Republic of Iraq
Ministry of Higher Education &
Scientific Research
University of Al-Qadisiyah
College of Administration & Economics
Department of Statistics

Model Selection in binary regression using nonlocal prior distributions with a practical application

A Thesis by

Maytham Saeed Kadhim

Submitted to the Council of the College of Administration & Economics / University of Al-Qadisiyah in Partial Fulfilment of the Requirements for Master Degree (MSc) in Statistical Sciences

Supervised by

Prof. Dr. Rahim J. Al-Hamzawi

2021 A.D. **1443 A.H.**

بِسْمِ اللّهِ الرَّحْمَنِ الرَّحِيمِ

وَإِن تَعُدُّواْ نِعْمَةَ ٱللَّهِ لَا تُحْصُوهَآ ۗ إِنَّ ٱللَّهَ لَغَفُورٌ رَّحِيمٌ

صدق الله العلي العظيم

سورة النحل الآية (١٨)

ACKNOWLEDGMENTS

First of all, I must thank Allah Almighty, Who enabled me to complete this thesis, which I hope will be of benefit to all who read it and seek help from it.

Also. I would like to express my deep thanks to my supervisor, Professor Dr. Rahim J. Al-Hamzawi, for his valuable observations and assistance throughout the period of study and writing the thesis.

Hence, A word of gratitude goes to the staff of Statistics Department for their help and assistance throughout my study.

I must mention my gratitude to the staff of statistics department, especially Prof. Dr. Mohammed Al-Sharoot, Prof. Dr. Ahmad Naeem, Prof. Dr. Ali Al-kenani, Dr. Muhannad Al-Saadony, Dr. Taha Mohamed, Dr. Fadhil Alhusseini, Dr. Bahar Kadhim, Prof. Dr. Tahir Dikheel and Dr. Hassan Sami for their help and assistance throughout my journey of knowledge.

Finally, I thank my dear parents, , brothers, family, sisters, all my friends and especially My MA Colleagues.

DEDICATION

To the person who did not see me in my graduation uniform

To the soul that will never leave me To the person I feel proud to have named after

Dear father

To the crown of my head, to my heaven in this world, to the one who still raises her palms to pray for me

My mother

To my support

Brothers & Sisters

To my soul mate and life partner

My wife

To my children, Ruqaya, Ali and Karrar

For you, I do my best

Maytham

ABSTRACT

The problem of identifying the active covariates within a linear regression model has received much attention over the recent years. Very recently, Bayesian model selection methods employing nonlocal priors have received considerable attention. One of these methods in linear regression is the simplified shotgun stochastic search with screening (S5), (Shin et al., 2018) proposed two simplified shotgun stochastic search with screening algorithms. The first one is based on the product inverse-moment (piMoM) prior density Johnson and Rossell (2012) and the second one is based on the product exponential moment (peMoM) prior density (Rossell et al., 2013).

In this thesis, we proposed a new Bayesian method for (S5) in linear regression model. Our method is based on the assigning inverse Laplace prior distributions for the regression parameters. Then we extend the idea to linear quantile regression and binary quantile regression.

We compared our (S5) method with other methods through simulation studies and real data analysis. Our methods indicated that the proposed approach performed well compared to other methods in selecting active predictors and estimating parameters.

List of Abbreviations

Abbreviations	Meaning
S5	Simplified Shotgun Stochastic Search with Screening
piMoM	Product Inverse Moment
peMoM	Product Exponential Moment
BLUE	Beast Linear Unbiased Estimator
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
MLE	Maximum Likelihood Estimation
VS	Variable Selection
RSS	Residual Sum of Squares
OLS	Ordinary Least Squares
Lasso	Least Absolute Shrinkage and Selection Operator
AIC	Akaike Information Criterion
MSE	Mean Square Errors
SD	Standard Deviation
SMN	Scale Mixture Normal
SMU	Scale Mixture of Uniform
MCMC	Markov Chain Monte Carlo
aLasso	Adaptive Least Absolute Shrinkage and Selection Operator
GG	Generalized Gaussian
NLP	Nonlocal Prior
BMA	Bayesian Model Average

List of Abbreviations

SCAD	Smoothly Clipped Absolute Deviation
GLM	Generalized Linear Model
SSS	shotgun stochastic search
SIS	Sure Independent Screening
QR	Quantile Regression
BiQR	Binary Quantile Regression
B nonlocal R	Bayesian Nonlocal Regression
B nonlocal QR	Bayesian Nonlocal Quantile Regression
BB nonlocal QR	Bayesian Nonlocal Binary Quantile Regression
MAE	Mean Absolute Errors
MMAD	Median of Mean Absolute Deviations
RMSE	Relative Mean Square Error
MMSE	Median of Mean Squared Errors

Table of Contents

Page

Table of Contents

Table of Contents

List of Tables

Table of Contents

List of Figures

Chapter One

Introduction

and

Literature Review

1.1 Introduction

It is well known that the regression analysis is one of the statistical procedures that explain the relationship between one or more of the explanatory variables and the response variable. The different types of data motivated the authors to research the appropriate regression model that well fit the underlying data. In the thesis , we focused on the binary regression model analysis in which there is a single dependent binary variable which takes only two values ($y = 0$, $y = 1$). The Binary regression model in general may be interpreted as latent variable ${y_i}^*$ model, The standard formula of the binary regression model is defined by:

$$y_i = \begin{cases} 1, & {y_i}^* = \boldsymbol{x}_i{}'\boldsymbol{\beta} + u_i \geq 0 \\ 0, & {y_i}^* = \boldsymbol{x}_i{}'\boldsymbol{\beta} + u_i < 0 \end{cases}, \qquad (1.1)$$

where y_i is the response variable and ${y_i}^*$ is a latent variable defined by:

$${y_i}^* = \boldsymbol{x}_i{}'\boldsymbol{\beta} + u_i. \qquad (1.2)$$

Where ${y_i}^*$ is $n \times 1$ vector of responses , $\boldsymbol{x_i}$ is the $n \times p$ matrix of predictors and u_i is the error term will, $\varepsilon_i \sim NI_n(0, \sigma^2)$, p is a number of independent variables and n is the numbers of observations. In many of the sciences fields, the datasets included large number of the explanatory variables, the regression model becomes more difficult to interpret. The model that fits the data well is the main goal in many data analysis studies. The key idea to select the best model depends on the trade-off between the bias and the variance of the parameter estimates through the variable selection procedure. So, in the cases of $n >> p$**,** the least squares estimates are beast linear unbiased estimator (BLUE). But, in cases of high dimensional data, when $p >> n$, the least squares estimates are not stable and have high variance and high bias. This motivates the authors to find the appropriate minimization procedure that has the properties of the trade-off between the bias and variance for the parameter estimates. Therefore, the model selection procedure in its mechanism works on removing the irrelevant explanatory variable from the model which is not affecting in the response variable, and then produced the predictive model that well fits the data.

Generally, this thesis focuses on the model selection of the binary regression in the scene of the Bayesian analysis under nonlocal prior densities. The first nonlocal prior density named the product exponential moment (peMoM), and the second nonlocal prior is named the product inverse moment (piMoM).

1.2 Literature review

Model selection methods make a trade-off between the bias and variance of the parameter estimates. Usually, high bias parameter estimates come with the low complex model, and high variance parameter estimates come with high complex model. In statistical theory there are a lot of methods have suggested for variable selection. Sub-selection methods, such as, all possible subsets that produced 2^p with p-predictor variables.

James et al. (2013) proposed efficient algorithm for computing the all possible subsets of method. Moreover, forward selection, backward elimination, and stepwise regression are widely developed and applied.

Akaike (1973) suggested Criterion for comparing models through model selection procedure. Akaike Information Criterion (AIC), is defined as:

$$AIC = -2\, ln\, l + 2p, \qquad (1.3)$$

where l be the maximum likelihood estimation function (MLE), p is the number of parameters, and which is one of the most popular methods that is used for variable selection (VS).

Also, the (AIC) produces the inconsistent model (Nishii, 1984). So, (AIC) is weak in selecting the best model when $p \gg n$ (Dziak et al., 2005).

Javed and Mantalos (2013) shows that when the sample size is large, the selected model using AIC is inconsistent.

Schwarz (1978) proposed the Bayesian information criterion (BIC), which is defined as:

$$BIC = -2\ ln\, l + p\ ln\, n, \tag{1.4}$$

where n is the sample size, this method overcomes the problem of (AIC) and chooses a model with good characteristics . However, when $p > n$, its performance does not work well.

Mallows, (1973) suggested Mallow's C_p which is defined as:

$$C_p = \frac{RSS(p)}{S^2} - n + 2p, \tag{1.5}$$

where S^2 is the estimation of the variance to every response in the model. $RSS(p)$ is a residual sum of squares, p is a number of covariates in the model and n is the samples size of data.

Lately, models with high dimensional data have been developed for the purposing of model selection. The regularization method, such as, Ridge method, Lasso method, and many other different methods have become widely desired model selection procedures. The formula of regularization methods can be defined as:

$$\hat{\beta}_{RIDGE} = \underset{\beta}{argmin} (y - \boldsymbol{X\beta})'(y - \boldsymbol{X\beta}) + \varphi_\lambda(\boldsymbol{\beta}) , \tag{1.6}$$

where $\varphi_\lambda(\boldsymbol{\beta})$ is the penalty function which controls the degree of penalty in terms of tuning parameter $\lambda > 0$. Ridge regression, one of the regularization methods proposed by (Hoerl and Kennard, 1970), Also, Ridge regression performs better than the ordinary least square (OLS) method in case of $p \gg n$ with presence of multicollinearity problem, Ridge gives bias estimates with minimum variance.

Tibshirani (1996) proposed the Least Absolute Shrinkage and Selection Operator model (LASSO) regression which automatically choose the variable by shrinking some not important coefficients to zero by forcing (L_1 norm) the least squares (OLS).

Zou and Hastie (2005) proposed the elastic net regression, which compromised between the lasso penalty (L_1) and the Ridge penalty (L_2). In the status of high correlations between variables or when even, this method carry out well in choosing the important variables and estimate the parameters of the model, but it complex needs to the high computational cost.

Zou (2006) proposed another method, which is adaptive Lasso regression. It adds different weights for different coefficients, which produces consistent and unbiased estimates.

Yuan and Lin (2006) proposed the Lasso Group, was expanded by (Kim et al.,2006) to include general loss functions. Although this group makes a selection at the group level, unimportant variables cannot be removed defectively because they define all the variables in the same group.

Wang and Leng (2008) suggested the adaptive group Lasso to overcome the problems of the Lasso Group and improve its performance by imposing a penalty on each coefficient. In addition, the group Lasso is destined to determine the real model and is consistent with the Oracle attribute, however unable to select the bi -level variable.

Huang et al. (2009) proposed a group Bridge regularization method, this newly proposed method provide the variable selection between and within group of predictor variables. Therefore, the proposed method is capable of selecting a bi-level variable with oracle property and sparsity.

Frank and Friedman (1993) proposed the Bridge regression which has attractive properties such as Oracle, unbiasedness. The sub-selection, Ridge, Lasso penalized methods are special cases of the Bridge method. The Bridge method contains two shrinkage parameters, one controls the amount of shrinkage and the other controls the estimate's rotation with concerning for coordinate axes.

Park and Casella (2008) explained the Bayesian approach can be used to estimate Lasso parameters by using a scale mixture of normal (SMN) to create a hierarchical model. Some researchers used the Bayesian approach in their methods.

Hans (2009) proposed Bayesian Lasso regression.

Sun et al.(2010) suggested Bayesian adaptive Lasso and iterative adaptive Lasso, by imposing different adaptive weights and repetitive updating of weights. Adaptive Lasso is more computationally effective than the commonly used regression methods.

Li and Line (2010) suggested the Bayesian elastic net to solve the problem of the elastic net problem by using a Gibbs sampler and avoiding the double shrinkage problem in the elastic net.

Chen et al. (2011) developed a Gibbs sampler for Bayesian Lasso via reversible jump MCMC.

Simon et al. (2013) suggested the sparse group Lasso, show this procedure is able to select bi-level variables using a combination of Lasso and Lasso group penalty on the parameter. However, the estimators are relatively biased due to the shrinkage resulting from the penalty imposed on each parameter.

Mallick and Yi (2014) proposed the new Bayesian Lasso regression, using scale mixture of uniform (SMU) represent to the Laplace density and introducing a new hierarchical model for Bayesian Group Lasso. The method performed well compared with the Bayesian Lasso method.

Mallick and Yi (2017) proposed a Bayesian group Bridge to select bi-level variables, by imposing multiple shrinkage coefficients, this leads to collecting information between the variables within the group.

Mallick and Yi (2018) proposed Bayesian Bridge regression, using (SMU) as new representation for Generalized Gaussian (GG) prior because it makes the Markov Chain Monte Carlo (MCMC) algorithm easy to do. This method has good estimates compared with other methods.

Rossell and Telesca (2015) proposed using the nonlocal prior (NLP) with the regular models and Bayesian model average (BMA) in case of the $p >> n$. Scale mixture of truncated distribution have used to represents the (NLP). The posterior distribution of the interested parameters have notable results in terms of reducing the computations cost . The proposed model given less estimation error comparing with hyper g-priors, a group smoothly clipped absolute deviation (SCAD) and LASSO methods.

