Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.1 Year 2019
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

Bayesian adaptive Lasso Tobit regression

Haider Kadhim Abbas **Rahim Jabbar Thaher**

Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq.
haiderhilali@gmail.com rahim.alhamzawi@qu.edu.iq

Recived : 27\6\2018 **Revised : 17\7\2018** **Accepted : 22\10\2018**

Available online : 25 /1/2019

DOI: 10.29304/jqcm.2019.11.1.471

Abstract :
In this paper, we introduce a new procedure for model selection in Tobit regression, we suggest the Bayesian adaptive Lasso Tobit regression (BALTR) for variable selection (VS) and coefficient estimation. We submitted a Bayesian hierarchical model and Gibbs sampler (GS) for our procedure. Our proposed procedure is clarified by means of simulations and a real data analysis. Results demonstrate our procedure performs well in comparison to further procedures.

Keywords: Tobit regression; Bayesian adaptive Lasso Tobit regression (BALTR); Variable selection (VS).

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.1 Year 2019
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

1.Introduction:

Tobit regression procedure (Tr) is proposed as a statistical model by Tobin (1958). This model is also known as left truncated regression. Tr has become important in many real-world applied sciences, such as econometric, agriculture, ecology, the environment and genetics. It is an excellent procedure to evaluate the relation along with outcome variable and a group of explanatory variables.

One of the most mainly important troubles in the regression when the number of explanatory variables is so large. It is then difficult to see which variables actually important. In addition to several problems appear when the statistical researchers are use some explanatory variables that are not important in regression. This leads to a regression model that will be unstable and so weak concerning of prediction. The selection process provides a perfect agent for estimating the parameters as well as the identification of important variables (Griffin and Brown, 2010). There occur several varieties of strategies for investigators to use in handling high dimensional data (very large of explanatory variables), including VS procedures, and data reduction techniques. Prior analysis has found that, in the existence of high dimensional data, these VS procedures can produce estimates with inflated errors for the coefficients (Hastie, Tibshirani, & Friedman, 2009). Some of the technique models that have proved beneficial in the condition of high dimensional data, these models known as regularization.

In 1996, Tibishrani suggested a procedure for VS and parameter estimation in linear models known be as Lasso model (Least Absolute Shrinkage and Selection Operator model). A lot of work has been devoted to the development of diverse of Bayesian organizational procedures for making VS in linear models. In 2006, Zou proposed the adaptive Lasso, who upgraded the Lasso way proposed by Tibshirani, permitting different penalty parameters to different regression coefficients. Zou proved that his proposed procedure had the characteristics of Oracle mentioned in Fan and Bing (2004) that Lasso does not have. Specifically, Zou indicates that his proposed procedure adopts the correct form of non-zero coefficients with the probability that he tends to one. Park and Casella suggested in 2008 the Lasso procedure based from a Bayesian point of sight. Likewise, Mallick and Yi (2014) suggested a new procedure known to be as new Bayesian Lasso regression for VS and coefficient estimation in linear regression.

In general, the last procedure observed results display that the Mallick procedure applied well compares with other Bayesian and non-Bayesian regression procedures.

The above results and good results reported in Mallick procedure motivate us to suggest a new Bayesian regression procedure. Subsequently, we submitted a Bayesian hierarchical for BALTR, and proposed a new Gibbs sampler (GS) for BALTR, that is set up on a theoretical derivation of the Laplace density (LD). Next, we implemented several simulated examples and analyzed real data by using BALTR with four Tobit regression procedures to compare the best results. These procedures include Tr, Bayesian Tobit regression (BTr), Tobit median regression, and BALTR. Both simulation and real analysis proved that BALTR results are excellent, and this procedure may be is a best of current procedures being compared.

2.Methods:

The Tobit regression is applied to estimate the relevance among an outcome variable (y_i) and explanatory variables ($\boldsymbol{X}$). Tobit regression assumes that there is a latent variable (y_i^*) depends linearly on the parameters ($\boldsymbol{\beta}$) which determines relevance between ($\boldsymbol{X}$) and (y_i^*), the formula of outcome variable is

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

$$\boldsymbol{y}^* = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad \cdots (1)$$

$$\boldsymbol{y}^* = (y_1, \cdots, y_n),$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix},$$

