JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

Bayesian and Non-Bayesian Quantile Regression Methods: A comparative study

Murtadha.J. Aloqaili
Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq.Email:murtadha.aluqealy@gmail.com

Rahim Alhamzawi
Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq.Email:rahim.alhamzawi@qu.edu.iq

Received: 2022 March 15; **Revised:** 2022 April 20; **Accepted:** 2022 May 10

Abstract

Since its introduction in 1978 by Koenker and Basset, quantile regression has grown to be a significant tool and is being accepted more widely in a variety of applications it is a technique for measuring the relationship between a predictor and a response variable.It also provides information more than linear regression. Recently, it was recommended to use the Bayesian quantile regression method to handle model uncertainty and unknown parameters. Quantile regression thereafter saw the emergence of numerous Bayesian techniques. The current work focuses on studying the methods (Bayesian lasso quantile regression, and Bayesian adaptive lasso quantile regression) in addition to quantile regression and Bayesian quantile regression, and offering a comparison between the aforementioned methods based on the value of MSE, after calculating the parameters and interpreting them in terms of medical logic. A simulation study was provided, and we also conducted a practical study of thalassemia patients' data in Iraq - Babylon city for the year 2021 applying the methods indicated above. The simulation and practical study indicate that the Bayesian Adaptive lasso quantile regression method outperforms the other methods studied.

***Keywords*:**Quantile regression, Linear regression, Bayesian quantile regression, Bayesian lasso quantile regression, Bayesian adaptive lasso quantile regression.

1. Introduction

The ordinary least squares method (OLS) has been the dominant method for decades used in statistical studies. Despite its good performance, it hides many problems, as it sets assumptions that not all studied data can be subject to, which leads to the penetration and failure of its hypotheses.

Koenker and Bassett (1978) originally proposed quantile regression(QReg)as a robust alternative to the OLS estimator for estimating the coefficients of the linear regression function (Buchinsky, M. 1994). The classical mean regression model may fail to explain an intriguing element of the connection between prediction and response variables if the distribution of the response variable is excessively skewed in the regression model. So, it can be understood that this heavy-tailed distribution and regression may not be a good representative of the mean (Das, P. 2016). Although QReg is more immune to outliers, and unusual error distributions. However, for a given distribution, the efficiency of its estimator depends on the quantile level (Al-hamzawi, R. 2016).

Quantiles are more helpful than mean values because they are less sensitive to skewed distributions and outliers. This is a central factor of QReg, which, prides itself on its flexibility and capacity to investigate the whole. the conditional distribution of the response variable distribution owing to its estimations. It has attracted a lot of theoretical attention as well as a lot of practical applications in medicine, growth scheme, econometrics, climate change, agriculture, ecology, engineering, etc. (Gosling, A., et al. 2000., Mazucheli, Josmar.2021., ..., and others).

2. Classical Estimation

QReg model is given by

$$y_{i=} x_i^{'} \beta + \varepsilon_i \qquad ;(\grave{\imath}=1,2,3,..,n) \tag{1}$$

where $y_i$is the response of the ìth sample, x_i is a $k \times 1$ vector of variables.β is the vector $k\times 1$ for the coefficients, and ξ_i is the independent error term, whose distribution is unknown*(Davino, C., et al.2013)*. As a result, the θ^{th} QReg is defined as any solution to the quantile minimization issue. Koenker and Bassett (1978) show that the regression coefficient vector may be determined regularly as the solution to the minimization of:

$$\min_{\beta} \sum_{i=1}^{n} \rho_\theta \left(y_i - x_i' \beta\right) \tag{2}$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

where $\rho_\theta(\cdot)$ is a loss (or check) function defined:

$$\rho_\theta(u) = u\ \{\theta - I\ (u<0)\} \tag{3}$$

Where I(.)is the indicator function. Because the loss function is not differentiable at zero, we cannot deduce improving existing processes to the minimization issue. *(Park, and Casella .2008).* As a result, we resort to other methods, which we will discuss in the next chapter to use to get QReg estimates for β.

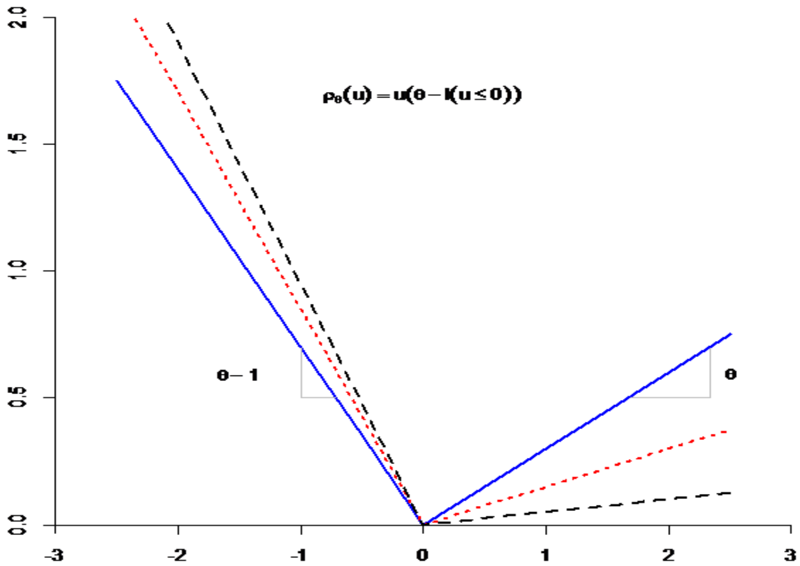


Figure 1: shows the loss (or check) function at θ = 0.30 (blue line), θ = 0.15 (red line) and θ = 0.05 (black line)

It is possible to estimate β using a function (1). However, this function is non-derivative *(Koenker and Bassett.1978)*. They demonstrate how to turn the issue of minimizing (2) into a linear program and provide details on how to solve it effectively for any or all θ ϵ (0,1) *(Koenker, and d'Orey .1993).*

Despite the good performance of QReg, it also showed some performance weakness in front of several types of data, so several researchers (such as Al-Hamzawi, Yu, Zou and Hastie,..., and others) resorted to introducing some improvements or additions to it to eventually produce new methods, each with its advantages and disadvantages, but the goal is one and that is to get more accurate results .

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

The rapid advancement of technology, particularly the creation of various programs and programming languages (for example, SPSS, R, Python...), has tremendously contributed to the evolution of these approachesEspecially after the introduction of Bayesian methods with their attractive specifications to quantile regression, this merger resulted in new methods with more attractive specifications. These methods have been covered in a lot of recent research.

3.Bayesian Quantile Regression (BQReg)

Yu and Moyeed (2001) used a Bayesian approach to QReg, which differs significantly from earlier work in this area. Bayesian inference for QReg begins with the formation of a probability function based on the ALD, regardless of the data's actual distribution. *(Phadkantha, R., et al .2019)*

$$f_{\theta}(y|\mu = 0, \sigma, \theta) = \frac{\theta(1-\theta)}{\sigma}\ exp\{-\rho_{\theta}(\frac{y-\mu}{\sigma})\}, \qquad (4)$$

where ($\rho_{\theta}(\xi)$ may also be written as:

$$\rho_{\theta}(\xi) = \frac{|\xi| + (2\theta - 1)\,\xi}{2} \qquad (5)$$

These utilities let you figure out how changes in variables impact the location, scale, and shape of a response variable's distribution.In general, any prior can be used for the QReg parameters. However, it has been shown that using incorrect uniform priors results in a team joining posterior distribution. Because QReg does not generally assume a likelihood for Y|X conditional distributions, Bayesian approaches employ a working likelihood *(Noufaily, A., and Jones, M. C. 2013).* The AL likelihood is a convenient choice since the mode of the generated posterior with a flat prior is the standard QReg estimate. The posterior conclusion, on the other hand, must be evaluated with caution. A posterior variance correction for valid inference was presented by Yang, Wang, and He (2016). Yang and He (2012) demonstrated that if the working likelihood is selected as the empirical likelihood, an asymptotically correct posterior inference may be obtained.

It is generally known that the conjugate prior distribution plays the most important role in Bayesian analysis since it is preferable to have conditional distributions as the prior in terms of the same functional form and similar features *(Chen and Ibrahim, 2003)*. The typical conjugate prior distribution does not exist for the parameters of the θ^{th} regression quantile, but you may

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

approximate the posterior distributions of the unknown parameters using Markov chain Monte Carlo (MCMC) techniques.

