CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES
https://cajmtcs.centralasianstudies.org/index.php/CAJMTCS
Volume: 06 Issue: 04 | October 2025 ISSN: 2660-5309

Article

Study Compression with Different New Prior distribution in Tobit Quantile Regression

Abbas Shaker Mohsen Al-Jashami[1], Fadel Hamid Hadi Alhuseeni[2]

1. Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Al-Qadisiyah, Iraq
2. Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Al-Qadisiyah, Iraq

* Correspondence: Shakrbas79@gmail.com, fadel.alhusiny@qu.edu.iq

Abstract: Bayesian estimation requires sampling from the posterior distributions. Where, the prior distributions are play a vital role in obtaining the simplifying the derivation of full conditional distributions, making Gibbs sampling algorithms more efficient. using the Laplace prior distribution (also known as the Double Exponential prior) is indeed a great choice in Bayesian Tobit quantile regression for both variable selection and parameter estimation simultaneously. The Laplace prior has become popular in regression models because of its ability to induce sparsity in the estimated coefficients, which is particularly beneficial for variable selection. However, directly using the Laplace prior distribution is a very complex task when building the hierarchical model. To overcome this issue, a set of transformations of the Laplace prior distribution has been used, which provide us with hierarchical models with more efficiency. In this paper, we will compare the transformations of the Laplace prior distribution that provide us with efficient estimators capable of generalization.

Keywords: Prior Distribution, Tobit Quantile Regression, Bayesian

Citation: Al-Jashami, A. S. M & Alhuseeni, F. H. H. Study Compression with Different New Prior distribution in Tobit Quantile Regression. Central Asian Journal of Mathematical Theory and Computer Sciences 2025, 6(4), 796-809

Received: 10th May 2025
Revised: 16th Jun 2025
Accepted: 24th Jul 2025
Published: 11th Aug 2025

1. Introduction

Tobit quantile regression (TQR) model is particularly useful when dealing with censored response variables while analyzing their conditional quantiles. Where, TQR model provides a more detailed view of the distribution of the dependent variable by estimating relationships at different quantile levels $\tau\epsilon(0,1)$. (TQR)model keep of main features of both the Tobit regression model ($y = \max(0, y_i^*)$, $y_i^* = x\beta + \epsilon_i$, $i = 1,2,\dots\dots.n$), which focuses on censoring, and the quantile regression model, which focuses the estimation of conditional quantiles $Q_{y_{i|x_i}}(\tau) = F^{-1}_{y_{i|x_i}}(\tau)$. Recently, the regularization process has become popular with regression models [1], [2]. Now, combining regression models with these methods provides us with good models for parameter estimation and high predictive capability. Some Tobit Quantile Regression models include numerous explanatory variables, each exhibiting different relationships with the left-censored response variable [3], [4], [5]. However, some of these variables contribute minimally to the model, making their inclusion unjustified. Identifying weak explanatory variables in this context is particularly challenging. Variable selection (VS) is a valuable statistical tool for addressing this issue. It plays a crucial role in constructing regression models by effectively identifying relevant explanatory variables while excluding irrelevant ones [6], [7]. Recently, researchers have introduced innovative methods to enhance the implementation of variable selection in regression models. These methods offer desirable properties and perform variable selection efficiently, as the process is automated and requires minimal time. One of the most important methods is LASSO (Least Absolute Shrinkage and Selection Operator). Researchers in the field of variable selection under Bayesian theory have primarily focused on the Laplace prior distribution due to its advantageous properties. This distribution aids the variable selection process by shrinking the estimates of insignificant explanatory variables to exactly zero. In this paper will involve an analysis and comparison of three distinct transformations applied to the prior Laplace distribution. We aim to determine which transformation is the most effective regarding efficiency and practical applicability [8], [9]. First transformation is Scale Mixture of Normals (SMN) that proposed by ((Andrews and Mallows)). Many researchers used this transformations with hierarchical Bayesian tobit quantile regression model such as, (Park & Casella), (R.Alhamzawi) and Fadel Hamid Hadi Alhusseini. Second transformation is Scale Mixture of uniform (SMU) that proposed by(Mallick, H. and Yi, N). The our paper has been organized as follows:

in Section 2, we present tobit quantile regression model briefly .In Section 3, we present the complete hierarchical Bayesian model, detailing likelihood specifications and prior distributions for each parameters [10]. Section 4, derives the full conditional posterior distribution, in section 5 ,introduced application side, including simulation studies and real-world data. In section 6 concludes important conclusion and future papers .

Tobit Quantile Regression

The Tobit model introduced by (James Tobin in), is a type of regression model specifically designed for dealing with right - or left-censored dependent variables. It is especially helpful when the data is censored at a specific threshold (usually zero), indicating that values past that point are not observed or constrained. The Tobit model is dealing with latent variable y_i^* which is based on the following linear regression model:

$$y_i = \max(y_i^*, 0) \qquad (1)$$

Where y is censored independent variable (observed variable), y_i^* is (latent variable) it have following mathematical model:

$$y_i^* = X\beta + u_i \text{ , } (i = 1,2, \dots . . n) \qquad (2)$$

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

The latent variable y_i^* is observed $if\ y_i^* > 0\ in\ this\ case\ \ y_i = y_i^*$ and the latent variable y_i^* isn't observed $if\ y_i^* \leq 0\ in\ this\ case\ \ y_i = 0$ (**Amemiya, T. (1984)**). Through its specifications, the Tobit regression model is highly flexible and useful in handling data constrained at the zero threshold. Where, this model accurately describes the impact relationship between the constrained response variable and a set of independent variables(Cameron, A. C., & Trivedi, P. K). The Tobit regression model depend on more normality assumptions for consistency and robust inference. These assumptions of normality are linked to the framework of latent variables and the structure of random errors, such as random errors is distributed normal distribution $u_i \sim N(0, \sigma^2)$, Homoskedasticity, autocorrelation and the Tobit regression model is highly sensitive to outliers values (Yohai, V. J., et al), and other problems. Tobit quantile regression (Powell) is a sophisticated statistical method aimed at overcoming the shortcomings of the conventional Tobit regression model. Tobit Q Reg offers a more comprehensive view by estimating the link between the censored response variable and a collection of explanatory variables across various quantile levels. Due to its robust modeling of censored data and ability to capture heterogeneous effects across quantiles, Tobit Quantile Regression has gained significant traction in various disciplines. , Such as Economics and Labor Studies(Buchinsky, M), Medical Expenditures (Kowalski, A), Survival Analysis (Portnoy, S) and Environmental Science (Wang, H. J., and Fygenson, M) etc., the Tobit Q Reg can be expressed as:

$$y_i = \max(0, y_i^*), \quad where\ y_i^* = x_i^T \beta_\tau + u_i \quad , Q_\tau \left({y_i^*}/{x_i^T} \right) = x_i^T \beta_\tau \qquad (3)$$

where

y_i^* is latent response variable (unobserved variable). x_i is vector of independent variables .β_τ is the vector of regression coefficients for specific-quantiles .$Q_\tau \left({y_i^*}/{x_i^T} \right)$ is conditional quantile function with level τ, $\tau \in (0,1)$ .u_i denotes the random error satisfying $Q_\tau\left({u_i}/{x_i^T}\right) = 0$ Chernozhukov, V., and Hong, H. The estimation of Tobit quantile regression coefficients requires the minimization of a weighted check function, which integrates the quantile regression model with the censoring mechanism of the Tobit model as following:

$$\min_{\beta_\tau} = \sum_{i=1}^{n} \rho_\tau\left(y_i - max\{0, y_i^*\}\right) \text{ ,} where\ y_i^* = x_i^T \beta_\tau + u_i \quad (4)$$

where $\rho_\tau(u)$ is called loss (check)function of (Koenker and Bassett). Unfortunately, the mathematical formula representing the loss function shown in Equation (4) is non-differentiable at zero. Therefore, minimization of equation (4) can be implemented via the linear programming approach is proposed by the (Koenker and D'Orey). in high-dimensional data the number of independent variables (p) is very large , the variable selection is becomes hard matter for interpretable , in tobit quantile regression , several penalization methods can be used to perform shrinkage and selection of relevant independent variables. The most important method used to regularize the Tobit quantile regression model is the LASSO method, which was proposed by the researcher (Tibshirani in). The penalized tobit quantile regression loss is

$$\min_{\beta_\tau} = \sum_{i=1}^{n} \rho_\tau\left(y_i - max\{0, y_i^*\}\right) + \lambda \parallel \beta_\tau \parallel \quad (5)$$

The term of $\lambda \parallel \beta_\tau \parallel$ represents the penalization penalty function is achieving the coefficients estimation and variable selection at same time , and λ is shrinkage parameter that controlled the Tobit quantile regression coefficients. Also , the equation (5) is non-differentiable at zero. The Bayesian approach incorporates prior distributions and likelihood functions to estimate tobit quantile regression coefficients and its variable selection,

2. Materials and Methods

Hierarchical Bayesian Method

One of the key ideas in Bayesian statistics is the conditional posterior distribution $f(\beta|y)$. The posterior distribution $f(\beta|y)$ provides us with comprehensive details regarding parameter estimate. The conditional posterior distribution is proportional to the product of likelihood function $f(y|\beta)$ and prior distribution $g(\beta)$. Therefore, the mathematical formula(Gelman et al):

$$f(\beta|y) \propto f(y|\beta)\, g(\beta) \quad (6)$$

$g(\beta)$ is gives us an idea of the estimated parameters β before the data observation , $f(y|\beta)$ is the likelihood function how credible the observed data y is for different values of the parameter

$f(\beta|y)$ is The final result of Bayesian analysis which, it gives a complete probabilistic description of β. In this paper we will focus on Bayesian framework for estimating Tobit quantile regression models with two sections its likelihood function for and Laplace prior distribution [11], [12].

Information of Likelihood Function

The loss function(check function) and the skew Laplace distribution (SLD) are equivalent in tobit quantile regression model(Yu and Moyeed). In tobit quantile regression the random error u_i belong to skew Laplace distribution . It is define with probability density function (pdf) as following:

$$f(u_i|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp{-\rho_\tau}\left\{\left(\frac{u_i - \mu}{\sigma}\right)\right\} \quad (7)$$

when μ *and* σ are equal 0 and 1 respectively, then (pdf) of u_i is become

$$f(u_i|\mu) = \tau(1-\tau) \exp{-\rho_\tau}\{(u_i)\} \quad (8)$$

$\rho_\tau(.)$ is the check function defined as in equation [4], The joint distribution of $y = (y_i)^T$ $given\ X = (x_i)^T$, $i = 1,2,\ldots..n$ is .

$$(y|X,\beta,\tau) = \tau^n(1-\tau)^n exp\left\{-\sum_{i=1}^{n}\rho_\tau(y_i - max\{0, x_i^T\beta_\theta + u_i\}\right\} \quad (9)$$

the equivalence between maximizing the likelihood function in a tobit quantile regression model (under the SLD) and minimizing the check function of tobit quantile regression is fundamental. However, it is noted that directly using the asymmetric Laplace distribution (ALD) leads to inefficient and computationally complex algorithms for estimating the parameters. To overcome this problem, transformation methods can be employed ,that proposed by (Kozumi and Kobayashi) [13].

In order to enable effective Gibbs sampling in Bayesian tobit quantile regression, (Kozumi and Kobayashi) showed that the (SLD) can be described as a scale mixture of normal (SMN) distributions.