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_k)\ ,$$

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n),$$

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

2.1 Bayesian adaptive Lasso Tobit regression (BALTR):

It is well known, that the Lasso procedure gives biased estimates of considerable coefficients, so it might be below the required optimal level in terms of estimation risk. In 2006, Zou evidenced that the Lasso opts the incorrect model with non-fade the probability, despite the sample size and how λ is chosen. The event requires that coefficients not in the model aren't representable by coefficients in the real model. But this event is simply suffering because of the collinearity case between the coefficients. On the opposite hand, that the Lasso technique does not have Oracle properties. So, Zou suggested the adaptive Lasso technique who gives a consistent model for VS. Therefore, we consider BALTR procedure in this paper, the adaptive Lasso enjoys the oracle properties by utilizing the adaptably weighted Lasso penalty parameter, and leads to a near minimax optimum estimator. Additionally, the adaptive Lasso technique needs to initial estimates of the regression coefficients, when a sample size is less than of the covariates number, which is mostly not available in the high dimensional data. The estimator of adaptive Lasso is given by

$$\widehat{\boldsymbol{\beta}}_{alasso} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} (\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta}) + \sum_{j=1}^{k} \lambda_j \left|\beta_j\right| \quad where\ \lambda_j \geq 0$$

where varied penalty parameters are utilized for the regression coefficients. Surely, for the not important explanatory variables, we must place larger penalty λ_j on their matching coefficients.

We propose a BALTR procedure in this paper for coefficient estimation and VS. We submit a new practice of the adaptive Lasso form by using the scale mixture of a uniform represent of the LD. Following (Mallick& Yi, 2014), the Laplace representation can adaptive as

$$\frac{\lambda_j}{2} e^{-\lambda_j|\beta_j|} = \int_{s_j>|\beta_j|} \frac{1}{2s_j} \frac{\lambda_j{}^2}{\Gamma 2} s_j{}^{2-1} e^{-\lambda_j s_j}\, ds_j \cdots (2)$$

$$\frac{\lambda_j}{2} e^{-\lambda_j|\beta_j|} = \int_{s_j>|\beta_j|} \frac{\lambda_j{}^2}{2} e^{-\lambda_j s_j}\, ds_j,\ \lambda_j > 0$$

In this paper, we modify the above formula as follows:

Let $v_j = \lambda_j s_j \quad \Rightarrow \quad dv_j = \lambda_j ds_j$ then

$$\frac{\lambda_j}{2} e^{-\lambda_j|\beta_j|} = \frac{\lambda_j}{2} e^{-|\lambda_j\beta_j|}$$

$$= \int_{v_j>|\lambda_j\beta_j|} \frac{\lambda_j}{2v_j} \frac{\lambda_j^2}{\Gamma 2} \left(\frac{v_j}{\lambda_j}\right)^{2-1} e^{-v_j} \frac{1}{\lambda_j} dv_j$$

$$= \int_{v_j>|\lambda_j\beta_j|} \frac{\lambda_j}{2} e^{-v_j} dv_j \cdots (3)$$

In practice, this formula produces more tractable and efficient Gibbs sampler than the formula in 2.

2.2 Model Hierarchy and Prior Distributions of BALTR:

By using equation (1) and equation (3), the Bayesian hierarchical model can be formulated as follows:

$$\boldsymbol{y}^*|\boldsymbol{X},\boldsymbol{\beta},\sigma^2 \sim N_n(\boldsymbol{X\beta},\sigma^2 I_n) \ldots (4)$$

$$\boldsymbol{\beta}|\lambda \sim \prod_{j=1}^{k} \text{Uniform}(-\frac{1}{\lambda_j},\frac{1}{\lambda_j}) \ldots (5)$$

$$\boldsymbol{v} \sim \prod_{j=1}^{k} \text{Exp}(1) \ldots (6)$$

$$\sigma^2 \sim \text{Inverse Gamma}(a,b) \ldots (7)$$

$$\lambda_j \sim \text{Gamma}(f,g) \ldots (8)$$

where $\boldsymbol{v} = (v_1,\cdots,v_k)$

2.3 Full Conditional Posterior Distributions of BALTR:

Firstly, we can express the joint posterior distribution of all our procedure parameters as follows

$$\pi(\boldsymbol{\beta},\boldsymbol{v},\lambda,\sigma 2|\boldsymbol{y}^*,\boldsymbol{X}) \propto \pi(\boldsymbol{y}^*|\boldsymbol{X},\boldsymbol{\beta},\sigma 2)\pi(\boldsymbol{\beta}|\lambda)\pi(\boldsymbol{v})\pi(\lambda_j)\pi(\sigma 2)$$

