Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 26, Issue 4 (2024) ISSN$_{Print}$: 1816-9171

Variable selection in composite Tobit Quantile regression by Bayesian new lasso technique

Fadel Hamid Hadi Alhusseini Asad Mahdi Faraj
Fadel.alhusiny@qu.edu.iq Assadmahdy94@gmail.com
University of AL-Qadisiyah

Received: 1/9/2024
Accepted: 8/10/2024
Available online: 25 /12 /2024

Corresponding Author : Asad Mahdi Faraj

Abstract : Our paper presents a new and simplified approach to variable selection and estimate coefficients in Tobit composite quantile regression using the new lasso technique. As an alternative to the Laplace distribution, it is the use of a new prior distribution made up of a combination of the Uniform distribution and the exponential distribution. This substitute appears to be a valuable method for increasing forecast accuracy. in this study, a new, simple, and effective Gibbs sampler algorithm computing method is derived for posterior distributions. Empirical and simulated data examples demonstrate that our proposed method outperforms the others in all scenarios.

Keyword:Laplace distribution, Composite tobit quantile, new lasso, Bayesian approach

Introduction: In order to describe the nature and structure of the relationship between these variables, modeling the relationship between the censored dependent variable(at zero point) and a set of independent variables is crucial (Li, G., & Wang, Q. H. (2003)). Therefore, when the data are censored at 0, choosing the Tobit regression model is an essential step in producing forecasting and generalizable estimations in the future(Amemiya, T. (1984)). To provide comprehensive coverage and explain all relationships between the censored dependent variable and a set of independent variables, the TQ regression model may be used (Yue, Y. R., & Hong, H. G. (2012)). Variable selection approaches are ways to find the most significant or pertinent variable in a TQ regression model. Consequently, an infinite collection of contour lines at infinitely quantile levels can be estimated by the TQ regression model. As such, choosing the best TQ regression model is seen to be an extremely difficult assignment. Composite quantile regression analysis is a useful tool for overcoming this challenge(Albadiri, A. H. J., & Alhusseini, F. H. H. (2024)).When an estimator obtained using the composite T Q regression model is more efficient than one obtained using the Tobit quantile at a single quantile level. Regardless of the residual distribution, the authors demonstrated that composite TQ regression has a relative efficiency of more than 70% when compared to the OLS method (Alhusseini, F. H. H., & Georgescu, V. (2018)). The most significant regularization technique is called Lasso (Least Absolute Shrinkage and Selection Operator), which selects variables by some coefficients being shrunk to exactly zero(Tibshirani 1996). Due to the Lasso technique's ability to balance the bias and variance of the coefficients, it is possible to obtain precise and broadly applicable estimates when utilizing a T Q regression model (Jacobson, T., & Zou, H. (2024)). Recently, a Bayesian approach has been used to estimate the parameters and perform variable selection for a T Q regression model using the Laplace distribution as prior distribution(Alhusseini, F. H. H., & Georgescu, V. (2018)). In this paper, we will apply a novel modification to the Laplace distribution formula: we will combine the exponential distribution with a uniform distribution to create a new formulation of the Laplace distribution. This new Laplace distribution method, when combined with the composite TQ regression model, gives us a reliable and effective hierarchical approach for choosing variables and estimating parameters in the composite TQ regression model.

Tobit quantile regression model

When it comes to determining the impact of a group of independent variables on a censored dependent variable, the TQ regression model has proven to have strong statistical features(Powell, J. L. (1986)). It gained popularity in a wide range of various applications such as Ecology and environmental studies , finance sciences, medicine and biological studies. In order to estimate quantile regression parameters in the presence of censored data, the TQ regression model is an extension of the traditional Tobit regression model. Compared to the normal Tobit regression model, the T Q regression approach is more robust against outliers and a normal assumption (Cunha, D. R (2022)). Given TQ regression model is defined as follows:

$$y_i = \begin{cases} y_i^* = \alpha_\tau + \beta_\tau x_i^T + u_i & if\ y_i^* > 0 \\ 0 & if\ y_i^* \le 0 \end{cases} \qquad (\ i = 1,2, \dots n) \qquad (1)$$

