Wasit Journal for Pure Science

Journal Homepage: https://wjps.uowasit.edu.iq/index.php/wjps/index
e-ISSN: 2790-5241 p-ISSN: 2790-5233

Simulation Study for Variable Selection in Quantile Regression

Hussein A. Hashem

Department of Mathematics, College of Science, University of Duhok, Kurdistan Region, IRAQ.

DOI: https://doi.org/10.31185/wjps.607
Received 20 November 2024; Accepted 02 December 2025; Available online 30 March 2025

ABSTRACT: Six essential techniques for variable selection in quantile regression models are thoroughly examined in this work. Quantile LASSO (qLASSO), gamma-divergence, quantile elastic net (qENet), quantile adaptive LASSO (qaLASSO), quantile SCAD (qSCAD), and quantile MCP (qMCP) are some of these techniques. The study's main objective is to compare and assess various approaches' performance, particularly in relation to their capacity to reduce median model error. An important part of the analysis includes situations in which there are more possible candidate variables than there are data. Through simulation analysis, the study also aims to identify the best practices for each and every linear regression situation.

Keywords: LASSO, SCAD; MCP, Gamma-Divergence, Regularization.

1. INTRODUCTION

In high-dimensional data analysis, variable selection is essential since it aids in determining the most significant aspects. Due to the substantial computing burden required, traditional variable selection techniques, such as best subset selection algorithms, present difficulties when used with high dimensional data [1].

Carefully choosing the predictors is crucial when dealing with an issue that has a lot of variables in the dataset. The use of sparse representation to address regression problems has been extensively studied in recent years. This study has its roots in Tibshirani's seminal work from 1996[2]. He presented the least absolute shrinkage and selection operator (LASSO) in his paper as a penalization technique that can carry out parameter estimation and variable selection at the same time. An L1-regularized least squares estimate of the model parameters is called LASSO estimation.

However, because LASSO tends to use individual dummy variables rather than the complete predictor, it might not produce adequate results when working with categorical predictors in the regression model. Moreover, it frequently chooses one variable from a group while ignoring the rest because it does not do grouped selection. Numerous extensions and solutions have been proposed in the literature to get over these restrictions, like the Elastic Net that was first presented by (Zou and Hastie, 2005) [3]. In general, the Elastic Net performs better than LASSO regularization, especially when correlated predictors are included, and it penalizes the model using both the L1 and L2 norms. Furthermore, the Elastic Net encourages a grouping effect, which tends to include or exclude highly correlated predictors from the entire model.

Conventional statistical techniques like likelihood-based approaches or simple least squares may produce skewed estimates and sometimes misleading conclusions when there are outliers or contamination in a dataset. Although variable selection for resilient linear models has been the subject of some research, little is known about this topic. Fan and Li (2001) [4], for instance, noted that the least squares estimate is not robust and recommended that more robust estimators for β be obtained by employing an outlier-resistant loss function, such as Huber's $\rho(.)$ function or L1 loss. Their simulated simulations showed that this method works well for robust regression. Furthermore, Li and Zhu (2008) [5] disclosed the solution route of the L1 penalized quantile regression, while Wu and Liu (2009) [6] illustrated the oracle qualities for the SCAD and adaptive LASSO penalized quantile regressions.

Roger Koenker developed the regression analysis technique known as Quantile Regression in 1978 [7]. By addressing the shortcomings of the conventional least squares approach in the presence of outliers and heteroscedasticity, it offers a more thorough and reliable representation of the data. Quantile Regression avoids biased or incorrect estimations, provides better logical interpretations, and captures the tails of the dependent variable's distribution. When used in conjunction with the least squares approach, it can improve statistical problem solving and offer insights on the applicability of least squares estimation.

*Corresponding author: hussein.hashem@uod.ac

Regularization in Quantile Regression has been investigated by scholars like, Yan and Song (2019) [8], Alhamzawi et al. (2012) [9] and Ajeel and Hashem (2020) [10] are other scholars that have investigated this subject from a Bayesian perspective.