Under the above posterior distribution, the posterior distribution of $\boldsymbol{\beta}$ is

$$\pi(\boldsymbol{\beta}/\boldsymbol{y}^*,\ \boldsymbol{X},\ \lambda) \propto \pi(\boldsymbol{y}^*/\boldsymbol{X},\boldsymbol{\beta},\sigma^2).\pi(\boldsymbol{\beta}|\lambda)$$

$$\propto exp\left\{-\frac{1}{2\sigma 2}(\boldsymbol{y}^*-\boldsymbol{X\beta})'(\boldsymbol{y}^*-\boldsymbol{X\beta})\right\}\prod_{j=1}^{k} I\left\{\left|\beta_j\right|<\frac{v_j}{\lambda_j}\right\}$$

$$\propto exp\left\{-\frac{1}{2\sigma 2}\left(-2\boldsymbol{y}^{*'}\boldsymbol{XB}+\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}\right)\right\}\prod_{j=1}^{k} I\left\{\left|\beta_j\right|<\frac{v_j}{\lambda_j}\right\}$$

$$\propto exp\left\{-\frac{1}{2\sigma 2}\left(-2\boldsymbol{y}^{*'}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{\beta}+\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}\right)\right\}\prod_{j=1}^{k} I\left\{\left|\beta_j\right|<\frac{v_j}{\lambda_j}\right\}$$

$$\propto exp\left\{-\frac{1}{2\sigma 2}\left(-2\widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X\beta}+\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}\right)\right\}\prod_{j=1}^{k} I\left\{-\frac{v_j}{\lambda_j}<\beta_j<\frac{v_j}{\lambda_j}\right\}$$

$$\boldsymbol{\beta}|\boldsymbol{y}^{*},\boldsymbol{X},\lambda \sim N_k(\widehat{\boldsymbol{\beta}}_{OLS},(\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma 2)$$
$$\text{x}\prod_{j=1}^{k} I\left\{-\frac{v_j}{\lambda_j}<\beta_j<\frac{v_j}{\lambda_j}\right\}\cdots(9)$$

As well, the posterior distribution of v_j is

$$\pi(\boldsymbol{v}|\boldsymbol{y}^{*},\boldsymbol{X},\boldsymbol{\beta},\lambda)\ \propto\ \pi(\boldsymbol{v})I\left\{v_j>\left|\lambda_j\beta_j\right|\right\}$$
$$\propto \prod_{j=1}^{k} e^{-v_j} I\left\{v_j>\left|\lambda_j\beta_j\right|\right\}$$
$$\boldsymbol{v}\sim\prod_{j=1}^{k} Exponential(1)\, I\left\{v_j>\left|\lambda_j\beta_j\right|\right\}\cdots(10)$$

Likewise, the posterior distribution of σ^2 is

$$\pi(\sigma^2|\boldsymbol{y}^{*},\boldsymbol{X},\boldsymbol{\beta})\ \propto \pi(\boldsymbol{y}^{*}|\ \boldsymbol{X},\boldsymbol{\beta},\sigma^2)\ \pi(\sigma^2)$$
$$\propto\ (\sigma^2)^{-\frac{n}{2}}exp\left\{-\frac{1}{2\sigma^2}(y^{*}-X\beta)'(y^{*}-X\beta)\right\}(\sigma^2)^{-a-1}exp\left\{-\frac{b}{\sigma^2}\right\}$$
$$\sigma^2|\boldsymbol{y}^{*},\boldsymbol{X},\boldsymbol{\beta}\sim$$
$$\text{InvGamma}\left(\frac{n}{2}+a,\frac{1}{2}(y^{*}-X\beta)'(y^{*}-X\beta)+b\right)\cdots(11)$$

Lastly, the posterior distribution of λ is

$$\pi\left(\lambda_j\middle|\beta_j\right)\ \propto\ \pi\left(\beta_j\middle|\lambda_j\right).\pi(\lambda_j)$$
$$\pi\left(\lambda_j\middle|\beta_j,v_j\right)\ \propto\ \pi(\lambda_j)\lambda_j I\left\{\lambda_j<\frac{v_j}{\left|\beta_j\right|}\right\}$$
$$\propto\ \lambda_j^{(f+1)-1}\,exp\{-g\lambda_j\}\,I\left\{\lambda_j<\frac{v_j}{\left|\beta_j\right|}\right\}$$
$$\propto\ \text{Gamma}(f+1,g)I\left\{\lambda_j<\frac{v_j}{\left|\beta_j\right|}\right\}\dots(12)$$

Where the $I(.)$ is an indicator function in equation (9) and equation (12).

2.4 Computation:

In the computation section, we outline our Gibbs sampler as follows

- Updating $\boldsymbol{\beta}$:

We simulate the β_j from a truncated multivariate normal distribution in equation (9), the mean of this distribution is $(\widehat{\boldsymbol{\beta}}_{OLS})$ and the variance is $((\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma 2)$.

- Updating $\boldsymbol{v}$:

We simulate the v_j from the left truncated exponential distribution in equation (10), by applying the inversion process, this simulate can be completed as follows:

1. Simulate $v_j{}^{*}$ from standard exponential distribution.
2. Set $v_j=v_j{}^{*}+\left|\lambda_j\beta_j\right|$

- Updating σ^2:

We simulate the σ^2 from Inverse Gamma distribution in equation (11), the shape parameter of this distribution is $(\frac{n}{2}+a)$ and the rate is

$$\left(\frac{1}{2}(y^{*}-X\beta)'(y^{*}-X\beta)+b\right)$$

- Updating λ :

We simulate λ_j from truncated Gamma distribution, the shape parameter of this distribution is $(f+1)$ and the rate parameter is (g).