Nikooienejad et al. (2016) employed the nonlocal priors into the Bayesian binary outcomes regression through the logistic regression .The variable selection procedure has been done in this work. The (PiMOM) in the parameter estimation for the logistic regression in terms of variable selection procedure . They applied the proposed are thud in a simulation study and real data to show the out per forms of the propped method in. The variable selection procedure by comparing the results with SCAD method.

Cao (2018) proposed novel Bayesian hierarchal by using the (PiMOM) in high- dimensional data sets. Variable selection precedence has applied. Also, new method for selecting the tuning parameter has developed based on the non-local (PiMOM) densities through simulative study.

Dwinatr et al .(2020) introduced the using of non-local priors in the generalized linear model (GLM) from a Bayesian point of view. Variable selection procedure has done best on the logistic regression by adaptor the product (piMoM) . Real data analysis conducted to test the (piMoM) in terms of variable selection through using the MCMC algorithm they compared the results with local priors and show that’s the non-local priors out per forms the local priors based on the values of AIC criterion.

Johnson and Rossell (2012) proposed two classes of non-local prior distribution , (peMoM) and (piMoM). The employed the proposed non local prior in the high dimensional data to test the model selection and variable selection precedence's in Bayesian methods provides more efficient and accurate estimates.

Johnson and Rossell (2008) interned the use of non-local prior distribution in the Bayesian hypothesis tests .The proposed the new priors distribution in testing the Bayesian null hypothesis through the Bayesian factor in linear model . some only test expressions for Bayes factors have blamed in large samples assumptions.

Shin et al . (2017) introduced the scalable Bayesian variable selection by using the non-local priors that proposed by Johnson and Rossell (2012) in settings of ultra-high dimensional datasets. Also, they express the connection between the non-local priors and the reciprocal lasso method. Simulation study was conducted to compare the results of (peMoM) and (piMoM) pars with g-prior, Lassoand scads method and based on so my quality criterions the results of the non-local priors are comparative and better mint .

Considered the model selection from the Bayesian model averaging (BMA) prospective with non-local priors (piMOM) and (peMOM). The regression model was the linear model in case of p is greater than sample size. They represented the nonlocal prior by using the scale mixture of truncated distribution which enables the posterior distribution algorithm to be simple and faster. Moreover, they compared the proposed model with benchmark and hyper-g priors, SCAD, lasso, the results illustrated that the prediction error of the proposed model less is than other methods.

In this thesis, we proposed a new Bayesian method for simplified shotgun stochastic search with screening in linear regression model. Our method is based on assigning inverse Laplace prior distributions for the regression parameters. Then we extend the idea to linear quantile regression and binary quantile regression.

Chapter Two

Nonlocal Prior Density for Linear Regression Model

Nonlocal Prior Density for Linear Regression Model

2.1. Introduction

Suppose that y_n is $n \times 1$ vector of response variable, X_n an $n \times p$ covariates matrix, and $\beta = (\beta_1, \dots, \beta_p)$ is a $p \times 1$, regression covariates parameters. The linear regression model is defined as:

$$y = X\beta + u_i. \tag{2.1}$$

Where $y_n \sim N(x_n\beta, \sigma^2 I_n), u_i \sim N(0, \sigma^2 I_n)$ Johnson and Rossell (2010) defined the nonlocal priors as densities that are exactly zero where as an interest of model parameter is equal to the null value θ_0 in setting of hypothec testing . Also, the nonlocal prior densities have attractive properties in the context of Bayesian variable selection, where the nonlocal prior densities removed the irrelevant explanatory variables faster as $n \to \infty$ see Johnson and Rossell (2012) for the Bayesian model selection under the nonlocal priors densities for the linear regression parameters thought the so called product moment (PMoM) and product inverse moment (PiMoM) prior distribution. Shin et al . (2017) introduced the scale by Bayesian variable selection by using the nonlocal priors that proposed by Johnson and Rossell (2012) in settings of ultrahigh dimensional data set. Also, they express the connection between the nonlocal priors and the reciprocal Lasso method. Simulative study was conducted to compare the results of (PiMoM) , (PeMOM) priors with g- prior. Lasso, the results of the nonlocal priors are comparative the other methods.

Rossell et al. (2013) introduced the so called product exponetional moment (PeMoM) prior density. Hence, the (PeMoM) and (PiMoM) priors density are defined by

$$\pi(\beta_p|\sigma^2,\tau,k) = C^{-|p|}\prod_{j=1}^{|p|}\exp\{-\beta^2{}_{p,j}/2\sigma^2\tau\} - \tau/\beta^2{}_{p,j}. \quad (2.2)$$

Where C is the normalizing constant define by,

$$C = \int_{-\infty}^{\infty}\exp\{-t^2/(2\sigma^2\tau) - \tau/t^2\}dt = (2\pi\sigma^2\tau)^{1/2}\exp\{-(2/\sigma^2)^{1/2}, \quad (2.3)$$

The product inverse-moment (piMoM) prior density (Johnson and Rossell, 2012) defined by:

$$\pi(\beta_p|\sigma^2,\tau,k) = C^{*-|p|}\prod_{j=1}^{|p|}[(\beta_{p,j})^{-2r}\exp\{-\tau/\beta^2{}_{p,j}\}, \quad (2.4)$$

where $C^* = \tau^{-r+\frac{1}{2}}\Gamma\left(r-\frac{1}{2}\right)$.

The following figure (1) illustrates the behaviors of the (PeMoM) and (PiMoM) prior densities with specific values for dispersion parameter (τ), the densities reach zero as the parameter β reach zero. See Shin et al. (2018) for more details.

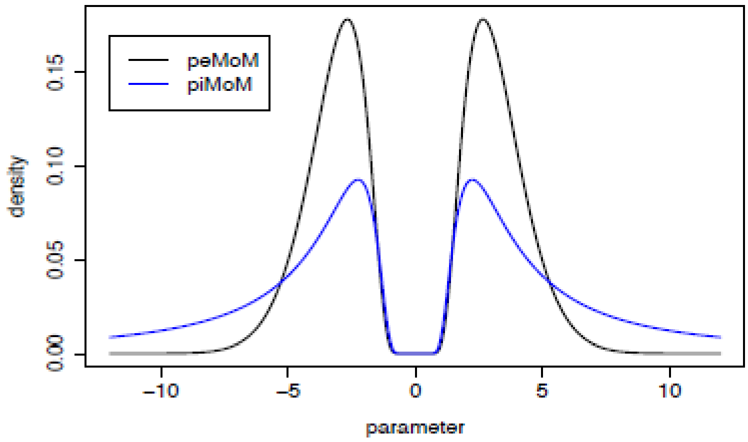


Figure (1) the nonlocal PeMoM and PeMoM) priors densities behaviors (Shin et al., 2018) explained that the (PeMoM) prior has the Gaussian tails, but (PeMoM) prior density has the inverse polynomial tails. Dwinata et al. (2021) stated that (PiMoM) prior density has the heavier fail compared with (PeMoM) and the smallest his to the estimated parameter which givens the less prediction error , Also they defined the (PiMoM) prior as follows:

$$\pi(\beta_p|\tau, \mathrm{r}) = \frac{\sigma^2 \tau^{\frac{rp}{2}}}{\Gamma\left(\frac{r}{2}\right)^p} \prod_{j=1}^{p} |\beta_j|^{-(r+1)} \; e^{\frac{-\tau\sigma^2}{\beta^2}}, \tag{2.5}$$

where $\tau > 0$ scale parameter is the hyper parameter that controls the dispersion of prior density (2.4) r is the order of density shape parameter of the inverse gamma distribution, and β_p are the vector of regression parameters.. Here, σ^2 distributed as invers gamma $(\frac{r+1}{2}, \lambda)$.

Nikooienejad et al. (2016) imposed the beta – binomial prior (subjective prior) that introduced by Scott et al. (2010) on the model space that defined as follows,

$$\rho(p) = \frac{\beta(a+k, b+p-k)}{\beta(a,b)}, \tag{2.6}$$

where $\beta(a, b)$ is the beta function. Castillo et al. (2015) suggested the sparsity inducing prior which decrease the size of a given model, let's say (model p) and sets zero prior probability for the parameters they are not include in model p .

There for , the prior for the parameter β given model p can be defined by:

$$\pi(\beta|p) \propto \pi \ (\beta_p \times g(\beta_{pc}). \tag{2.7}$$

Where $\pi(\beta|p)$ is the non-local prior for the non-zero regression parameter and $(\beta_{c,p}$) the function of probability zero for $(\beta = 0)$. From Bayesian perspective , the marginal likelihood of the model k is defined by:

$$m_p\ (g) = \int L\ (y|\beta, p\ )\pi\ \ (\beta|p) d\beta, \qquad (2.8)$$

where $L(y|.)$ is the likelihood parameter for model k .

Consequently, based on the nonlocal prior (2.7) and the marginal likelihood (2.8) the posterior Probability for model k defined by

$$\pi(k|y_n\ ) = \frac{\pi\ (k)}{\pi(y_n)} \times m_{p.}(y_n), \qquad (2.9)$$

where $\pi(y_n)$ is the marginal of the response y_n , and $m_p(y_n)$ is marginal density of respos of model k , where

$$m_p\ (y_n) = \int \pi\ (y_n|\beta_k) \times \pi_p(\ \beta_p) d\beta_p, \qquad (2.10)$$

2.2. Bayesian Hierarchical Models for Nonlocal Prior in Linear Regression

Following Johnson and Rossell (2012) and cao and Ghosh (2018), we formulate the following hierarchical model representation based on the linear model (2.1) as follows:

First for the nonlocal (peMoM) prior

$$\left.\begin{array}{c} y_n|\beta_k,\sigma^2,k \sim N(x_k^t\,\beta_k\,,\sigma^2\,I_n\,), \\ \pi(\beta_k|\tau,\sigma^2,k) = \left[(2\pi\sigma^2\,\tau)^{\frac{1}{2}}\,e^{-(\frac{2}{\sigma^2})^{\frac{1}{2}}}\right]^{-k} \prod_{j=1}^{k} exp\left[-\frac{\beta^2}{kj}\times(2\sigma^2\tau)-\frac{\tau}{\beta^2{}_{kj}}\right], \\ \pi(\tau)\sim inverse\ gamma\ (\frac{1}{\Gamma_{\frac{1}{2}}}\ \ (\frac{n}{2})^{\frac{1}{2}},\tau^{-\frac{3}{2}}\ \ e^{-\frac{n}{2\tau}}), \\ \pi\ (\sigma^2)\sim inverse\ gamma\ (\frac{1}{\Gamma_{\alpha_1}}(\alpha_2\,)^{\alpha_1},(\sigma^2)^{-(\alpha_1+1)}e^{-\frac{\alpha_2}{\sigma^2}}). \end{array}\right\} \quad (2.11)$$

Second, for the (piMoM) prior , the hierarchical model is defined as:

$$\left.\begin{array}{c} y_n|\beta_k,\sigma^2,k \sim N(x_k^t\,\beta_k,\sigma^2 I_n\,), \\ \pi(\,\beta_p|\tau\,,\sigma^2,p\,) = \left[\tau^{\frac{-(r+1)}{2}}\ \Gamma_{r-1/_2}\right]^{-p} \prod_{j=1}^{p}\left[(\beta_{pj})^{-2r}\ \exp\left(\frac{-\tau}{\beta_{pj}{}^2}\right)\right], \\ \pi(\tau)\sim inverse\ gamma\ (a_1\,=1/2\,,a_2\,=\,n/2), \\ \pi(\sigma^2\,)\sim inverse\ gamma\ (b_1\,,b_2\,). \end{array}\right\} \quad (2.12)$$

Chapter Three

Model selection in quantile regression using inverse Laplace prior density

Model selection in quantile regression using inverse Laplace prior density

3.1 Introduction

Since it is introduction in Koenker and Bassett (1978), quantile regression (QR) models have been studied in-depth. It is insensitive to outliers which are unusual values in the data. QR is able to accommodate non-normal errors, which are common in a lot of real applications (Benoit et al., 2013). The θth quantile of a specific distribution is interpreted as the value such that there is 100θ% of mass on its left side. Compared to the conditional expectation, quantiles are more robust to outliers.

Model selection is important for sparse high dimensional data analysis in many fields of modern science such us economics, genetics, genomics, tomography and tumor classifications. A great body of work exists on model selection in the literature from both frequentist and Bayesian standpoints, such as the least absolute shrinkage and selection operator (LASSO) Tibshirani, (1996), smoothly clipped absolute deviation (SCAD) Fan and Li (2001), the adaptive LASSO (Zou, 2006), the elastic net (Zou and Hastie, 2005), the adaptive elastic net (Zou and Zhang, 2009), the Bayesian LASSO (Hans, 2009; Park and Casella, 2008), the Bayesian adaptive LASSO (Alhamzawi et al., 2012) and the Bayesian elastic net (Li et al., 2010). However, the performance of these methods is usually discounted as the dimensionality grows fast. To overcome this problem, Hans et al. (2007) proposed a Bayesian method for variable selection, with a simple and efficient shotgun stochastic search (SSS) algorithm to explore subsets of covariates that are in the same neighbourhood. Fan and Lv (2008) proposed a sure independent screening (SIS) method to select active set of covariates in ultrahigh-dimensional linear models by considering only those covariates which have a large correlation with the residuals of the current model.

Recently, Shin et al. (2018) proposed a Simplified Shotgun Stochastic Search with Screening (S5) algorithm to explore the enormous model space and reduces the computing time by using the idea of SIS. Specifically, Shin et al. (2018) proposed S5 algorithms based on the product exponential moment prior densities (Rossell et al., 2013) and the product inverse-moment prior densities (Johnson and Rossell, 2012) for the regression coefficients. By using simulation studies and real data analysis, they show that their algorithm is effective in model selection and able to accelerate the computation speed under a variety of scenarios.