We will provide a comprehensive overview of the different approaches and techniques used for variable selection in the context of linear regression below in this paper.

2. METHODS

Consider the linear regression model

$$y = \boldsymbol{X}\beta + \epsilon \quad (1)$$

The error terms ϵ have a mean of zero and finite variance σ^2, and they are distributed independently and identically.
A model fitting procedure produces the vector of estimated coefficients

$\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)^T$.

The ordinary LSE are obtained by minimizing the residual sum of squared errors

$$\hat{\beta}_{LSE} = \underbrace{min}_{\beta} = \{(Y - X\beta)^T(Y - X\beta)\} \quad (2)$$

Although the least squares estimates are simple to compute, Tibshirani (1996) [2] identified two primary drawbacks. First, even if only a subset of predictors may have the greatest impact on the response variable Y, all least squares estimations are not zero. Second, least squares estimates frequently result in poor prediction accuracy due to their propensity for low bias and large variance. A considerable overall gain in prediction accuracy can be achieved by accepting a little increase in bias in order to lower the variance of the anticipated values. In addition to these drawbacks, when the number of linear predictors (p) surpasses the sample size (n), least squares estimates are totally ineffective.

2.1 Ridge Regression

Originally introduced by Hoerl and Kennard in 1970 [11], ridge regression is well recognized as a practical alternative to ordinary least squares (OLS) regression. This method works particularly well when predictor variables have a high degree of correlation. To get more accurate estimates for the regression parameters, it applies an L2 penalty to the coefficients and minimizes the sum of squared residuals. The following formula is used to calculate the ridge estimate:

$$\hat{\beta}_{ridge} = \min_{\beta}\{\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda\sum_{j=1}^{p}\beta_j^2\} \quad (3)$$

The positive scalar $\lambda \geq 0$ is a regularization parameter that determines the level of shrinkage, and the penalty function is defined by the L2-norm.
Hoerl and Kennard (1970) [20] recommended the utilization of all accessible variables to acquire estimates.

$\hat{\beta}_{ridge} = (X^T X + \lambda I_p)^{-1} X^T y$.

where I is the $p \times p$ identity matrix. By adding λI_p to $X^T X$, this results in a regular and invertible matrix. The intercept β_0 is usually not included in the penalty.

The ridge estimator adds bias to ordinary least squares (OLS) estimation. This approach, however, permits the inclusion of a moderate degree of bias in order to reduce variance and mean squared error, which may result in increased prediction accuracy. The ridge trace, a graphical tool for figuring out the ideal value of the regularization parameter lambda (λ), was created by Hoerl and Kennard in 1970 [11].

Ridge regression shrinks regression coefficients to produce more reliable estimates. Ridge regression does not decrease coefficients to exactly zero like some other techniques do, which makes the final model more difficult to understand.

2.2 LASSO Regression

The LASSO penalty was introduced by Tibshirani (1996) [2] as a method of regularization for carrying out simultaneous estimation and variable selection in large datasets. The LASSO estimate $\hat{\beta}$ is defined by:

$$\hat{\beta}_{lasso} = \min_{\beta}\left\{\sum_{i=1}^{n}\left(y_i - \sum_j \beta_j x_{ij}\right)^2 + \lambda\sum_j|\beta_j|\right\}, \quad (4)$$

or

$$\hat{\beta}_{lasso} = \min_{\beta}\|y - x\beta\|_2^2 + \lambda\|\beta\|_1$$

The parameter lambda determines the trade-off between minimizing the penalty term, which is the sum of the absolute values of the coefficients, and the residual sum of squares (RSS).

The LASSO method makes sure that the total of the coefficients' absolute values is less than a constant while minimizing the sum of squared residuals. Regression models with a lot of variables and not many observations frequently employ it. Choosing variables while fitting the regression line to the data is LASSO's main goal. This is accomplished by setting certain coefficients to zero and shrinking others. LASSO adds a penalty to the optimization goal in order to do L1 regularization.