3.Simulation Studies:

The performance of our procedure is evaluates in a simulation study in which the procedure for a BALTR is compared with, Tr procedure through using R language within package AER (Christian Kleiber, Achim Zeileis 2017), Bayesian Tobit regression procedure (BTr) through using R language within package MCMCpack (Jong Hee Park, 2018), and Bayesian Analysis of Quantile Regression Models (Bayesian Tobit quantile regression BTqr,and Bayesian adaptive Lasso Tobit quantile regression BALTqr); and Tau=0.5 by estimating the median through using R language within package Brq (Alhamzawi, R., & Alhamzawi, M. R., 2017) . For comparison, we draw 11,000 iterations of the GS, the first 1000 were ruled out as burn-in. The procedures are evaluated based on the median of mean absolute deviations (MMAD). The formula of MMAD is

$$\text{MMAD}=\text{median}(\text{mean}(|\boldsymbol{X}\widehat{\boldsymbol{\beta}}-\boldsymbol{X}\boldsymbol{\beta}^{true}|))$$

where $\widehat{\boldsymbol{\beta}}$ is the posterior mean of $\boldsymbol{\beta}$.

3.1 Independent and identically distributed random errors:

Here, simulation examples consider three cases (dense case, sparse case, and very sparse case), eight predictors $x_1.\cdots.x_8$ were simulated independently from a multivariate normal distribution with mean 0, and two values of the variance σ^2, the σ^2 is 1and 4.

3.1.1 Simulation example 1:

This example considers a dense case model, the true regression coefficients is

$$\boldsymbol{\beta}=(0,\underbrace{0.75,\dots,0.75}_{8})'$$

The response variable was generated according to the model

$$y_i^* = \beta_0 + 0.75x_{1i} + 0.75x_{2i} + 0.75x_{3i} + 0.75x_{4i} + 0.75x_{5i} + 0.75x_{6i} + 0.75x_{7i} + 0.75x_{8i} + \varepsilon_i$$

We simulate 100 observations and $\beta_0 = 0$, the pair wise correlations between x_i and x_j is $0.5^{|i-j|}$.

Method	σ^2	MMAD	SD
BALTR	1	0.36193	0.10830
Tr		0.36779	0.12457
BTr		0.39495	0.15897
BTqr		0.41936	0.19493
BALTqr		0.38301	0.13498
BALTR	4	0.56246	0.11160
Tr		0.57281	0.12713
BTr		0.63695	0.19218
BTqr		0.63797	0.17940
BALTqr		0.60005	0.11972

Table 1: MMAD and SD for the dense case example

3.1.2 Simulation example 2:

This example considers a sparse case model, the setup is the same in simulation 1, except the number of observations is 150, and the true regression coefficients is

$$\boldsymbol{\beta} = (0, 2, 1, 0, 0, 2, 0, 0, 0)'$$

The response variable was generated according to the model

$$y_i^* = \beta_0 + 2x_{1i} + x_{2i} + 2x_{5i} + \varepsilon_i$$

Method	σ^2	MMAD	SD
BALTR	1	0.25350	0.08704
Tr		0.27300	0.08761
BTr		0.28327	0.10003
BTqr		0.29655	0.11001
BALTqr		0.25856	0.09530
BALTR	4	0.50916	0.12261
Tr		0.53753	0.14044
BTr		0.57080	0.17250
BTqr		0.59050	0.18781
BALTqr		0.51686	0.12951

Table 2: MMAD and SD list for the simulation 2

3.1.3 Simulation example 3:

This example considers a very sparse case model with high correlation. We simulate 200 observations and the pair wise correlations between x_i and x_j equals to 0.75, and the true regression coefficients is

$$\boldsymbol{\beta} = (0, 4, 0, 0, 0, 0, 0, 0, 0)'$$

The response variable was generated according to the model

$$y_i^* = \beta_0 + 4x_{1i} + \varepsilon_i$$

and intercept coefficient is 0 .

The response variable was generated according to the model

$$y_i^* = \beta_0 + 4x_{1i} + \varepsilon_i$$

and intercept coefficient is 0 .