Motivated by their empirical finding, we extend the S5 algorithm to quantile regression using independent inverse Laplace prior densities for the regression coefficients. Over the current decade, model selection in quantile regression has received considerable attention (for example see, Alhamzawi and Yu, 2014; Belloni et al., 2011; Bradic et al., 2011; Chen et al., 2013; Lamarche, 2010; Li et al., 2010; Tang et al., 2013; Zheng et al., 2013). However, variable selection in quantile regression by using S5 algorithm (or in short, S5-QR) has not been proposed, yet. Instead of using the product exponential moment prior densities and the product inverse moment prior densities, we use the inverse Laplace prior densities for the regression coefficients. Under this prior, the Bayesian posterior mode estimate is equivalent to reciprocal Lasso estimate (Mallick et al., 2019), which is not proposed yet in quantile regression. As demonstrated later by simulations, S5-QR provides more accurate estimates and better prediction accuracy than other existing methods in quantile regression.

3.2 Methods

3.2.1 Quantile Regression with reciprocal LASSO penalty

In a linear quantile regression setup, we have the following model:

$$\boldsymbol{y} = X\beta + \boldsymbol{\epsilon}\,, \qquad \theta \in (0,1),$$

where $\boldsymbol{y}$ is the $n \times 1$ vector of centered responses, $X = (x_1, \dots, x_n)\,'$is the $n \times p$ matrix of standardized regressors, β is the $p \times 1$ vector of quantile coefficients to be estimated, and ϵ is the $n \times 1$ vector of errors whose distribution is restricted to have the θth quantile equal to zero. The regression coefficient vector β and the error term $\boldsymbol{\epsilon}$ should be indexed by θ, i.e. β_θ and ϵ_θ For sake of simplicity, however, we will omit θ in the rest of this paper. The unknown parameter vector β is estimated by minimizing (Koenker and Bassett, 1978).

$$\underset{\beta}{min} \sum_{i=1}^{n} \rho_\theta\left(y_i - x'_i\beta\right), \tag{3.1}$$

where $p_\theta(w) = w\{\theta - I(w < 0)\}$ and $I(.)$ is the indicator function. The prediction accuracy of the ckeck function (3.1) can often be improved by selecting an active subset of covariates. In this chapter, to improve the prediction accuracy we consider the reciprocal Lasso quantile regression (rLasso-QR) which has not been proposed yet, that results from the following regularization problem:

$$\underset{\beta}{min} \sum_{i=1}^{n} \rho_\theta(y_i - x'_i\,\beta) + \lambda \sum_{j=1}^{p} \frac{1}{|\beta j|} I\{\beta_j \neq 0\}, \tag{3.2}$$

where $I(.)$ denotes an indicator function and $\lambda > 0$ is the tuning parameter that controls the degree of penalization.

The rLasso penalty $\lambda \sum_{j}^{p} \frac{1}{|\beta j|} I\{\beta_j \neq 0\}$ (Song and Liang, 2015), is decreasing in$(0, \infty)$, discontinuous at 0, and converge to ∞ when the regression parameters approach zero. It shares the same oracle property and same rate of estimator error with other Lasso-type penalties. Compared to traditional penalization functions (e.g., Lasso and SCAD) that are give nearly zero coefficients nearly zero penalties, the rLasso penalty give nearly zero coefficients infinity penalties, which makes it very attractive for variable selection. In this chapter, rather than minimizing (3.2), we solve the problem by constructing S5-QR algorithm via a Gibbs sampler which involves constructing a Markov chain having the joint posterior for β as its stationary distribution.

3.2.2 Posterior inference

Since quantile regression does not equipped with a parametric likelihood, to proceed a Bayesian analysis we model the errors by the asymmetric Laplace distribution (ALD) (Alhamzawi et al., 2012; Chen et al., 2013; Kozumi and Kobayashi, 2011; Yu and Moyeed, 2001).The density function of an ALD is:

$$f(y|\mu,\theta) = \theta(1-\theta)\exp\{-\rho_\theta(y-\mu)\}, \qquad (3.3)$$

where, μ is a location parameter. In our model setup, the conditional distribution for the observations is:

$$f(y|\beta,\theta) = \theta^n(1-\theta^n)\exp\{-\sum_{i=1}^{n}\rho_\theta(y_i - x'_i\,\beta)\}. \qquad (3.4)$$

Maximizing the joint likelihood function over β is equivalent to minimizing the usual quantile check function $\sum_{i=1}^{n}\rho_\theta(y_i - x'_i\beta)$. However, direct use of this likelihood is rather unsuitable for posterior computation because the posterior distribution of β does not have a closed form. In this context, Kozumi and Kobayashi (2011) show that ALD can be written as a location-scale mixture representation, i.e.,

$$\epsilon_i = (1-2\theta)w_i + \sqrt{2w_i z_i}, \qquad (3.5)$$

where wi and zi are mutually independent, $wi \sim \exp(\theta(1-\theta))$ and $z_i \sim N(0,1)$. Marginally, the error distribution ϵ_i maintains its ALD form. However, conditional on the latent variable $w_i\epsilon_i$ follows a normal distribution. Thus, posterior inference can be suitably carried out using Gibbs sampler.

Following Mallick et al. (2019), a Bayesian solution for the minimization problem in (3.2) can be obtained by placing appropriate priors on the regression coefficients that will mimic the effects of the rLasso penalty. As apparent from (3.2), this choice of prior would be an independent inverse Laplace density on each of the coefficients

$$\pi(\beta) = \prod_{j=1}^{p} \frac{\lambda}{2\beta_j^{\,2}} \exp\left\{\frac{\lambda}{|\beta_j|}\right\} I\left\{\beta_j \neq 0\right\}. \tag{3.6}$$

Hence, Gibbs sampling algorithm for the rLasso-QR is constructed by sampling β and $\boldsymbol{w} = (w_1, \dots, w_n)'$ from their full conditional distributions. However, because no point mass at zero is assigned in this regularization problem, the samples of the regression parameters for the inactive set of covariates would not be exactly zero. To overcome this problem, we propose an efficient Simplified Shotgun Stochastic Search with Screening in Quantile Regression (S5-QR) to explore the enormous model space.

3.3 Model Setup

To fix the terminology, let $\boldsymbol{p} = \{p_1, \ldots, p_{|\boldsymbol{p}|}\}$ denote a model, where$1 \leq p_1 <, \ldots, < p_{|\boldsymbol{p}|}$, with $\beta_{pl} \neq 0 \; for \; l = 1, \ldots, |\boldsymbol{p}|$ and all other components of β are 0. Let X_p and $\beta = \{\beta_{p,1}, \ldots, \beta_{p,|\boldsymbol{p}|}\}$ are the design matrix and the regression coefficients of the model **k** only including the predictors with $\beta_{pl} \neq 0$. Let $\boldsymbol{t}$ denote the true model and the cardinality of model $\boldsymbol{t}$ is denoted by$t = |t|$. Under each model $\boldsymbol{p}$, the sampling density for the observations is

$$y_n|\beta_p \sim N(X_p\, \beta_p + (1 - 2\theta)w, W_n), \qquad (3.7)$$

where $W_n = dig\;(2w_n, \ldots, 2w_n)$. Given a model k, the inverse Laplace prior on the regression coefficients is defined as:

$$\pi(\beta_p|\lambda, \boldsymbol{p}) = \prod_{j=1}^{|\boldsymbol{p}|} \frac{\lambda}{2\beta_{p,j}^2} \exp\left\{- \frac{\lambda}{|\beta_{p\,,j}|}\right\} I\{\beta_{p,j} \neq 0\}. \qquad (3.8)$$

Following Shin et al. (2018), we put a uniform prior on the model space of the form $\pi\;(p) \propto I\;(|\boldsymbol{p}| \leq q_n)\; with\;\; q_n < n\,,\;\; where\; I(.)$ denotes the indicator function. The basic idea in calculating the posterior probabilities of each model is to get the marginal likelihood of the observations $m_p\;(y)$ under model $\boldsymbol{p}$ by integrating out the model parameters. Under model $\boldsymbol{p}$, the marginal likelihood of the observations $m_p\;(y)$ can be obtained by integrating out β_p , resulting in

$$m_p(y) = (4\pi w_i)^{-\frac{n}{2}} Q^*{}_p \exp\{-R^*{}_p/4\}, \qquad (3.9)$$

Where

$$\left.\begin{aligned} R^*{}_p &= u'\big(I_n W X_p (X'{}_P W X_P)^{-1} X' W_P\big)u, \\ u &= y - (1 - 2\theta)w, \end{aligned}\right\} \qquad (3.10)$$

$$Q^*{}_p = \int \prod_{j=1}^{|p|} \frac{\lambda}{2\beta^2{}_{p,j}} \exp\{-(\beta_p - \widehat{\beta_p})' \textstyle\sum_p^{*-1} (\beta_p - \widehat{\beta_p})/4 - \sum_{j=1}^{|p|} \frac{\lambda}{|\beta_{p,j}|}\}, \quad (3.11)$$

$$\left.\begin{array}{c} \widehat{\beta_p} = ((X'{}_P W X_P)^{-1} X' W_P)u, \\ \textstyle\sum_p^* (X'{}_P W X_P)^{-1}\,. \end{array}\right\} \quad (3.12)$$

To estimate $\widehat{\boldsymbol{\beta}_{\boldsymbol{p}}}$, we assume that the size of $\widehat{\boldsymbol{\beta}_{\boldsymbol{p}}}$ is p and $\lambda \sim \lambda^{c-1} \exp(-d\lambda)$. We follow the Gibbs sampler of Alhamzawi and Mallick (2020). This Gibbs sampler is described with some modifications in Algorithm 1.

3.4 Simplified Shotgun Stochastic Search with Screening in QR

Shin et al. (2018) proposed a simplified shotgun stochastic search with screening (S5) algorithm in an attempt to reduces the computing time of the SSS algorithm without losing the capacity to search the interesting region in the model space. They introduced temperature parameter" to explore a broader spectrum of models. The Simplified Shotgun Stochastic Search with Screening (S5) algorithm (Shin et al.,2018)

Let $t_1 > t_2 > \cdots > 0$ is a set of temperature schedule and nbd $(\boldsymbol{p}) = \{\Lambda^+, \Lambda^-, \Lambda^0\}, where\ \Lambda^+ = \{\boldsymbol{p} \cup \{j\}: j \in \boldsymbol{p}^c\}, \Lambda^- = \{\boldsymbol{p}\ \{j\}: j \in \boldsymbol{p}\}, and\ \Lambda^0 = [\{\boldsymbol{p} \backslash j\}] \cup \{l\}: l \in \boldsymbol{p}^c, j \in \boldsymbol{p}\}$.

This method is described in Algorithm (2).

<u>Algorithm 1</u>: MCMC sampling for the Bayesian reciprocal LASSO quantile regression (Alhamzawi and Mallick, 2020).

Input: (y, X)

Initialize: $(\beta_{\boldsymbol{p}}, w, u, \lambda)$

for $t = 1, \ldots, (t_{max} + t_{burn-in})$ ***do***

1. Sample $w^{-1} \mid . \sim \prod_{i=1}^{n} Inverse - Gaussian\left(\frac{1}{2}, \frac{1}{|y_i - \acute{x}_i \beta_p}, \frac{1}{2}\right)$
2. Sample $u \mid . \sim \prod_{j=1}^{p} Exponential(\lambda) I \left\{u_j > \frac{1}{|\beta_j|}\right\}$
3. Sample $\beta_p \mid .$ from a truncated multivariate normal proportional to

$$N_p((X'W^{-1}X)^{-1}X'W^{-1}(y - \theta w)), 2(X'W^{-1}X)^{-1} \prod_{j=1}^{p} I\left\{|\beta_j| > \frac{1}{u_j}\right\}$$

4. Sample $\lambda \mid . \sim Gamma(c + 2p \,, d + \sum_{j=1}^{p} \frac{1}{|\beta_j \;|}\;)$

end for

Algorithm 2: Simplified Shotgun Stochastic Search with Screening (S5)

Input:$(y\,, X)$

Initialize:$\boldsymbol{p}^{(1\,,1)}$

Select: a set of predictors X^{*}corresponding to the initial model $\boldsymbol{p}^{(1,1)}$

Select: a subset of predictors from X^{*} after the first screening step $\boldsymbol{S^{(1,1)}}_{\boldsymbol{p}}$

for$l = 1, \dots, L$ **do**

for$i = 1, \dots, (j-1)$ **do**

1. Compute all $\pi\;(\boldsymbol{p}|\boldsymbol{y})$for all $\boldsymbol{p} \in n\;bd_{scr}\left(p^{(i,j)}\right)$
2. Update $\boldsymbol{p}^{+}$from Λ_{scr}^{+} with probability proportional to $\pi\;(\boldsymbol{p}\;|\boldsymbol{y}\;)^{1\backslash ti}$
3. Update $\boldsymbol{p}^{-}$from Λ^{-} with probability proportional to $\pi\;(\boldsymbol{p}\;|\boldsymbol{y}\;)^{1\backslash ti}$
4. Update $\boldsymbol{p}^{(i+1,l)}$from $\{\boldsymbol{p}^{+}, \boldsymbol{p}^{-}\}$ with probability proportional to $\left\{\pi\;(\boldsymbol{p}^{+}|\boldsymbol{y}\;)^{1\backslash tj}\;, \pi\;(\boldsymbol{p}^{-}|\boldsymbol{y}\;)^{1\backslash ti}\right\}$
5. Update $\boldsymbol{S}_{\boldsymbol{p}(i+1,l)}$ according to $\left|\boldsymbol{r}^{T}{}_{p(i+1,l\,)}\,X_j\right|: j = 1\,, \dots, p$

end for

end for

Chapter Four

Binary Quantile Regression

4.1 Binary Quantile Regression

Binary Quantile Regression (BiQR) is important special case of QR which is widely used in genetics, engineering farming, finance, medicine, and other fields of knowledge. Manski (1975) developed methods to estimate BiQR models within the traditional framework and Benoit and Van den Poel (2012) propose a Bayesian framework to BiQR Kordas (2002) proposed binary QR for the aim of classification employing Quantiles. The standard binary quantile regression problem for the θth quantile can be defined as:

$$y_i{}^* = \boldsymbol{x}_i{}'\beta + u_i, \qquad \theta \in (0,1) \tag{4.1}$$

$$y_i = \begin{cases} 1, & if\ y_i{}^* > 0 \\ 0, & otherwise' \end{cases} \tag{4.2}$$

where y_i is the observed response of the subject determined by the latent unobserved response $y_i{}^*$, $\boldsymbol{x}_i$ is the $p \times 1$ vector of regressors, β is the $p \times 1$ vector of quantile coefficients to be estimated, and u_i is the $n \times 1$ vector of errors whose distribution is restricted to have the θth quantile equal to zero. For an overview, we refer to Algamal et al. (2018); Alhamzawi (2015); Benoit et al. (2013); Benoit and Van den Poel (2012); Bottai et al. (2010); Hashem et al. (2016); Ji et al. (2012); Li and Miu(2010); Rahman and Vossmeyer (2019); Wei et al. (2019).