2.3 Adaptive LASSO

A modified version of the LASSO technique was created by Zou (2006) [12] to estimate and choose variables at the same time. This sophisticated technique, known as the Adaptive LASSO, applies penalties to various coefficients inside the L1 penalty by integrating adaptive weights. The coefficient estimates (β_js) provided by the Adaptive LASSO technique is modified according to the particular characteristics of the data.

$$\hat{\beta}_{\text{adaptive LASSO}} = \underbrace{min}_{\beta} (Y - X\beta)^T (Y - X\beta) + \lambda \sum_{j=1}^{p} w_j |\beta_j|, \quad (6)$$

$w = \{w_1, w_2, \ldots, w_p\}$is a known weights vector used in the weighted LASSO method. If the weights are correctly chosen depending on the data, the weighted LASSO can display oracle features. This implies that it can function as well as though we were already aware of the actual underlying model. Moreover, the oracle technique is optimal since the adaptive LASSO solution is continuous by definition. Finally, the estimate generated by the Adaptive LASSO shrinkage is nearly minimax-optimal.

2.4 Elastic Net

The elastic net approach, a kind of regularization technique used in statistical modeling and machine learning, was proposed by Zou and Hastie (2005) [3]. By minimizing a particular objective function that incorporates the $L_1$and L_2 penalties, the elastic net estimator is produced, encouraging both group selection and sparsity. When dealing with multicollinearity and locating significant predictors in high-dimensional data, this approach is quite helpful. The definition of the elastic net criterion is

$$\hat{\beta}_{elastic\ net} = \min_{\beta} \{ \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \} , \quad (7)$$

Which depends on two regularized parameters $\lambda_1, \lambda_2 > 0$.

A convex combination of LASSO and Ridge regression is known as the Elastic Net. The Elastic Net reduces to the LASSO when $\alpha = 0$. Two steps must be taken when α is between 0 and 1. Prior to using the LASSO, the ridge regression coefficients are determined for every fixed λ_2.

Using predictor variables, the ordinary least squares (OLS) regression calculates the average response. Nevertheless, the median function is estimated using a different technique called least absolute deviation (LAD) regression. Due to its enhanced robustness, LAD regression is especially helpful when response outliers and heavy-tailed errors are present.

2.5 Quantile Regression

Quantile Regression (QR) was created by Koenker and Bassett (1978) [7] as an expansion of LAD regression. By estimating the response's conditional quantile function, QR provides a thorough understanding of the response variable's conditional distribution. QR provides a more informative model overall while retaining the advantageous features of LAD regression.

By adjusting the quantile parameter θ, Quantile Regression provides a versatile and thorough method for simulating the relationship between response variables and predictors. Notably, quantile regression, which is renowned for its resistance to outliers, is equal to least absolute deviation regression or median regression when θ is 0.5. This method is favored in these circumstances since it is well known for its resilience to outliers and for estimating the conditional quantiles of a response variable. Since both median and least absolute deviation (LAD) regression aim to reduce the absolute differences between the data's actual values and its predicted values, they are practically interchangeable. Whereas median regression seeks to identify the line (or hyperplane in higher dimensions) that minimizes the absolute deviations of the data points from a central point, the median, LAD regression seeks to minimize the sum of the absolute values of the residuals (differences between predicted and actual values).