The censored dependent variable, y_i, has limited information at the censored point, quantitative to zero. The latent variable, y_i^*, has free information that is quantitative to quantitative data (Greene, W. (1999)). Another formula is written as an alternative to equation (1) as follows:
$y_i = \max(0, y_i^*)$, The latent variable, $y_i^* = \alpha_\tau + \beta_\tau x_i^T + u_i$. α_τ is intercept term, $\beta_\tau = (\beta_{1\tau}, \beta_{2\tau}, \dots . \beta_{k\tau})^T$ are vectors of unknown parameters of the T Q regression model, τ is (0,1) as $0 < \theta < 1$ Tobit quantile level. The estimation of T Q regression coefficients is obtained by minimizing the subsequent formula:

$$\underset{\alpha_\tau, \beta_\tau}{min} = \sum_{i=1}^{n} \rho_\tau \left(y_i - max\{0, y_i^*\}\right) \qquad (2)$$

The equation (2) is not differentiable at (0). Therefore, linear programming solves the minimization of equation (2). Numerous algorithms have been proposed to solve equation (2). The Lasso method can be utilized for the estimation of parameters and variable selection in the TQ regression model as the following equation:

$$\underset{\alpha_\tau, \beta_\tau}{min} = \sum_{i=1}^{n} \rho_\tau \left(y_i - max\{0, y_i^*\}\right) + \lambda \parallel \beta_\theta \parallel, \qquad (3)$$

where $\rho_\tau(u)$, $\rho_\tau(u) = u\{\tau - \mathrm{I}(u \le 0)\}, and$ where $I(u < 0)$ is the indicator function and $\lambda(\ (\lambda \ge 0)$ is the shrinkage parameter. Numerous algorithms have also been proposed to solve the equation (3) . G are various Tobit quantile levels $0 < \tau_1 < \tau_2 < \cdots < \tau_G < 1\ and\ g = 1, \dots .., G,. \tau_g = \frac{g}{G+1} for\ g = 1, \dots ., G.$
The estimation of composite Tobit quantile regression (C T Q regression) coefficients is obtained by minimizing the following formula:

$$\left(\hat{\alpha}\tau 1, \hat{\alpha}\tau 2, \dots, \widehat{\alpha\tau}\mathrm{G}, \hat{\beta}\right) = \underset{\hat{\alpha}\tau 1, \hat{\alpha}\tau 2, \dots, \widehat{\alpha\tau}\mathrm{G}, \hat{\beta}}{Min} \sum_{g=1}^{G} \left\{ \sum_{i=1}^{n} \rho_{\tau_g} \left(y_i - \max(0, \alpha_g + x_i^T \beta\right) \right\}, \quad (4)$$

Unfortunately, the equation 4 also is not differentiable at zero. To achieve variable selection and coefficient estimation in C T Q regression as follows:

$$\left(\hat{\alpha}\tau 1, \hat{\alpha}\tau 2, \dots, \widehat{\alpha\tau}\mathrm{G}, \hat{\beta}\right) = \underset{\hat{\alpha}\tau 1, \hat{\alpha}\tau 2, \dots, \widehat{\alpha\tau}\mathrm{G}, \hat{\beta}}{Min} \sum_{g=1}^{G} \left\{ \sum_{i=1}^{n} \rho_{\tau_g} \left(y_i - \max(0, \alpha_g + x_i^T \beta\right) + \lambda \parallel \beta_\theta \parallel \right\}, \qquad (5)$$

Equation 5 is also not differentiable at zero. Therefore, minimizing equation (5) is solved by the Bayesian method (Yu and Stander (2007)).
They propose using the Tobit quantile regression (T Q regression), assuming an asymmetric Laplace distribution for the error distribution. Based on the linear dependence between the random error term and the dependent variable, we find that the dependent variable will be distributed according to the asymmetric Laplace distribution. Therefore, maximizing the likelihood function for below (6), is equivalent to minimizing equation (5)

$$f(y|X, \alpha, \beta) = \prod_{g=1}^{G} (\tau^n (1-\tau)^n\ exp \left\{ - \sum_{i=1}^{n} \rho_\tau (y_i - \max(0, \alpha + x^T \beta_\tau)) \right\} \qquad (6)$$