In a way similar to (Benoit et al., 2013; Benoit and Van den Poel, 2012), Binary QR estimation may proceed by the solution to the following minimization problem:

$$\underset{\beta}{min} \sum_{i=1}^{n} \rho_\theta\left(y_i - g(x'_i\beta)\right), \tag{4.3}$$

where $g(x'_i\beta)I\{x'_i\beta > 0\}$.

A serious challenge in BiQR lies in the identification of the active regressors in regression. Here, we improve the prediction accuracy of BiQR by proposing the reciprocal Lasso binary quantile regression (rLasso-BiQR) which has not been proposed yet, that results from the following regularization problem:

$$\min_{\beta} \sum_{i=1}^{n} \rho_\theta\left(y_i - g(x'_i\beta)\right) + \lambda \sum_{j=1}^{p} \frac{1}{|\beta_j|} I\left\{\beta_j \neq 0\right\}. \qquad (4.4)$$

In this chapter, rather than minimizing (4.3), we solve the problem by constructing S5-QR algorithm via a Gibbs sampler which involves constructing a Markov chain having the joint posterior for β as its stationary distribution.

4.2 Methods

4.2.1 Binary Quantile Regression with reciprocal LASSO penalty

In this section, we follow Kozumi and Kobayashi (2011) and use the following mixture representation:

$$u_i = (1-2\theta)w_i + \sqrt{2\,w_i z_i}\,, \tag{4.5}$$

Where w_i and z_i are mutually independent, $w_i \sim \exp(\theta(1-\theta)$ and $z_i \sim N(0,1)$. We use the same prior distributions in the previous section. Under each model **p**, the sampling density for the observations is:

$$y_n|\beta_p \sim N\big(X_p\beta_p + (1-2\theta)\,w, W_n\big), \tag{4.6}$$

where $w_i = diag(2w_1, \ldots, 2w_n)$. Again, we assume the inverse Laplace prior on the regression coefficients. Then the full conditional distribution of $y_i{}^*$ is given by:

$$y_i{}^*|y_i, \beta, \theta \sim N(\boldsymbol{x}_i{}'\beta + (1-2\theta)w_i\; 2w_i)\; truncated\; at\; the\; left\; by\; 0, if\; y_i = 1, \tag{4.7}$$

$$y_i{}^*|y_i, \beta, \theta \sim N(\boldsymbol{x}_i{}'\beta + (1-2\theta)w_i\; 2w_i)\; truncated\; at\; the\; left\; by\; 0, if\; y_i = 0, \tag{4.8}$$

4.2.2 Simplified Shotgun Stochastic Search with Screening in BiQR

Under model p, the marginal likelihood of the observations $m_p(y)$ can be obtained by integrating out β_p, resulting in

$$m_p(y) = (4\pi w_i\ )^{-1}\ Q_p \exp\{-R^*{}_p/4\}, \tag{4.9}$$

$$\left.\begin{array}{c} R^*{}_p = u'\ (I_n - WX_p\big(X'{}_pWX_p)^{-1}X'{}_pW\big)u, \\ u = y - (1-2\theta)w, \end{array}\right\}, \tag{4.10}$$

$$Q^*{}_p = \int \prod_{j=1}^{|p|} \frac{\lambda}{2\beta^2{}_{p,j}} \exp\left\{-\ (\beta_p - \hat{\beta}_p\ )'\ \Sigma_p^{*-1}\big(\beta_p - \hat{\beta}_p\ \big)/4 - \sum_{j=1}^{|p|} \frac{\lambda}{|\beta_{p,j}|}\right\}, \tag{4.11}$$

$$\hat{\beta}_p = (X'{}_p\ WX_p)^{-1}X'{}_pW_u,\ \Sigma_p^* = (X'{}_pWX_p)^{-1}, \tag{4.12}$$

Chapter Five

Simulation Study
and
Real Data

5.1 Simulation Study

5.1.1 Introduction

In this section , we want to investigate our proposed methods through a simulation approach .Three proposed methods are used to study three different cases, first method will be used to coefficients estimation and variables selection in the traditional regression model which is named Bayesian nonlocal regression model, denoted by (B nonlocal R).Second method will be used for coefficients estimation and variables selection in quantile regression model which is named Bayesian nonlocal quantile regression model, denoted by (B nonlocal QR). Third method will be used to coefficients estimation and variables selection in traditional regression model which is named Bayesian nonlocal binary quantile regression model, denoted by (BB nonlocal QR). Each method will be compared with other methods in the same filed. In this simulation study , we will employ five example, first three examples belong to Bayesian nonlocal regression model. This proposed method is also compared with four other method (Lasso, Adaptive Lasso, Bayesian Lasso and B Reg N) through sample size (n=25, n=50, n=100, n=150 and n=250), three criteria are used in this study are relative mean square error, denoted by (RMSE), $RMSE = \frac{\|X^t\hat{\beta} - X^t\beta^R\|^2}{\sigma}$

where β^t is true parameters , $\hat{\beta}$ is estimation parameters , σ is stander deviation of random error. Median of mean absolute deviations denoted by (MMAD).$MMAD = median(mean(|X^t\hat{\beta} - X^t\beta^r|)$. where β^t is true parameters , $\hat{\beta}$ is estimation parameters. Mean absolute error denoted by (MAE) .

<u>Example 1</u> (sparse case)

In this example, the true parameters are $\beta^t = (0,1,0,0,2,0,1.5,0,1)^t$. Therefore , We generate data set from traditional regression model, as follow :

$$y = x_{2i} + 2x_{5i} + 1.5x_{7i} + x_{9i} + \varepsilon_i \qquad [i=1,\ldots,100]$$

where $\varepsilon_i \sim N(0, \sigma^2)$

We generate Nine explanatory variables from a multivariate normal with mean 0, and cov-variance $(x_i\,, x_j) = 0.5^{|i-j|}$.

The RMSE, MMAD,MAE and standard deviations (SD) are inserted in table 1. It is clearly observed that via all the sample size under study. Our proposed method (B nonlocal R) generate smaller RMSE, MMAD,MAE and (SD) comparison to other method (LASSO, Adaptive Lasso, Bayesian Lasso and B Reg N). This means our proposed method is very accurate. It is a very good method to achieving variables selection and coefficient estimation. In general , our proposed method has a good performance via small sample size and also ,it has a good performance via large sample size. Finally, the method (B nonlocal R) is not different its performance via all levels of sample size with sparse models:

Table 1: Show results (RMSE) , (MMAD) and MAE via averaged over 50 replications

n	Methods	RMSE	MMAD	MAE
25	LASSO	0.436(0.372)	0.418 (0.383)	0.342(0.315)
	Adaptive Lasso	0.428(0.341)	0.519 (0.421)	0.410(0.311)
	Bayesian Lasso	0.354(0.308)	0.357 (0.314)	0.412(0.342)
	B Reg N	0.362(0.307)	0.293 (0.295)	0.389(0.317)
	B nonlocal R	0.256(0.063)	0.243 (0.216)	0.351(0.224)
50	LASSO	0.315(0.286)	0.363 (0.273)	0.405(0.351)
	Adaptive Lasso	0.406(0.371)	0.368 (0.364)	0.351(0.351)
	Bayesian Lasso	0.383(0.325)	0.337 (0.268)	0.373(0.262)
	B Reg N	0.362(0.385)	0.392 (0.306)	0.312(0.279)
	B nonlocal R	0.302(0.275)	0.238 (0.218)	0.227(0.085)
100	LASSO	0.364(0.293)	0.384 (0.318)	0.343(0.277)
	Adaptive Lasso	0.436(0.382)	0.323 (0.326)	0.297(0.267)
	Bayesian Lasso	0.373(0.374)	0.353 (0.248)	0.294(0.092)
	B Reg N	0.268(0.163)	0.220 (0.083)	0.239(0.138)
	B nonlocal R	0.219(0.077)	0.2288 (0.092)	0.219(0.118)
150	LASSO	0.347(0.318)	0.368 (0.269)	0.326(0.235)
	Adaptive Lasso	0.327(0.284)	0.316 (0.254)	0.307(0.326)
	Bayesian Lasso	0.269(0.231)	0.263 (0.232)	0.257(0.253)
	B Reg N	0.282(0.203)	0.314 (0.251)	0.272(0.263)
	B nonlocal R	0.226(0.095)	0.253 (0.183)	0.243(0.228)
250	LASSO	0.452(0.353)	0.416 (0.375)	0.427(0.329)
	Adaptive Lasso	0.362(0.284)	0.462 (0.386)	0.376(0.304)
	Bayesian Lasso	0.428(0.368)	0.402 (0.392)	0.451(0.307)
	B Reg N	0.384(0.362)	0.382 (0.328)	0.327(0.172)
	B nonlocalR	0.284(0.107)	0.253 (0.192)	0.304(0.096)

Note: In the parentheses are SDs of the MAD.

We can see the regression parameters estimates via our proposed method is very closed from normal distribution through histogram graphs. Also, it convergence to stationary this clearly from trace plot (when sample size n=50). This means the MCMC sampler is easy and effective.