When the assumptions of least squares regression are not met or when a more thorough comprehension of the relationship between variables across several regions of the conditional distribution is required, quantile regression's potent utility is a significant benefit. However, its interpretation and computational elements need careful thought. In reality, a minimization problem may be solved to reliably estimate the coefficients, yielding accurate parameter estimates across different quantiles of interest.

$$\min_{\beta} \sum_{i=1}^{n} \rho_\theta (y_i - x_i^T \beta)$$

Where $\rho(.)$ the objective function refers to an outlier–resistant loss function used in various optimization problems. It is particularly robust in the presence of outliers, making it a valuable tool for tasks such as regression and machine learning.

$$\rho_\theta(t) = \begin{cases} \theta t & if \quad t \geq 0 \\ -(1-\theta)t & if \quad t < 0 \end{cases}, \text{ where } 0 < \theta < 1. \quad (8)$$

Under Koenker's direction, regularization was applied to Quantile Regression for the first time in 2004 [13]. In order to manage random effects in a mixed-effect Quantile Regression framework, the LASSO penalty was introduced in this groundbreaking study. The objective was to use the LASSO method's regularization properties to lessen the random effects in the direction of zero. This novel approach offered a fresh approach to managing model complexity and

improving estimation accuracy in mixed-effect Quantile Regression models, marking a substantial advancement in the field.

2.6 Gamma-Divergence

In their paper, Fujisawa and Eguchi (2008) [14] first proposed the idea of gamma divergence for regression. The variance difference between two conditional probability density functions is assessed using this metric. In their alternative version of gamma divergence for regression, Kawashima and Fujisawa (2017) [15] made changes to the way the base measure on the explanatory variable is handled. A detailed review of gamma divergences for regression and an investigation of the associated parameter estimation as outlined by Fujisawa and Eguchi (2008) [14] will be covered in the next section.

2.7 Smoothly Clipped Absolute Deviation (SCAD)

The best way to comprehend the SCAD penalty, as presented by Fan and Li (2001) [4], is to look at its first derivative.

$$p'_\lambda(\beta) = \lambda\left\{I\{\beta \le \lambda\} + \frac{(a\lambda-\beta)+}{(a-1)\lambda} I\{\beta > \lambda\}\right\} \text{ for some } a \text{ and } \beta > 0, \quad (9)$$

The letter I stands for the indicator function in the SCAD technique, β for a vector of unknown parameters, and λ for the regularization parameter.

The approach described in this response involves setting all less significant variables to zero in order to produce simpler, easier-to-manage models. Fan and Li (2001) [4] demonstrated that the SCAD penalty may produce estimates with the required oracle quality. This suggests that non-zero coefficient estimate is as accurate as it would be if the correct model had been known beforehand. Moreover, a true parameter is roughly equal to zero, and its probability is much guaranteed when it is zero.

Numerous search criteria, such as the Bayesian Information Criterion (BIC), generalized cross-validation, and cross-validation, can be used to locate the two tuni0ng parameters (λ,a).

According to Fan and Li (2001) [4], a = 3.7 should be selected as a suitable value for one of the tuning parameters.

According to Fan and Li (2001) [4], a good penalty function should yield an estimator with three crucial characteristics. The first feature is unbiasedness, which ensures that there is no unnecessary modeling bias by avoiding an excessive penalty of big parameters in the final estimator. Additionally, the estimator should indicate sparsity by automatically setting unimportant parameters to 0. Last but not least, continuity is essential because if the final estimate can demonstrate continuity in the data, model prediction instability may be prevented.

2.8 Minimax Concave Penalty (MCP) Method

Regression analysis uses the Minimax Concave Penalty (MCP), a statistical method, to solve the problem of bias in sparse models. Zhang (2010) [16] established the MCP technique, which uses the MCP penalty function to choose variables in linear regression models. Since it lessens the issue of inconsistent variable selection that LASSO frequently has, it is thought to be an improvement over LASSO. The following formula can be used to produce the MCP estimator.

$$\hat{\beta}_j^{MCP} = \underset{p}{\underbrace{argmin}} \left[\|Y - X\beta\|^2 + \sum_{j=1}^{p} P_{\lambda,\gamma}^{MCP}\right] \quad (10)$$

Where: $\sum_{j=1}^{p} P_{\lambda,\gamma}^{MCP}$ the MCP penalty function.