Method	σ^2	MMAD	SD
BALTR	1	0.21634	0.06547
Tr		0.23154	0.06120
BTr		0.23800	0.07200
BTqr		0.26720	0.07970
BALTqr		0.22145	0.07329
BALTR	4	0.42391	0.12948
Tr		0.45243	0.11599
BTr		0.48257	0.12186
BTqr		0.52948	0.12676
BALTqr		0.43534	0.11626

Table 3: MMAD and SD list for the simulation 3

3.2 Simulation example 4:

This example considers a Difficult case model. We simulate 100 observations, four predictors $x_1 . \cdots . x_4$ were simulated independently from a multivariate normal distribution with mean zero and variance σ^2. We consider three values of σ^2 (1, 4 and 9), and the pair wise correlations between x_i and x_j equal to (-0.4), the true regression coefficients is

$$\boldsymbol{\beta} = (0, 5.5, 5.5, 5.5, 0)'$$

The response variable was simulated according to the model

$$y_i^* = 0 + 0.55x_{1i} + 0.55x_{2i} + 0.55x_{3i} + 0.55x_{4i} + \varepsilon_i$$

Method	σ^2	MMAD	SD
BALTR	1	0.22283	0.13737
Tr		0.23864	0.13851
BTr		0.23633	0.14456
BTqr		0.30271	0.17220
BALTqr		0.27722	0.15230
BALTR	4	0.49759	0.32372
Tr		0.51872	0.31529
BTr		0.53169	0.35028
BTqr		0.59351	0.40493
BALTqr		0.52709	0.33300
BALTR	9	0.62072	0.42722
Tr		0.65780	0.39645
BTr		0.67667	0.47057
BTqr		0.77867	0.48305
BALTqr		0.68023	0.34077

Table 4: MMAD and SD list for the simulation 4

From above tables 1, 2, 3 and 4, we noted that the BALTR procedure performs better than the other procedures in terms the median of mean absolute deviations.

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.1 Year 2019
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

3.3 Simulation example 5 (Heterogeneous random errors):

In this section, errors are considered to demonstrate the performance of our proposed procedure for VS. We simulated 100 observations from the model

$$y_i^* = x_i{}'\boldsymbol{\beta} + (1 + x_{3i})\varepsilon_i,$$
$$\varepsilon_i \sim \mathrm{N}(0,1) \text{ and } \boldsymbol{\beta} = (0,1,1,1,1,,0,0,0,0,0)'$$
$$\text{where } x_{1i} \sim \mathrm{N}(0,1),$$
$$x_{3i} \sim \mathrm{Uniform}[0,1],$$
$$x_{2i} = x_{1i} + x_{3i} + z_i, z_i \sim \mathrm{N}(0,1)$$

this process is often used to simulate data in the VS context (example of Wu and Liu, 2009 and Li et al., 2010). In this simulation, added 5 independent standard normal noise variables, $x_4 \cdots x_8$, were simulated. In this paper, we set $y_i = \max\{y_i^*, 0\}$

Method	MMAD	SD
BALTR	0.26923	0.06925
Tr	0.27969	0.06596
BTr	0.27911	0.07437
BTqr	0.32919	0.07278
BALTqr	0.29920	0.06916

Table 8: MMAD and SD list for the simulation 5

Table (8) reports MMADs and SDs of simulation example 5. The performance of BALTR procedure is excellent compared to the other procedures (Tr, BTr, BTqr, BALTqr).

4.Real Data Analysis:

In data analysis section, we implement our proposed procedure on wheat production data, we apply the four Tobit regression procedures in this data to compare in terms of the coefficient's estimation accuracy. The real data used for this study is taken from the national program for the development of wheat cultivation in Iraq - Qadisiyah governorate branch (2017). This real data contains 584 observations and are based on 10 explanatory variables. The outcome of interest in this dataset is (Percentage increase of wheat yield per dunam "2500 m^2").

The other ten variables (covariates) include fertilize the field with Urea (numeric variable coding the quantity of fertilizer in kilogram; "U"), the date of sowing wheat seeds (numeric variable coding date: 1 the ideal date, 2 early date, 3 late date; "Ds"), the quantity of sowing wheat seeds (numeric variable coding the quantity of sowing seeds in kilogram; "Qs"), laser field leveling technique (numeric variable coding date: 2 if there are used this technique; 1 otherwise; "LT"), fertilize the field with compound fertilizers "NPK" (numeric variable coding the quantity of fertilizer in kilogram; "NPK"), seed sowing machine technique (numeric variable coding date: 2 if there are used this technique; 1otherwise; "SMT"), planting successive mung bean crops (numeric variable coding type: 2 planting mung bean, 1 otherwise; "SC"), used herbicide for weed control (numeric variable coding the quantity of herbicide in milliliter; "H"), high Potassium fertilizer "Potash" (numeric variable coding the quantity of fertilizer in kilogram; "K") and Micro-Element fertilizer (numeric variable coding the quantity of fertilizer in gram; "ME").

Method	MSE
BALTR	0.4617
Tr	0.4784
BTr	0.4795
BTqr	0.4724
BALTqr	0.4685

Table 9: wheat production data analysis: Mean squared prediction errors (MSE) based on a test set with 584 observations.