Figure 1 . Histogram of (B nonlocalR) parameter estimation

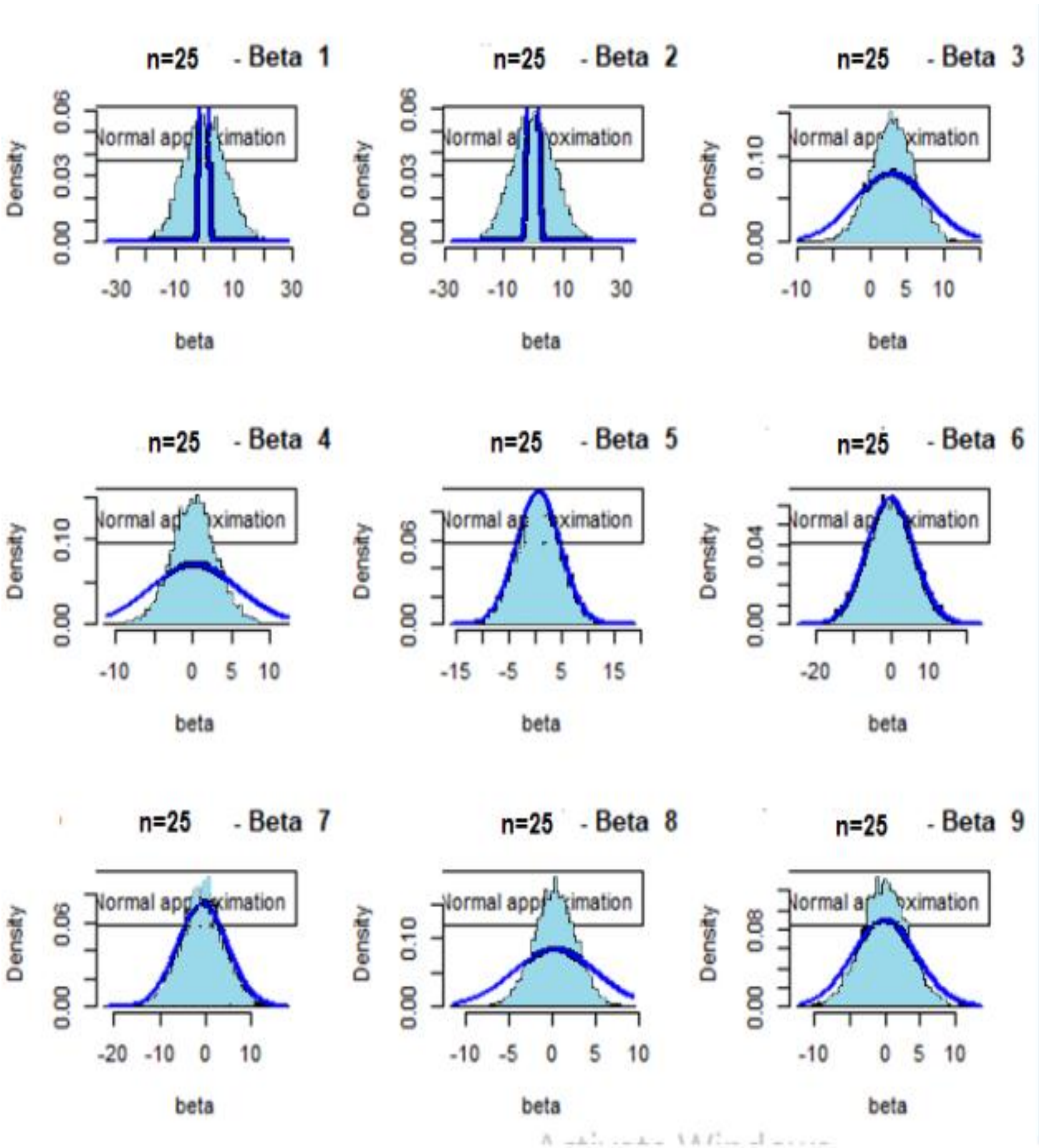
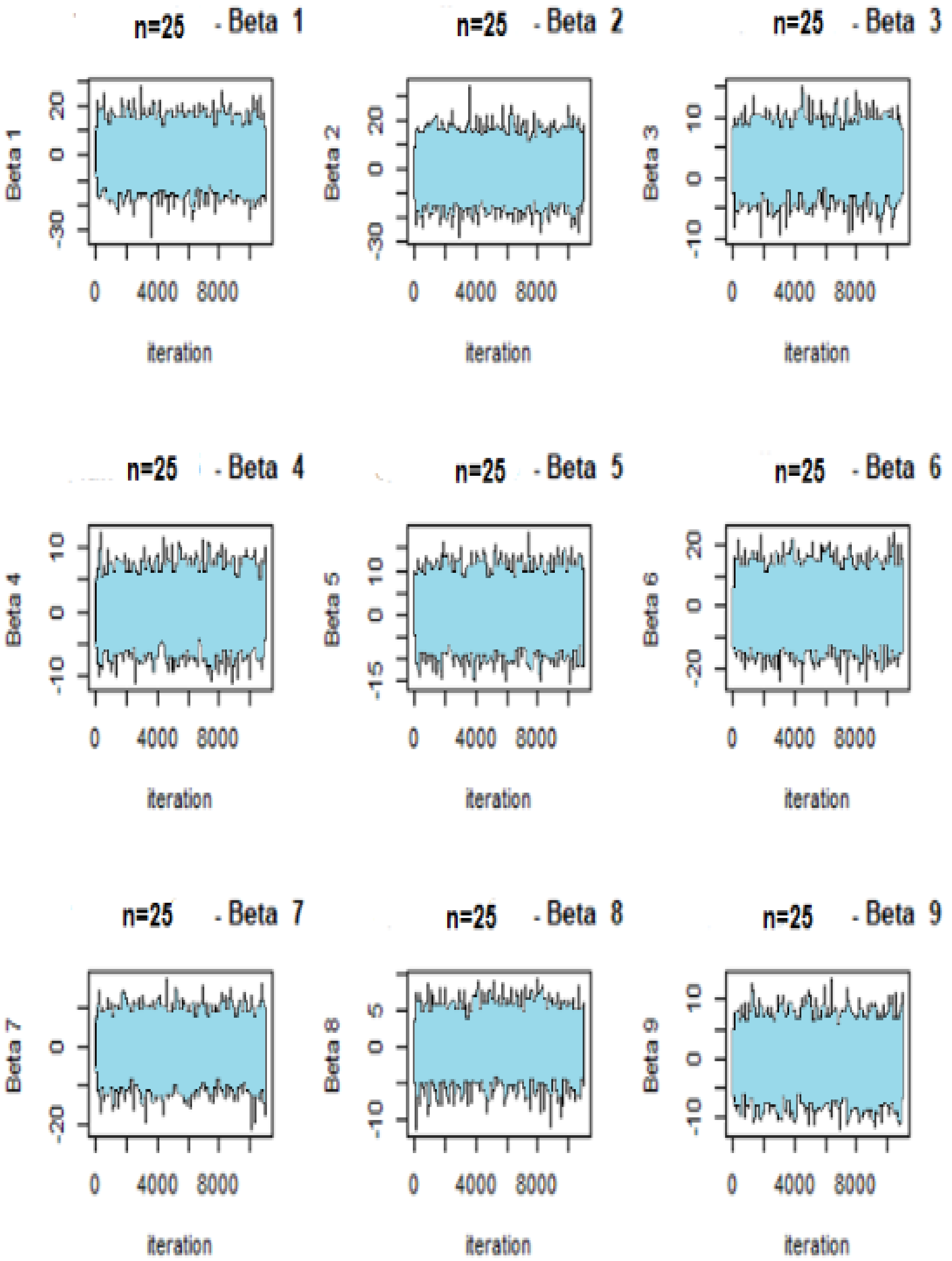


Figure 2. Trace plots of (B nonlocal R) with n=25



<u>Example 2</u>. (very sparse case)

In this example, the true parameters are $\beta^t = (0,1,0,0,0,0,0,0,0)^t$. Therefore , We generate data set from traditional regression model, as follow formula

$$y = x_{2i} + \varepsilon_{i,} \qquad [i=1,\ldots,100]$$

where $\varepsilon_i \sim N(0, \sigma^2)$

We generate Nine explanatory variables from a multivariate normal with mean 0, and cov-variance $(x_i , x_j) = 0.5^{|i-j|}$.
The RMSE, MMAD,MAE and standard deviations (SD) are inserted in table 2. It is clearly observed that via all the sample size under study. Our proposed method (B nonlocal R) generate smaller RMSE, MMAD,MAE and (SD) comparison to other method (LASSO, Adaptive Lasso, Bayesian Lasso and B Reg N). This means our proposed method is very accurate. It is a very good method to achieving variables selection and coefficient estimation. In general , our proposed method has a good performance via small sample size and also ,it has a good performance via large sample size. In finally, the method (B nonlocal R) is not different its performance via all levels of sample size with sparse models.

Table 2: Show results (RMSE), (MMAD) and (MAE) via averaged over 50 replications

n	Methods	RMSE	MMAD	MAE
25	LASSO	0.561(0.482)	0.521 (0.472)	0.421(0.204)
	Adaptive Lasso	0.525(0.439)	0.472 (0.342)	0.408(0.293)
	Bayesian Lasso	0.603(0.429)	0.463 (0.371)	0.528(0.301)
	B Reg N	0.492(0.459)	0.426 (0.361)	0.473(0.361)
	B nonlocal R	0.203(0.145)	0.296 (0.182)	0.238(0.083)
50	LASSO	0.382(0.289)	0.340 (0.232)	0.351(0.205)
	Adaptive Lasso	0.381(0.269)	0.374 (0.304)	0.391(0.283)
	Bayesian Lasso	0.362(0.284)	0.462 (0.386)	0.269(0.231)
	B Reg N	0.327(0.284)	0.462 (0.386)	0.427(0.329
	B nonlocal R	0.212(0.153)	0.263(0.147)	0.216(0.147)
100	LASSO	0.382(0.289)	0.340 (0.232)	0.382(0.245)
	Adaptive Lasso	0.382(0.289)	0.357 (0.314)	0.351(0.205)
	Bayesian Lasso	0.463 (0.371)	0.426 (0.361)	0.412(0.342)
	B Reg N	0.335 (0.172)	0.330 (0.845)	0.521 (0.832)
	B nonlocal R	0.243 (0.130)	0.217 (0.092)	0.234 (0.170)
150	LASSO	0.345 (0.263)	0.547 (0.382)	0.529 (0.353)
	Adaptive Lasso	0.482 (0.334)	0.518 (0.334)	0.465 (0.332)
	Bayesian Lasso	0.382 (0.351)	0.317 (0.283)	0.520 (0.393)
	B Reg N	0.397 (0.280)	0.232 (0.245)	0.336 (0.398)
	B nonlocal R	0.256 (0.145)	0.201 (0.091)	0.235 (0.134)
250	LASSO	0.521 (0.432)	0.503 (0.372)	0.612 (0.492)
	Adaptive Lasso	0.434 (0.370)	0.352 (0.257)	0.334 (0.322)
	Bayesian Lasso	0.434 (0.332)	0.346(0.299)	0.403(0.243)
	B Reg N	0.464 (0.874)	0.469 (0.382)	0.492 (0.382)
	B nonlocalR	0.229 (0.153)	0.246 (0.182)	0.246 (0.183)

Note: In the parentheses are SDs of the MAD

We can see the regression parameters estimates via our proposed method is very closed from normal distribution through histogram graphs. Also, it convergence to stationary this clearly from trace plot (when sample size n=100). This mean the MCMC sampler is easy and effective.

Figure 3 . Histogram of (B nonlocalR) parameter estimation

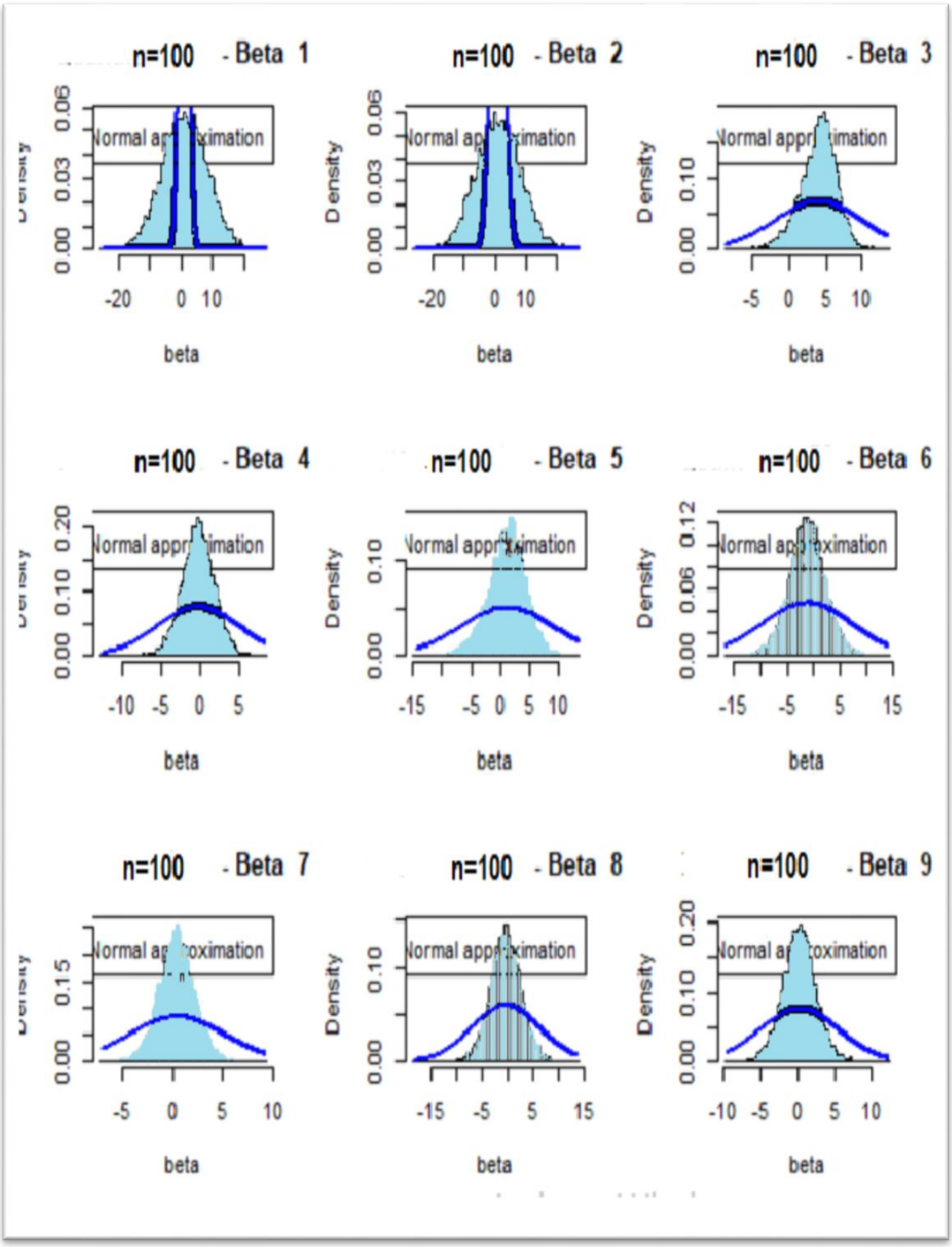
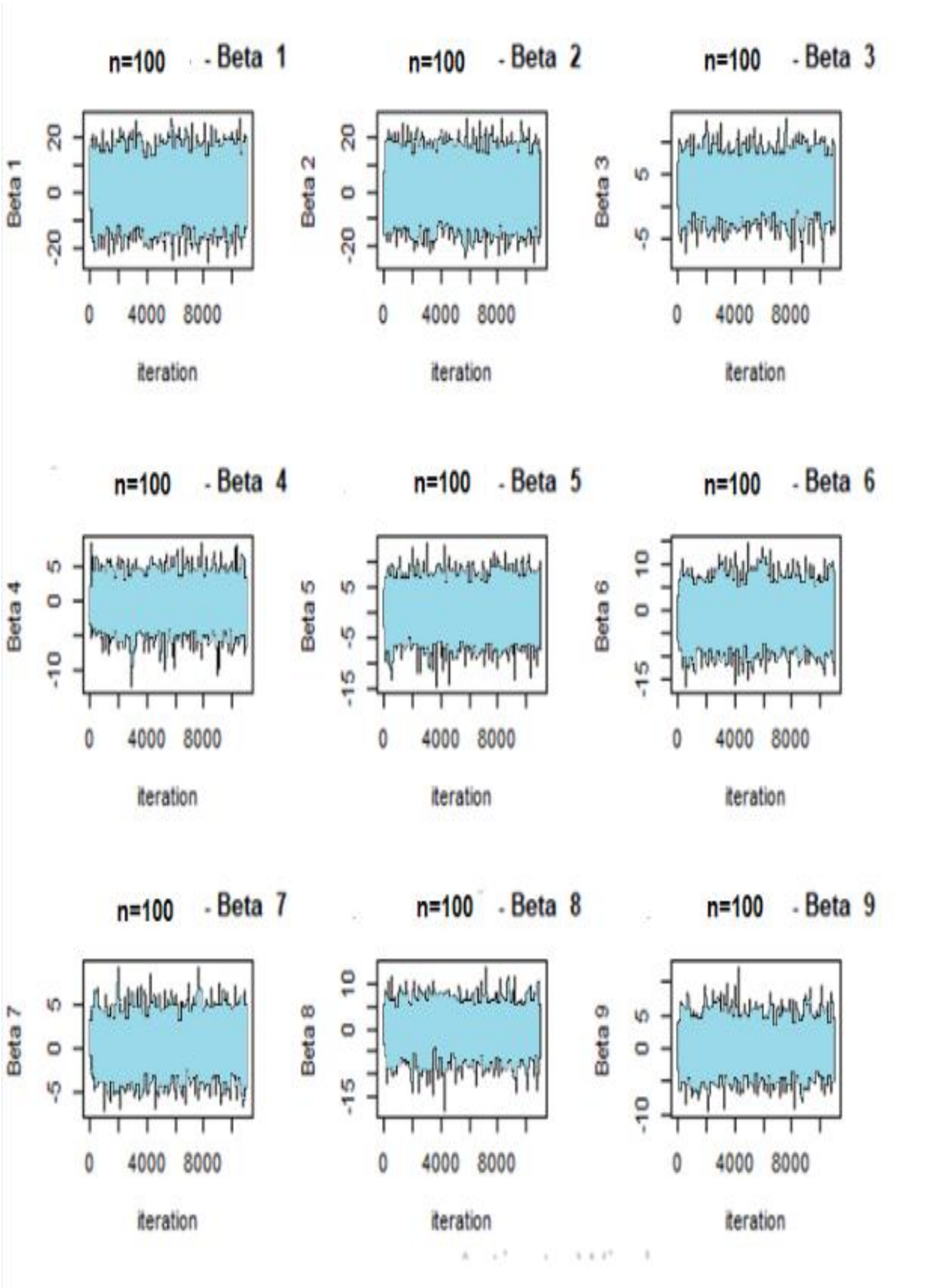