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

$$y_i^* = x_i^T\beta_\tau + (1-2\tau)w_i + \sqrt{2w_i}\,\varepsilon_i \quad (10)$$

where w_i is belong to experiential distributed with rate parameter $\tau(1-\tau)$, ε_i is belong to standard normal. Therefore , the latent variable $y_i^* \sim N(x_i^T\beta_\tau + (1-2\tau)w_i, 2w_i)$ via this information the (p.d.f) of latent variable y_i^* is

$$f(y_i^*|x_i^T,\tau,\beta_\tau,w_i) = \frac{1}{\sqrt{4\pi w_i}} e^{\frac{-\left(y_i^* - x_i^T\beta_\tau - (1-2\tau)w_i\right)^2}{4w_i}} \quad (11)$$

Then the Likelihood Function of tobit quantile regression according to proposed Koizumi and Kobayashi (2011) is become as following:

$$f(y_i^*|x_i^T,\tau,\beta_\tau,w_i) = \left[\frac{1}{\sqrt{4\pi w_i}}\right]^n e^{-\sum\frac{\left(y_i^* - x_i^T\beta_\tau - (1-2\tau)w_i\right)^2}{4w_i}} \quad (12)$$

The above equation is very important for estimating the coefficients in tobit quantile regression

Prior of Laplace Distribution

Variable selection(V.S) is a major step in statistical modeling .Where, aiming to determine the most relevant independent variable for a regression model while exclude irrelevant . The aim (V.S) is to improve regression model performance, , generalization and interpretability, also reducing overfitting. Tibshirani noted for the researchers in variable selection within the Bayesian approach that the suitable prior distribution is a Laplace distribution. The Laplace prior distribution is defined as following formula:

$$p(\beta_j|\lambda) = {}^{\lambda}\!/_{2}\ exp\{-\lambda|\beta_j|\} \quad (13)$$

However, directly using a Laplace prior distribution leads to inefficient and unstable Gibbs sampling algorithms when constructing full posterior distributions. To overcome this problem, a set of Laplace transforms can be used, as demonstrated below:

Scale Mixture of Normal (SMN)

Andrews and Mallows discuss (SMN) distributions, that provide us a flexible method to generate distributions from the normal distribution. One main result is that the Laplace distribution can be expressed as a (SMN), when the mixing distribution is exponential distribution. Therefore, the (SMN) is take the mathematical formula

$$\frac{\lambda}{2}e^{-\lambda|\beta_j|} = \int_0^\infty \frac{1}{\sqrt{2\pi v_j}} e^{\frac{-\beta^2}{2v_j}}\ \frac{\lambda^2}{2} e^{\frac{-\lambda^2}{2}v_j}\ dv_j \quad (14)$$

The right side of the equation above represents (SMN) which consists of two parts, first part is belong to normal distribution with mean 0 and latent variance v_j as following: $\frac{1}{\sqrt{2\pi v_j}} e^{\frac{-\beta^2}{2v_j}}$ and second part is belong to exponential distribution that have rate parameter $\frac{\lambda^2}{2}$. In equation (14) the integral combines between two parts, for prove the equation (14) we will change some variables let $t = v_j$,therefore the integral becomes:

$$\frac{\lambda^2}{2\sqrt{2\pi}}\int_0^\infty t^{-\frac{1}{2}}\ exp\left(\frac{-\beta^2}{2t} - \frac{\lambda^2}{2}t\right)\ dt \qquad (15)$$

We will the modified second kind ($K_V(Z)$) of Bessel function. where, $s = -\frac{1}{2}$,a=$\frac{\beta^2}{2}$ and b=$\frac{\lambda^2}{2}$the equation (15) is become $\int_0^\infty t^{s-1}\ exp\left(\frac{-a}{t} - \ bt\right)\ dt = 2\left(\frac{a}{b}\right)^{\frac{v}{2}} K_V\left(2\sqrt{ab}\right)$ from identity $K_{-\frac{1}{2}}(Z) = K_{\frac{1}{2}}(Z) = \sqrt{\frac{\pi}{2z}}e^{-z}$, by steps of integral ,we will obtained $\int_0^\infty t^{-\frac{1}{2}}\ exp\left(\frac{-\beta^2}{2t} - \frac{\lambda^2}{2}t\right)\ dt = \frac{2\pi}{\sqrt{2}\lambda|\beta_j|}e^{-\lambda|\beta_j|}$ more detail see (Andrews and Mallows (1974)) and Olver, F. W. (Ed.). (2010).

$$\int_0^\infty \frac{1}{\sqrt{2\pi v_j}} e^{\frac{-\beta^2}{2v_j}}\ \frac{\lambda^2}{2}e^{\frac{-\lambda^2}{2}v_j}\ dv_j = \frac{\lambda}{2}e^{-\lambda|\beta_j|}$$

Scale Mixture Uniform (SMU)

(Mallick and Yi) are proposed another transformation formula of Laplace prior distribution as shown

$$\frac{\lambda}{2}e^{\{-\lambda|\beta_j|\}} = \int_{v_j>|\beta_j|}^\infty \frac{1}{2v_j}\frac{\lambda^2}{\Gamma(2)}\ {v_j}^{2-1}\ exp\{-\lambda v_j\}\, dv_j \qquad (16)$$

within the integral is a product of two parts, first part is $\frac{1}{2v_j}$ uniform distribution -like term for parameter of (β_j) conditional on v_j. The second part is $\frac{\lambda^2}{\Gamma(2)}\ {v_j}^{2-1}\ exp\{-\lambda v_j\}$ is special case of Gamma distribution for v_j when gamma is 2. Let β_j given v_j is belong to Conditional Uniform Distribution with interval$[-v_j, v_j]$ .Therefore, $f\{\beta_j|v_j\} = \frac{1}{2v_j}\mathrm{I}(|\beta_j| \le v_j)$, $\mathrm{I}(.)$ is indicator function. The v_j is special case of Gamma distribution with shape parameter 2 and rate parameter λ $f\{v_j|\lambda\} = \frac{\lambda^2}{\Gamma(2)}$ ${v_j}^{2-1}\ exp\{-\lambda v_j\} = \lambda^2 v_j exp\{-\lambda v_j\}$,where $\Gamma(2) = (2-1)! = 1$.