Table (9) reports the mean squared errors for five Tobit regression procedures. We can observe that mean squared errors of BALTR procedure is lower than that of Tr, BTr, BTqr and BALTqr, that means BALTR procedure produces the lowest prediction errors.
that means BALTR procedure produces the lowest prediction errors.

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.1 Year 2019
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

	β_0	U	Ds
	Estimate (25%, 95%)	Estimate (25%, 95%)	Estimate (25%, 95%)
BALTR	-0.039 (-0.402, 0.285)	0.021 (0.020, 0.023)	-0.672 (-0.749, -0.620)
Tr	-0.085 (-0.872, 0.702)	0.021 (0.014, 0.028)	-0.664 (-0.786, -0.541)
BTr	-0.082 (-0.899, 0.720)	0.021 (0.014, 0.028)	-0.666 (-0.791, -0.546)
BTqr	-1.228 (-1.815, -0.546)	0.024 (0.018, 0.031)	-0.654 (-0.806, -0.505)
BALTqr	-1.072 (-1.712, -0.263)	0.024 (0.017, 0.030)	-0.649 (-0.797, -0.498)
	Qs	LT	NPK
BALTR	-0.022 (-0.025, -0.020)	1.333 (1.012, 1.648)	0.005 (0.003, 0.007)
Tr	-0.022 (-0.035, 0.009)	1.357 (0.681, 2.034)	0.005 (-0.008, 0.017)
BTr	-0.022 (-0.035, -0.008)	1.358 (0.658, 2.035)	0.005 (-0.008, 0.018)
BTqr	-0.006 (-0.018, 0.004)	1.428 (0.459, 2.343)	-0.005 (-0.017, 0.007)
BALTqr	-0.008 (-0.022, 0.002)	1.441 (0.493, 2.181)	-0.004 (-0.016, 0.008)
	SMT	SC	H
BALTR	-0.090 (-0.409, 0.161)	0.925 (0.841, 1.003)	0.004 (0.004, 0.005)
Tr	-0.143 (-0.838, 0.553)	0.933 (0.611, 1.255)	0.004 (0.003, 0.006)
BTr	-0.148 (-0.840, 0.559)	0.931 (0.601, 1.259)	0.004 (0.003, 0.006)
BTqr	0.248 (-0.631, 1.204)	0.991 (0.651, 1.313)	0.005 (0.004, 0.007)
BALTqr	0.192 (-0.433, 1.132)	0.967 (0.622, 1.293)	0.005 (0.004, 0.007)
	K	ME	
BALTR	0.033 (0.032, 0.034)	0.006 (0.006, 0.006)	
Tr	0.033 (0.026, 0.040)	0.006 (0.005, 0.008)	
BTr	0.033 (0.026, 0.040)	0.006 (0.005, 0.008)	
BTqr	0.024 (0.014, 0.036)	0.008 (0.005, 0.010)	
BALTqr	0.025 (0.014, 0.036)	0.007 (0.005,0.0104)	

Table 10: Coefficients estimation and Credible intervals CIs (25%, 95%)

Although, our CIs in table (10) are narrower than the other methods, it is including all the estimations of other procedures.