Figure 4. Trace plots of (B nonlocalR) with n=100



<u>Example 3</u> (density case)

In this example, the true parameters are $\beta^t = (0.85,0.85,0.85,0.85,0.85,0.85,0.85,0.85,0.85)^t$. Therefore , We generate data set from traditional regression model, as follow formula

$$y = 0.85x_{1i} + 0.85x_{1i} + 0.85x_{1i} + 0.85x_{1i} + 0.85x_{1i} + 0.85x_{1i} + 0.85x_{1i} + 0.85x_{1i} + 0.85x_{1i} + \varepsilon_i \qquad [i=1,\ldots,100]$$

where $\varepsilon_i \sim N(0, \sigma^2)$

We generate Nine explanatory variables from a multivariate normal with mean 0, and cov-variance $(x_i , x_j) = 0.5^{|i-j|}$.
The RMSE, MMAD,MAE and standard deviations (SD) are inserted in table 3. It is clearly observed that via all the sample size under study. The our proposed method (B nonlocal R) generate smallest RMSE, MMAD,MAE and (SD) comparison to other method (LASSO, Adaptive Lasso, Bayesian Lasso and B Reg N). This means our proposed method is very accurately. It is a very good method for achieving variables selection and coefficient estimation. In general , our proposed method has a good performance via small sample size and also ,it has a good performance via large sample size. Finally, the method (B nonlocal R) is not different from its performance via all levels of sample size with sparse models.

Table 3: Show results (RMSE), (MMAD) and (MAE) via averaged over 50 replications

n	Methods	RMSE	MMAD	MAE
25	LASSO	0.365 (0.232)	0.334 (0.242)	0.482 (0.371)
	Adaptive Lasso	0.420 (0.293)	0.330 (0.234)	0.406 (0.385)
	Bayesian Lasso	0.336 (0.298)	0.373 (0.340)	0.339 (0.232)
	B Reg N	0.435 (0.334)	0.310 (0.227)	0.320 (0.270)
	B nonlocal R	0.234 (0.103)	0.278 (0.096)	0.238 (0.150)
50	LASSO	0.520 (0.411)	0.564 (0.483)	0.405 (0.383)
	Adaptive Lasso	0.433 (0.304)	0.426 (0.302)	0.452(0.320)
	Bayesian Lasso	0.407 (0.332)	0.323(0.353)	0.352(0. 234)
	B Reg N	0.343 (0.278)	0.354 (0.204)	0.476 (0.328)
	B nonlocal R	0.281 (0.105)	0.230 (0.104)	0.282 (0.087)
100	LASSO	0.456 (0.345)	0.461 (0.391)	0.435 (0.334)
	Adaptive Lasso	0.462 (0.392)	0.420 (0.307)	0.434 (0.303)
	Bayesian Lasso	0.469 (0.327)	0.372 (0.375)	0.520 (0.384)
	B Reg N	0.423 (0.332)	0.471 (0.321)	0.433 (0.404)
	B nonlocal R	0.255(0.158)	0.246 (0.152)	0.194 (0.032)
150	LASSO	0.364 (0.263)	0.432 (0.332)	0.543 (0.478)
	Adaptive Lasso	0.241 (0.132)	0.230 (0.082)	0.281 (0.105)
	Bayesian Lasso	0.337 (0.264)	0.276 (0.164)	0.370 (0.253)
	B Reg N	0.304 (0.250)	0.338 (0.264)	0.354 (0.263)
	B nonlocal R	0.222 (0.192)	2.127 (0.071)	0.234 (0.092)
250	LASSO	0.473 (0.332)	0.321 (0.232)	0.321 (0.232)
	Adaptive Lasso	0.336 (0.298)	0.373 (0.240)	0.339 (0.332)
	Bayesian Lasso	0.369 (0.227)	0.452 (0.375)	0.320 (0.211)
	B Reg N	0.455 (0.358)	0.446 (0.352)	0.407 (0.232)
	B nonlocal R	0.234 (0.103)	0.278 (0.196)	0.138 (0.050)

Note: In the parentheses are SDs of the MAD

We can see the regression parameters estimations via our proposed method is very closed from normal distribution through histogram graphs. Also, it convergence to stationary. This clearly from trace plot (when sample size n=250). This mean the MCMC sampler is easy and effective.

Figure 5 . Histogram of (B nonlocalR) parameter estimation

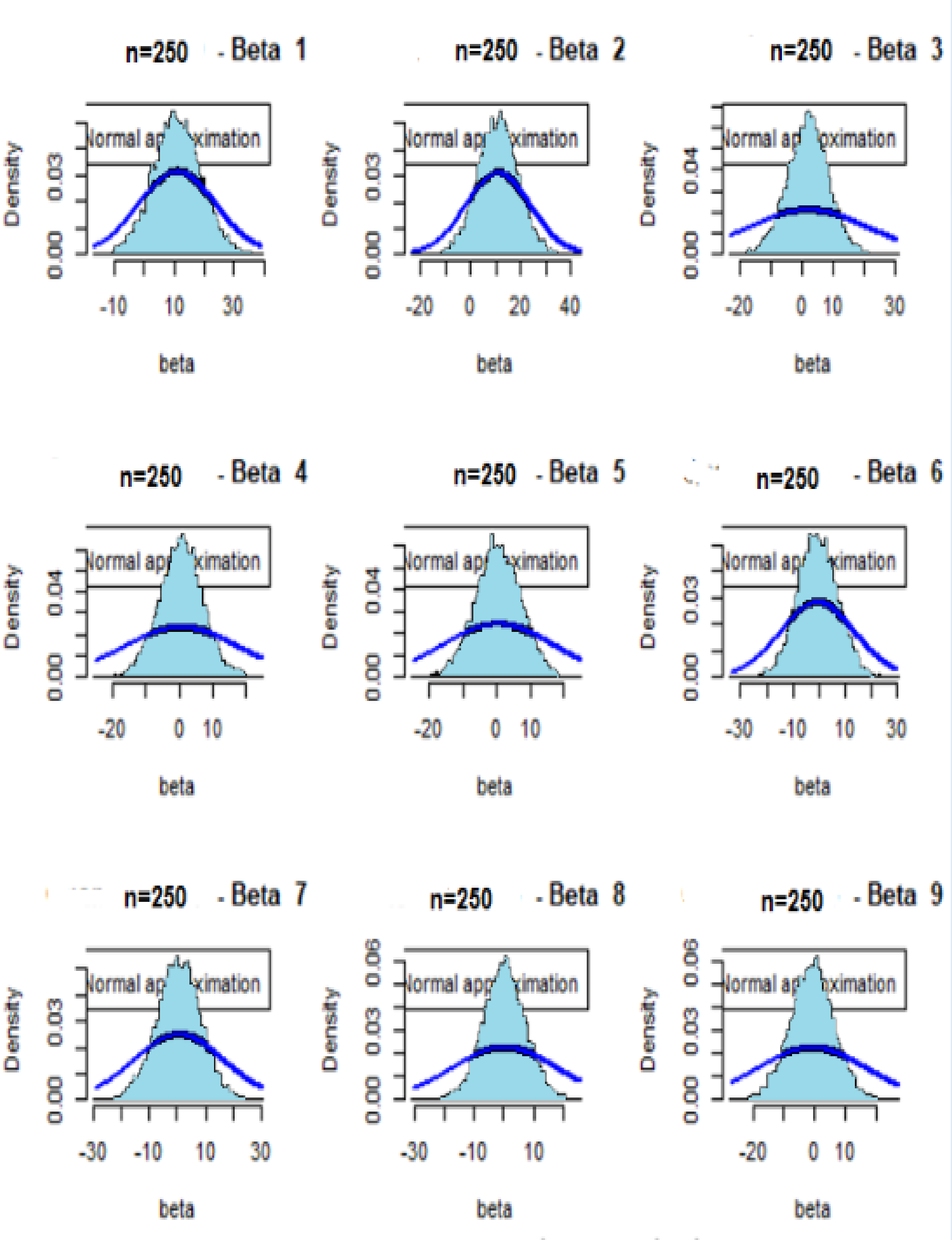
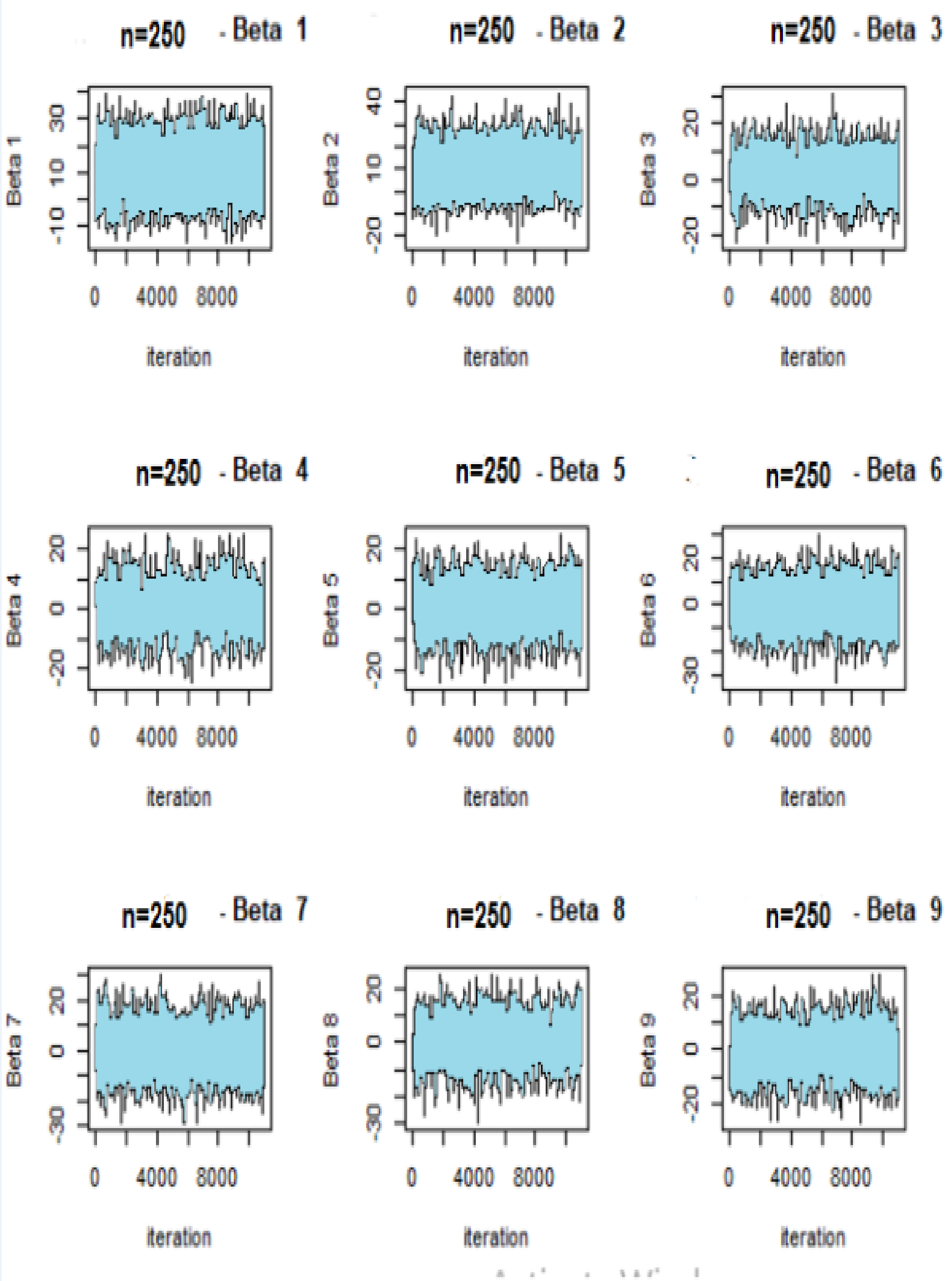


