Using Simulation study to evaluate Bayesian Elastic Net Tobit Regression Model

Mohammed Rasool ALsafi [1]

Prof Dr. Ahmad N. Flaih [2]

[1] Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah // Iraq – 0069 – Wasit //stat.post26@qu.edu.iq

[2] Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah // Iraq – 0069 – diwaniyah //ahmad.flaih@qu.edu.iq

Abstract

The elastic net method has become very common regularization method, where the elastic net have two main advantages, first advantage is the shrinkage procedure for the parameters estimates through the ridge penalty function, and second advantage is the variable selection procedure through the lasso penalty function. Employing the Bayesian estimation depends on the choosing the prior distribution carefully for the interested parameter. In this paper we employed the scale mixture of normal mixing with truncated gamma distribution that proposed by as double exponential prior distribution for the tobit regression parameters (β_j). We proposed we hierarchical model for the Bayesian tobit regression and new MCMC Gibbs sampling. The posterior distribution estimates are comparable with classical.

Keywords: Tobit regression, Bayesian elastic net regression model, MCMC Gibbs sampler algorithm, Simulation scenarios.

استخدام دراسة محاكاة لتقييم انحدار توبت للشبكة المرنة

محمد رسول محسن الصافي [1]

أ.د أحمد نعيم. فليح [2]

[1]**قسم الإحصاء ، كلية الإدارة والاقتصاد ، جامعة القادسية //العراق - 0069 – واسط//** stat.post26@qu.edu.iq

[2]**قسم الإحصاء ، كلية الإدارة والاقتصاد ، جامعة القادسية// العراق – 0069- الديوانية// ahmad.flaih@qu.edu.iq**

المستخلص:

أصبحت طريقة الشبكة المرنة طريقة تنظيم شائعة جدًا ، حيث تتمتع الشبكة المرنة بميزتين رئيسيتين ، الميزة الأولى هي إجراء الانكماش لتقديرات المعلمات من خلال وظيفة عقوبة رج ، والميزة الثانية هي إجراء الاختيار المتغير من خلال وظيفة عقوبة لاسو .يعتمد استخدام تقدير بايز على اختيار التوزيع المسبق بعناية للمعامل المعني. في هذا البحث ، استخدمنا خليط المقياس للخلط العادي مع توزيع غاما المقتطع كتوزيع مسبق أسي مزدوج لمعلمات الانحدار توبت .(β_j) لقد اقترحنا نموذجًا هرميًا لانحدار بايز توبت وأخذ عينات MCMC كبس الجديدة. تقديرات التوزيع اللاحق قابلة للمقارنة مع نموذج الانحدار الكلاسيكي.

الكلمات المفتاحية : انحدار Tobit ، نموذج انحدار الشبكة المرنة Bayesian ، خوارزمية MCMC ، Gibbs sampler ، سيناريوهات المحاكاة.

1. Introduction

Regression methods are the most common methods in different fields of sciences, such as ecology, physical, social sciences, and in economics. Regression models are very important in forming the function of the two or more variables. Prediction is the most reason for finding the regression analysis. So finding the more interpretable regression model is the main aim for most of the scientists. The best model is the model that has only the relevant covariates. So removing the irrelevant covariates (variable selection) from the regression model is the second main aim. The ordinary least squares (OLS) method produced BLUE, but under some violated for the assumptions of the OLS, the OLS produced high variances estimators with some biased. In case of the covariates greater than or equal to sample size, or with the presents of the high correlated covariates, the OLS produced unreliable estimates. Using the regularization method is the only way to deals with such that circumanterc.

In [2] proposed a ridge regression method that estimates the coefficients of a multiple regression model based on adding a small positive amount to the diameter of the coefficient of (x′x) Studies reached when this positive amount is added increases the possibility of making the data perpendicular and thus to obtain better estimates of the regression model coefficients. In [3] proposed a new method for estimating parameters of linear models called Lasso, which are least absolute shrinkage and selection operators, that reduce the sum of squares of residuals subject to the sum of the absolute value of the coefficients, it tends to produce some coefficients that are equal to zero. Tibshirani concluded that Lasso has better properties than the sequential step method and the ridge regression method. [4] Suggested the elastic net,a new regularization and variable selection method. Reality world data and a simulation study show that the elastic net often surpass the lasso, whilst enjoy a similar sparsely of representation. Moreover, the elastic net encourages a grouping effect, where robustly correlated predictors tend to be in or out of the model together. The elastic net in particular usefully when the number of predictors (p) is much bigger than the number of observations (n). By contrast, the lasso is not a very favorable variable selection method in the $p \gg n$ case. An algorithm called LARS-EN is suggestion for computing elastic net regularization paths efficiently, much like algorithm LARS does for the lasso. In [1], N, proposed Bayesian method to solve the elastic net model using a Gibbs sampler. The Bayesian elastic net has two major advantages. Firstly, as a Bayesian method, the distributional results on the estimates are straightforward, Secondly, it chooses the two penalty parameters simultaneously, avoiding the double shrinkage problem" in the elastic net method. Real data examples and simulation studies show that the Bayesian elastic net it performs well compared to other methods.

In [5] introduced the Tobit quantile regression model using the adaptive lasso penalty function new hierarchical model and new Gibbs sample algorithm have developed through employing of the location – scale mixture of normal as formula for the skewed Laplace prior distribution. The proposed model performs well comparing with other regularization method .In [6] introduced the Bayesian Tobit quantile regression model by employing the g-prior density additionally to using the ridge parameter. In this paper adding ridge parameter was to deal with some challenges that comes with censored data, like collinearity between the covariates. This work also deal with variable selection procedure Basel on the g-prior. The results of simulation and real data analysis illustrated the outperformance of the proposed model. In [7] introduced the Bayesian elastic net for the Tobit quantile regression model. The new regularization method deals with the variable selection procedure and parameters estimation for the Tobit quantile model by using the elastic net penalty function through employing the gamma priors. In this work Alhamzawi treated the hyper parameters of the proposed gamma priors. The results of simulation and real data analysis were comparable with some exists methods.

In [8] introduced new hierarchal model for the Tobit regression by using lasso penalty function. In this work the scale mixture for uniforms mixing with special case of gamma distribution as representation of the Laplace prior distribution employed for develop. New Gibbs samples algorithm. Parameter estimation and variable selection were performs. Simulation example, and real data analysis have been showed that the proposed model performs well comparing with some other methods. In [9] and Haithem suggested a new Bayesian elastic net (EN) approach for variable selection and coefficient estimation in Tobit regression. Mostly, we present a new hierarchical formularization of the Bayesian EN by utilizing the scale mixture of truncated normal distribution (with exponential mixing distributions) of the Laplace density part. The Proposal method is an alternate method to Bayesian method of the EN problem. It is shown up that the model performs well Comparison with old elastic net representation. In [10], introduced new regularization method by using transformation for the scale mixture of Laplace prior distribution that proposed by Ma lick and yi (2014). Also, new Gibbs sampling algorithm proposed for the Bayesian adaptive lasso Tobit regression. The results of simulation and practiced side were very promising.