$$f\{\beta_j|\lambda\} = \int_{v_j>|\beta_j|}^\infty \frac{1}{2v_j}\frac{\lambda^2}{\Gamma(2)}\ v_j\ exp\{-\lambda v_j\}\, dv_j$$

$$f\{\beta_j|\lambda\} = \frac{\lambda^2}{2}\int_{v_j>|\beta_j|}^\infty exp\{-\lambda v_j\}\, dv_j$$

$$\int_{v_j>|\beta_j|}^\infty exp\{-\lambda v_j\}\, dv_j = \frac{exp\{-\lambda|\beta_j|\}}{\lambda} = \frac{\lambda^2}{2}\frac{exp\{-\lambda|\beta_j|\}}{\lambda} = \frac{\lambda}{2}exp\{-\lambda|\beta_j|\}$$

3. Results and Discussion

Conditional Posterior Distribution with SMN prior

From the equation in(12) and the equation (14), we will obtained the Conditional Posterior Distribution with SMN prior. our Bayesian hierarchical model consists of multiple levels of randomness distributions. To summary , our Bayesian hierarchical given by

$y_{i=max\{0,y_i^*\}},\quad i=1,\dots,n,$ censored at 0 $if y_i^* > 0$

$y_i^* \sim N(x_i^T\beta_\tau + (1-2\tau)w_i, 2w_i)$

$w_i \sim Exp\ (\tau(1-\tau))$,

$\beta_j \sim N(v_j)$

$V_j \sim Exp\left(\frac{\lambda_j}{2}\right)$,

$\lambda_j \sim Gamma(a, b)$

$a\ and\ b$ are fixed hyperparameters

From the our Bayesian hierarchical show in the above ,we will obtain a good Gibbs sampler, it produced efficient algorithm via the following conditional posterior distribution

Application Said

We use Tobit quantile regression (Toqureg) model to analyze the liquefaction data by two methods as shown in theoretical section. The liquefaction data are collecting from Diwaniyah hospital [15]. The sample size of our data 130 observations .The response variable is known according to following:

$$y_i = \begin{cases} y_i^* & if\ y_i^* > 15 \\ 0 & if\ y_i^* \le 15 \end{cases}$$

The response variable y_i is censored at zero point, if the time of Liquefaction more than 15 minute the response variable is positive values ,but when the time of Liquefaction less than 15 minute the response variable is zero values. There are 20 independent variables are: x_1 is (Ph Basic and Acidic), x_2 is (Undescended testicle), x_3 is (Sperm Antibodies), x_4 is (Count Sperm) , x_5 is (Active Sperm Count) , x_6 is (Alcohol Consumption) , x_7 is (Weak Motility Sperm Count) , x_8 is (Dead Sperm Count) , x_9 is (Amount of Testosterone), x_{10} is (Amount of Polactin), x_{11} is (Blood group) , x_{12} is (Physical stimulants) , x_{13} is (Smoking) , x_{14} is (varicocele) , x_{15} is (R.B.sugar) , x_{16} is White Blood Cell (WBC) , x_{17} is (Weight) , x_{18} is Erythrocyte Sedimentation Rate (ESR) , x_{19} is Procalcitonin (**PCT**) , x_{20} is (Testicular abscess). In this part, we compare two Bayesian lasso tobit quantile regression methods with different prior distribution formulations (New Bayesian Lasso in Tobit Quantile Regression(new B Tobit Q Reg, Scale Mixture of Normal (SMN))). To assess the performance of the methods under study, the mean squared error (MSE) are computed at tobit quantile levels $\tau\epsilon(0.15\ ,0.35,0.55,0.75\ and\ 0.95$. The mean squared error results are presented in Table.1

Table 1. Mean squared error (MSEs) for the liquefaction time data.

Methods	$\tau = 0.15$ MSE	$\tau = 0.35$ MSE	$\tau = 0.55$ MSE	$\tau = 0.75$ MSE	$\tau = 0.95$ MSE
SMN	0.6248	0.6742	0.7614	0.8537	0.9721
new B Tobit Q Reg	0.5147	0.5938	0.7218	0.8009	0.9214

From the results listed in above table, The new B Tobit Q Reg method demonstrates a lower mean squared error relative to the other method, which reflects its a good performance. Based on this result, we find that this method outperformed the comparison method. Therefore, we will adopt this approach for both point estimation and confidence interval in 90% estimation at tobit quantile levels $\tau\epsilon(0.15\ ,0.35,0.55,0.75\ and\ 0.95$of the independent variables, as shown below.

At first tobit quantile level ($\boldsymbol{\tau = 0.15}$)

The parameters of the Tobit regression model for the liquidity data can be estimated as follows

Table 2. Show point estimates and 95% confidence intervals for the 0.15 tobit quantile level by (new B Tobit Q Reg) method of the liquefaction data.

Variable name	Variable Simple	$\hat{\boldsymbol{\beta}}$	Lower bound	Upper bound
--------------	Intercept	1.525	0.659	3.837
Ph Basic and Acidic	x_1	0.082	-0.110	0.016
Undescended testicle	x_2	0.000	-0.007	0.009
Sperm Antibodies	x_3	0.000	-0.018	0.081

Count Sperm	x_4	0.859	6.197	1.343
Active Sperm Count	x_5	0.804	0.238	0.915
Alcohol Consumption	x_6	0.000	-0.051	0.091
Weak Motility Sperm Count	x_7	0.775	0.531	1.422
Dead Sperm Count	x_8	0.093	-0.318	0.186
Amount of Testosterone	x_9	0.506	0.322	0.890
Amount of Polactin	x_{10}	0.751	-0.081	0.812
Blood group	x_{11}	0.335	-0.720	0.591
Physical stimulants	x_{12}	-0.215	-0.432	0.541
Smoking	x_{13}	0.074	-0.090	0.418
varicocele	x_{14}	0.000	-0.008	0.001
R.B.sugar	x_{15}	2.020	1.389	2.582
WBC	x_{16}	0.792	0.964	1.378
Weight	x_{17}	0.000	-0.001	0.009
E . S .R	x_{18}	0.000	-0.018	0.045
PCT	x_{19}	0.000	-0.012	0.012
Testicular abscess	x_{20}	-1.0015	-1.372	0.421

From the results presented in the table above, it is evident that seven variables can be excluded from constructing the predictive model of the Tobit quantile regression model at the 0.15 quantile level. (Table 2) This is because their estimated coefficients are equal to zero exactly, indicating that these variables have no effect on the time of liquefaction. Therefore, they can be excluded from the model, allowing us to focus on the remaining variables that have either a positive or negative effect on the liquefaction time variable. A more detailed view of the results is provided in the following figure.