The MCP function has the following format:

$$p_\lambda(\lceil\beta\rceil) = \begin{cases} \lambda\left(|\beta| - \frac{|\beta|^2}{2\lambda\gamma}\right), & \lceil\beta\rceil < \lambda\gamma \\ \frac{\lambda^2\gamma}{2}, & \lceil\beta\rceil \ge \lambda\gamma \end{cases} \quad (11)$$

Where: $\gamma > 1$

Numerous penalty functions are used in statistical regression models, and these functions are usually concave. Typically, they rely on a tuning parameter (λ) and furthermore incorporate another tuning parameter (γ) that regulates the penalty's level of concavity. The penalty's rate of decline depends critically on this parameter (γ).

The adaptive LASSO (least absolute shrinkage and selection operator) and MCP approaches enable the estimated coefficients to develop faster than with the classic LASSO approach, especially for nonzero coefficients. While the goal of most of these methods is to shrink coefficients towards zero, the adaptive LASSO and MCP methods reduce the amount of shrinkage that is applied to nonzero coefficients, suggesting reduced bias in the estimate process. Numerous penalty functions are used in statistical regression models, and these functions are usually concave.

Typically, they rely on a tuning parameter (λ) and furthermore incorporate another tuning parameter (γ) that regulates the penalty's level of concavity. The penalty's rate of decrease is largely dependent on this parameter (γ).

The MCP (minimally concave penalty) function has an intriguing property in that it covers a range of values where all estimations stay constant. Interestingly, the estimates are exactly the same inside this range as when derived with the least squares regression technique.

3. SIMULATION STUDY

In the upcoming section, we will conduct a comprehensive comparison of various Quantile Regression methods in different scenarios. To start with, we will delve into small-dimensional scenarios featuring both sparse and non-sparse parameters. Subsequently, we will explore large-dimensional settings with parameters. In this setting, the relationship is represented by the equation $y = \beta_0 + x\beta + \varepsilon$, where y is the dependent variable, x represents the independent variables, β_0 is the intercept, β denotes the parameters, and ε is the error term. To simulate independent variables x, we will draw them from a multivariate normal distribution, $N(0, \Sigma_x)$. Furthermore, the pairwise covariance between x_i and x_j will be set to $(\Sigma_x)_{ij} = r^{|i-j|}$.To evaluate the resilience of the techniques against deviations from normality, we will utilize different error distributions, such as the standard normal distribution $\mathrm{N}(0,1)$ and a mixed normal distribution with significant outliers. Laplace distribution, t_3, $\mathrm{G}(3,1)$ and chi-square distribution .

The Quantile Regression methods to be compared include quantile LASSO, gamma-divergence, quantile elastic net, quantile adaptive LASSO, quantile SCAD, and quantile MCP. The R package gamreg will be utilized for the gamma-divergence method, while for the rest of the methods, the R package rqPen will be used.

In addition, we will conduct experiments with different correlations ($\mathrm{r} = 0.5$ and $\mathrm{r} = 0.95$) and consider three cases for the β values:

If $\mathrm{p} = 50$:

Case 1. β_1=3, β_2=1.5, $\beta_3 = 0$, $\beta_4 = 0$, $\beta_5 = 2$, $\beta_6 = 0$, $\beta_7 = 0, \ldots, \beta_{50} = 0$

Case 2. β_1 =1, β_2 =0, $\beta_3 = 0$, $\beta_4 = 0$, $\beta_5 = 5$, $\beta_6 = 0$, $\beta_7 = 1$, $\beta_8 = 0$, $\beta_9 = 0$, $\beta_{10} = 5$, $\beta_{11} = 0$, $\beta_{12} = 1$, $\beta_{13} = 0$, $\beta_{14} = 0, \ldots$, $\beta_{50} = 0$

Case 3. β_1=0.1, β_2=0.1, $\beta_3 = 0.1$, $\beta_4 = 0.1$, $\beta_5 = 0.1$, $\beta_6 = 0.1$, $\beta_7 = 0.1, \ldots$, $\beta_{50} = 0.1$