3.1Asymmetric Laplace Distribution (ALD)

Following Yu and Moyeed (2001), we consider the linear regression model

$$y_i = b_\theta + x_i\beta + \xi_i, \quad i = 1, \cdots, n \quad and \quad \theta \in (0,1), \tag{6}$$

and assume that the error term ϵ_i has the ALD with density:

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

The mean is known to be:

$$E(\xi_i) = \frac{1-2\theta}{\theta(1-\theta)} \tag{8}$$

and variance

$$Var(\xi_i) = \frac{1-2\theta+2\theta^2}{\theta^2(1-\theta)^2} \tag{9}$$

whereσisthescaleparameterandμisthelocationparameter.Underthisassumption,minimizing thecheckfunction(2)isequivalenttomaximizingthelikelihoodfunctionoftheaboveALD *(Alhamzawi,2016).*

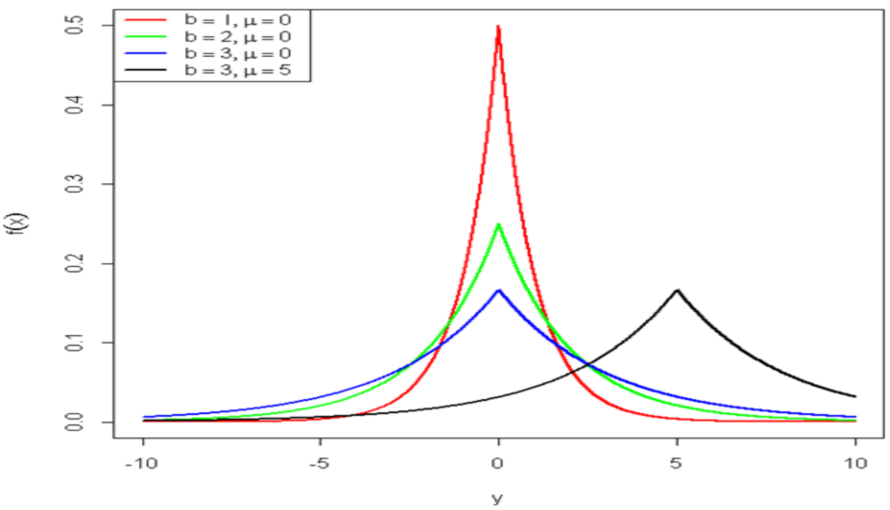


Figure 2: Graph illustrate Probability density function of ALD

4.Bayesian Least Absolute Shrinkage and Selection Operator Quantile Regression (BLQReg)

Lasso has generated a lot of attention in the literature since its debut by Tibshirani (1996), because it may be used to do variable selection and shrinkage, which can increase the model's predictive performance and interpretability.

The BLQReg offers the benefit of providing a more detailed study of the relationship between variables while also automatically selecting variables. When there are many irrelevant noise variables.It can also produce accurate predictions as well as variable selection or compact models from an interpretation viewpoint at least *(Peter and Gareth .2008).*Within the framework of QReg, the BLQReg may be regarded as an augmentation of a Lasso regression. Its estimations are as follows*(Li and Zhu, 2008)*:

$$\min_{\beta_0 \beta} \sum_{i=1}^{n} \rho_\theta(y - \beta_0 - \acute{x}_i \beta) + \lambda \parallel \beta \parallel_1 \quad (10)$$

Where λ is a nonnegative regularizationparameter, and the second term is the so-called $\|.\|_1$ penalty QReg (the L_1-norm), which is critical to the Lasso method's success *(Alhamzawi, et al.2012).*In BLQReg,λ is the most essential parameter since it causes Lasso to shrink the QReg coefficients towards zero as it rises.

4.1 Park – Casella (2008)

The Gibbs sampler for the Bayesian Lasso employs the Laplace distribution (LD)as a normal scale mixture (with an exponential mixing density), as shown below:

$$\pi(\beta|\sigma^2) = \prod_{i=1}^{\theta} \frac{\lambda}{2\sqrt{\sigma^2}} \exp - \frac{\lambda|\beta_i|}{\sqrt{\sigma^2}} \quad (11)$$

$$\frac{c}{2}\, e^{-c|m|} = \int_0^\infty \frac{1}{\sqrt{2\pi w}} e^{-m^2/(2w)} \frac{c^2}{2} e^{-c^2 w/2} dw\,, \qquad c > 0 \quad (12)$$

Has the desired shape after including the conditional prior on (12). The inappropriate prior density $\pi(\sigma^2) = 1/\sigma^2$ is used by Park and Casella (2008), however, any inverse-gamma prior for σ^2would likewise maintain conjugacy. Analytically, averaging from the posterior of the joint under its

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

independent, flat prior is easy. We exclude it for the sake of simplicity and convenience since it is rarely of interest. It can be presented with a completely normal conditional distribution, including mean and variance if desired. σ^2/n.

These complete conditionals serve as the foundation for a fast Gibbs sampler that updates the block of β.

4.2 The Hierarchical model

A new scale mixture based on a uniform distribution mixing with the density of a specific gamma distribution to the LD was represented by (Mallick, H., and, Yi, N. 2014.). In the other word:

$$\frac{\lambda}{2}e^{-\lambda|\beta_j|} = \int_{-s_j<\beta_j<s_j} \frac{1}{2s_j}\frac{\lambda^2}{\Gamma(2)} s_j^{2-1} exp\{-\lambda s_j\}$$

$$\beta|s \sim Unif\ (-s_j, s_j)$$

$$s|\lambda \sim Ga\ (2, \lambda)$$

$$\lambda \sim Ga(a_1, b_1)$$

$$\sigma \sim Ga(a_2, b_2)$$

Wherea_1, b_1, a_2 and b_2 are hyperparameters. In summary, our Bayesian hierarchical model is given by

$$y_i = x_i^{'}\beta_p + \theta z_i + \sqrt{2\sigma z_i}\xi_i$$

$$y_i|\beta_p, \sigma, z_i \sim N(x_i^{'}\beta_p + \theta z_i, 2\sigma z_i)$$

$$z_i|\ \sigma \sim Exp(\tfrac{p(1-p)}{\sigma})$$

$$\beta|s \sim Unif\ (-s_j, s_j)$$

$$s|\lambda \sim Ga\ (2, \lambda)$$

$$\lambda \sim Ga(a_3, b_3)$$

$$\sigma \sim Ga(a_4, b_4)$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

5.Bayesian Adaptive Lasso Quantile Regression (BALQReg)

The Bayesian adaptive lasso models presented by Leng et al. (2014) and Alhamzawi et al. (2012) are hierarchical models based on conventional linear regression and linear QReg, respectively, with β_j $(1 \leq j \leq k)$ following an LD.Almost all of the literature's known variable selection (13) techniques for QReg are predicated on the assumption of a large sample size or even a little amount of data.Zou (2006) demonstrated that adaptive Lasso regression (BALReg) has oracle qualities that Lasso does not have.As well,Permitting alternative penalization values for different regression coefficients *(Alhamzawi, and Ali .2018).*

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

Where $\lambda_j \geq 0$. Unimportant predictors, on the other hand, are frequently substantially associated with certain important predictors in many real-world applications. Gibbs sampling methods employing an enlarged hierarchy with conjugate normal priors for coefficients and separate exponential priors for variances have been reported. Full conditional distributions are tractable with this enlarged hierarchy.

5.1 The Hierarchical model

BALQReg is a Bayesian hierarchical model given by (Alhamzawi, et al.2012):

$$y_i = x_i'\beta_p + \theta z_i + \sqrt{2\sigma z_i}\xi_i$$

$$p(\xi_i) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{1}{2}\xi_i^2\right\}$$

$$z_i | \, \sigma \sim Exp(\frac{p(1-p)}{\sigma})$$

$$p(\sigma, \lambda_j^2) = \frac{1}{\sqrt{2\pi s_j}} exp\left\{-\frac{\beta_j^2}{2s_j}\right\} \frac{\sigma}{2\lambda_j^2} exp\left\{-\frac{\sigma s_j}{2\lambda_j^2}\right\}$$

$$p(\gamma, \theta) = \frac{\theta^{\gamma}}{\Gamma(\gamma)} (\lambda_j^2)^{-1-\gamma} exp\left\{-\frac{\theta}{\lambda_j^2}\right\}$$

$$p(\sigma) = \sigma^{a_5 - 1} exp(b_5\sigma)$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

$$p(\theta,\gamma) = \theta^{-1}$$

6.Simulation study

In this section, simulation was used to compare the studied strategies to find the optimum estimation method between them. Four approaches were compared (QReg, BQReg, BLQReg, and BALQReg). The initial step is to estimate the parameters. Has been usedmean square errors (MSE) as a metric to evaluate the performance of each approach.

A sample of 150 views of over 100 repeats was used. The number of significant factors that vary from one example to another is referred to throughout this study as (S). The correlation coefficient is denoted by the symbol (ρ), and in this study, we used five values of ρ) that are (0.05, 0.15, 0.50, 0.70, 0.90). The algorithms under study were programmed in the R language, and the outcomes and their interpretation were shown Excel is used to display graphs for additional illustration.

6.1 Simulation 1 (Very Sparse)

In this simulation study, we generated 150 views of the model

$$y_{i=} x_i' \beta + \xi_i$$

Where x_i represents a vector of (20) common variables. Where S=1 Represents the number of influencing or important variablesβ(5,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0). In Simulation 1 we focus on the effect of correlation (ρ) on the performance of the studied methods (QReg, BQReg, BLQReg, and BALQReg). Five levels (0.05, 0.15, 0.50, 0.70, 0.90) of (ρ) have been adopted, as shown in Table (1) below.