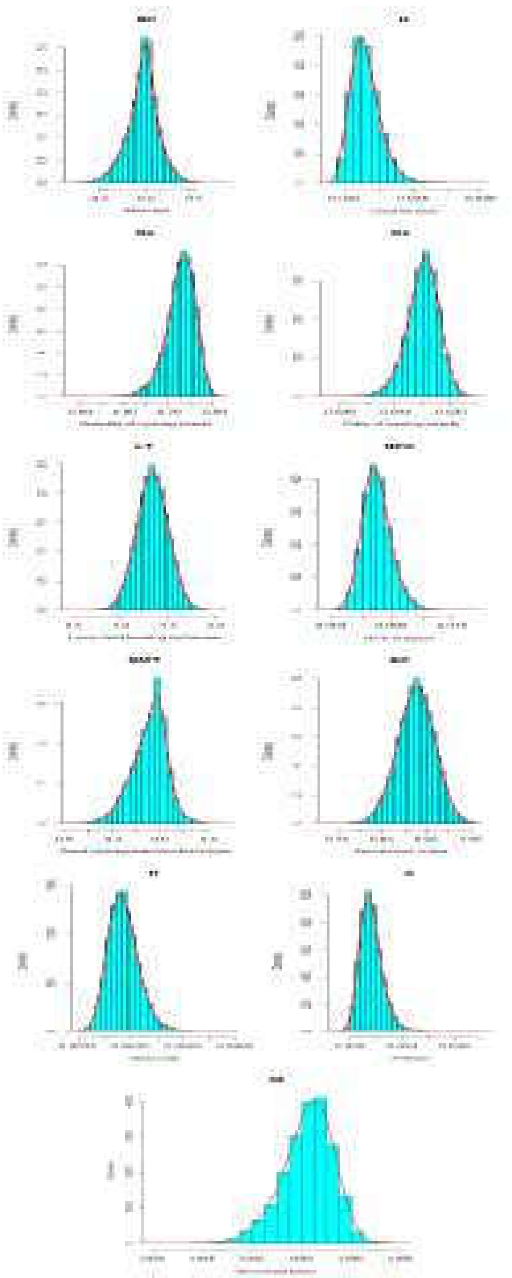


Figure 1: BALTR predictors histograms of wheat production data

Journal of AL-Qadisiyah for computer science and mathematics Vol.11 No.1 Year 2019
ISSN (Print): 2074 – 0204 ISSN (Online): 2521 – 3504

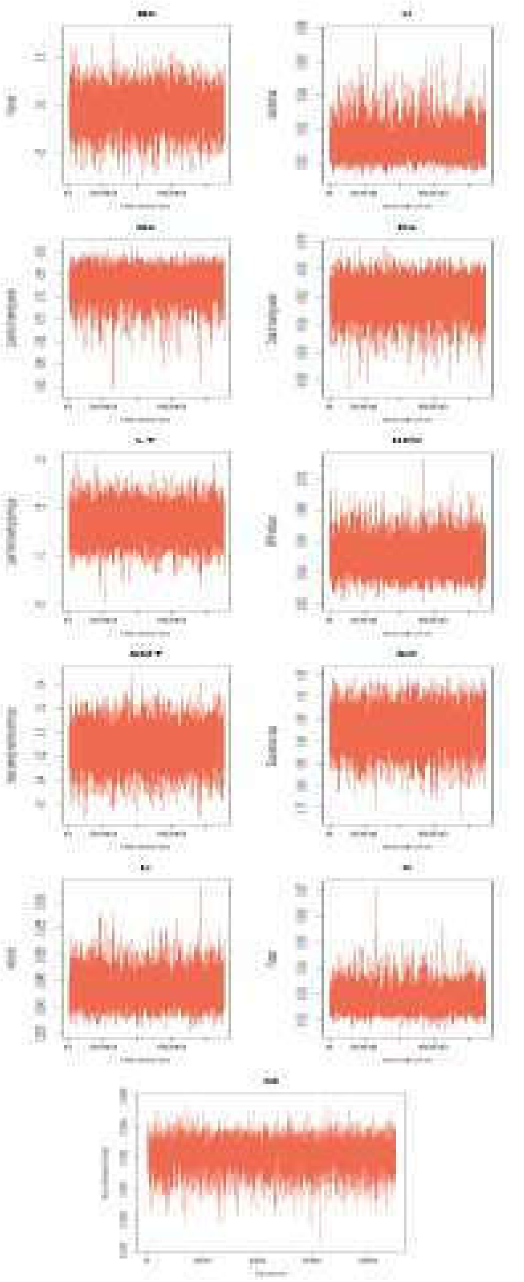
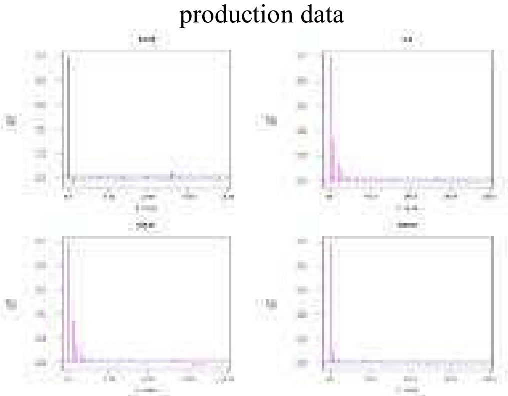


Figure 2: BALTR predictors trace plots of wheat production data



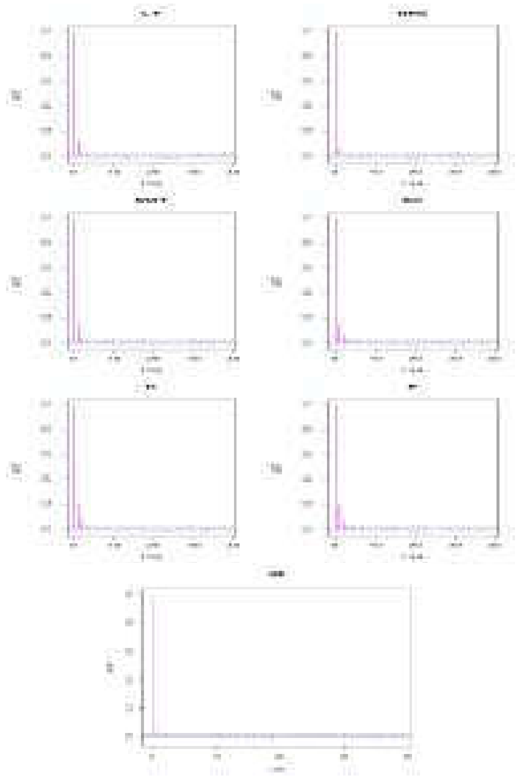


Figure 3: BALTR predictors autocorrelations of wheat production data

The predictors histograms of the wheat production based on posterior samples of 11,000 iterations are point up in figure 1, these histograms displayed that the conditional posteriors of wheat production data predictors are the preferred stationary truncated normal.