3.1 Example 1: Small-dimensional scenario with very sparse parameters (Case 1)

We are analyzing low-dimensional data with very few parameters, where there are 50 predictors (p) and 100 observations (n). The results of the trial can be found in Table 1A, Table 1B, and Figure 1. We are exploring both low correlation (r = 0.5) and high correlation (r = 0.95) among the predictors. The top sections show the median model error computed over 500 iterations. The model error is calculated as $(\hat{\beta} - \beta)^T S_x (\hat{\beta} - \beta)$, where $\hat{\beta}$ represents the estimated parameters and S_x is the sample covariance. The bottom sections illustrate the number of true positives, which indicate the accurately identified non-zero coefficients. A value of three corresponds to all non-zero coefficients being correctly detected.

Our results suggest that the Quantile LASSO (qLASSO) and quantile elastic net (qENet) techniques exhibit poor performance when handling highly correlated predictors. In contrast, the quantile SCAD (qSCAD) and quantile MCP (qMCP) methods demonstrate superior performance compared to all other methods across most error distributions.

Table1A: Average Median Model Error across 500 iterations for the scenario where $\mathrm{p} = 50$, $\mathrm{n} = 100$, $\mathrm{r} = 0.5$, and β values are the same as in example 1.

	qLASSO	Gamma Divergence	qENet	qaLASSO	qSCAD	qMCP
N(0,1)	0.190	0.113	0.392	0.064	0.049	**0.048**
Laplace	0.155	0.155	0.424	0.054	**0.041**	**0.041**
t_3	0.207	0.177	0.458	0.083	**0.062**	0.063
G(3,1)	0.430	0.321	0.913	0.201	0.135	**0.133**
Normal. M	0.336	0.224	0.688	0.135	**0.084**	0.091
Chi(3)	0.727	0.541	1.332	0.372	**0.240**	0.252

Table1B: Average Median Model Error across 500 iterations for the scenario where $p = 50, n = 100, r = 0.95$, and β values are the same as in example 1.

	qLASSO	Gamma Divergence	qENet	qaLASSO	qSCAD	qMCP
N(0,1)	0.132	0.080	0.231	0.089	**0.073**	0.081
Laplace	0.108	0.118	0.238	0.068	**0.047**	**0.047**
t_3	0.147	0.126	0.293	0.120	0.091	**0.084**
G(3,1)	0.310	0.202	0.478	0.237	0.207	**0.194**
Normal. M	0.214	0.138	0.382	0.162	0.136	**0.134**
Chi(3)	0.503	0.335	0.703	0.431	**0.336**	0.338