2.Hierarchical Model and Prior Distributions for Bayesian Elastic Net Tobit Regression (BENTR) Model

We proposed a BENTR analysis in this thesis for the parameters estimation and the variable selection procedure. We employed the prior distribution of $\boldsymbol{\pi}(\boldsymbol{\beta}|\boldsymbol{\sigma}^{2},\boldsymbol{\lambda}_{1},\boldsymbol{\lambda}_{2})$, which is defined by

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

Then, in general the posterior marginal distribution of the parameter β of the Tobit regression model conditioning on latent variable y^* is

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

Where y^0 is a censoring point. We exploits the above formulas (2 and 1) to setup the Bayesian elastic net Tobit regression through the following general posterior marginal density of β,

$$f(\beta/y) = \int_0^\infty h\,(2\pi\sigma^2)^{-n/2}\ exp\left\{-\frac{1}{2\sigma^2}\Big((y_i - max(y^0,y^*)) - \frac{1}{2\sigma^2}(\lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p} {\beta_j}^2\Big)\right\} \pi\sigma^2\, d\sigma^2 \ldots\ldots\ldots\ldots (3)$$

Where h is the normalizing constant of λ_1 , λ_2 , and σ^2.
By using the structure equation of Tobit regression (2.6) and the prior proposed by [1]

$$\exp\{-\frac{1}{2\sigma^2}(\lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p} {\beta_j}^2)\} \propto$$
$$K \prod_{j=1}^{p} \int_1^\infty \sqrt{c}\, \exp\{-\frac{\beta_j^2}{2}(\frac{\lambda_2}{\sigma^2}c)\} w^{\frac{-1}{2}}\ \exp(-\frac{1}{2\sigma^2}\frac{\lambda_1^2}{4\lambda_2}w)\, dw \ldots\ldots\ldots\ldots (4)$$

where K is the normalizing constant and $c = \frac{w}{w-1}$. The prior formula (6) represent the scale mixture of normal mixing with truncated gamma. Suppose that $y^0 = 0$, then we list the following proposed hierarchical model for the Bayesian elastic net regression model.

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

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

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

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

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

$$\sigma^2 \sim \frac{1}{\sigma^2}.$$

Where TG is the truncated gamma supported on (1,0). Our contribution is employing the hierarchy model (3) to develop new Bayesian computation for the elastic net Tobit regression.

3. Conditional Posterior Distributions

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

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

Where i = 1, 2, …, n

Following [6] and [1] and conditioning on y^*, β the posterior distribution of β is

$$\pi(\ \beta/y^*,\ \sigma^2, \gamma) \propto\ \pi(\ y^*/\beta,\ \sigma^2, \gamma)\ \pi(\beta/\sigma^2)$$
$$\propto \exp\{-\frac{1}{2\sigma^2}(y^*-x'\beta)'(y^*-x'\beta)\}\ \exp\{-\frac{1}{2\sigma^2}\beta' Q_\gamma \beta)$$
$$\text{Where } Q = \text{diag}\ (\ \frac{\gamma_1}{\gamma_2-1}\ , \dots ,\frac{\gamma_p}{\gamma_p-1})$$
$$= -\frac{1}{2\sigma^2}[\ \beta'(x'x)\ \beta - 2y^*x\beta + y'^*y^* + \beta' Q_\gamma \beta\ ]$$
$$= -\frac{1}{2\sigma^2}[\ \beta'\ (x'x - Q_\gamma\ )\ \beta - 2y^*x\beta + y^{*'}y^*]$$
$$\text{Let } s = x'x + \lambda_2 Q_\gamma \text{ , then}$$
$$= -\frac{1}{2\sigma^2}[\ \beta' s\beta - 2y^*x\beta + y^{*'}y^*]$$
$$= -\frac{1}{2\sigma^2}\ (\beta - s^{-1}x'y^*\ )'c\ (\beta - s^{-1}x'y^*) \dots \dots \dots \dots \dots \dots \quad (5)$$

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

$$\boldsymbol{\beta/y^*\ ,\ \sigma^2,\ \gamma \sim N\ (s^{-1}x'y^*,\ \sigma^2 s^{-1})} \dots \dots \dots \dots \dots \dots \dots \quad \mathbf{(6)}$$

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

$$\boldsymbol{\pi(\sigma^2/y^*,\ \beta,\gamma) \propto\ \pi(y^*/\beta,\sigma^2,\gamma)\ \pi(\beta/\sigma^2)\ \pi\ (\sigma^2)}$$
$$\boldsymbol{\propto (\sigma^2)^{\frac{-n}{2}-p-1}\{\Gamma z\ (\ \frac{1}{2}\ ,\frac{\lambda_1^2}{8\sigma^2\lambda_2})\}^{-p}\ exp[\ -\frac{1}{2\sigma^2}\{(y^*-x'\beta)'(y^*-x'\beta) + \lambda_2\sum_{j=1}^{p}\frac{\gamma_j}{\gamma_j-1}\ \beta_j^2}$$
$$\boldsymbol{+\frac{\lambda_1^2}{4\lambda_2}\sum_{j=1}^{p}\gamma_j\}],\dots \dots \dots \dots \dots} \quad \mathbf{(7)}$$

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

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

$$(\gamma - 1_P)/y^*, \sigma^2, \beta \sim \prod_{j=1}^{p} \text{GIG}\left(\lambda = \frac{1}{2}, \varphi = \frac{\lambda_1}{4\lambda_2\sigma^2}, \chi = \frac{\lambda_2\beta_j^2}{\sigma^2}\right), \dots\dots\dots\dots \textbf{(8)}$$

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

$$f(x/\lambda, \varphi, \chi) = \frac{(\varphi/\chi)^{\lambda/2}}{2k_\lambda(\sqrt{\varphi\chi})} x^{\lambda-1} \exp\left\{-\frac{1}{2}(\chi x^{-1} + \varphi x)\right\}, \dots\dots\dots\dots \textbf{(9)}$$

Where $\boldsymbol{x} > \mathbf{0}$, $\boldsymbol{k_\lambda}$ **(.)** is the Basel function of the third Kind with order $\boldsymbol{\lambda}$. So, we can easily say that

With the following pdf,

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

See [11] for more details.[12] and [13] suggested the empirical Bayes estimates for the shrinkage parameters $\boldsymbol{\lambda_1}$ and $\boldsymbol{\lambda_2}$ by using the marginal maximum likelihood of the data and use the Monte Carlo Expectation- maximization (**MCEM**) algorithm. Following [1] we treated $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\sigma^2}$ as missing data and $(\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2})$ as fixed parameters, the likelihood is

$$\lambda_1^P \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+p+1} \left\{\Gamma_u\left(\frac{1}{2}, \frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\right\}^{-p} \prod_{j=1}^{p} \left(\frac{1}{\gamma_j - 1}\right)^{1/2} exp\left[-\frac{1}{2\sigma^2}\left\{(y^* - x'\beta)(y^* - x'\beta) + \lambda_2 \sum_{j=1}^{p} \frac{\gamma_j}{\gamma_j - 1}\beta_j^2 \frac{\lambda_1^2}{4\lambda_2} \sum_{j=1}^{p} \gamma_j\right\}\right], \dots\dots\dots\dots\dots\dots \textbf{(10)}$$

and the we can take the log for the function (10) and the solve the maximization problem by MCMC algorithm. One can see [1] for more details.