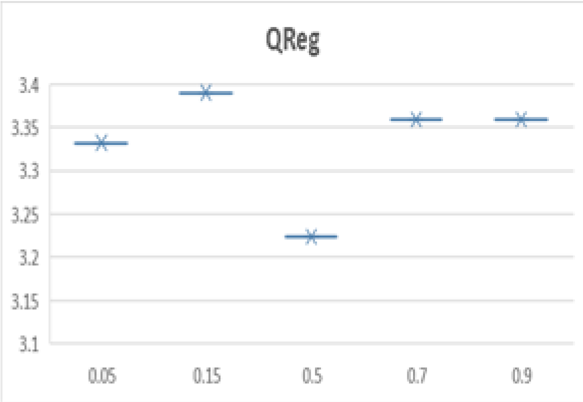
??	*QReg*	*BQReg*	*BLQReg*	*BALQReg*
0.05	3.3319	3.1369	1.6245	1.7696
0.15	3.3900	3.0436	1.4936	1.8439
0.50	3.2240	2.9993	1.4222	1.6913
0.70	3.3587	3.1348	1.4970	1.9585
0.90	3.3587	3.0907	1.6874	1.7908

Table (1) : Shows MSE for methods (QReg, BQReg, BLQReg, and BALQReg) when S= 1,ρ is (0.05, 0.15, 0.50, 0.70, 0.90).

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

At first glance in Table (1), almost all of the studied methods (QReg, BQReg, BLQReg, and BALQReg,) use ascending values from MSE with increasing ρ, which is usual. To assess the data thoroughly, we must first recognize that the technique with the lowest MSE with varied values of ρ is the best-performing method among the others.The BLQReg method recorded the lowest MSE values in all levels ρ, while BALQReg method was tracking it and nearly close to it on many occasions. In contrast, QReg was ranked last since it had the greatest MSE value across all ρ levels. In simulation 1, the BLQReg technique performed the best.

The figures below depict the spread of MSE values for each method alone.



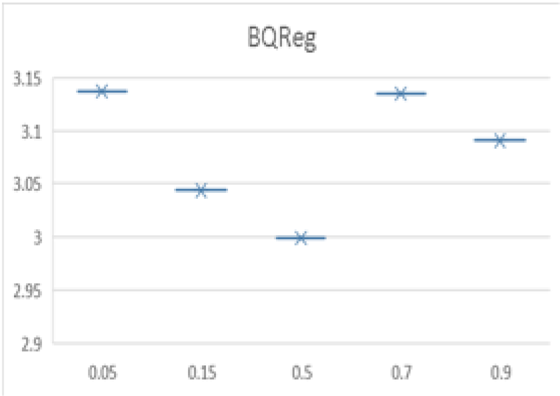
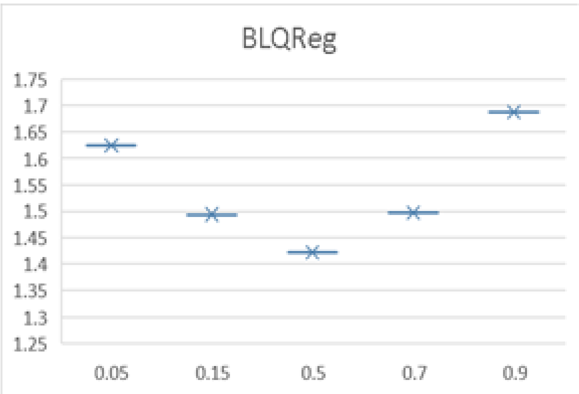


Figure (3):QReg method when S=1 andρ = (0.05, 0.15, 0.50, 0.70, 0.90)

Figure (4): BQReg method when S=1 andρ = (0.05, 0.15, 0.50, 0.70, 0.90)



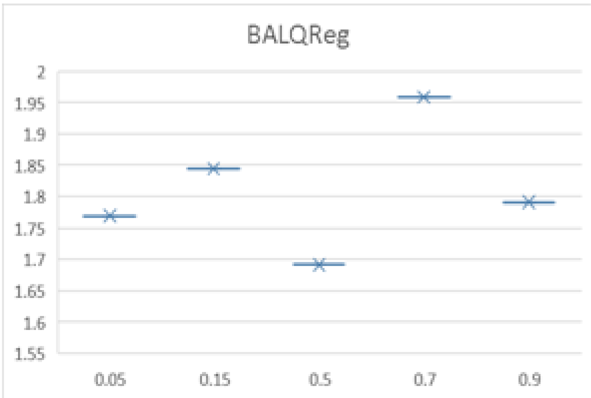


Figure (5): BLQReg method when S=1 andρ = (0.05, 0.15, 0.50, 0.70, 0.90)

Figure (6): BALQReg method when S=1 andρ = (0.05, 0.15, 0.50, 0.70, 0.90)

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

Figures (3),(4),(5), and (6) reveal that the spread points appeared to be unstable and erratic, that is, there is no specific pattern for the MSE values for each method independently, whether ascending or descending. The BALQReg method, on the other arm, generated looks like a letter (V) pattern, and while it was not the greatest way, it was the most stable and balanced of the methods evaluated.The figure below takes a comprehensive look at all the studied methods combined.

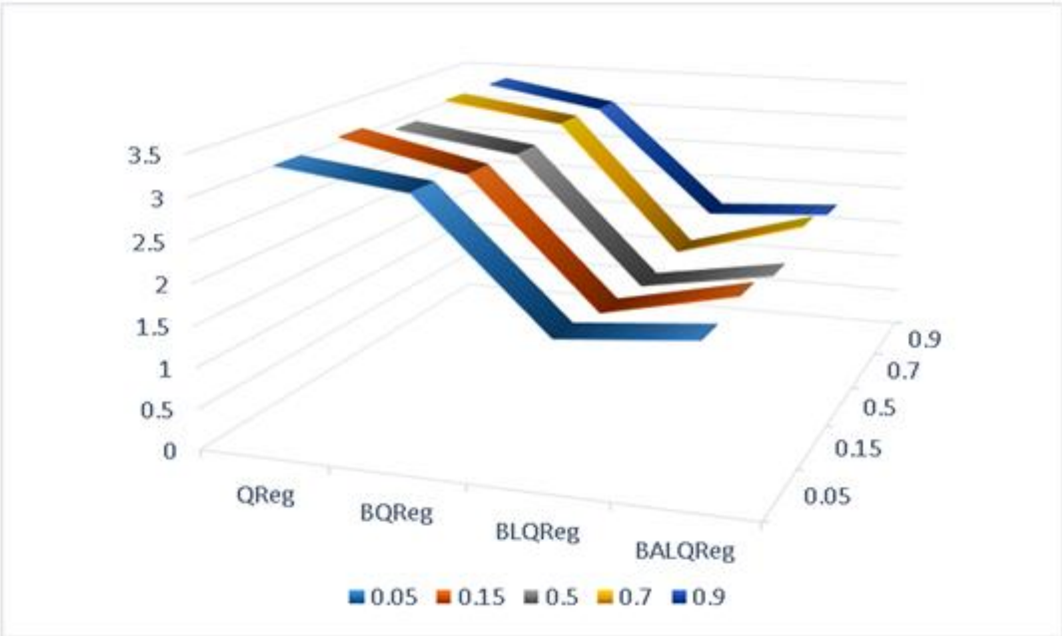


Figure (7): Shows whenMSE for methods (QReg, BQReg, BLQReg, and BALQReg) S = 1, and ρ is (0.05, 0.15, 0.50, 0.70, 0.90).

Figure (7) above demonstrates shows BLQReggot the lowest MSE values at all levels ρ among the (QReg, BQReg, BLQReg, and BALQReg) approach, making it the top performing method when S = 1, and ρ is (0.05, 0.15, 0.50, 0.70, 0.90).

6.2 Simulation 2 (Sparse)

In this simulation study, we generated 150 views of the model

$$y_{i=} x_i^{'} \beta + \xi_i$$

Where x_i represents a vector of (20) common variables. Where S=5 Represents the number of influencing or important variablesβ (5, 0, 0, 0, 3, 0, 0, 0, 0, 5, 0, 0, 0, 0, 0, 2, 0, 0, 0, 7). In Simulation 2 we focus

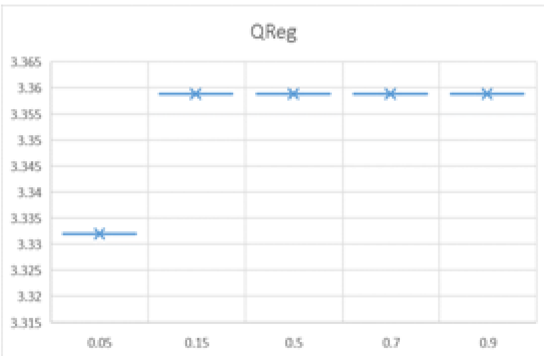
JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

on the effect of correlation (ρ)Five levels (0.05, 0.15, 0.50, 0.70, 0.90) of (ρ) have been adopted on the performance of the studied methods (QReg, BQReg, BLQReg, and BALQReg). As shown in Table (2).