Because of the mixing of H components, solving Equation (6) directly is exceedingly challenging (Huang and Chen, 2015). We employ a cluster assignment matrix C, wherein the $(i,g)^{th}$ element C_{ig} belongs to one if the $(i)_{it}$ subject is associated with the g_{th} cluster, belong to 0 otherwise. We treat the element C_{ig} as a missing value. Consequently, our probability looks like this:

$$f(y|X, \alpha, \beta) = \prod_{g=1}^{G} \left[(\tau^n (1-\tau)^n\ exp \left\{ - \sum_{i=1}^{n} \rho_\tau (y_i - \max(c, \alpha + x^T \beta_\tau)) \right\} \right]^{C_{ig}} \qquad (7)$$

The direct use of Equation 7 in estimating the parameters of the compound Tobit regression model becomes very difficult, and to overcome this difficulty, the (Kozumi and Kobayashi (2011)) transformations can be used.

$$f(y_i^*|\alpha_g,\beta,z_i)=\prod_{i=1}^{n}\frac{1}{\sqrt{4\pi v_i}}\,e^{\frac{C_{ig}\ \sum_{g=1}^{G}\sum_{1}^{n}\left(T_i^*-\alpha_g-x_i^T\beta_\tau-(1-2\theta_h)v_i\right)^2}{2z_i}} \qquad (8)$$

Where v_i are independent composition, therefore marginalizing over v_i gives us $y_i|\alpha_h,\beta\sim ALD(max\{c,\alpha_g+x_i^T\beta_\tau\},1,\tau_g$

Equation number [8] is considered the first part of the Bayesian approach, which represents the maximum likelihood component. To obtain the posterior distributions, it is necessary to employ the appropriate and suitable prior distribution.

appropriate prior distribution

Tibshirani (1996) provided a comment stating the academics interested in variable selection and users of Bayesian theory that the Laplace distribution is the only distribution that achieves the variable selection task in the Bayesian approach. From that time on, all academics interested in variable selection have focused on the Laplace distribution as the appropriate prior distribution for this purpose. However, the direct use of the Laplace distribution is considered a very difficult task, and it will result in an inefficient and unstable algorithm for estimating the parameters. The Laplace distribution is often represented by researchers as a scale combination of well-known statistical distributions. In this paper, we use one of the transformations of the Laplace distribution, which is a scale mixture of Uniform distribution mixing with standard exponential distribution as in the following proposition; see Mallick and Yi (2014) for more details.

$$f(\beta|\theta,\lambda)=\frac{\theta\lambda}{2}\,e^{-\theta\lambda|\beta|}=\int_{g>\theta\lambda|\beta|}^{,}\frac{\theta\lambda}{2}\,e^{-g}\quad dg \qquad (\mathbf{9})$$

The Laplace transform shown in the equation above is considered a simplified formulation in developing an efficient and good algorithm for estimating the parameters of the model being studied. Therefore ,the , our Bayesian hierarchical model given by

$y_{i=max\{0,y_i^*\}},\qquad i=1,\ldots,n,$

$y_i^*|\alpha_g,\beta,z_i\sim\left[N\left(\alpha_g+x_i^T\beta_\tau+\left(1-2\theta_g\right)v_i,2v_i\right)\right]^{C_{ig}},$

$p(\alpha_g)\propto 1$

$m_i\sim Exp\ \left(\theta_g\left(1-\theta_g\right)\right)$,

$\beta|\ \theta,\lambda\sim$ uniform$\left(-\frac{1}{\theta\lambda},\frac{1}{\theta\ \lambda}\right)$ (10)

$m\sim$ standard exponential

$\theta\sim\quad\theta^{a-1}\exp^{(-b\theta)}$

$\lambda\sim\ \lambda^{c-1}e^{-d\lambda}$

$w\sim$ Dirichlet $\left(d_1,\ldots,d_g\right)$

(a,b,c and d) are hyperparameter

Inference posterior distribution

Estimating the parameters of composite tobit quantile regression models using a Bayesian approach involves multiplying the maximum likelihood function with the appropriate prior distribution to obtain the good posterior distributions. Therefore , multiplying equation (8) by equation (9) and the Bayesian hierarchical model shown in set equation (10). We will obtain the posterior distributions as follows

Let $\Upsilon(.)$ denoted to a degenerate distribution .the latent variable $T_i^*, i=1,\ldots\ldots,n,$ has a full conditional posterior distribution given by

$$y_i^*|y_i,v_i,\alpha_g,\beta_\tau\sim\begin{cases}\{\Upsilon(y_i), & if\ y_i>0;\\ \left\{\prod_{g=1}^{G}\left[N\left(\alpha_g+x_i^T\beta_\tau+(1-2\tau_h)v_i,2v_i\right)\right]^{c_{ig}}\right\}I(y_i^*\le 0), otherwise\end{cases} \qquad [11]$$