4.Simulation Study

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

$$MMAD = median\,[mean|x'\hat{\beta} - x'\beta^{true}|] \dots\dots\dots\dots\dots\dots \textbf{(11)}$$

The overall efficiency of each estimation method can be compared by the total MMAD. MCMC (Gibbs sampling) algorithm have used with **20000** iterations to reach the stationary for the posterior distributions of the interested parameters, we burned in the first 1000 iterations, moreover I have generated the observations of predictor variables x1, … , x9 from Normal distribution, Nn=9(0 , ∑) , where the variance covariance matrix ∑ij = $\rho^{|i-j|}$ under four different distributions of the i.i.d errors. For each simulation study, we run 300 simulations.

4.1. Simulation Scenario One

In this simulation scenario, I assumed the true vector of coefficients Q = (0, 3, 0, 0, 0, 0, 0, 0, 0)^t which is the case of very sparse vector with error terms followed ci∼$N(\mu = 0, \sigma^2 = 1)$ci∼$N(\mu = 0, \sigma^2 = 5)$$e$_i∼$N$(0, 1) +(0, 1)$e$_i∼$t$_((4)) . As well, I generated the observations of x1, … , x9 predictor variables through Nn=9(0 , ∑), where ∑ is the **variance covariance** matrix defined as ∑ ij = 0. $5^{|i-j|}$. Consequently I have simulated the following regression model,yi = 3x_2 under different samples sizes (n= 25,50,100,150,200,250) and different estimation methods (Tobit, BAnet, Crq, our method). The censored point was equal to zero **(c=0)** to figure out the behavior of the estimation methods. as shown in table (1)

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

	Methods		$e_i \sim N(0,1))$	$e_i \sim N(0,5)$	$e_i \sim N(0,1) + N(0,1)$	$e_i \sim t_{(4)}$
Sim1	n=25	Tobit	3.30982 (0.75746)	4.7487 (1.25884)	2.37132 (2.10976)	2.9044 (1.10347)
		BAnet	2.07033 (1.13683)	6.72492 (3.06104)	4.5450638 (3.11594)	4.94623 (0.54026)
		BCrq	3.01247 (0.58850)	4.94020 (1.44609)	3.07189 (2.06448)	3.41190 (0.44052)
		Our method	0.37287 (0.22559)	1.08482 (0.69666)	0.33702 (0.70987)	0.36363 (0.17223)
	n=50	BTobit	2.48780 (0.14205)	3.1066 (0.64449)	1.84961 (0.54231)	1.82237 (0.40552)
		BAnet	4.32501 (0.46857)	5.49438 (0.76207)	4.16627 (0.57398)	4.36752 (0.70298)
		BCrq	3.26266 (0.11311)	4.12225 (0.77484)	2.73711 (0.50161)	2.78015 (0.51844)
		Our method	0.14582 (0.04867)	0.63749 (0.19990)	0.20283 (0.06488)	0.18982 (0.04529)
	n=100	Tobit	1.95808 (0.22672)	1.76343 (0.46198)	1.74778 (0.29782)	1.90115 (0.58042)
		BAnet	3.82422 (0.16980)	4.64568 (0.15125)	3.87999 (0.418029)	3.72509 (1.01583)
		Crq	2.80118 (0.22277)	2.66914 (0.60335)	2.61293 (0.317344)	2.68085 (0.59662)
		Our method	0.12408 (0.05791)	0.58881 (0.14659)	0.18863 (0.061547)	0.14853 (0.03733)

	n=150	Tobit	1.94378 (0.22304)	1.83543 (0.38865)	1.78748 (0.44371)	1.93511 (0.24592)
		BAnet	3.70282 (0.35578)	5.07402 (0.49252)	3.93621 (0.17507)	3.77861 (0.33838)
		Crq	2.86402 (0.21994)	2.79554 (0.26790)	2.64137 (0.27084)	2.74846 (0.24919)
		Our method	0.12257 (0.01294)	0.43897 (0.13995)	0.12061 (0.05273)	0.14641 (0.02494)
	n=200	Tobit	1.57864 (0.45740)	1.70312 (0.33581)	1.82346 (0.27705)	1.71068 (0.20790)
		BAnet	3.76246 (0.29654)	4.31168 (0.746763)	3.56677 (0.15152)	3.6962 (0.41481)
		Crq	2.62427 (0.39261)	2.50376 (0.66136)	2.70301 (0.17888)	2.62185 (0.19342)
		Our method	0.10162 (0.01697)	0.42092 (0.07651)	0.1364 (0.032570)	0.11030 (0.02014)
	N=250	Tobit	1.69367 (0.25736)	1.8003 (0.44779)	1.68130 (0.096407)	1.75113 (0.28881)
		BAnet	3.85296 (0.24572)	4.95694 (0.50716)	3.64999 (0.43214)	3.60149 (0.19433)
		Crq	2.72156 (0.17765)	2.4298 (0.64758)	2.55005 (0.239980)	2.68496 (0.26166)
		Our method	0.08116 (0.01489)	0.40163 (0.13975)	0.12660 (0.011466)	0.12872 (0.03229)

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

4.2. Simulation Scenario Two

In this simulation scenario, I assumed the true vector of coefficients $\beta = (0,3,0,0,0,1,0,0,0)^t$ which is the case of sparse vector with error terms followed $\epsilon_i \sim N(\mu = 0, \sigma^2 = 1)\epsilon_i \sim N(\mu = 0, \sigma^2 = 5)e\_i \sim N(0,1) + N(0,1)e\_i \sim t\_((4))$. As well, I generated the observations of $x_1, \dots, x_9$ predictor variables through $N_{n=9}(0\,, \Sigma)$, where Σ is the variance covariance matrix defined as $\sum ij = 0.5^{|i-j|}$. Consequently I have simulated the following regression model under different samples sizes (n= 25,50,100,150,200,250) and different estimation methods (Tobit, BAnet, Crq, our method). The censored point was equal to zero (c=0) to figure out the behavior of the estimation methods. as shown in table (2)