??	*QReg*	*BQReg*	*BLQReg*	*BALQReg*
0.05	3.3319	3.1395	2.6799	2.0100
0.15	3.3588	3.0721	2.1159	1.6904
0.50	3.3588	3.0243	2.4045	1.8401
0.70	3.3588	3.0967	2.5255	1.7589
0.90	3.3588	3.1192	2.2737	2.1277

Simulation 2 is significantly different from its predecessor. When we look closely at Table (2), the discrepancy becomes evident, since QReg seems to be the constant value in all circumstances save one (0.05). In terms of the method that produced the lowest MSE, the BALQReg prevailed, since it achieved the lowest in all levels of ρ.

At the same time, the BLQReg technique was extremely close toBALQReg becoming the second-best method in terms of MSE. Whereas BQReg stays bound to QReg and refuses to stray from it.The dispersion of points for each approach is depicted individually in the figures below.



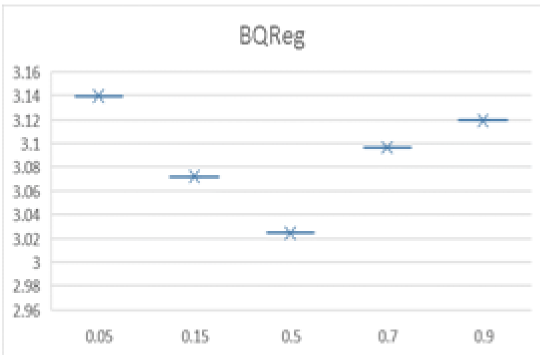
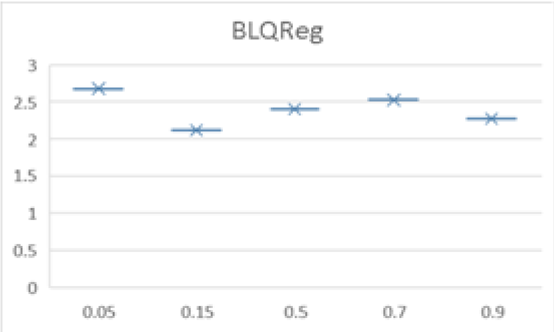


Figure (8): QReg method when ρ=(0.05, 0.15, 0.50, 0.70, 0.90)

Figure (9): BQReg method whenρ=(0.05, 0.15, 0.50, 0.70, 0.90)



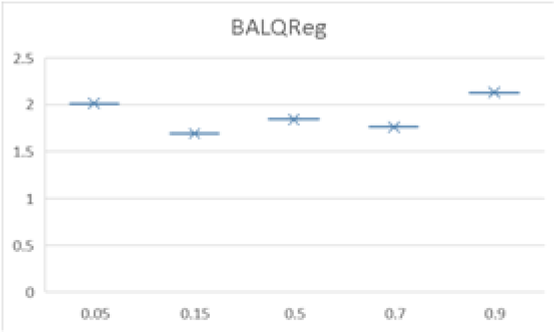


Figure (10): BLQReg method when ρ=(0.05, 0.15, 0.50, 0.70, 0.90)

Figure (11): BALQReg method when ρ=(0.05, 0.15, 0.50, 0.70, 0.90)

Figures (8), (9), (10), and (11) show that the spread points were unstable and erratic, that is, there is no definite pattern for the MSE values for each technique separately, whether ascending or descending. The BQReg approach, on the other hand, produces what seems to be a letter (V) pattern.

We can see that the QReg form began at a low level, climbed somewhat, and then settled stably in a straight line, unaffected by changes in MSE value. However, it was the weakest performer. Method BALQReg has the lowest MSE. As a result, it outperforms the other strategies evaluated.

Figure (12) gives a comprehensive overview of all methods together.

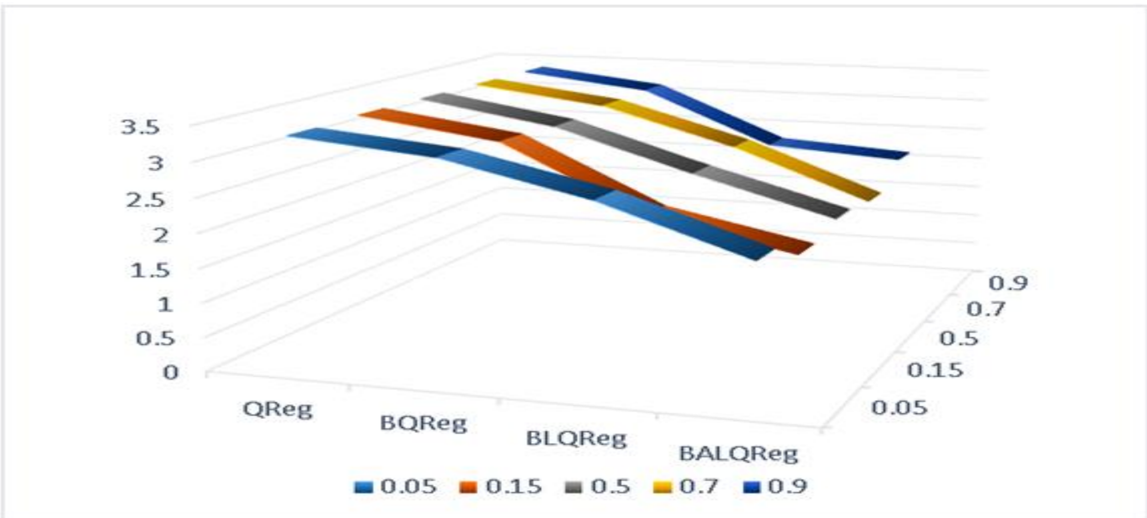


Figure (12) : Shows when S = 5, ρ is (0.05, 0.15, 0.50, 0.70, 0.90) ,and MSE for methods (QReg, BQReg, BLQReg, and BALQReg).

Among the four methods (QReg, BQReg, BLQReg, and BALQReg), Figure (12) above reveals that BALQReg had the lowest MSE values at all levels, making it the best performing strategy when S = 1 and ρ is (0.05, 0.15, 0.50, 0.70, 0.90).

6.3 Simulation 3 (Sparse)

We generated 150 views of the model in this simulation study.

$$y_{i=}x_i^{'}\beta + \xi_i$$

Where x_i represents a vector of (20) common variables. Where S=10Represents the number of influencing or important variables.= β (0, 3, 6, 0, 1, 0, 7, 0, 2, 4, 0, 3, 0, 6, 0, 0, 0, 7, 0, 4).In Simulation 3, we concentrate on correlation (ρ) impacts the examined methods study (QReg, BQReg, BLQReg, and BALQReg). As stated in Table (3), five levels (0.05, 0.15, 0.50, 0.70, and 0.90) of ρ have been implemented.

??	*QReg*	*BQReg*	*BLQReg*	*BALQReg*
0.05	3.3319	3.1367	2.4820	1.9307
0.15	3.3588	3.0712	2.6275	1.5688
0.50	3.3588	3.0701	2.5924	1.9196
0.70	3.3588	3.1238	3.0706	2.6939
0.90	3.3588	3.1219	2.4976	1.7078

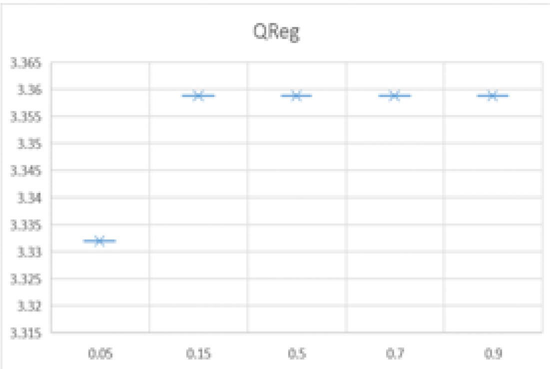
Table (3): Shows MSE for methods (QReg, BQReg, BLQReg, and BALQReg) when S= 10,ρ is (0.05, 0.15, 0.50, 0.70, 0.90).

Simulation 3 was no different from its predecessor. Where the BALQReg method is incomparably superior at all levels ρ. This becomes clear to us just by looking at Table (3). QReg, on the other hand, maintains a consistent attitude, refusing to leave the bottom of the rankings because of having

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

the highest MSE among the techniques under consideration. BQReg insists on remaining close to QReg, resulting in a near-uniform position at the bottom of the list. BLQReg slipped to the second position, well behind the competition as the best method, allowing the BALQReg method to take first place as the best performing method.

The MSE value spread for each approach is displayed individually in the graphs below.