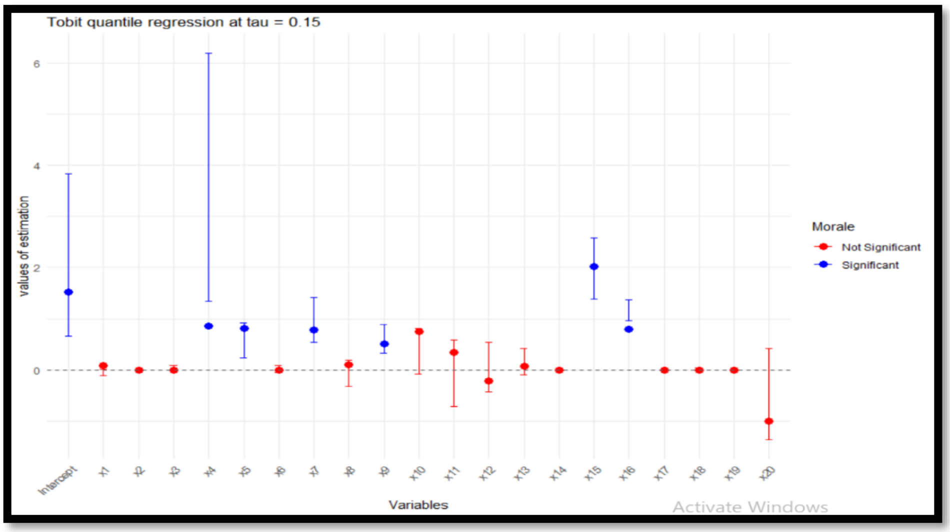

Figure 1. Show the coefficients estimation and 95% confidence intervals for the 0.15 tobit quantile level.

From the figure 1 above, which provides a comprehensive visualization of the impact of the independent variables on the liquefaction time variable, it can be observed that the horizontal dashed line lies along the zero point. Accordingly, the estimated variable values that fall on this line are effectively zero and statistically insignificant, and thus can be excluded from the construction of the predictive model. In contrast, the remaining variables exhibit an influence on the the liquefaction time variable. Some of these variables have statistically insignificant effects, as indicated in red, while others show statistically significant effects, as indicated in blue.

At second tobit quantile level ($\boldsymbol{\tau} = \mathbf{0.35}$)

The parameters of the Tobit regression model for the liquidity data can be estimated as follows

Table 3. Show point estimates and 95% confidence intervals for the 0.35 tobit quantile level by (new B Tobit Q Reg) method of the liquefaction data.

Ariable name	**Variable simple**	$\hat{\boldsymbol{\beta}}$	**Lower bound**	**Upper bound**
--------------	Intercept	1.211	0.982	2.122
Ph Basic and Acidic	x_1	0.117	0.013	0.617
Undescended testicle	x_2	0.000	-0.247	0.057
Sperm Antibodies	x_3	0.081	-0.172	0.284
Count Sperm	x_4	0.418	0.326	0.744
Active Sperm Count	x_5	0.277	0.129	0.764
Alcohol Consumption	x_6	0.341	-0.081	0.364
Weak Motility Sperm Count	x_7	1.005	0.229	1.384
Dead Sperm Count	x_8	0.290	-0.784	0.592
Amount of Testosterone	x_9	0.791	0.540	1.090
Amount of Polactin	x_{10}	0.092	-0.046	0.715
Blood group	x_{11}	1.015	-1.816	1.137
Physical stimulants	x_{12}	-0.442	-0.832	0.541
Smoking	x_{13}	0.074	-0.073	0.252
varicocele	x_{14}	0.000	-0.067	0.194
R.B.sugar	x_{15}	1.157	0.726	1.324
WBC	x_{16}	0.884	0.561	1.178
Weight	x_{17}	0.000	-0.041	0.241
E . S .R	x_{18}	0.000	-0.029	0.185
PCT	x_{19}	0.000	-0.095	0.082
Testicular abscess	x_{20}	-1.045	-1.137	0.681

From the results presented in the table above, it is evident that five variables can be excluded from constructing the predictive model of the Tobit quantile regression model at the 0.15 quantile level. (Table 3) This is because their estimated coefficients are equal to zero exactly, indicating that these variables have no effect on the time of liquefaction. Therefore, they can be excluded from the model, allowing us to focus on the remaining variables that have either a positive or negative effect on the liquefaction time variable. A more detailed view of the results is provided in the following figure.

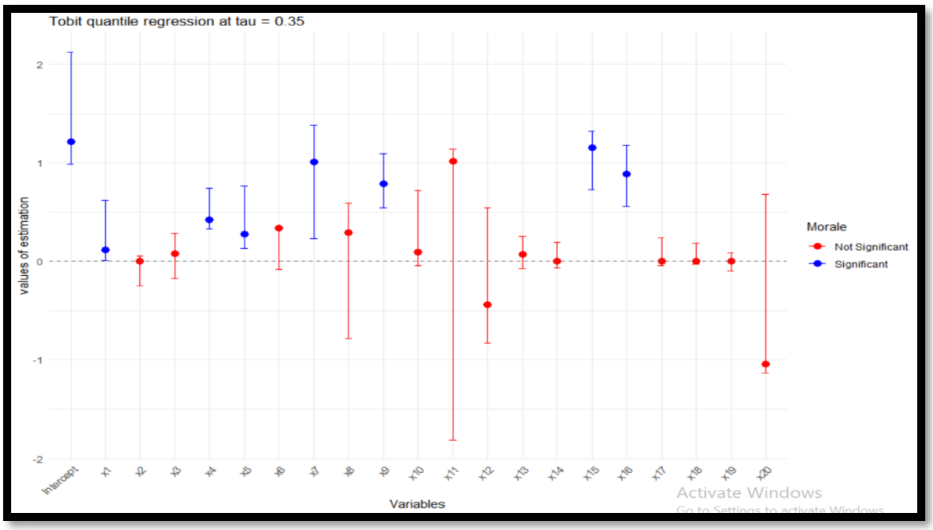

Figure 2. Show the coefficients estimation and 95% confidence intervals for the 0.35 tobit quantile level.