-The inference distribution of $\boldsymbol{\beta_j}$ is normal distribution. where

$$\boldsymbol{pdf\ \left(\beta_j|.\right)\ \sim N\left(\frac{\sum_{1=1}^{n}x_{ij}\,h_{1i}}{\sum_{i=1}^{n}x_{ij}^2},\frac{1}{\sum_{i=1}^{n}x_{ij}^2/h_{2i}}\right)}$$

$\boldsymbol{h_{1i}=\sum_{g=1}^{G}\ \left(y_i-\ \alpha_{\tau_g}\ -\ \tau_{1g}\,v_i\right)}$ **, and** $\boldsymbol{h_{2i}=\ \sum_{g=1}^{G}\ \theta^{-1\ \tau_{1q}^2 v_i}}$

-The inference distribution of $\boldsymbol{\alpha_{g_\tau}}$ is normal distribution. where

$$\boldsymbol{\alpha_{g_{\tau}\sim N(\gamma,\delta)}}$$

$$\boldsymbol{\gamma} = \frac{\frac{\sum_{i=1}^{n} c_n\theta\, y^*{}_{ik}/\tau_{2g}^2 v_i}{\sum_{i=1}^{n} c_{ig}/\tau_{2g}^2 v_i}}{\frac{1}{\sum_{i=1}^{n} c_{ig}/\tau_{2g}^2 v_i}} \quad , \ \boldsymbol{\delta} = \frac{1}{\sum_{i=1}^{n} c_{ig}\theta/\, Q_{2g}^2\theta v_i}$$

-The inference distribution of C_i is multinomial distribution. where

p $\left(c_{ig}\middle|.\right) \propto$ **multinomial distribution** $\left(1, \hat{p}_1, \ldots, \hat{p}_g\right)$,

where

$$\hat{p}_g = \frac{\frac{w_g}{G_{2g}}\left[exp\left(-\left(y_i - \alpha_{\tau_g} + x_i^T\beta_\tau - \tau_{1q} v_i\right)^2 \middle| 2\,\theta^{-1}\ \tau_{1g}^2 v_i\right)\right]}{\sum_{g=1}^{G}\frac{w_g}{Q_{2g}}\left[exp\left(-\left(y_i - \alpha_{\tau_q} + x_i^T\beta_\tau - \tau_{1q} v_i\right)^2 \middle| 2\,\theta^{-1}\ \tau_{1q}^2 v_i\right)\right]}$$

Where $\sum_{g=1}^{G} \hat{p}_h = 1$

Then $c_i = (c_{i1}, c_{i2} \ldots, c_{iG})^T$ *is Multinomial distribution*

-The inference distribution of w is Dirichlet distribution. where

$\propto \prod_{g=1}^{G} w_q^{ng+dg} \propto$ **Dirichlete (n_1+d_1, … ,$n_g + d_g$).**

-The inference distribution of θ is inverse gamma distribution ,where

$$\propto gamma\left(\frac{3n}{2} + a + 1, \left[\frac{exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{g=1}^{G} c_{ig}\ \left(y_i - \alpha_{\tau_g} + x_i^T\beta_\tau - \tau_{1q} v_i\right)^2\ \ \sum_{i=1}^{n} v_i + b\ \right\}\theta}{\theta^{-1}\ \varphi_{2q}^2 v_i}\right]\right)$$

-The inference distribution of λ is gamma distribution

with shape parameter $(a+1)$ and scale parameter ($\left(b + \frac{V_i}{2}\right)$) . $\lambda_j \sim Gam\left[(a+1), \left(b + \frac{v_i}{2}\right)\right]$

We can derive an effective and straightforward Gibbs sampler with a good and efficient algorithm for variable selection and parameter estimation of the composite Tobit regression model from the posterior distributions presented above.