$$y_i = 3x_{2i} + x_{6i} \dots \dots \dots \dots (12)$$

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

Methods			$e_i \sim N(0,1))$	$e_i \sim N(0,5)$	$e_i \sim N(0,1) + N(0,1)$	$e_i \sim t_{(4)}$
Sim2	n=25	Tobit	4.9820 (1.01681)	7.84054 (2.27518)	4.46189 (1.58553)	4.43623 (0.34469)
		Banet	5.3221 (0.52401)	7.6939 (1.07881)	5.12101 (0.77253)	5.53461 (0.49935)
		Crq	5.22560 (1.05874)	7.71490 (1.84375)	4.73255 (1.38319)	4.52299 (0.43189)
		Our method	0.50118 (0.08693)	1.63418 (0.33883)	0.65546 (0.16668)	0.67112 (0.07113)
	n=50	Tobit	3.12874 (0.62296)	3.90474 (0.90037)	2.90393 (0.60466)	2.54021 (0.40373)
		Banet	4.52996 (0.21200)	6.06506 (0.67733)	4.85503 (0.41606)	4.72445 (1.24641)
		Crq	3.77596 (0.50462)	4.52070 (0.75565)	3.77654 (0.45691)	3.47951 (0.15341)
		Our method	0.42141 (0.07036)	1.15171 (0.15558)	0.50533 (0.08443)	0.43870 (0.13008)
	n=100	Tobit	2.51882 (0.18451)	2.77265 (0.72729)	2.42383 (0.460615)	2.53287 (0.40271)
		Banet	4.29508 (0.19913)	5.46653 (0.93122)	4.12029 (0.38019)	4.54986 (0.62229)
		Crq	3.12670 (0.22563)	3.54866 (0.58411)	3.09834 (0.37150)	3.32728 (0.52900)
		Our method	0.21551 (0.01403)	0.78428 (0.14250)	0.32320 (0.08093)	0.27606 (0.06570)
	n=150	Tobit	2.01061 (0.16882)	2.87802 (0.54578)	2.40299 (0.38323)	2.34290 (0.16806)
		Banet	4.14242 (0.19149)	5.51047 (0.58329)	4.37114 (0.47051)	3.86158 (0.36087)
		Crq	2.98740 (0.12283)	3.46618 (0.46774)	3.24750 (0.21941)	3.13223 (0.16784)
		Our method	0.19261 (0.02928)	0.74683 (0.13671)	0.30538 (0.05840)	0.21724 (0.01709)
	n=200	Tobit	2.19052 (0.25246)	2.25748 (0.16924)	1.86645 (0.11085)	2.00493 (0.22569)
		Banet	3.74736 (0.23581)	5.076705 (0.99285)	3.94052 (0.24737)	3.80818 (0.30723)
		Crq	2.91260 (0.17236)	2.89879 (0.29618)	2.70515 (0.17736)	2.79608 (0.20286)
		Our method	0.16439 (0.03634)	0.58875 (0.10006)	0.22428 (0.05414)	0.21633 (0.03776)
	N=250	Tobit	2.11974 (0.27044)	2.06982 (0.58897)	1.95753 (0.26718)	1.83596 (0.29351)
		Banet	3.92183 (0.16112)	4.96555 (0.62139)	3.93794 (0.17582)	3.80586 (0.15093)
		Crq	2.90824 (0.18913)	2.91075 (0.45084)	2.86795 (0.26016)	2.76428 (0.33360)
		Our method	0.14347 (0.01214)	0.55799 (0.07386)	0.17744 (0.03823)	0.18510 (0.03892)

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

4.3. Simulation Scenario Three

In this simulation scenario, I assumed the true vector of coefficients $\beta = (0,0.85,0.85,0.85,0.85,0.85,0.85,0.85,0.85)^t$ which is the case of density vector with error termsfollowed $\epsilon_i \sim N(\mu = 0, \sigma^2 = 1) \epsilon_i \sim N(\mu = 0, \sigma^2 = 5) e\_i \sim N(0,1) + N(0,1) e\_i \sim t\_((4))$. As well, I generated the observations of $x_1, \dots, x_9$ predictor variables through $N_{n=9}(0\,, \Sigma)$, where Σ is the variance covariance matrix defined as $\sum ij = 0.5^{|i-j|}$. Consequently I have simulated the following regression model under different samples sizes (n= 25,50,100,150,200,250) and different estimation methods (Tobit, BAnet, Crq, our method). The censored point was equal to zero (c=0) to figure out the behavior of the estimation methods. as shown in table (3)

$$y_i = \sum_{i=1}^{8} 0.85 X_i \dots \dots \dots \dots (13)$$

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

	Methods		$e_i \sim N(0,1))$	$e_i \sim N(0,5)$	$e_i \sim N(0,1) + N(0,1)$	$e_i \sim t_{(4)}$
Sim3	n=25	Tobit	4.99173 (0.64713)	5.15854 (0.91697)	5.84224 (1.00871)	4.80568 (1.44157)
		BAnet	5.90819 (0.79945)	7.74524 (0.30492)	5.98727 (1.41810)	5.73861 (0.38307)
		Crq	4.96285 (0.53977)	5.78119 (0.68264)	6.61679 (2.38345)	4.92277 (1.33457)
		Our method	0.67795 (0.16489)	1.21295 (0.28481)	1.02232 (0.16457)	0.72957 (0.15808)
	n=50	Tobit	3.46772 (0.42519)	3.14952 (0.78548)	4.40413 (0.65500)	3.03778 (0.35735)
		BAnet	4.83583 (0.18644)	5.86021 (0.94120)	4.84948 (0.56408)	4.52154 (0.73988)
		Crq	4.04724 (0.22848)	4.22451 (0.75211)	4.62629 (0.62147)	3.33893 (0.33316)
		Our method	0.38497 (0.08232)	0.74404 (0.11085)	0.75890 (0.21315)	0.33860 (0.04648)
	n=100	Tobit	2.77326 (0.37594)	2.78412 (0.28275)	2.34919 (0.37167)	2.39887 (0.51498)
		BAnet	4.44287 (0.51859)	5.67594 (1.11220)	4.35189 (0.48969)	4.18720 (0.34679)

		Crq	3.28826 (0.50175)	3.55983 (0.48211)	3.11690 (0.42883)	3.21952 (0.28978)
		Our method	0.24549 (0.06827)	0.73532 (0.18677)	0.44187 (0.06088)	0.20285 (0.03367)
	n=150	Tobit	2.44676 (0.25534)	2.58381 (0.52537)	2.48412 (0.25806)	2.45422 (0.29177)
		BAnet	4.01119 (0.44898)	5.09955 (0.80961)	4.38368 (0.61308)	4.01516 (0.35297)
		Crq	3.26777 (0.32659)	3.06652 (0.28603)	3.22498 (0.45663)	3.11895 (0.28316)
		Our method	0.17205 (0.03548)	0.62611 (0.09641)	0.40948 (0.06592)	0.21682 (0.02663)
	n=200	Tobit	2.36452 (0.18308)	2.14097 (0.44484)	2.30627 (0.39299)	2.176333 (0.31181)
		BAnet	3.93020 (0.20159)	4.95660 (0.31110)	4.01492 (0.33161)	3.75485 (0.15923)
		Crq	3.00732 (0.09246)	3.08527 (0.43955)	2.87015 (0.31929)	2.98302 (0.20379)
		Our method	0.13473 (0.02470)	0.56940 (0.13748)	0.31753 (0.06180)	0.18114 (0.03056)
	N=250	Tobit	2.38328 (0.20155)	2.19319 (0.83867)	1.97403 (0.27295)	2.26859 (0.39627)
		BAnet	3.92374 (0.16314)	5.42051 (0.45303)	3.95325 (0.38264)	3.92732 (0.23367)
		Crq	3.03154 (0.12195)	3.10431 (0.83456)	2.68799 (0.27220)	2.95460 (0.16223)
		Our method	0.15859 (0.04718)	0.56458 (0.10610)	0.31516 (0.05041)	0.16410 (0.02288)