From the figure 2 above, which provides a comprehensive visualization of the impact of the independent variables on the liquefaction time variable, it can be observed that the horizontal dashed line lies along the zero point. Accordingly, the estimated variable values that fall on this line are effectively zero and statistically insignificant, and thus can be excluded from the construction of the predictive model. In contrast, the remaining variables exhibit an influence on the the liquefaction time variable. Some of these variables have statistically insignificant effects, as indicated in red, while others show statistically significant effects, as indicated in blue.

At third tobit quantile level ($\boldsymbol{\tau = 0.55}$)

The parameters of the Tobit regression model for the liquidity data can be estimated as follows

Table 4. Show point estimates and 95% confidence intervals for the 0.55 tobit quantile level by (new B Tobit Q Reg) method of the liquefaction data.

Variable name	Variable Simple	$\hat{\beta}$	Lower bound	Upper bound
--------------	Intercept	1.542	0.829	1.682
Ph Basic and Acidic	x_1	0.367	0.274	0.726
Undescended testicle	x_2	0.000	-0.374	0.241
Sperm Antibodies	x_3	0.153	-0.211	0.523
Count Sperm	x_4	0.139	0.091	0.684
Active Sperm Count	x_5	0.418	0.282	0.973
Alcohol Consumption	x_6	0.215	-0.524	0.691
Weak Motility Sperm Count	x_7	1.451	0.765	1.682
Dead Sperm Count	x_8	0.368	-0.374	0.754
Amount of Testosterone	x_9	1.111	0.784	1.325
Amount of Polactin	x_{10}	0.104	-0.124	0.447
Blood group	x_{11}	1.611	-1.754	1.821
Physical stimulants	x_{12}	-0.524	-0.775	0.485
Smoking	x_{13}	0.443	-0.137	0.683
varicocele	x_{14}	0.000	-0.341	0.552

R.B.sugar	x_{15}	1.247	0.976	1.763
WBC	x_{16}	0.927	0.487	1.965
Weight	x_{17}	0.000	-0.153	0.525
E . S .R	x_{18}	0.000	-0.173	0.384
PCT	x_{19}	0.000	-0.324	0.318
Testicular abscess	x_{20}	-0.153	-0.357	0.455

From the results presented in the table above, it is evident that five variables can be excluded from constructing the predictive model of the Tobit quantile regression model at the 0.15 quantile level. (Table 4). This is because their estimated coefficients are equal to zero exactly, indicating that these variables have no effect on the time of liquefaction. Therefore, they can be excluded from the model, allowing us to focus on the remaining variables that have either a positive or negative effect on the liquefaction time variable. A more detailed view of the results is provided in the following figure.

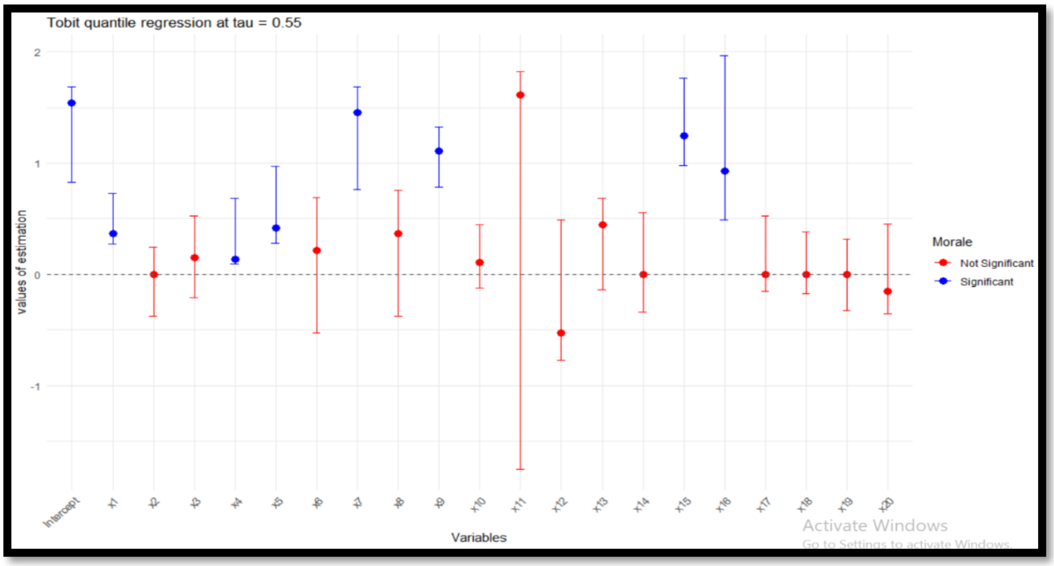

Figure 3. Show the coefficients estimation and 95% confidence intervals for the 0.55 tobit quantile level.

From the figure above, which provides a comprehensive visualization of the impact of the independent variables on the liquefaction time variable, it can be observed that the horizontal dashed line lies along the zero point. Accordingly, the estimated variable values that fall on this line are effectively zero and statistically insignificant, and thus can be excluded from the construction of the predictive model. In contrast, the remaining variables exhibit an influence on the the liquefaction time variable. Some of these variables have statistically insignificant effects, as indicated in red, while others show statistically significant effects, as indicated in blue.

At forth tobit quantile level ($\boldsymbol{\tau = 0.75}$)

The parameters of the Tobit regression model for the liquidity data can be estimated as follows

Table 5. Show point estimates and 95% confidence intervals for the 0.75 tobit quantile level by (new B Tobit Q Reg) method of the liquefaction data.

