11 (2011): 1565-1578 .**

**50) Lancaster, T. and Jun, S. J. (2010). "Bayesian quantile regression methods." Journal of Applied Econometrics, 25: 287–307. MR2758636. doi: http://dx.doi.org/10.1002/ jae.1069. 162**

**51) Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, *54*(3), 217-224.**

52) Lopes, H. F. and Polson, N. G. (2014). "Bayesian instrumental variables: priors and likelihoods." Econometric Reviews, 33: 100–121. MR3170842.

**53) Mallick, Himel, and Nengjun Yi. "A new Bayesian lasso." Statistics and its interface 7.4 (2014): 571-582**

**54) Minjung Kyung, Jeff Gill, Malay Ghosh, George Casella, et al. Penalized regression, standard errors, and bayesian lassos. Bayesian Analysis, 5(2):369–411, 2010**

55) Park, Trevor, and George Casella. "The bayesian lasso." Journal of the American Statistical Association 103.482 (2008): 681-686.

56) Petrella, Lea, and Valentina Raponi. "Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress." Journal of Multivariate Analysis 173 (2019): 70-84

57) Qing Li, Rubin Xi & Nan Lin "Bayesian Regularized Quantile regression" Bayesian Analysis (2010) ,5.Number 3,pp. 533-556.

**58) Rahim Alhamzawi and Haithem Taha Mohammad Ali. A new gibbs sampler for bayesian lasso. Communications in Statistics-Simulation and Computation, pages 1–17, 2018**

**59)Rahim Alhamzawi and Haithem Taha Mohammad Ali. Bayesian tobit quantile regression with penalty Buchinsky, M. (1998). "Recent advances in quantile regression models: a practical guide for empirical research." Journal of Human Resources, 33: 88–126. 161**

60)Rahim Alhamzawi, Ahmed Alhamzawi, and Haithem Taha Mohammad Ali. New gibbs sampling methods for bayesian regularized quantile regression. Computers in biology and medicine, 110:52–65, 2019.

61)Reed, William J. "The normal-Laplace distribution and its relatives." Advances in distribution theory, order statistics, and inference. Birkhäuser Boston, 2006. 61-7.

62)Samuel Kotz, Tomasz J. Kozubowski, Krzysztof Podg´orski(2001) “The Laplace Distribution and Generalizations” A Revisit with New Applications to Communications ,Economics,Boston:Engineering and Finance Birkauser.

63)Sharma, Sanjib. "Markov chain Monte Carlo methods for Bayesian data analysis in astronomy." Annual Review of Astronomy and Astrophysics 55 (2017): 213-259.

64)Smith, A.F.M., and Roberts, G.O. (1992), Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, Journal of the Royal Statistical Society Series B, Vol. 55, No. 1, pp. 3-23.

65)Sohail Chand (2012) “On Tuning Parameter Selection of Lasso-Type Methods - A Monte Carlo Study” Applied Sciences and Technology (IBCAST). INSPEC Accession Number: 12640393, DOI:10.1109/IBCAST.2012.6177542 ,Publisher: IEEE.

66) Sriram, K., Ramamoorthi, R. V., and Ghosh, P. (2013). “Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density.” Bayesian Analysis, 8: 479–504. MR3066950. doi: http://dx.doi.org/10.1214/ 13-BA817. 164

67) Tibshirani, Robert. "Regression shrinkage and selection via the lasso." Journal of the Royal

**Statistical Society: Series B (Methodological) 58.1 (1996): 267-288.**

68)Tierney, L. (1991), Exploring posterior distributions using Markov chains, Computer Science and Statistics 23rd Symposium at the Interface Edited by Keramidas, E.M., pp. 56

**69) Tobin, James. "Estimation of relationships for limited dependent variables." Econometrica: journal of the Econometric Society (1958): 24-36.**

70)Tsionas, E. G. (2003). Bayesian quantile inference. Journal of statistical computation and simulation, 73(9), 659-674.

71)Wooldridge, J. (2002). Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal*, *1*(2), 117-139.

**72) Wu, Yichao, and Yufeng Liu. "Variable selection in quantile regression." Statistica Sinica (2009): 801-817.**

**73) Young, D. H. & Al-Saadi, S. DH.(1983) "Statistical Theory and Method ", .Vol.1, Al-Risala Presses, Al-Kwait**

74)Yu, Keming, and Julian Stander. "Bayesian analysis of a Tobit quantile regression model." Journal of Econometrics 137.1 (2007): 260-276.