Simulation approach

Through simulation studies, the effectiveness of our proposed method composite TQ regression model. (C.T.Q Regs) is examined. Through simulation studies, the performance of our proposed method composite TQ regression model (C.T.Q Regss) is examined. We compare our proposed model with two current approaches: the Bayesian composite tobit quantile regression reported in (Fadel Al Husseini (2018)) symbolizes it (CTQReg) and Bayesian adaptive elastic net T.Q Regss reported in (Alhamzawi(2014)) symbolizes it (BAnet). Four options for quantile levels were used: $\tau = 0.20, 0.40, 0.60$and 0.80. For every simulation study, therefore G=4 then quantile levels are $\frac{g}{G+1}$. Three random error distributions can be used are normal distribution $u \sim N(0,25)$, chi-square distribution with 5-degree freedom $u \sim \chi_{(5)}^2$ and t-student with 5-degree freedom $u \sim t_{(5)}$. The first 2000 iterations of our proposed method were eliminated as burn-in, that it runs for 12000 iterations. The mean absolute error (MAE) and standard division (S.D) are used to evaluate the approaches under study. MAE is computed via the following formula

$$MAE = \frac{\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|}{r}$$

Five sample size (50,100,150,200,250) have been used. With two Simulation scenarios, sparse case and dense case :

Simulation one example

In this simulation we will use sparse case**,** the tobit quantile regression with sparse case is defined with the following model **:**

$$y_i = \begin{cases} y_i^* = \alpha_\tau + \beta_\tau x_i^T + u_i & if\ y_i^* > 0 \\ 0 & if\ y_i^* \le 0 \end{cases}$$

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

$y_i^* = 5 + 3x_{1i} + 1x_{3i} + 1x_{6i} + u_i \qquad i = 1,2, \ldots .100$

We generated seven explanatory variables through a multivariate normal distribution $\underline{X} \sim MN\left(\underline{0}, \Sigma_X\right), Where\ \Sigma_X = (\frac{1}{2})^{|i-j|}$. In this simulation approach, the Vector initial parameters are $\beta = (5,3,0,1,0,0,1,0)$ with an initial intercept term

Table -1- MAE and SD for simulation one example

n	τ	BAnet	CTQReg G=4	C.T.Q Regss G=4
50	$\tau = 0.20$	1.830 (0.472)	1.291 (0.410)	**0.780 (0.387)**
	$\tau = 0.40$	1.691 (0.585)		
	$\tau = 0.60$	1.784 (0.762)		
	$\tau = 0.80$	1.781 (0.687)		
100	$\tau = 0.20$	1.982 (0.765)	1.268 (0.472)	**0.759 (0.381)**
	$\tau = 0.40$	1.782 (0.692)		
	$\tau = 0.60$	1.579 (0.572)		
	$\tau = 0.80$	1.762 (0.672)		
150	$\tau = 0.20$	1.643 (0.677)	1.103 (0.475)	**0.617 (0.278)**
	$\tau = 0.40$	1.634 (0.623)		
	$\tau = 0.60$	1.732 (0.582)		
	$\tau = 0.80$	1.605 (0.452)		
200	$\tau = 0.20$	1.427 (0.485)	1.095 (0.385)	**0.549 (0.282)**
	$\tau = 0.40$	1.284 (0.241)		
	$\tau = 0.60$	1.176 (0.375)		
	$\tau = 0.80$	1.130 (0.352)		
250	$\tau = 0.20$	1.242 (0.434)	1.101 (0.714)	**0.523 (0.268)**
	$\tau = 0.40$	1.184 (0.372)		
	$\tau = 0.60$	1.023 (0.357)		
	$\tau = 0.80$	1.006 (0.318)		

In the parentheses are SD values

The table above displays the results of our proposed method. It is evident from these that the mean absolute error (MAE) and standard deviation calculated using our method are significantly lower than those calculated using the comparison methods employed in this investigation. This result leads us to the conclusion that, we can conclude that our proposed method has good performance in variable selection and parameter estimation for the composite Tobit quantile regression model, according to the Bayesian approach.