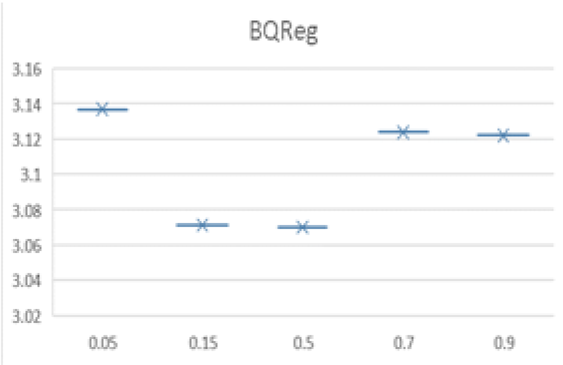
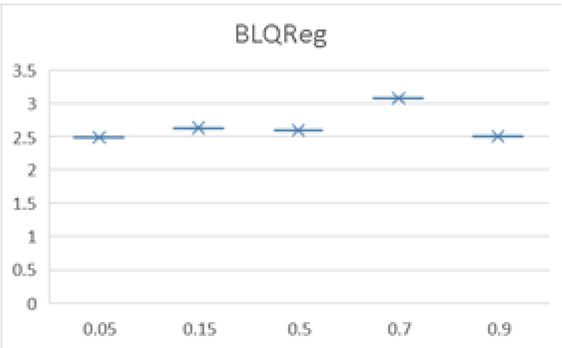


Figure (13): QReg method when ρ=(0.05, 0.15, 0.50, 0.70, 0.90)

Figure (14): BQReg method when ρ=(0.05, 0.15, 0.50, 0.70, 0.90)



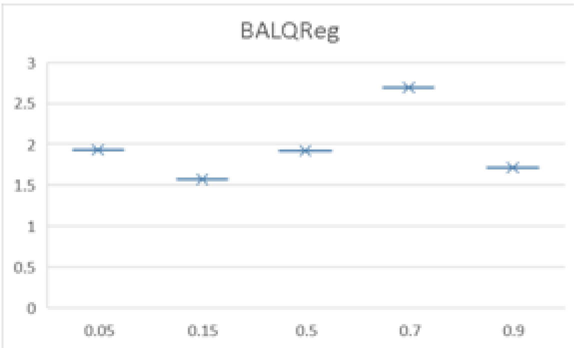


Figure (15): BLQReg method when ρ=(0.05, 0.15, 0.50, 0.70, 0.90)

Figure (16): BALQReg method when ρ=(0.05, 0.15, 0.50, 0.70, 0.90)

Figures (13),(14),(15), and (16) demonstrate that the MSE values' spread points were unstable, there was no clear trend for the MSE values of each technique individually, whether it was ascending or descending.The QReg model began at a low level, increased somewhat, and then gradually settled

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

in a straight line, unaffected by changes in the MSE value it is the worst performer. Although the BLQReg technique is the most stable, it wasn't the finest among its peers. BALQReg has the lowest MSE and is the highest-performing method overall.

Figure (17) depicts a complete overview of all the methods.

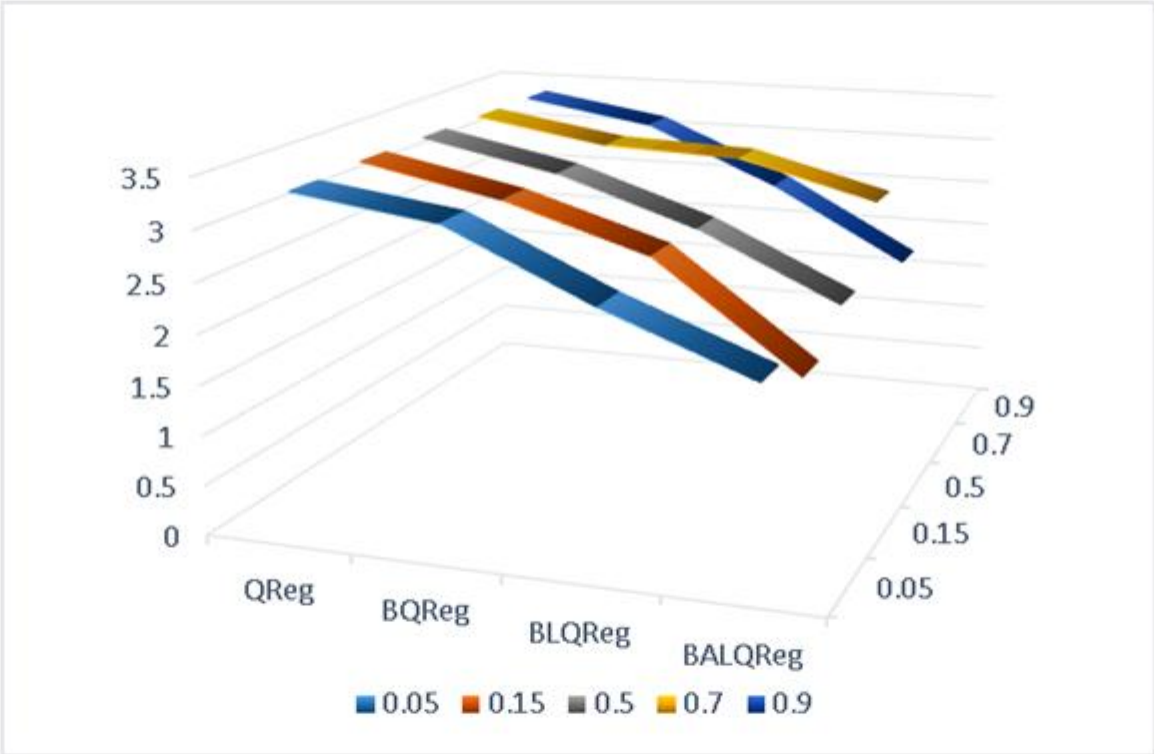


Figure (17) : Shows whenMSE for methods (QReg, BQReg, BLQReg, and BALQReg),S = 10,and ρ is (0.05, 0.15, 0.50, 0.70, 0.90).

Figure (17) indicates that among the (QReg, BQReg, BLQReg, and BALQReg) methods, BALQReg had the lowest MSE values at all levels, making it the best performing technique when S = 10 and ρ = (0.05, 0.15, 0.50, 0.70, 0.90).

6.4 Simulation 4 (Without Sparse)

We created 150 views of the model in this simulation study.

$$y_{i=} x_i^{'}\beta + \xi_i$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

Where x_i is a vector of (20) predictor. Where S=20 represents the number of affecting or relevant variables (1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1). In Simulation 4, we investigate the effect of correlation (ρ) on the performance of the techniques under examination (QReg, BQReg, BLQReg, and BALQReg).

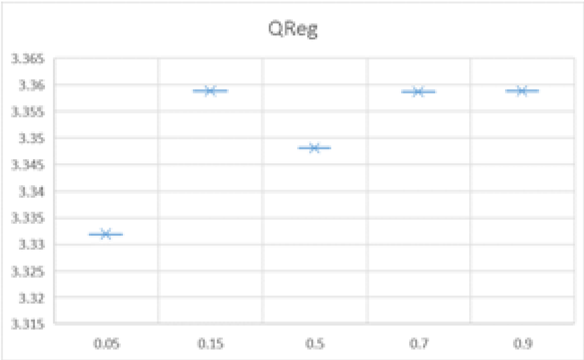
??	*QReg*	*BQReg*	*BLQReg*	*BALQReg*
0.05	3.3319	3.1177	1.6089	1.4958
0.15	3.3588	3.0289	1.9811	1.5944
0.50	3.3481	3.0088	2.1266	1.8940
0.70	3.3587	3.0719	2.0263	1.9849
0.90	3.3588	3.0878	2.1213	2.0853

Table (4) Shows MSE for methods (QReg, BQReg, BLQReg, and BALQReg) when S= 1,ρ is (0.05, 0.15, 0.50, 0.70, 0.90).

Looking at Table (4), we can see that the researched methods have adopted two positions in dealing with the percentage increase ρ. Because the MSE values for the two approaches (QReg, and BQReg) are irregular, they generated a semi-uniform position by not following a precise pattern such as growing or decreasing with the change of ρ values.

The two approaches (BLQReg, and BALQReg), on the other hand, chose a different position because they were impacted by the increase of ρ and took a semi-regular pattern directly related to it. As is customary, the technique (BALQReg) took first place as the best method since it lowest MSE.

The Graphs below demonstrate the distribution of MSE values for each method separately.