Figure 6. Trace plots of (B nonlocal R) with n=250



Example 4 (very sparse case)

In this example ,we will discuss another proposed method Bayesian nonlocal quantile regression model), denoted by (B nonlocal Q R) . B nonlocal Q R is compared with four other method in the same filed : classical quantile regression denoted by (rq) , Bayesian Lasso quantile

Regression denoted by (**BQReg N**), denoted by **MCMCquantreg** and Bayesian new lasso quantile regression denoted by (**BQReg U**) . In this example , we will use five quantile levels ($\tau = 0.15, \tau = 0.35, \tau = 0.55, \tau = 0.75$ *and* $\tau = 0.90$).Also, we will used three criterions are relative mean square error, denoted by (RMSE), Median of mean absolute deviations denoted by (MMAD). Mean absolute error denoted by (MAE) . In this example , the performance of our proposed method is evaluated via very sparse model

$$y = x_{2i} + \varepsilon_i \qquad [i=1,\dots,100] \ , \qquad 0 < \tau < 1$$

where $\varepsilon_i \sim N(0, \sigma^2)$

We generate Nine explanatory variables from a multivariate normal with mean 0, and cov-variance $(x_i \, , x_j) = 0.5^{|i-j|}$.

From the results in table 4,we can see our method outperformed the other methods ,because our method has smallest RMAE,MMAD,MAE and SD compared other methods. Our proposed method is a very good for achieving variables selection and coefficient estimation in quantile regression model via all quantile levels.

Table 4: Show results (RMSE), (MMAD) and (MAE) via averaged over 50 replications

Comparison Methods		RMSE	MMAD	MAE
τ=0.15	Rq	0.576(0.433)	0.602(0.453)	0.547(0.492)
	MCMCquantreg	0.413(0.359)	0.579(0.483)	0.447(0.310)
	BQRegU	0.377(0.245)	0.492(0.323)	0.419(0.346)
	BQReg N	0.347(0.229)	0.342(0.261)	0.403(0.379)
	B nonlocalQR	0.252(0.119)	0.173(0.036)	0.119(0.023)
τ=0.35	Rq	0.454(0.371)	0.462(0.383)	0.364(0.278)
	MCMCquantreg	0.484(0.384)	0.478(0.367)	0.594(0.437)
	BQRegU	0.378(0.278)	0.369(0.284)	0.333(0.205)
	BQReg N	0.383(0.465)	0.451(0.312)	0.469(0.315)
	B nonlocalQR	0.229(0.122)	0.250(0.124)	0.287(0.141)
τ=0.55	Rq	0.424(0.354)	0.407(0.395)	0.423(0.290)
	MCMCquantreg	0.458(0.383)	0.374(0.206)	0.338(0.207)
	BQRegU	0.363(0.293)	0.483(0.302)	0.411(0.396)
	BQReg N	0.482(0.389)	0.472(0.385)	0.414(0.352)
	B nonlocalQR	0.232(0.111)	0.242(0.144)	0.236(0.141)
τ=0.75	Rq	0.562(0.483)	0.567(0.482)	0.528(0.392)
	MCMCquantreg	0.482(0.353)	0.452(0.352)	0.434(0.355)
	BQRegU	0.472(0.347)	0.482(0.384)	0.492(0.382)
	BQReg N	0.445(0.346)	0.345(0.145)	0.487(0.491)
	B nonlocalQR	0.255(0.143)	0.248(0.147)	0.334(0.173)
τ=0.90	Rq	0.456(0.356)	0.446(0.395)	0.473(0.272)
	MCMCquantreg	0.484(0.359)	0.473(0.384)	0.583(0.396)
	BQRegU	0.505(0.353)	0.484(0.445)	0.445(0.359)
	BQReg N	0.584(0.454)	0.479(0.303)	0.445(0.345)
	B nonlocalQR	0.363(0.135)	0.383(0.273)	0.234(0.106)

Example 5 (sparse case)

In this example ,we will discuss another proposed method Bayesian nonlocal binary quantile regression model), denoted by (B nonlocal BQ R) . B nonlocal BQ R is compared with two other method in the same filed : binary regression quantiles denoted by (BRQ), Bayesian lasso binary quantile regression denoted by (BBRQL) . In this example , we will use five quantile levels ($\tau = 0.15, \tau = 0.35, \tau = 0.55, \tau = 0.75$ *and* $\tau = 0.90$).Also, we will use three criterions are relative mean square error, denoted by (RMSE), Median of mean absolute deviations denoted by (MMAD). Mean absolute error denoted by (MAE)**.**

In this example, the data are generated from the following model

$$y_i = \max\{c = 0, y_i{}^*\}, \quad i = 1,2,\dots,100,$$

$$y_i{}^* = \boldsymbol{x_i}'\boldsymbol{\beta_\tau} + \varepsilon_i,\ \mathbf{0} < \boldsymbol{\tau} < 1 \quad , \varepsilon_i \sim N(0,\sigma^2).$$

We consider a very sparse model , We set the true regression coefficients, including the intercept term, $\boldsymbol{\beta}$ = $(2,3,0,0,0,0,0,0,0)^t$. We generate Nine explanatory variables from a multivariate normal with mean 0, and cov-variance $(x_i\,, x_j) = 0.5^{|i-j|}$.

The (RMSE) ,(MMAD) ,(MAE) (SD) are inserted in Table 5. We can clearly noticed via all the quantile levels, the proposed method (B nonlocal BQ R) has a good performance compared to other two methods (BRQ)(BBRQL).

In general, the results in Table (5) show that the (RMSE) ,(MMAD) ,(MAE) (SD) for the our proposed method (B nonlocal BQ R) are smallest than that for existing other methods.

Table 5: Show results (RMSE), (MMAD) and (MAE) via averaged over 50 replications

Comparison Methods		RMSE	MMAD	MAE
τ=0.15	BRQ	0.532 (0.334)	0.475 (0.383)	0.473 (0.374)
	BBRQL	0.526 (0.435)	0.517 (0.420)	0.427 (0.363)
	B nonlocalBQR	0.352 (0.293)	0.246 (0.135)	0.237 (0.182)
τ=0.35	BRQ	0.431 (0.371)	0.434 (0.273)	0.263 (0.223)
	BBRQL	0.345 (0.233)	0.371 (0.231)	0.447(0.333)
	B nonlocalBQR	0.234 (0.118)	0.264 (0.137)	0.227 (0.094)
τ=0.55	BRQ	0.528 (0.394)	0.506 (0.427)	0.483 (0.384
	BBRQL	0.496 (0.346)	0.537 (0.386)	0.383 (0.218)
	B nonlocalBQR	0.337 (0.248)	0.293 (0.163)	0.237 (0.173)
τ=0.75	BRQ	0.531 (0.428)	0.592 (0.438)	0.571 (0.442)
	BBRQL	0.530 (0.471)	0.581(0.436)	0.427(0.337)
	B nonlocalBQR	0.382 (0.234)	0.333(0.219)	0.241(0.182)
τ=0.90	BRQ	0.524 (0.461)	0.430 (0.320)	0.418 (0.371)
	BBRQL	0.487 (0.316)	0.464 (0.328)	0.438 (0.373)
	B nonlocalBQR	0.318 (0.186)	0.252 (0.173)	0.222 (0.162)

Note: In the parentheses are SDs of the MAD

5.2 Real data

To study the behavior of our method in another approach , real data will be used, to clarification proposed methods for (**B nonlocal R**, **B nonlocal QR**, and **B nonlocal BQR**) , we will use two dataset of real data , firstly, This dataset(Air Pollution Data) was measured by the public roads administration in norway . The sample size of this dataset is 500 observations, and seven explanatory variables and one response variable. This dataset used to evaluation our two methods(**B nonlocal lR**)(**B nonlocal QR**) . Secondly, we will use real dataset collected from Basrah hospital , which this dataset contains one response variable and explanatory variables . This dataset used to evaluation our two methods(**B nonlocal BQR**) as following:

5.2.1 Air Pollution Data

5.2.1.1 Air Pollution Data with Bayesian nonlocal regression

This dataset (Air Pollution Data) is existing in R package , this dataset contains one response variable, log (concentration of NO2per hour), and 7 explanatory variables are temperature (x_1), temperature difference (x_2) , time of day in hours (x_3), day number (x_4), the log number of cars per hour (x_5), wind speed in meters per second (x_6) and wind direction (x_7).

As like section, simulation study.in this section , we will compare four methods (Lasso, Adaptive Lasso, Bayesian Lasso ,B Reg N) with our proposed method (B nonlocal R), the methods under this study are evaluated through three criterions (MSE,MAE and SD).

From the results are listed in Table 6, the MSE,MAE and SD of our proposed method are 0.426,0.324 and 0.283,213 respectively. Where, the MSE,MAE and SD computed by our proposed method is much smaller than MSE,MAE and SD computed by others methods (**Lasso**, **Adaptive Lasso, Bayesian Lasso ,B Reg N**). So, our proposed method has performance better than (**Lasso**, **Adaptive Lasso, Bayesian Lasso ,B Reg N**).

Table 6: MSEs, MAE and (SD) for dataset of the air pollution data.

Methods	MSE	MAE
Lasso	0.746(0.579)	0.635(0.529)
Adaptive Lasso	0.868(0.669)	0.626(0.459)
Bayesian Lasso	0.734(0.559)	0.617(0494)
B Reg N	0.629(0.536)	0.443(394)
B nonlocal R	0.426(0.283)	0.324(213)

5.2.1.2 Air Pollution Data with Bayesian nonlocal quantile regression

As like section, simulation study.in this section , we will compared four methods (rq,BQReg N, MCMCquantreg,BQReg U) with our proposed method (B nonlocal QR) via all quantile levels $\tau \in (0.15,0.35,0.55,0.75$ and 0.90 the methods under this study are evaluated through four criterions (MSE, MMAD,MAE and SD).

From the results are listed in Table 7, the MSE,MMAD,MAE and SD computed by our proposed method is much smaller than MSE,MMAD,MAE and SD computed by others methods (rq,BQReg N, MCMCquantreg,BQReg U). This means, our proposed method has performance better than (rq,BQReg N, MCMCquantreg,BQReg U).

Table 7: MSEs, MMAD, MAE and (SD) for dataset of the air pollution data

Comparison Methods		MSE	MMAD	MAE
τ=0.15	Rq	0.527(0.442)	0.702(0.554)	0.552(0.432)
	MCMCquantreg	0.514(0.453)	0.523(0.374)	0.452(0.310)
	BQRegU	0.422(0.325)	0.532(0.424)	0.513(0.437)
	BQReg N	0.452(0.313)	0.452(0.242)	0.504(0.423)
	B nonlocalQR	0.221(0.143)	0.223(0.147)	0.313(0.125)
τ=0.45	Rq	0.555(0.421)	0.572(0.474)	0.475(0.227)
	MCMCquantreg	0.575(0.455)	0.527(0.472)	0.535(0.412)
	BQRegU	0.427(0.327)	0.473(0.275)	0.444(0.205)
	BQReg N	0.474(0.575)	0.551(0.412)	0.573(0.415)
	B nonlocalQR	0.223(0.122)	0.250(0.125)	0.272(0.151)
τ=0.55	Rq	0.525(0.455)	0.502(0.435)	0.524(0.230)
	MCMCquantreg	0.557(0.474)	0.425(0.207)	0.447(0.202)
	BQRegU	0.474(0.234)	0.574(0.402)	0.511(0.437)
	BQReg N	0.572(0.473)	0.522(0.475)	0.515(0.452)
	B nonlocalQR	0.242(0.111)	0.252(0.155)	0.247(0.151)
τ=0.75	Rq	0.572(0.574)	0.572(0.572)	0.527(0.432)
	MCMCquantreg	0.572(0.454)	0.552(0.452)	0.545(0.455)
	BQRegU	0.522(0.452)	0.572(0.475)	0.532(0.472)
	BQReg N	0.555(0.457)	0.455(0.155)	0.572(0.531)
	B nonlocalQR	0.255(0.154)	0.257(0.152)	0.445(0.124)
τ=0.90	Rq	0.557(0.457)	0.557(0.435)	0.524(0.222)
	MCMCquantreg	0.575(0.453)	0.524(0.475)	0.574(0.437)
	BQRegU	0.505(0.454)	0.575(0.555)	0.555(0.453)
	BQReg N	0.575(0.555)	0.523(0.404)	0.555(0.455)
	B nonlocalQR	0.474(0.145)	0.474(0.224)	0.245(0.107)

5.2.2 Children Cancer Diseases

This dataset collected from Children's Specialist Hospital in Bsarah city. This dataset contain one response variable (take chemotherapy dose or no) and 12 explanatory variables are age (x_1), Gender (x_2), The number of sisters and brothers (x_3), Weight (x_4), Height (x_5), Body mass index (BMI) (x_6), Liver disease (x_7), Kidney disease(x_8), Family History (x_9), disease diagnosis(x_{10}), Father's age(x_{11}), mother's age(x_{12}).