Simulation two example

In this simulation, we will use dense case**,** the tobit quantile regression with dense case is defined with the following model **:**

$$y_i = \begin{cases} y_i^* = \alpha_\tau + \beta_\tau x_i^T + u_i & if\ y_i^* > 0 \\ 0 & if\ y_i^* \leq 0 \end{cases}$$

$y_i = \max(o, y_i^*)$
$\mathrm{y}_\mathrm{i}^* = 5 + 0.85\mathrm{x}_{1\mathrm{i}} + 0.85\mathrm{x}_{2\mathrm{i}} + 0.85\mathrm{x}_{3\mathrm{i}} + 0.85\mathrm{x}_{4\mathrm{i}} + 0.85\mathrm{x}_{5\mathrm{i}} + 0.85\mathrm{x}_{6\mathrm{i}} + 0.85\mathrm{x}_{7\mathrm{i}} + \mathrm{u}_\mathrm{i}$ $\quad$ $\mathrm{i} = 1,2, \ldots .100$
We generated seven explanatory variables through a multivariate normal distribution $\underline{X} \sim MN(\underline{0}, \Sigma_X), Where\ \Sigma_X = (\frac{1}{2})^{|i-j|}$. In this simulation approach, the Vector initial parameters are $\beta = (5,0.85,0.85,0.85,0.85,0.85,0.85,0.85)$ with an initial intercept term

Table -2- MAE and SD for simulation one example

n	**τ**	BAnet	CTQReg G=4	C.T.Q Regss G=4
50	$\boldsymbol{\tau = 0.20}$	1.934 (0.653)	1.187 (0.658)	**0.883 (0.419)**
	$\boldsymbol{\tau = 0.40}$	1.711 (0.745)		
	$\boldsymbol{\tau = 0.60}$	1.723 (0.762)		
	$\boldsymbol{\tau = 0.80}$	1.634 (0.672)		
100	$\boldsymbol{\tau = 0.20}$	1.756 (0.865)	1.114 (0.435)	**0.732 (0.472)**
	$\boldsymbol{\tau = 0.40}$	1.653 (0.584)		
	$\boldsymbol{\tau = 0.60}$	1.673 (0.467)		
	$\boldsymbol{\tau = 0.80}$	1.552 (0.522)		
150	$\boldsymbol{\tau = 0.20}$	1.562 (0.677)	1.008 (0.381)	**0.693 (0.491)**
	$\boldsymbol{\tau = 0.40}$	1.573 (0.623)		
	$\boldsymbol{\tau = 0.60}$	1.533 (0.607)		
	$\boldsymbol{\tau = 0.80}$	1.532 (0.451)		
200	$\boldsymbol{\tau = 0.20}$	1.652 (0.652)	1.102 (0.329)	**0.546 (0.452)**
	$\boldsymbol{\tau = 0.40}$	1.482 (0.572)		
	$\boldsymbol{\tau = 0.60}$	1.362 (0.562)		
	$\boldsymbol{\tau = 0.80}$	1.452 (0.352)		
250	$\boldsymbol{\tau = 0.20}$	1.173 (0.342)	1.236 (0.452)	**0.467 (0.373)**
	$\boldsymbol{\tau = 0.40}$	1.105 (0.424)		
	$\boldsymbol{\tau = 0.60}$	0.974 (0.352)		

	$\tau = 0.80$	0.806 (0.456)		

In the parentheses are SD values

The table above displays the results of our proposed method. It is evident from these that the mean absolute error (MAE) and standard deviation calculated using our method are significantly lower than those calculated using the comparison methods employed in this investigation. This result leads us to the conclusion that, we can conclude that our proposed method has good performance in variable selection and parameter estimation for the composite Tobit quantile regression model, according to the Bayesian approach.

Real data

In order to compare the proposed method for Bayesian C.T.Q Regss with previous techniques, a medical dataset will be used. Out of the 75 observations in the sample size, 24 are censored at zero in this medical dataset. In this paper, we focus on studying the effect between the censored dependent variable, which represents Interleukin-10 (IL-10). The response variable Y:is a Censored variable(IL-10) that define

$$(Y)= \begin{bmatrix} y^* & if y^* > 46.6 \\ 0 & if y^* \leq 46.6 \end{bmatrix}$$

The sixteen independent variables are mean corpuscular hemoglobin (MCH),hematocrit (HCT),mean corpuscular volume(MCV),hemoglobin, red blood cells,platelet(PLT),neutrophils(NEU),monocytes(MON),lymphocytes(LYM),White Blood cells (WBC),C-X-Cmotifchemokine ligand 13(CXCL13),interferon (IFN), Interleukin-27 (IL-27) , Interleukin-17 (IL-17). The same simulation conditions used in the current study and the same model structure will be applied to a four-level quantile are ($\boldsymbol{\tau = 0.20, \tau = 0.40, \tau = 0.60, \tau = 0.80}$). Therefore, the G=4