Variable name	**Variable Simple**	$\hat{\boldsymbol{\beta}}$	**Lower bound**	**Upper bound**
--------------	Intercept	2.274	1.341	2.356
Ph Basic and Acidic	x_1	0.728	0.374	1.127

Undescended testicle	x_2	0.000	-0.549	0.374
Sperm Antibodies	x_3	0.246	0.187	0.409
Count Sperm	x_4	0.388	0.218	0.756
Active Sperm Count	x_5	0.845	0.627	1.247
Alcohol Consumption	x_6	1.315	-0.384	1.453
Weak Motility Sperm Count	x_7	1.218	1.021	0.784
Dead Sperm Count	x_8	0.242	-0.118	0.685
Amount of Testosterone	x_9	1.483	1.134	1.821
Amount of Polactin	x_{10}	0.719	-0.227	0.937
Blood group	x_{11}	0.906	-1.672	1.196
Physical stimulants	x_{12}	-0.820	-1.112	-0.329
Smoking	x_{13}	0.618	-0.422	0.922
varicocele	x_{14}	0.000	-0.516	0.223
R.B.sugar	x_{15}	1.264	0.655	1.464
WBC	x_{16}	1.288	0.631	1.627
Weight	x_{17}	0.000	-0.541	0.524
E . S .R	x_{18}	0.000	-0.083	0.408
PCT	x_{19}	0.000	-0.149	0.246
Testicular abscess	x_{20}	-0.153	-0.497	0.391

From the results presented in the table above, it is evident that five variables can be excluded from constructing the predictive model of the Tobit quantile regression model at the 0.15 quantile level. (Table 5) This is because their estimated coefficients are equal to zero exactly, indicating that these variables have no effect on the time of liquefaction . Therefore, they can be excluded from the model, allowing us to focus on the remaining variables that have either a positive or negative effect on the liquefaction time variable. A more detailed view of the results is provided in the following figure.

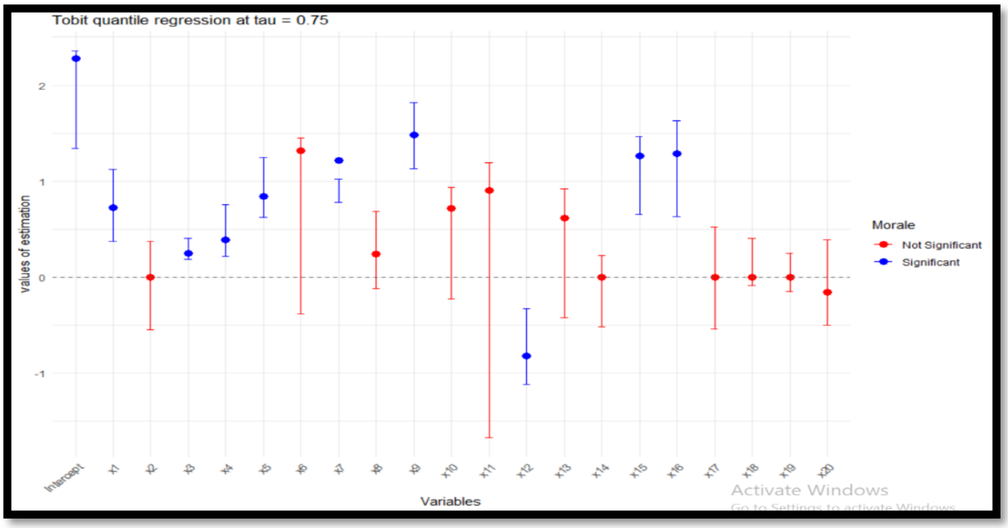

Figure 4. Show the coefficients estimation and 95% confidence intervals for the 0.75 tobit quantile level.

From the figure 4 above, which provides a comprehensive visualization of the impact of the independent variables on the liquefaction time variable, it can be observed that the horizontal dashed line lies along the zero point. Accordingly, the estimated variable values that fall on this line are effectively zero and statistically insignificant, and thus can be excluded from the construction of the predictive model. In contrast, the remaining variables exhibit an influence on the liquefaction time variable. Some of these variables have statistically insignificant effects, as indicated in red, while others show statistically significant effects, as indicated in blue.

At forth tobit quantile level ($\boldsymbol{\tau = 0.95}$)

The parameters of the Tobit regression model for the liquidity data can be estimated as follows

Table 6. Show point estimates and 95% confidence intervals for the 0.95 tobit quantile level by (new B Tobit Q Reg) method of the liquefaction data.

Variable name	Variable Simple	$\hat{\boldsymbol{\beta}}$	Lower bound	Upper bound
--------------	Intercept	1.548	0.675	1.857
Ph Basic and Acidic	x_1	0.924	0.584	1.283
Undescended testicle	x_2	0.000	-0.181	0.267
Sperm Antibodies	x_3	0.306	0.259	0.586
Count Sperm	x_4	0.573	0.325	1.126
Active Sperm Count	x_5	0.766	0.364	0.958
Alcohol Consumption	x_6	0.635	0.253	0.815
Weak Motility Sperm Count	x_7	1.581	0.674	1.697
Dead Sperm Count	x_8	0.464	-0.474	0.512
Amount of Testosterone	x_9	1.184	0.434	1.314
Amount of Polactin	x_{10}	0.812	-0.157	1.114
Blood group	x_{11}	0.375	-0.524	0.751
Physical stimulants	x_{12}	-0.274	-0.674	0.184
Smoking	x_{13}	0.413	-0.245	0.854
varicocele	x_{14}	0.000	-0.225	0.124
R.B.sugar	x_{15}	1.418	0.808	1.700
WBC	x_{16}	0.652	0.452	1.114
Weight	x_{17}	0.591	0.374	0.867
E . S .R	x_{18}	0.000	-0.185	0.310
PCT	x_{19}	0.000	-0.099	0.176
Testicular abscess	x_{20}	-0.427	-0.725	0.543

From the results presented in the table above, it is evident that five variables can be excluded from constructing the predictive model of the Tobit quantile regression model at the 0.15 quantile level. (Table 6) This is because their estimated coefficients are equal to zero exactly, indicating that these variables have no effect on the time of liquefaction. Therefore, they can be excluded from the model, allowing us to focus on the remaining variables that have either a positive or negative effect on the liquefaction time variable. A more detailed view of the results is provided in the following figure.