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

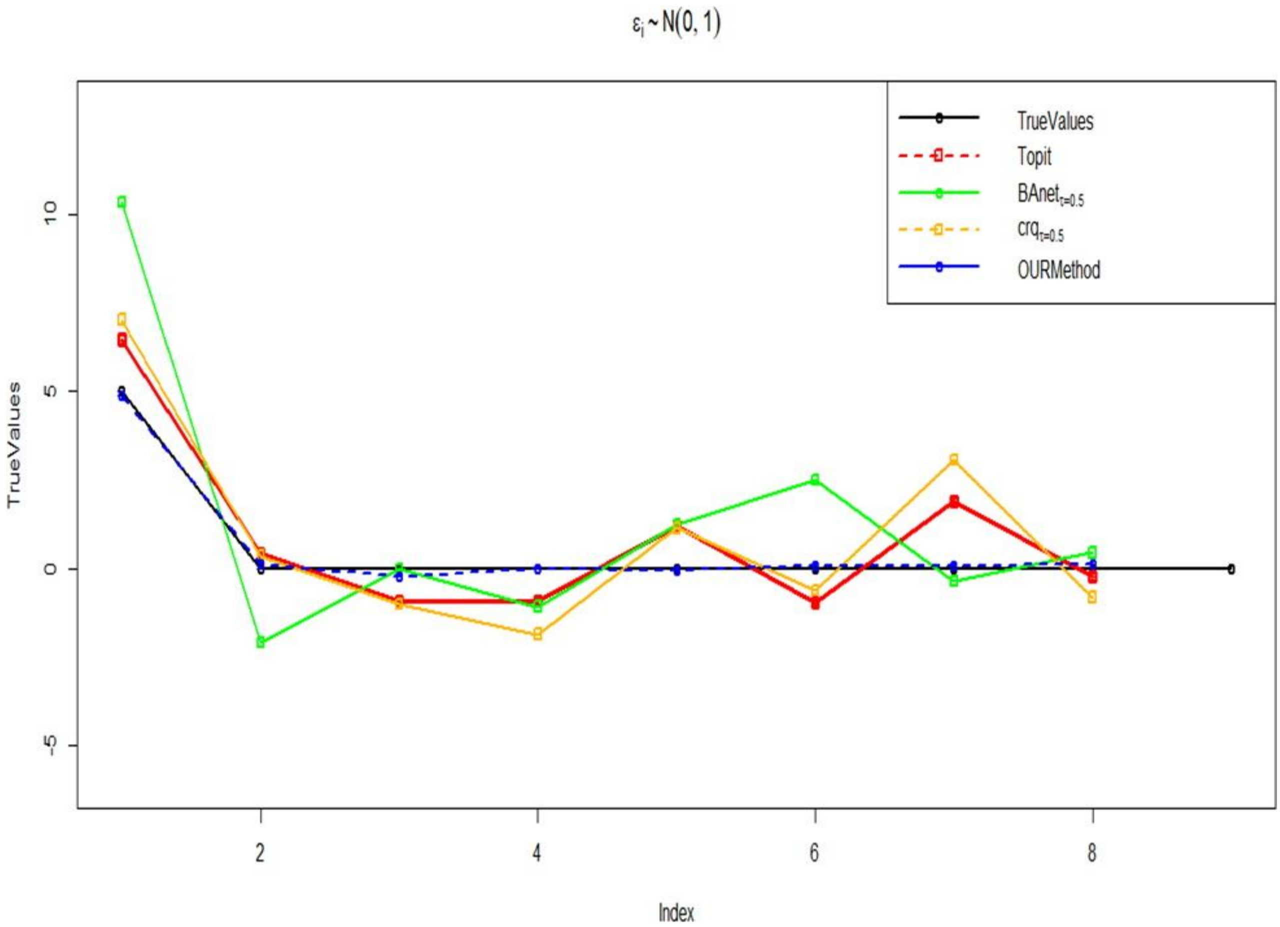

Figure (1). Lines plot for the different estimation methods with ei~(0, 1) and n=25

Figure (1) Contain the results of simulation scenario one, where the error $e_i \sim N(0,1)$. The figure contains the sparse line (black) in the middle . The vertical line represents the true vector furthermore the Blue line represents the parameters estimates based the proposed model using sample size (n=25) the red line is the Tobit model results, the orange line is the (Crq=0.5) results , and the green line (BA net) results. From figure (1) it is very clearly to observe that the blue line is the closed line to the standard line (sparse) and matching some points . But the Tobit model parameters estimates come next. as shown in figure (2)