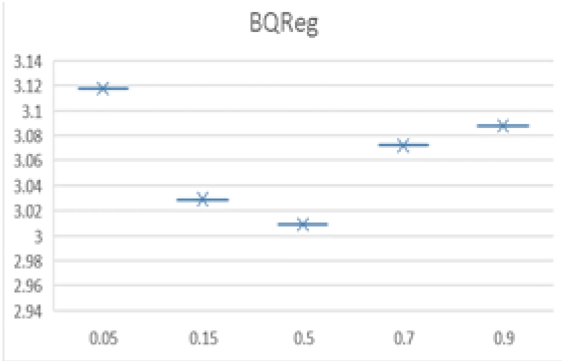
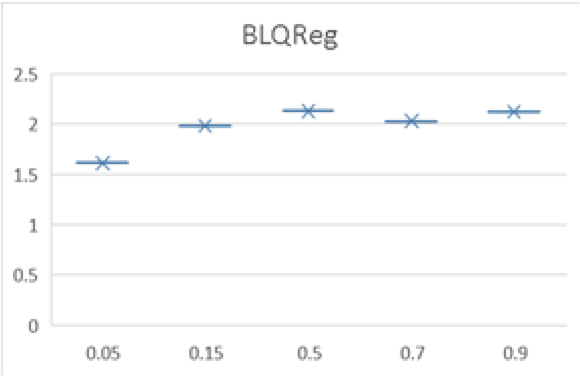


Figure (18): QReg method when ρ=(0.05, 0.15, 0.50, 0.70, 0.90)

Figure (19): BQReg method when ρ=(0.05, 0.15, 0.50, 0.70, 0.90)



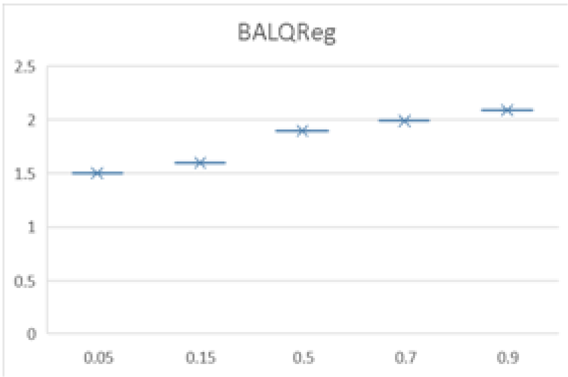


Figure (20): BLQReg method when ρ=(0.05, 0.15, 0.50, 0.70, 0.90)

Figure (21): BALQReg method when ρ=(0.05, 0.15, 0.50, 0.70, 0.90)

Figures (18), and (19) illustrate that the spread points in the two methods (QReg, BQReg) were irregular, as they did not create a precise pattern by which we may explain their impact by raising or lowering the proportion of ρ.

Figure (21) demonstrates that the MSE of the BALQReg approach increases when the ratio ρ increases, despite having the lowest MSE. The BLQReg approach is nearly identical to the prior method, with a substantial convergence of MSE values. The methods(QReg,BQReg,BLQReg, and BALQReg) are illustrated in the graph below for a more detailed look.

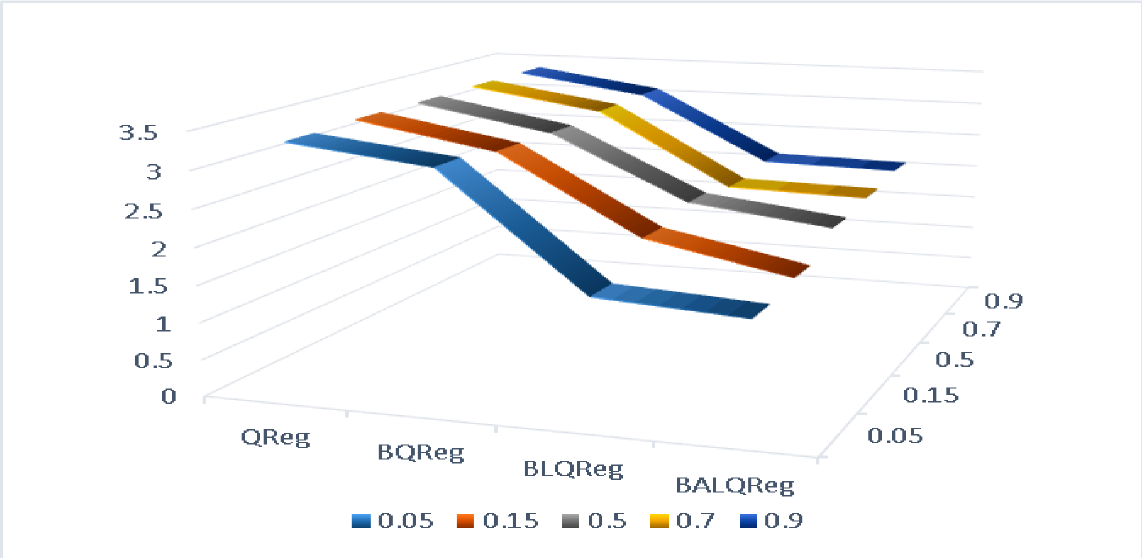


Figure (22): Shows whenMSE for methods (QReg, BQReg, BLQReg, and BALQReg) S =20, and ρ is (0.05, 0.15, 0.50, 0.70, 0.90).

Figure (22) depicts that the QReg method had the highest MSE values at all levels ρ, making it the least effective of the methods tested (BQReg, BLQReg, and BALQReg). BALQReg technique, on the other hand, recorded the lowest MSE values at all values of ρ, demonstrating to be the best performing when S = 20,ρ=(0.05, 0.15, 0.50, 0.70, 0.90).

Through what was presented in the simulation study, we discovered that BALQReg approach is the best performing method among the evaluated ways based on the simulation results. It was followed by BLQReg, which had the best approach in simulation 1 but fell to second place in the remaining cases.

In all circumstances, BQReg technique moved away from its peers to get extremely near to the final rank, whilst QReg method stayed at the last position with the greatest MSE in all instances.It is worth noticing that the MSE value of QReg method has stayed constant on several occasions, which leads to one of two interpretations: either it was unaffected by the change in proportions ρ, or it was unable to deal with the issue.

7 .Thalassemia disease

Thalassemia is a genetic disorder condition characterized by low hemoglobin production. Hemoglobin is known to be the oxygen-carrying component of red blood cells. Where it is made up

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

of two proteins, alpha, and beta. If the body does not produce enough of one or both of these proteins, red blood cells do not develop properly and are unable to carry enough oxygen, resulting in anemia that begins in childhood and lasts a lifetime. Iron overload is one of the disease's complications as well as Splenomegaly, osteoporosis, heart and liver problems. It is reported that thalassemia is a hereditary condition, which means that at least one parent must be a carrier. It is caused by a genetic mutation or the loss of specific essential genetic components *(Bajwa, H., and Basit, H. 2019).*

Blood tests, such as a complete blood count, specific hemoglobin tests, and genetic testing, are frequently used to make a diagnosis. It can also Prenatal testing may be used to make a diagnosis before delivery. The pre-marital tests are the most important to prevent the transmission and spread of the disease because it detects early whether a person who is about to marry is infected or a carrier of the infection *(Alswaidi, F. M., and O'brien, S.2009).*

As of 2015, around 280 million individuals have thalassemia, with approximately 439,000 critically sick. It is more frequent among persons of Italian, Greek, Turkish, Middle Eastern, South Asian, and African origin. Males and females have equal infection rates. This resulted in 16,800 fatalities in 2015, down from 36,000 in 1990 *(Unissa, R., et al.2018).*

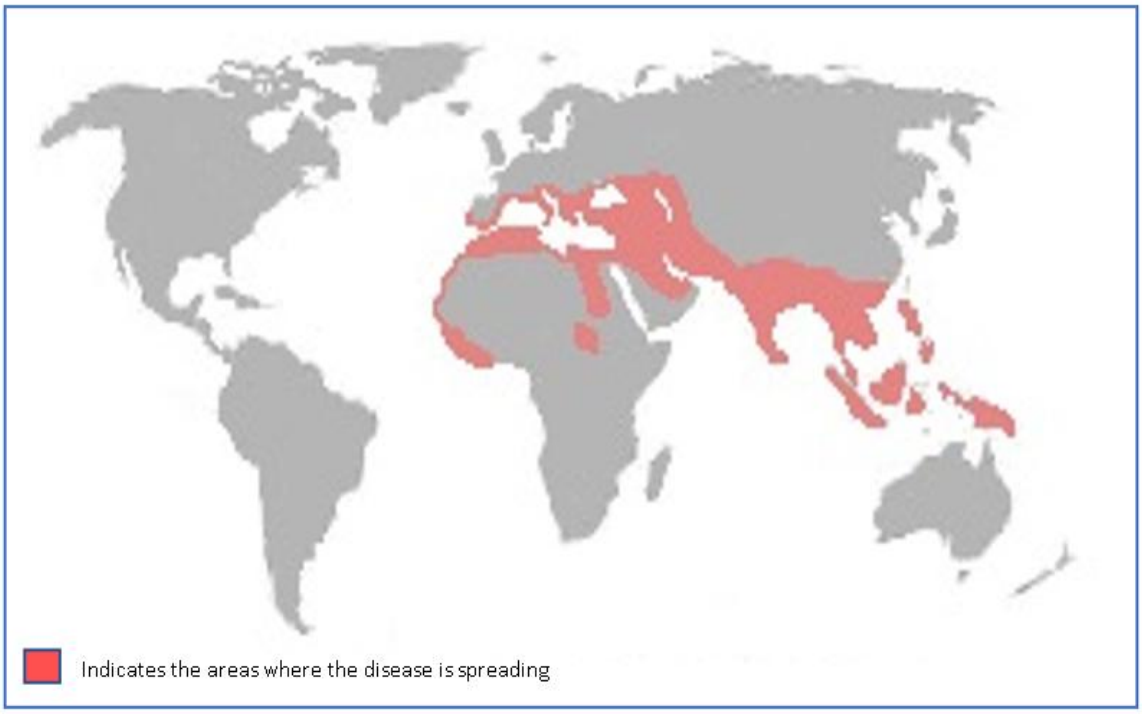


Figure (23): Shows the areas in which thalassemia is common in the world

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

7.1 Real Data

To compare the methods used in our study using real-world data. Thalassemia statistics for the year 2021 were obtained from individuals getting treatment at the Center for Genetic Blood Diseases in Babylon Governorate. the study sample contained data from 150 patients based on (25) variables gathered. As indicated in Table (5).