As like section, simulation study.in this section , we will compared two

methods (BRQ, BBRQL) with our proposed method (B nonlocalBQR), the methods under this study are evaluated through three criterions (MSE,MAE and SD).

From the results are listed in Table 8, the MSE,MMAD,MAE and SD generated by our proposed method is much smaller than MSE,MMAD,MAE and SD generated by others methods (BRQ, BBRQL). This means, our proposed method has performance better than (BRQ, BBRQL).

Table 8: MSEs, MMAD, MAE (SD)for dataset of children cancer disease

Comparison Methods		RMSE	MMAD	MAE
τ=0.15	BRQ	0.412(0.334)	0.424(0.343)	0.483(0.374)
	BBRQL	0.435(0.364)	0.426(0.340)	0.466(0.333)
	B nonlocalBQR	0.242(0.173)	0.245(0.184)	0.236(0.192)
τ=0.35	BRQ	0.481(0.341)	0.474(0.243)	0.243(0.233)
	BBRQL	0.364(0.293)	0.311(0.221)	0.486(0.363)
	B nonlocalBQR	0.283(0.126)	0.248(0.182)	0.282(0.154)
τ=0.55	BRQ	0.453(0.353)	0.441(0.326)	0.543(0.363)
	BBRQL	0.439(0.314)	0.488(0.325)	0.352(0.204)
	B nonlocalBQR	0.286(0.144)	0.262(0.139)	0.229(0.159)
τ=0.56	BRQ	0.431(0.424)	0.452(0.434)	0.461(0.442)
	BBRQL	0.430(0.461)	0.44(0.435)	0.426(0.336)
	B nonlocalBQR	0.242(0.234)	0.233(0.215)	0.241(0.142)
τ=0.65	BRQ	0.424(0.451)	0.430(0.320)	0.414(0.361)
	BBRQL	0.446(0.315)	0.454(0.324)	0.434(0.363)
	B nonlocalBQR	0.214(0.145)	0.242(0.163)	0.222(0.152)

Chapter Six

Conclusions and Recommendation

6.1 Conclusions

In this thesis, we will be propose three methods in regression models by using Bayesian approach . The first method (**B nonlocal R**) is to focus by adding a new contributions to achieving variables selection and coefficients estimation in traditional regression model. **B nonlocal R** has a good performance compared with other methods, this clears through results of simulation and real data study.

The second method (**B nonlocal QR**) is to focus by adding new contribution to achieving variables selection and coefficients estimation in quantile regression model. **B nonlocal QR** has a good performance compared with other methods via all quantile levels , this clear through results of simulation and real data study.

The three method (**B nonlocal BQR**) is to focus by adding new contribution to achieving variables selection and coefficients estimation in binary quantile regression model. **B nonlocal BQR** has a good performance compared with other methods via all quantile levels , this clears through results of simulation and real data study.

6.2 Recommendation

We recommend the use of suggested extinction regularization hierarchical model under the nonlocal prior distribution to Bayesian Tobit regression model, Bayesian Tobit quantile regression model and Bayesian principal component regression model. Also, we recommended by employing our proposed methods in analyzing the real dataset in filed medical, economic, and others fields. Because our proposed method is efficient in coefficients estimation and variables selection.

Bibliography

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716-723.

Alhamzawi, R., Yu, K., & Benoit, D. F. (2012). Bayesian adaptive Lasso quantile regression. *Statistical Modelling*, *12*(3), 279-297.

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

Alhamzawi, R. (2015). Model selection in quantile regression models. *Journal of Applied Statistics*, *42*(2), 445-458.

Alhamzawi, R., & Mallick, H. (2020). Bayesian reciprocal LASSO quantile regression. *Communications in Statistics-Simulation and Computation*, 1-16.

Algamal, Z. Y., Alhamzawi, R., & Ali, H. T. M. (2018). Gene selection for microarray gene expression classification using Bayesian Lasso quantile regression. *Computers in biology and medicine*, *97*, 145-152.

Bottai, M., Cai, B., & McKeown, R. E. (2010). Logistic quantile regression for bounded outcomes. *Statistics in medicine*, *29*(2), 309-317.

Belloni, A., & Chernozhukov, V. (2011). ℓ1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, *39*(1), 82-130.

Benoit, D. F., & Van den Poel, D. (2012). Binary quantile regression: a Bayesian approach based on the asymmetric Laplace distribution. *Journal of Applied Econometrics*, *27*(7), 1174-1188.

Benoit, D. F., Alhamzawi, R., & Yu, K. (2013). Bayesian lasso binary quantile regression. *Computational Statistics*, *28*(6), 2861-2873.

Bradic, J., Fan, J., & Wang, W. (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 325-349.

Cao, X., Khare, K., & Ghosh, M. (2020). High-dimensional posterior consistency for hierarchical non-local priors in regression. *Bayesian Analysis*, *15*(1), 241-262.

Chen, X., Wang, Z. J., & McKeown, M. J. (2011). A Bayesian Lasso via reversible-jump MCMC. *Signal Processing*, *91*(8), 1920-1932.

Chen, C. W., Dunson, D. B., Reed, C., & Yu, K. (2013). Bayesian variable selection in quantile regression. *Statistics and its Interface*, *6*(2), 261-274.

Dwinata, A., Notodiputro, K. A., & Sartono, B. (2021, March). Penalized Logistic Regression Model to Predict a Results of RT-PCR by Using Blood Laboratory Test. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1115, No. 1, p. 012087). IOP Publishing.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348-1360.

Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(5), 849-911.

Frank, L. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, *35*(2), 109-135.

Hans, C., Dobra, A., & West, M. (2007). Shotgun stochastic search for “large p” regression. *Journal of the American Statistical Association*, *102*(478), 507-516.

Hans, C. (2009). Bayesian lasso regression. *Biometrika*, *96*(4), 835-845.

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.

Huang, J., Ma, S., Xie, H., & Zhang, C. H. (2009). A group bridge approach for variable selection. *Biometrika*, *96*(2), 339-355.

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

Javed, F., & Mantalos, P. (2013). GARCH-type models and performance of information criteria. *Communications in Statistics-Simulation and Computation*, *42*(8), 1917-1933.

James, G. M., Paulson, C., & Rusmevichientong, P. (2013). Penalized and constrained regression. *Unpublished manuscript, http://www-bcf. usc. edu/gareth/research/Research. html*.

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.

Johnson, V. E., & Rossell, D. (2008). Non-local prior densities for default Bayesian hypothesis tests.

Johnson, V. E., & Rossell, D. (2010). On the use of non- local prior densities in Bayesian hypothesis tests. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(2), 143-170.

Johnson, V. E., & Rossell, D. (2012). Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association*, *107*(498), 649-660.

Kim, Y., Kim, J., & Kim, Y. (2006). Blockwise sparse regression. *Statistica Sinica*, 375-390.

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

Kordas, G. (2002). Credit scoring using binary quantile regression. In *Statistical data analysis based on the l1-norm and related methods* (pp. 125-137). Birkhäuser, Basel.

Kozumi, H., & Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of statistical computation and simulation*, *81*(11), 1565-1578.

Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics*, *157*(2), 396-408.

Li, M. Y. L., & Miu, P. (2010). A hybrid bankruptcy prediction model with dynamic loadings on accounting-ratio-based and market-based information: A binary quantile regression approach. *Journal of Empirical Finance*, *17*(4), 818-833.

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

Mallick, H., Alhamzawi, R., Paul, E., & Svetnik, V. (2021). The reciprocal Bayesian lasso. *Statistics in medicine*, *40*(22), 4830-4849.

Mallows, C. L. (1973). Bounds on distribution functions in terms of expectations of order-statistics. *The Annals of Probability*, 297-303.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of econometrics*, *3*(3), 205-228.

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

Mallick, H., & Yi, N. (2017). Bayesian group bridge for bi-level variable selection. *Computational statistics & data analysis*, *110*, 115-133.

Mallick, H., & Yi, N. (2018). Bayesian bridge regression. *Journal of applied statistics*, *45*(6), 988-1008.

Nikooienejad, A., Wang, W., & Johnson, V. E. (2016). Bayesian variable selection for binary outcomes in high-dimensional genomic studies using non-local priors. *Bioinformatics*, *32*(9), 1338-1345.

Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *The Annals of Statistics*, 758-765.

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

Rahman, M. A., & Vossmeyer, A. (2019). Estimation and applications of quantile regression for binary longitudinal data. In *Topics in Identification, Limited Dependent Variables, Partial Observability,* *Experimentation, and Flexible Modeling: Part B*. Emerald Publishing Limited.

Rossell, D., Telesca, D., & Johnson, V. E. (2013). High-dimensional Bayesian classifiers using non-local priors. In *Statistical Models for Data Analysis* (pp. 305-313). Springer, Heidelberg.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.

Shin, M., Bhattacharya, A., & Johnson, V. E. (2018). Scalable Bayesian variable selection using nonlocal prior densities in ultrahigh-dimensional settings. *Statistica Sinica*, *28*(2), 1053.

Scott, J. G., & Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, 2587-2619.

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2013). A sparse-group lasso. *Journal of computational and graphical statistics*, *22*(2), 231-245.

Song, Q., & Liang, F. (2015). High-dimensional variable selection with reciprocal l 1-regularization. *Journal of the American Statistical Association*, *110*(512), 1607-1620.

Sun, W., Ibrahim, J. G., & Zou, F. (2010). Genomewide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression. *Genetics*, *185*(1), 349-359.

Tang, Y., Wang, H. J., & Zhu, Z. (2013). Variable selection in quantile varying coefficient models with longitudinal data. *Computational Statistics & Data Analysis*, *57*(1), 435-449.

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

Wang, H., & Leng, C. (2008). A note on adaptive group lasso. *Computational statistics & data analysis*, *52*(12), 5277-5286.

Wei, Y., Kehm, R. D., Goldberg, M., & Terry, M. B. (2019). Applications for quantile regression in epidemiology. *Current Epidemiology Reports*, *6*(2), 191-199.

Yu, K., & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, *54*(4), 437-447.

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

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301-320.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418-1429.

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

المستخلص

في السنوات الاخيرة , حظيت مشكلة تحديد المتغيرات المستقلة الفعالة في نموذج الانحدار الخطي اهتمام كبير . وايضاً في الآونة الاخيرة , نجد ان طرق اختيار النموذج البايزي التي تستخدم التوزيعات الغير محلية المسبقة حظيت باهتمام كبير .واحدى هذه الطرق التي تهتم بتقدم للبحث والفرز العشوائي مبسط في الانحدار الخطي اقترحت من (Shin et al., 2018).
في الحقيقة (Shin et al., 2018) , اقترحوا خوارزميتين للبحث والفرز العشوائي. الاولى تعتمد على تقديم العزم المعكوس (piMoM) لدالة كثافة التوزيع المسبقة, والثانية تعتمد على تقديم العزم التجريبي (piMoM) لدالة كثافة التوزيع المسبقة. في هذه الدراسة , سوف نقترح طريقة بيزية جديدة تعتمد على البحث والفرز العشوائي المبسط في نموذج الانحدار . طريقتنا تعتمد على توظيف التوزيعات السابقة لمعكوس لابلاس لمعالم الانحدار ثم نوسع هذه الفكرة الى الانحدار القسيمي الخطي والانحدار الثنائي القسيمي . وسوف نقارن طريقة البحث والفرز العشوائي المبسط مع طرق اخرى من خلال دراسة المحاكاة وتحليل البيانات الحقيقية , طرقنا المقترحة تشير الى انها تمتلك أداءً جيدًا مقارنة بالطرق الأخرى في اختيار التنبؤات الفعالة وتقدير المعلمات.

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة القادسية / كلية الإدارة والاقتصاد
قسم الإحصاء / الدراسات العليا

اختيار نموذج في الانحدار الثنائي باستخدام التوزيعات السابقة غير المحلية مع تطبيق عملي

رسالة مقدمة الى
مجلس كلية الإدارة والاقتصاد - جامعة القادسية
جزءاً من متطلبات نيل درجة الماجستير
في علوم الإحصاء

من الطالب

ميثم سعيد كاظم الخميس

إشراف

أ.د. رحيم جبار الحمزاوي

١٤٤٣ هـ ٢٠٢١ م