From figure 2, the trace plot indicates reasonably good convergence, and the noise does not appear to drift majorly. The chain has reached stable and the mean keeps relatively constant. it is mean that the chain is mixed well and converged.

From figure 3, the explanatory variables (covariates) in this real data are highly correlated and the mixing of the MCMC chain was reasonably good.

5.Conclusions:

This paper has introduced a new procedure for model selection of Tobit regression, we proposed BALTR for VS and coefficient estimation. Our proposed procedure depends on the scale mixture uniform as prior distribution. We advanced new Bayesian hierarchical models for BALTR. In addition, we introduced a Gibbs sampler for BALTR method. We clarified the features of the new procedure on both simulation studies and real data analysis. Results displayed that BALTR method performs very well in terms of VS and coefficient estimation.

References:

[1] Alhamzawi, R., & Yu, K. (2012). Variable selection in quantile regression via Gibbs sampling. *Journal of Applied Statistics*, *39*(4), 799-813.

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

[3] Alhamzawi, R., & Yu, K. (2013). Conjugate priors and variable selection for Bayesian quantile regression. *Computational Statistics & Data Analysis*, *64*, 209-219.

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

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

[6] Alhamzawi, R., & Taha Mohammad Ali, H. (2018). A new Gibbs sampler for Bayesian lasso. *Communications in Statistics-Simulation and Computation*, 1-17.

[7] Alhamzawi, R., & Ali, H. T. M. (2018). Bayesian tobit quantile regression with penalty. *Communications in Statistics-Simulation and Computation*, *47*(6), 1739-1750.

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

[9] Andrews, D. F., & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 99-102.

[10] Andrew D. Martin, Kevin M. Quinn, Jong Hee Park (2011). MCMCpack: Markov Chain Monte Carlo in R. Journal of Statistical Software. 42(9): 1-21.

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

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

[13] Chib, S. (1992). Bayes inference in the Tobit censored regression model. *Journal of Econometrics*, *51*(1-2), 79-99.

[14] Choy, S. T. B., & Smith, A. F. M. (1997). Hierarchical models with scale mixtures of normal distributions. *Test*, *6*(1), 205-221.

[15] Christian Kleiber and Achim Zeileis (2008). Applied Econometrics with R. New York: Springer-Verlag. ISBN 978-0-387-77316-2.

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

[17] 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.

[18] Fan, J., & Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics, 32*(3), 928-961.

[19] Griffin, J. E., & Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, *5*(1), 171-188.

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

[21] Hastie, T., Tibshirani, R., & Friedman, J. (2009). Linear methods for regression. In The elements of statistical learning (pp. 43-99). Springer, New York, NY.

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

[23] Lykou, A., & Ntzoufras, I. (2013). On Bayesian lasso variable selection and the specification of the shrinkage parameter. *Statistics and Computing, 23*(3), 361-390.

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

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

[26] Rahim Alhamzawi (2018). Brq: An R package for Bayesian Quantile Regression. *Working Paper*.

[27] 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.

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

[29] Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*. *26* (1): 24–36.

[30] Wang, H., & Leng, C. (2007). Unified LASSO estimation by least squares approximation. *Journal of the American Statistical Association*, *102*(479), 1039-1048.

[31] West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, *74*(3), 646-648.
[32] 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.
[33] Yu, K., & Stander, J. (2007). Bayesian analysis of a Tobit quantile regression model. *Journal of Econometrics*, *137*(1), 260-276.
[34] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418-1429.

انحدار adaptive Lasso Tobit البيزي

حيدر كاظم عباس رحيم جبار ظاهر
كلية الادارة والاقتصاد ، جامعة القادسية

المستخلص :

في هذا البحث، نقدم طريقة جديدة لاختيار النموذج في انحدار Tobit ، حيث اقترحنا انحدار Lasso Tobit البيزي لاختيار المتغيرات (VS) وتقدير المعلمات لنموذج الانحدار. ونقدم في هذا البحث نموذجًا هرميًا جديدا و عينات Gibbs (GS) لطريقتنا المقترحة. وقد تم توضيح طريقتنا المقترحة عن طريق المحاكاة وتحليل حقيقي للبيانات. وقد ثبتت النتائج أن طريقتنا تحقق أداءً جيدًا بالمقارنة مع الطرق الأخرى.