**75) Yu, Keming, and Rana A. Moyeed. "Bayesian quantile regression." Statistics & Probability Letters 54.4 (2001): 437-447.**

**76) Yue, Y. R., & Hong, H. G. (2012). Bayesian Tobit Q Reg model for medical expenditure panel survey data. *Statistical Modelling*, *12*(4), 323-346.**

77) Zhao, K. and Lian, H. (2015). "Bayesian Tobit quantile regression with singleindex models." Journal of Statistical Simulation and Computation, 85: 1247–1263. MR3299347. doi: http://dx.doi.org/10.1080/00949655.2013.873041. 162

78) Zhu, D. and Zinde-Walsh, V. (2009). "Properties and estimation of asymmetric exponential power distribution." Journal of Econometrics, 148: 86–99. MR2494820. doi: http://dx.doi.org/10.1016/j.jeconom.2008.09.038. 166

## الخلاصـــــــــــة:

تناولت هذه الرساله دراسه نموذج انحدار توبيت القسيمي وفق داله الجزاء لأسو .حيث يعتبر تحليل نماذج الانحدار القسيمي احد اهم النماذج في تحليل الانحدار وفي كثير من المجالات العلميه والتي يكون فيها المتغير المعتمد من النوع المراقبه من اليسار والتي تسمى في احيان كثيره ( بيانات توبيت ). ويمكن للباحث في دراسه هذا النوع من البيانات للتعرف على القسيمات الشرطيه (دوال القسيمات ) للمشاهدات التي تتصف بكونها قيما غير مشاهده (قيم كامنه )او قيما شاذه, الهدف الرئيسي من هذه الرساله لوصف الصيغ الشرطيه من قسيميات انحدار توبت . من خلال تقدير معالم نموذج الانحدار مستخدمين طريقه لآسو البيزيه كذلك تم تقديم النموذج الهرمي البيزي للتوزيعات المسبقه من اجل تنفيذ خوارزمية معاينه (Gibbs). اضافه الى ذلك وضحنا التمثيل المختلط بمعلمة القياس من خلال خلط التوزيع المنتظم مع التوزيع الاسي القياسي من خلال اشتقاق تمثيل التوزيع المسبق الاسي المضاعف ودرسنا تقدير بيز لانحدار القسيمي لنموذج توبت من خلال خوارزميه لسلسله ماركوف- منت كارلو / معاينه (Gibbs) بافتراض تجربه محاكاه سناريوهين حيث اظهرت النتائج ان نموذج الجزاء لأسو المقترح هو نموذج قابل للمقارنة تحت مستويات قسيمات مختلفة وحجوم عينات مختلفة ,علاوتا على ذلك وضحنا اداء طريقه لآسو الجزائية المقترحة من خلال تحليل بيانات حقيقيه التي تمثل (200) مشاهده لمعدل ترسب كريات الدم الحمراء الذي يمثل متغير الاستجابة لهذه الدراسة لوجود (20) متغير تفسيري .وبناءا على كل ذلك اظهرت النتائج ان النموذج المقترح يوفر قدره بأسلوب الاختيار للمتغيرات مقارنه مع نماذج الانحدار الاخرى و تحت مستويات قسيمات مختلفة معتمدين على معيار متوسط الخطأ المطلق (MAE) ومتوسط الخطأ التربيعي (MSE ) حيث ان النموذج الافضل هو من يمتلك اقل قيمه لهذه المعاير .

.

جمهوريه العراق
وزارة التعليم العالي والبحث العلمي
جامعه القادسية
كليه الإدارةوالاقتصاد-قسم الاحصاء

# طريقة لاسو بيزي الجديدة في انحدار توبيت القسيمي

رساله مقدمة الى مجلس كلية الادارة والاقتصاد في جامعة القادسية وهي جزء من متطلبات نيل درجة الماجستير في علوم الاحصاء

قدمت بها

اسراء حاتم عبد جبر الشمري

بإشراف

أ. م. د. فاضل حميد هادي الحسني

2023 م　　　　1444 هـ