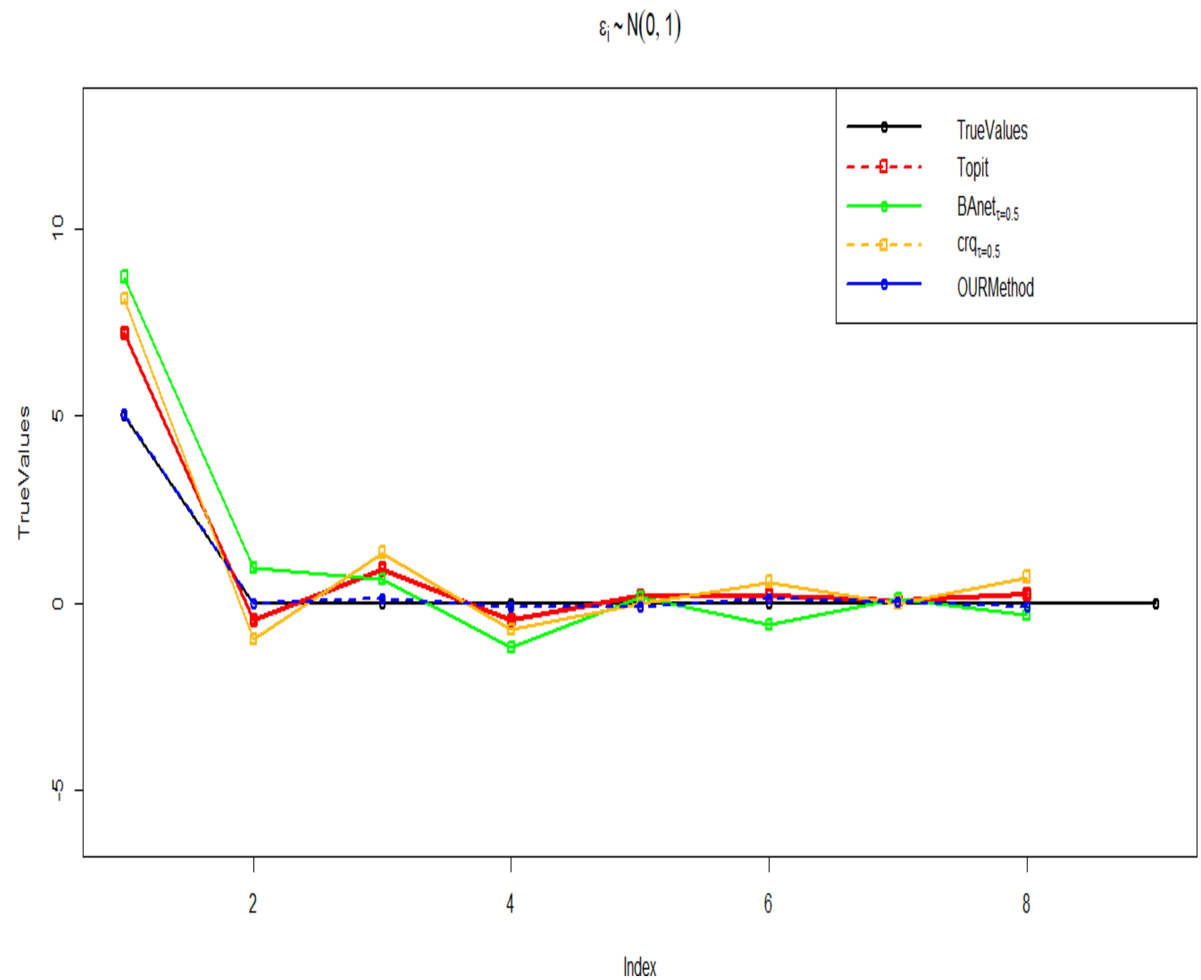

Figure (2). Lines plot for the different estimation methods with e_i~N(0, 1) and (n=50).

In figure (2) , we draw the results of simulation scenario tow with e_i~N(0, 1) and (n=50) . The result represent the parameters estimates for the different models. We observed that the parameter estimates of the proposed mode (blue line) are very close and matching in some points the standard line (sparse) .Also, for the other model results are closed to each other and matching the sparse line in some points. as shown in figure (3)

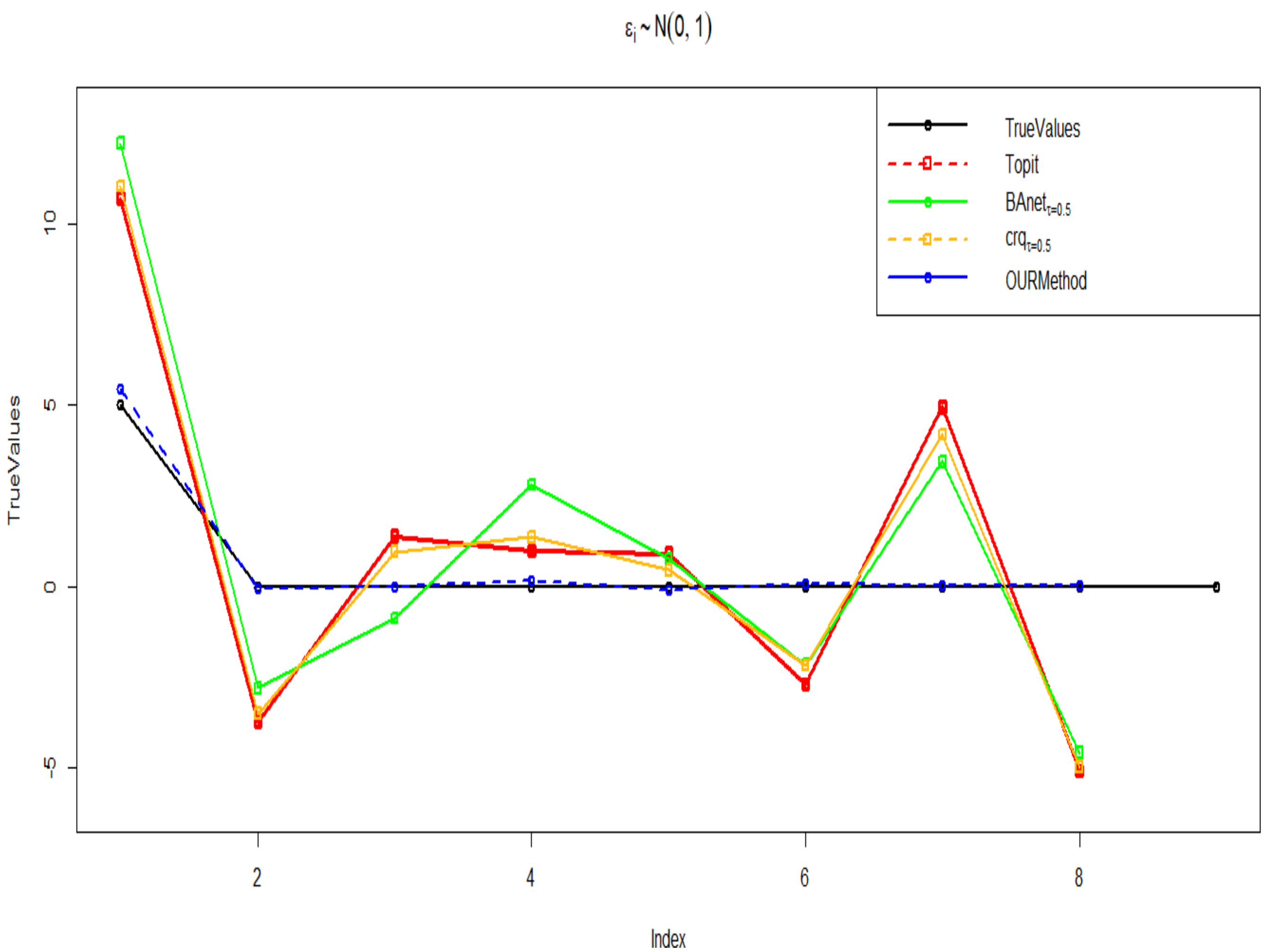

Figure (3). Lines plot for the different estimation methods with e_i~N(0, 1) and (n=100).