Table (5): Shows the studied variables and their details.

Variables	*Variables description*	*Rank*	*Rank description*
X_1	*Age*	Numeral	*years*
X_2	*Blood Group (BG)*	1.	*O*
		2.	*A*
		3.	*B*
		4.	*AB*
X_3	*Rhesus factor (Rh)*	1.	*Positive*
		2.	*Negative*
X_4	*Housing environment*	1	*City*
		2	*Rural*
X_5	*prothrombin time test* (*PT*)	Numeral	
X_6	*partial thromboplastin timetest* (*PTT*)	Numeral	
X_7	*Glutamate- pyruvic transaminase (GPT*)	Numeral	
X_8	*Glutamic-oxaloacetic transaminase (GOT)*	Numeral	
X_9	*Ferritin*	Numeral	
X_{10}	*Hemoglobin Blood Test (HB)*	Numeral	
X_{11}	*White Blood Cell (WBC)*	Numeral	
X_{12}	*Red Blood Cell (RBC)*	Numeral	
X_{13}	*Blood Urea Test (B.Urea)*	Numeral	
X_{14}	*Packed Cell Volume (PCV)*	Numeral	
X_{15}	*Alkaline Phosphatase (ALP)*	Numeral	

X_{16}	*Platelet Count Test (PLT)*	Numeral	
X_{17}	*Alanine aminotransferase (ALT)*	Numeral	
X_{18}	*Hepatitis Test(HBV)*	1.	*Negative*
		2.	*Positive*
X_{19}	*Hepatitis C Virus (Anti HCV)*	1.	*Negative*
		2.	*Positive*
X_{20}	*Human Immunodeficiency Virus Test (HIV)*	1.	*Negative*
		2.	*Positive*
X_{21}	*Platelet Count (plat. C)*	Numeral	
X_{22}	*Total Serum Bilirubin Test (TSB)*	Numeral	
X_{23}	*Calcium Test (S. Ca)*	Numeral	
X_{24}	*One or both parents is a carriers of the genetic trait*	Numeral	
X_{25}	*The number of blood transfusions per month*	Numeral	

7.2 Real data results:

By representing the estimations in Table (6), which were carried out by the four studied methods (QReg, BQReg, BLQReg, and BALQReg). We reached results that differ from one method to another, some with a positive result and the other with a negative sign. To conclude the best method in terms of Applied (practical), we have adopted the medical opinion in determining the results closest to reality, and therefore the way that reaps the most results close to reality is the best. The following outcomes were achieved by employing the packages for each method in the programming language (R).

Table (6): Parameters estimated by the studied methods (QReg, BQReg, BLQReg, and BALQReg)

Covariate	Methods			
	QReg	BQReg	BLQReg	BALQReg
X_1	0.7535	0.3568	-1.2754	-0.3797
X_2	-0.8922	-0.7496	1.2566	1.4514

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

X_3	-0.2393	-0.1699	-0.3732	-0.9683
X_4	-1.9895	-1.7169	2.1163	4.3636
X_5	1.7710	1.9382	1.8824	1.6198
X_6	0.5186	0.8195	-0.1742	-0.2686
X_7	0.6999	0.8534	0.1364	0.0479
X_8	2.6095	1.7247	0.1332	0.1186
X_9	-1.3949	-1.5308	-0.0010	-0.0017
X_{10}	0.1315	0.5600	-0.1161	-1.2365
X_{11}	-1.3982	-1.0283	-0.1049	-0.2842
X_{12}	-0.9992	-1.2082	1.2433	1.4255
X_{13}	1.9572	2.2144	-2.9063	-4.4455
X_{14}	0.5373	0.2181	0.6609	-0.1398
X_{15}	-0.4026	-0.0908	0.0234	0.0371
X_{16}	0.5721	0.8887	0.0163	0.0136
X_{17}	-0.7015	-0.4871	0.1910	0.2011
X_{18}	0.0842	0.0829	0.8983	3.1949
X_{19}	-0.6005	-0.3743	-0.2195	-5.1802
X_{20}	-0.2277	-0.2227	0.2223	-0.9752
X_{21}	0.3382	-0.1197	-0.0005	0.0018
X_{22}	-0.0422	-0.0962	0.3611	-0.1157
X_{23}	0.7959	0.3107	1.6516	3.6780
X_{24}	-0.3399	-0.4879	-0.2929	-1.0956
X_{25}	-1.3268	-1.2202	1.9492	2.0466

At first glance, the results of methods (BLQReg, BALQReg) appear to act similarly in Table (6). Where it took a uniform behavior about the increase or decrease in the values of the variables, while the methods (QReg, BQReg) are on the opposite side of (BLQReg, BALQReg). These situations were documented when the variables (X_1, X_2, X_4,X_6,X_{10},X_{12},X_{13}, X_{15}, X_{17},X_{25}).In (11) caseswhen variables(X_3, X_5, X_4,X_7,X_8,X_9,X_{11}, X_{16}, X_{18},X_{19},X_{23}, X_{24}), the similarity of all the studied methods was indicated in terms of their impact on the response variable, whether all of them had a direct effect or

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

of them had a reverse effect.The situation is different with the remaining variables (X_{14}, X_{20}, X_{21}, X_{22}) since there is no definite path that the methods agreed on in their behavior.

In general, when the results of the estimates utilizing the examined approaches (QReg, BQReg, BLQReg, and BALQReg) are compared to the medical opinion, it is obvious that the Bayesian quantile regression methods outperformed the non-Bayesian quantile regression methods. In a more precise sense, the BALQRegapproach outperformed the other investigated methods.

The above is illustrated by the graph (24).

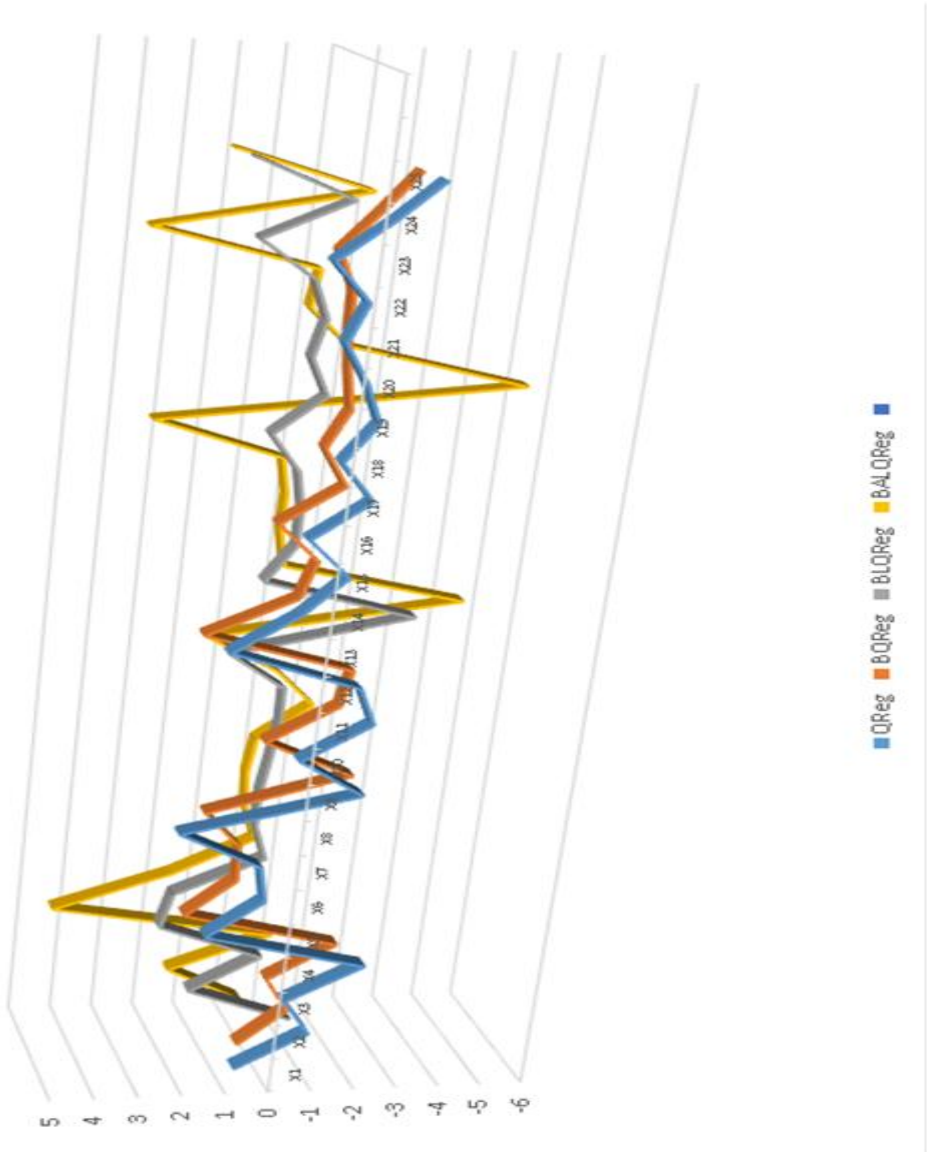


Figure (24) Shows estimated parameters of variables using methods (QReg, BQReg, BLQReg, and BALQReg)

From Figure (24), we notice that the two methods (QReg, and BQReg) behave convergently, that is, they are in a semi-uniform position in front of the relationship with the response variable, either of them being a direct or inverse relationship. While the methods (BLQReg, BALQReg) behave similarly, there is a difference between them and differences in some cases, especially for the BALQReg method, which often tweets alone outside the flock and is unique in its distinguished position in most cases.