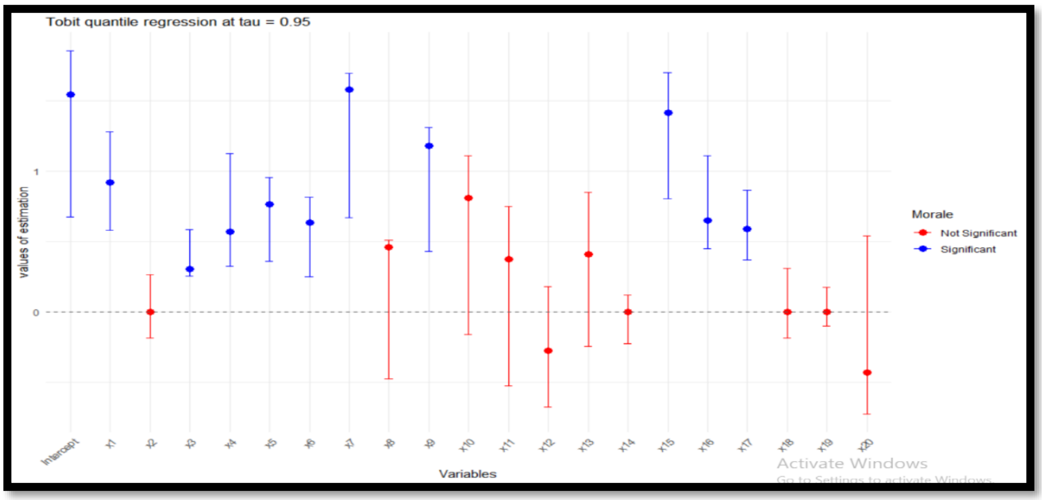

Figure 5. Show the coefficients estimation and 95% confidence intervals for the 0.95 tobit quantile level.

From the figure 5 above, which provides a comprehensive visualization of the impact of the independent variables on the liquefaction time variable, it can be observed that the horizontal dashed line lies along the zero point. Accordingly, the estimated variable values that fall on this line are effectively zero and statistically insignificant, and thus can be excluded from the construction of the predictive model. In contrast, the remaining variables exhibit an influence on the liquefaction time variable. Some of these variables have statistically insignificant effects, as indicated in red, while others show statistically significant effects, as indicated in blue.

4. Conclusion

It can be concluded that this method represents the optimal approach for analyzing the liquefaction time data, as evidenced by the mean squared error(MSE) values across all levels of the Tobit quantile regression model. It is also observed that there are five variables that are not significant in constructing the predictive model for analyzing the liquefaction time data across all levels of the Tobit quantile regression model. This is due to the fact that their estimated coefficients are zero, indicating that they have no meaningful effect on the response variable. The results indicate the presence of approximately 15 variables that influence the response variable. Among them, 13 variables exhibited a positive (direct) effect, suggesting that an increase in these variables leads to an increase in the response variable. In contrast, two variables demonstrated an inverse effect, meaning that their increase is associated with a decrease in the response variable.

Recommendations

In analyzing the real data used in this study, it is essential to employ an appropriate analytical method that provides the researcher with accurate and objective results in interpreting the studied phenomenon. Therefore, it is recommended to compare the applied methods and adopt the most suitable one for the analysis. In the analysis of liquefaction time data, it is important to investigate additional variables that may have a direct impact on the response variable. We recommend analyzing the data under study using more effective regularization methods that can better identify and select significant variables while excluding irrelevant ones. Finally, we propose extending the current study to include new regularization methods with strong statistical properties. Among the suggested methods are the Adaptive LASSO and the Elastic Net, as these techniques are considered regularization methods that possess the oracle property.

REFERENCES

[1] T. Amemiya, "Tobit models: A survey," Journal of Econometrics, vol. 24, no. 1-2, pp. 3-61, 1984.

[2] M. Buchinsky, "Changes in the US wage structure 1963-1987: Application of quantile regression," Econometrica: Journal of the Econometric Society, pp. 405-458, 1994.

[3] C. Cameron and P. K. Trivedi, Microeconometrics: Methods and Applications, Cambridge University Press, 2005.

[4] V. Chernozhukov and H. Hong, "Three-step censored quantile regression and extramarital affairs," Journal of the American Statistical Association, vol. 97, no. 459, pp. 872-882, 2002.

[5] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," Journal of the American Statistical Association, vol. 96, no. 456, pp. 1348-1360, 2001.

[6] Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, Bayesian Data Analysis, Boca Raton, FL: CRC Press, 2013.

[7] P. D. Hoff, A First Course in Bayesian Statistical Methods, vol. 580, New York: Springer, 2009.

[8] Kowalski, "Censored quantile instrumental variable estimates of the price elasticity of expenditure on medical care," Journal of Business & Economic Statistics, vol. 34, no. 1, pp. 107-117, 2016.

[9] F. W. Olver, Ed., NIST Handbook of Mathematical Functions Hardback and CD-ROM, Cambridge University Press, 2010.

[10] S. Portnoy, "Censored regression quantiles," Journal of the American Statistical Association, vol. 98, no. 464, pp. 1001-1012, 2003.

[11] H. J. Wang and M. Fygenson, "Inference for censored quantile regression models in longitudinal studies," 2009.

[12] V. J. Yohai, R. A. Maronna, and R. D. Martin, Robust Statistics, West Sussex, England: John, 2006.

[13] K. Yu and R. A. Moyeed, "Bayesian quantile regression," Statistics & Probability Letters, vol. 54, no. 4, pp. 437-447, 2001.

[14] R. K. Gupta, "Quantile regression methods," International Journal of Statistical Sciences, vol. 10, no. 2, pp. 223-235, 2021.

[15] H. H. Ho, "Estimation of quantile regression parameters for censored data," Biometrika, vol. 101, no. 1, pp. 13-27, 2014.