Figure (3) shows that the blue line is the closed line to the sparse line under e_i~N(0, 1) and with sample size (n=100). Also, we observed the matching of blue line points (parameters estimates) with the black line. For the other models, clearly all the lines (red, orange , green) are for away from the sparse that, but it matching each other in some points. as shown in figure (4)

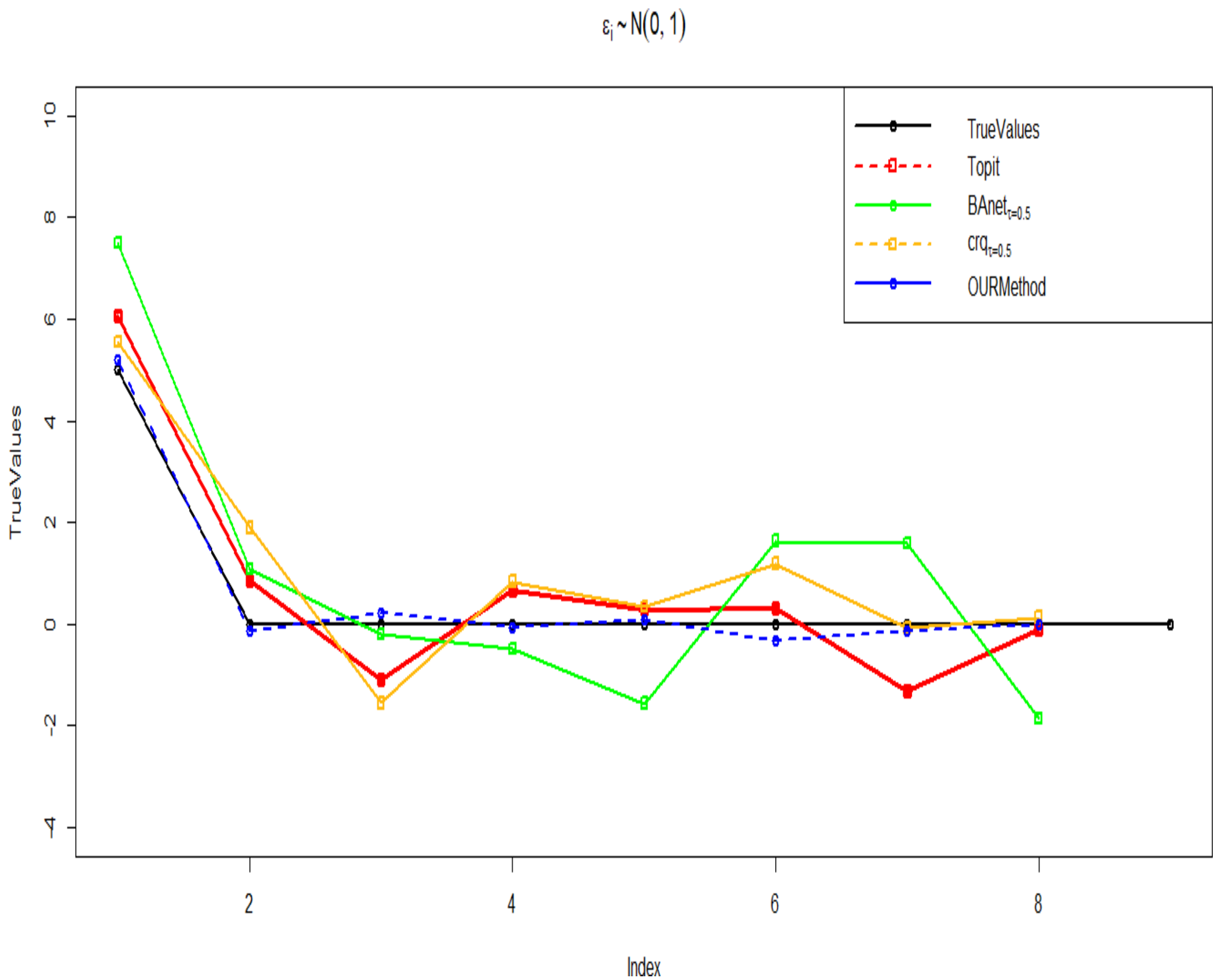

Figure (4). Lines plot for the different estimation methods with e_i~N(0, 1) and (n=150).

In figure (4) displayed the results of parameters estimates for the simulation scenario one under e_i~N(0, 1) and sample size (n=150). Very clearly, the blue line is the closed line to the sparse vector of true parameters estimates comparing with the other lines. as shown in figure (5)