Table -3- MAE and SD for Medical dataset

τ	BAnet	CTQReg G=4	C.T.Q Regss G=4
$\tau = 0.20$	1.274 (0.463)	1.054 (0.532)	**0.745 (0.592)**
$\tau = 0.40$	1.542 (0.475)		
$\tau = 0.60$	1.1562 (0.534)		
$\tau = 0.80$	1.135 (0.452)		

In the parentheses are SD values

The table above displays the results of our proposed method. It is evident from these that the mean absolute error (MAE) and standard deviation calculated using our method are significantly lower than those calculated using the comparison methods employed in this investigation. This result leads us to the conclusion that, we can conclude that our proposed method has good performance in variable selection and parameter estimation for the composite Tobit quantile regression model, according to the Bayesian approach. After we have demonstrated that our proposed method has clear importance in variable selection and parameter estimation for the studied regression model, we will rely on it to estimate the parameters of the studied model.

Table-4-shown the point estimation and confidence interval for parameter of estimation

Variables	Upper bound	Point estimation	Lower bound
MCH	0.647	0.135	0.076
HCT	0.683	0.184	0.124
MCV	0.163	0.000	-0.154
HGB	0.851	0.487	0.292
RBC	0.248	0.062	0.008
PLT	0.170	0.102	0.086
NEU	0.693	0.270	0.156
MON	0.217	0.152	0.094
LYM	0.106	0.087	0.009
WBC	0.371	0.185	0.027
CXCL13	0.128	0.000	-0.284
IFN	0.419	0.215	0.134
IL27	0.697	0.361	0.164
IL17	0.319	0.112	0.097
AGE	0.238	0.210	0.117
SEX	0.034	0.000	-0.342

From the results presented in the table above, we find that there are three independent variables that do not have any effect or importance on the response variable. Therefore, we can exclude these variables from the structure of the studied model, as they have very weak explanatory power due to coefficient values being exactly zero. However, the rest of the variables had a high explanatory power, and these remaining independent variables are considered the main variables in the construction of our studied model.

Conclusions and Future Research

Based on the results obtained from the simulation aspect, we conclude that our proposed method (C.T.Q Regss) distinguishes itself from the comparative methods, as it demonstrated good performance compared to the two previous methods. We can also observe that our proposed method exhibits a directly proportional relationship with the increase in sample size. Additionally, our proposed method was also effective and outperformed the previous comparative methods when applied to real-world data. From observing the estimated values of the coefficients for the independent variables, we find that our proposed method successfully achieved its objective in estimating the model parameters and selecting the variables. Specifically, our proposed method was able to set the coefficients of three independent variables to close to zero, effectively excluding these variables from the final predictive model of the relationship between the set of independent variables and the response variable. We can expand our current study into promising future studies with distinguished results by utilizing other organizational tools such as the Elastic Net or Adaptive Lasso, and so on.

References

Albadiri, A. H. J., & Alhusseini, F. H. H. (2024). Bayesian new lasso composite quantile regression with updated Hierarchical. المجلة العراقية للعلوم الاقتصادية, 22(80. S), 1247-1258.

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

Amemiya, T. (1984). Tobit models: A survey. Journal of econometrics, 24(1-2), 3-61.

Cunha, D. R., Divino, J. A., & Saulo, H. (2022). On a log-symmetric quantile tobit model applied to female labor supply data. Journal of Applied Statistics, 49(16), 4225-4253.

Greene, W. (1999). Marginal effects in the censored regression model. Economics Letters, 64(1), 43-49.

Jacobson, T., & Zou, H. (2024). High-dimensional censored regression via the penalized tobit likelihood. Journal of Business & Economic Statistics, 42(1), 286-297.

Koenker, R. (2005). Quantile Regression. Cambridge Books. Cambridge University Press.

Li, G., & Wang, Q. H. (2003). Empirical likelihood regression analysis for right censored data. Statistical Sinical, 51-68.

Powell, J. L. (1986). Censored regression quantiles. Journal of econometrics, 32(1), 143-155.

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

Yue, Y. R., & Hong, H. G. (2012). Bayesian TQ regression model for medical expenditure panel survey data. Statistical Modelling, 12(4), 323-346.