Returning to the medical opinion that we adopted as a measurement by which we determine the accuracy of our study, in which it shows the relationship between the independent variables and the response variable in terms of being a direct or inverse relationship. It shows us that the (BALQReg) method is the most efficient and accurate, and therefore it is the best way to express that. While QRegmethod was the worst and least accurate method for expressing the relationship between the independent variables and the dependent variable, as it moved away from reality in many cases.

7.3 Mean Squared Error (MSE)

MSE has been used to evaluate the efficacy of the methods under study (QReg, BQReg, BLQReg, and BALQReg). The method that achieves the least MSE is the method with the least errors and therefore the most accurate than others, then it will be the best method. The MSE for each approach is explained in the table below (7).

Measurement	Methods			
	QReg	BQReg	BLQReg	BALQReg
MSE	1.5271	1.4671	1.5196	0.9566

Table (7): Displays the MSE for methods (QReg, BQReg, BLQReg, and BALQReg)

An illustrated graph has been created (25)to explain the abovetable.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5078-5105
https://publishoa.com
ISSN: 1309-3452

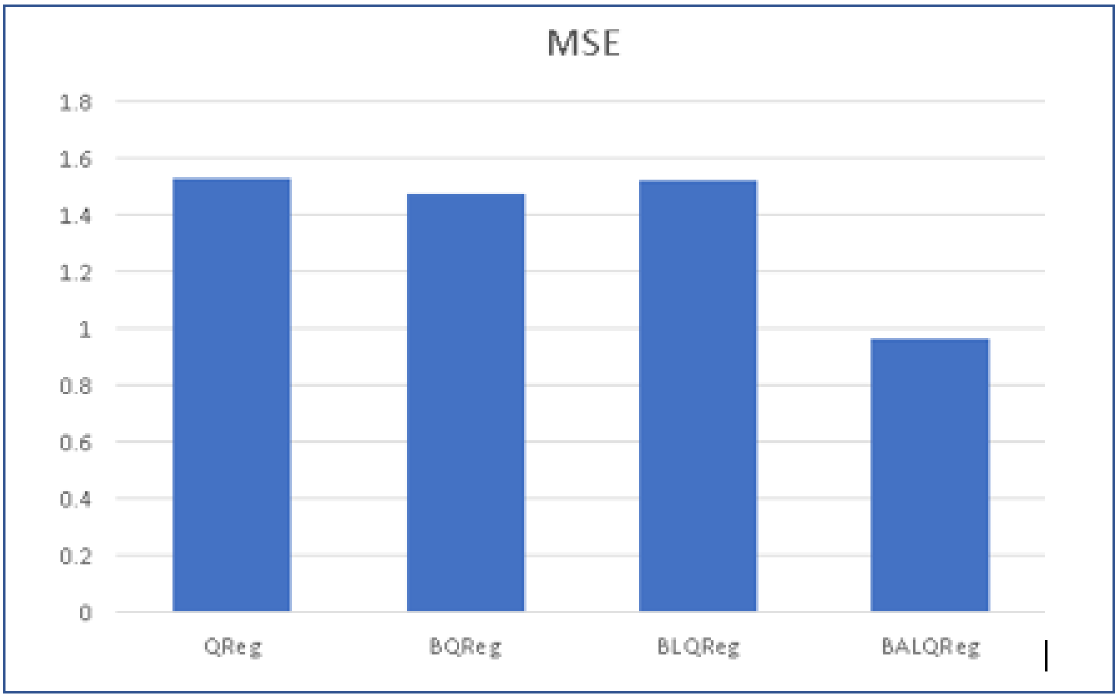


Figure (25): Shows MSE for methods (QReg,BQReg,BLQReg, andBALQReg)

Through Table (7) and Figure (25), which show the MSE for each method, it is very clear that BALQRegapproach has an attractive performance, as it is the method with the lowest MSE and therefore it is the most accurate and efficient method among the studied methods. Followed by the method BQReg by a relatively large margin. While the other approaches (QReg, BLQReg) performed poorly.

8. Conclusion

Four methods for selecting the appropriate number of iterations for sparse, very sparse, and Without Sparse selection are discussed with the methods addressed in our study (QReg, BQReg, BLQReg, and BALQReg) through the simulation study. We also compared the studied methods in terms of estimating parameters and mean error squares MSE it was applied to real-world data.In addition, we based our assessment on medical opinion.

Bayesian methods in QReg can be a compelling option to obtain more accurate results and as well expand the set of methodological tools for regression analysis in biomedical research.

Where they appear both the simulation studies and the data analysis demonstrate that the BALQRegworks well and may be chosen over the present Bayesian and non-Bayesian methods. We suggest the use of the BALQReg method in the analysis of clinical data or other fields and we recommend its use in subsequent studies.

References:

1- Alhamzawi, R., Yu, K., and Benoit, D. F. (2012). "Bayesian adaptive Lasso quantile regression." Statistical Modelling, 12(3): 279–297.

2- Alhamzawi, R. (2016). Bayesian analysis of composite quantile regression. Statistics in Biosciences 8 (2), 358–373.

3- Alhamzawi, R., & Ali, H. T. M. (2018). The Bayesian adaptive lasso regression. Mathematical biosciences, 303, 75-82.

4- Alswaidi, F. M., &O'brien, S. J. (2009). Premarital screening programs for hemoglobinopathies, HIV and hepatitis viruses: review and factors affecting their success. Journal of medical screening, 16(1), 22-28.

5- Bajwa, H., & Basit, H. (2019). Thalassemia.

6- Buchinsky, M. (1994). Changes in the U.S. wage structure 1963-1987: Applications of quantile regression. Econometrica, vol. 62, pp. 405-58

7- Chen, M.-H. and Ibrahim, J. G. (2003). "Conjugate Priors for Generalized Linear Models." StatisticaSinica, 13: 461–476. 585, 586, 587, 588, 608.

8- Das, P. (2016). Bayesian Quantile Regression.

9- Davino, C., Furno, M., &Vistocco, D. (2013). Quantile regression: theory and applications (Vol. 988). John Wiley & Sons.

10- Gosling, A., Machin, S., and Meghir, C. (2000). The changing distribution of male wages in the U.K., 1966-1992. Review of Economic Studies, vol. 67, pp. 635-66.

11- Koenker, R. and Bassett, G. (1978). Regression quantiles.Econometrica,46(1):35–50.

12- Koenker, R. and d'Orey, V. (1993). Computing regression quantiles. J. Roy. Statist. Soc. Ser. C (Appl. Statist.), 43, 410-414.

13- Mazucheli, J., Leiva, V., Alves, B., & Menezes, A. F. (2021). New quantile regression for modeling bounded data under a unit Birnbaum–Saunders distribution with applications in medicine and politics. Symmetry, 13(4), 682.

14- Noufaily, A., & Jones, M. C. (2013). Parametric quantile regression based on the generalized gamma distribution. Journal of the Royal Statistical Society: Series C (Applied Statistics), 62(5), 723-740.

15- Park, Trevor,and Casella, George (2008). The Bayesian Lasso, Journal of the American Statistical Association, 103:482, 681-686.

16- Phadkantha, R., Yamaka, W., &Sriboonchitta, S. (2019, January). Analysis of herding behavior using Bayesian quantile regression. In International Econometric Conference of Vietnam (pp. 795-805). Springer, Cham.

17- Peter Radchenko& Gareth M James (2008) Variable Inclusion andShrinkage Algorithms, Journal of the American Statistical Association, 103:483,1304-1315.

18- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B 58, 267–288.

19- Unissa, R., Monica, B., Konakanchi, S., Darak, R., Keerthana, S. L., & Kumar, S. A. (2018). Thalassemia: A Review. Asian Journal of Pharmaceutical Research, 8(3), 195-202.

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