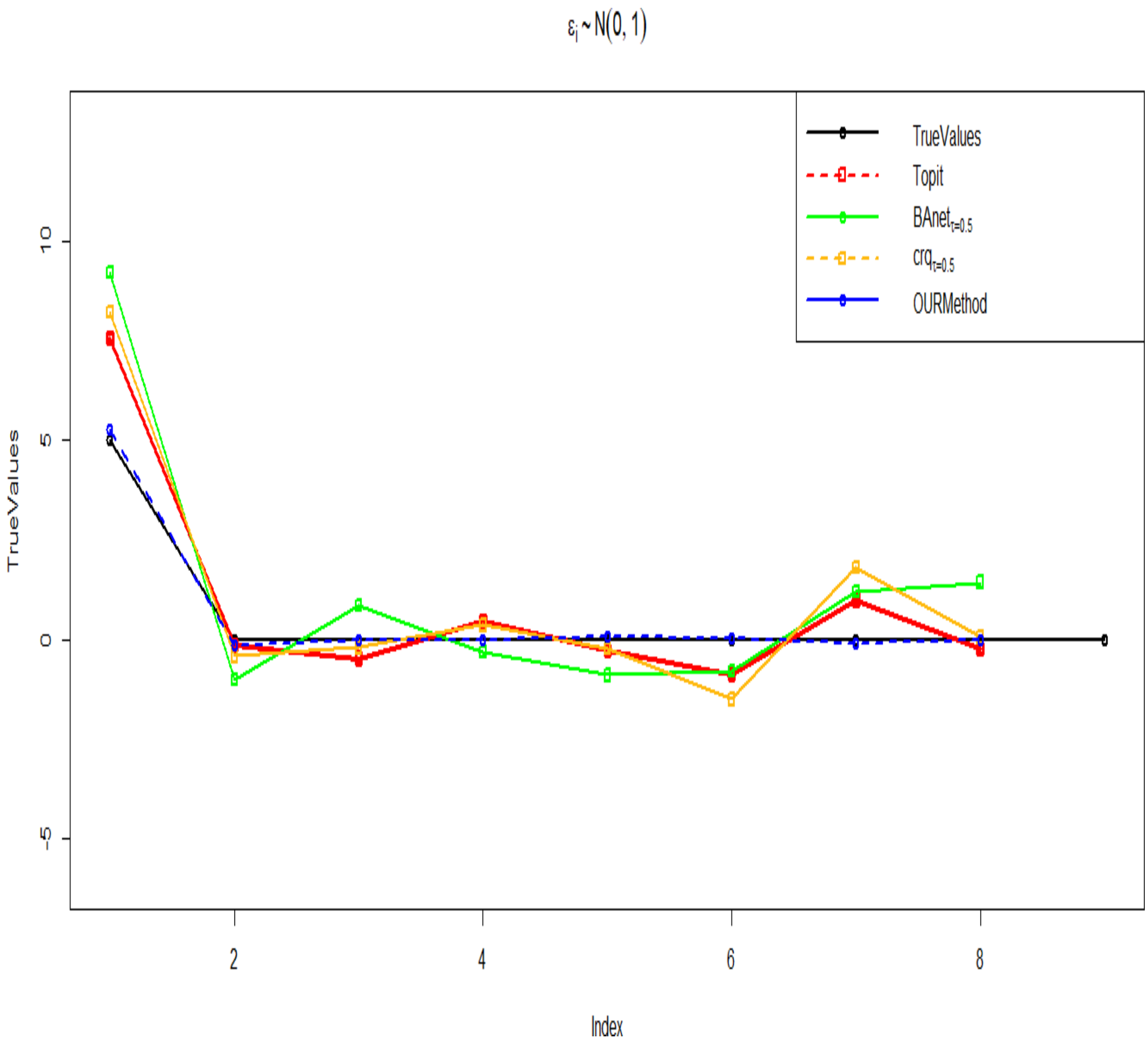

Figure (5). Lines plot for the different estimation methods with $e_i \sim \mathrm{N}(0, 1)$ and (n=200).

In figure (5) the lines drawn for the simulation scenario one with e_i~N(0, 1) and (n=200).Also, it is very clear that the parameters estimates that computed from the proposed posterior distribution for $B^{'s}$ are very closed to the sparse line (Black line) and matching in some points. The other lines are very close to each other and close to sparse line. as shown in figure (6)

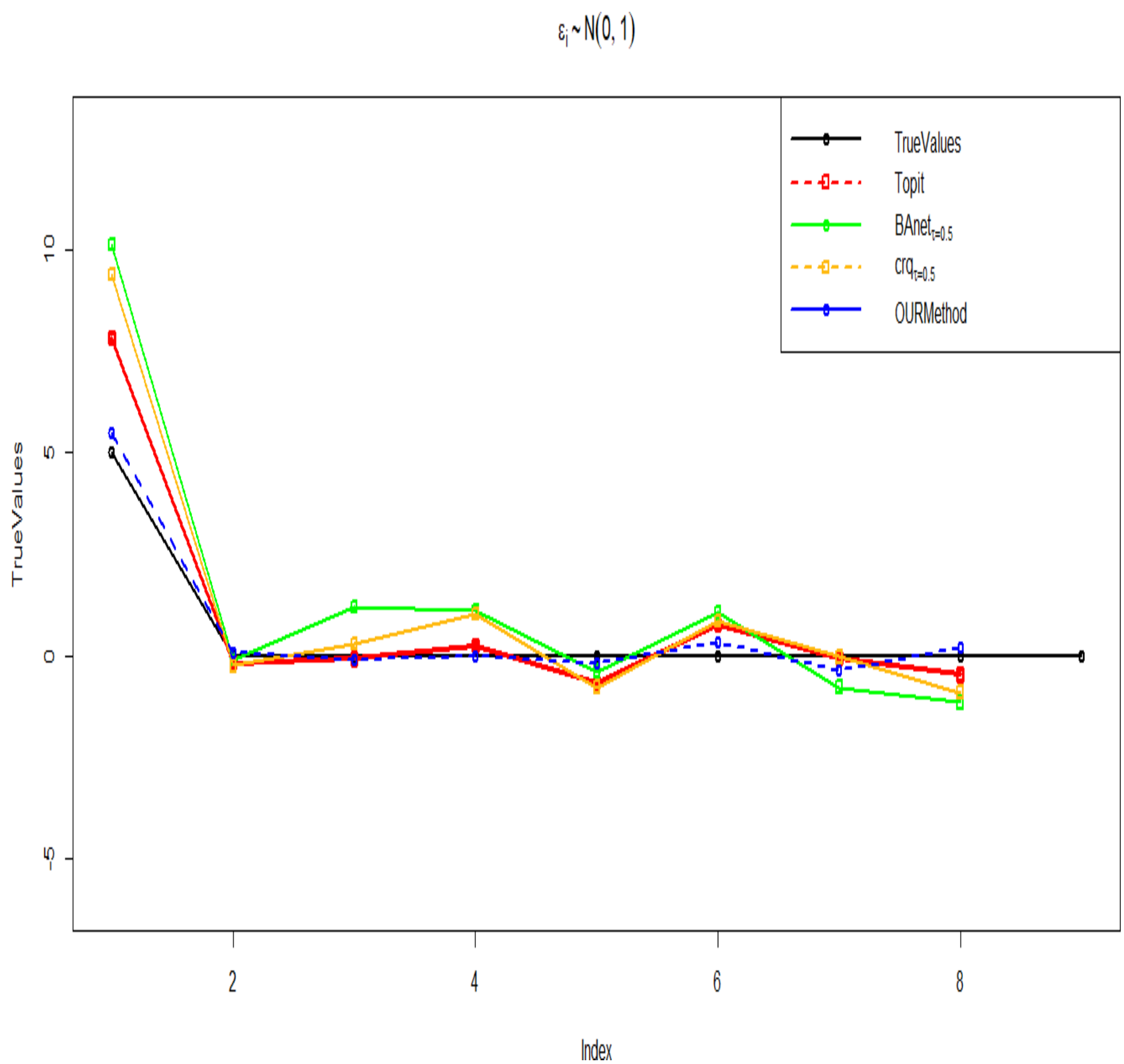

Figure (6). Lines plot for the different estimation methods with e_i~N(0, 1) and (n=250).

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

5. Conclusions

In this paper we presented the Bayesian elastic net tobit regression models by employing the scale mixture of normal mixing with truncated gamma distribution that proposed by [1] as double exponential density of parameter (β). We proposed new hierarchical model, also we employed new MCMC Gibbs sampler algorithm for the proposed posterior distribution based on the above scale mixture. I illustrated the behavior of the proposed model in the simulation analysis. The results shown that the proposed model performs well comparing with some models based on the MMAD values, different sample sizes, and different types of error distribution. Furthermore, I illustrated the results of simulation scenarios by using some figure that emphasized the results of the MMAD values.

Bibliography

[**1**] - Li, Q., and Lin, N. (2010). The Bayesian Elastic Net. Bayesian Analysis, 5(1), 151-170.

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

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

[**4**] - Zou, H. and 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.

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

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

[**7**] - Alshaybawee, T. Midi, H. & Alhamzawi, R. (2016). Bayesian elastic net single index quantile regression.

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

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

[**10**] - Hilali, H. K. A. (2019). Bayesian adaptive Lasso Tobit regression with a practical application.

[**11**] - Chhikara, R. and Folks, L. (1988). The Inverse Gaussian Distribution Theory: Method-ology, and Applications. Marcel Dekker, Inc, 1st edition. 156

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

[**13**] - Casella, G. (2001). "Empirical Bayes Gibbs sampling." Biostatistics, 2: 485–500.