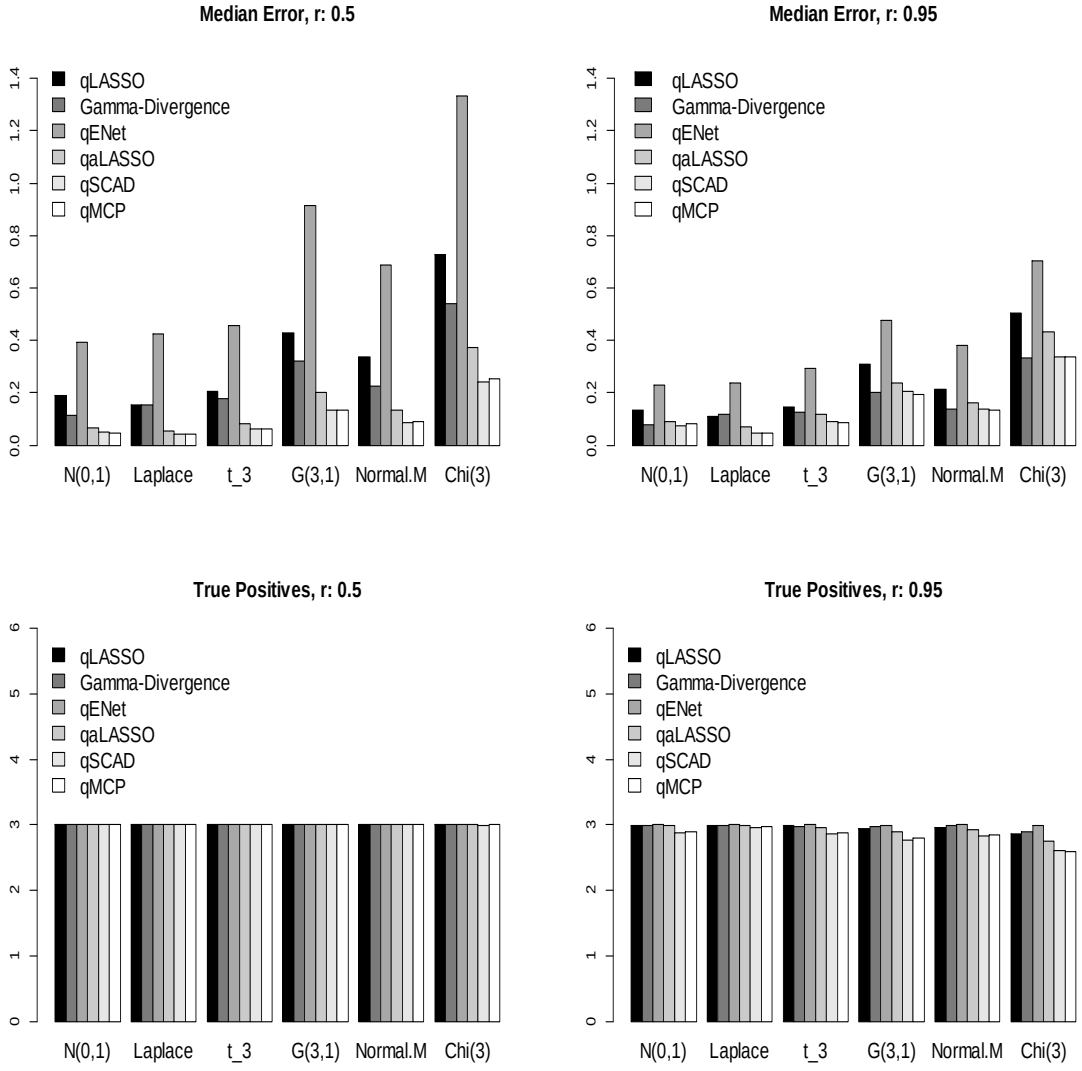

Figure1: Quantile Regression results for different error distributions with small and large correlated predictors. The top panels show the median model error over 500 iterations for example 1, while the bottom panels depict the average true positives with 50 predictors (p = 50) and 100 samples (n = 100).

3.2 Example 2: Large-dimensional scenario with very sparse parameters (Case 1)

In a manner reminiscent of example 1, our current study involves exploring a comparable scenario, but with a varied sample size and multiple predictors. Specifically, we are examining a large-dimensional example featuring extremely sparse coefficients, with p = 100 (representing the number of predictors) and n = 50 (representing the sample size). The results of our study are showcased in Table 2A, Table 2B, and Figure 2, which present the median model error from iterations. The model error is computed similarly to the method used in Figure 1.

Table 2A: Average Median Model Error across 500 iterations for the scenario where: $p = 100, n = 50$,r=0.5, and β values are the same as example 2.

	qLASSO	Gamma Divergence	qENet	qaLASSO	qSCAD	qMCP
N(0,1)	0.404	0.293	0.820	0.180	0.118	**0.114**
Laplace	0.456	0.399	1.198	0.163	**0.101**	0.110
t_3	0.607	0.533	1.334	0.262	0.180	**0.157**
G(3,1)	0.995	0.682	1.999	0.436	0.367	**0.366**
Normal. M	0.753	0.508	1.440	0.318	0.259	**0.225**
Chi(3)	1.727	1.719	2.970	1.080	**1.017**	1.081

Table2B: Average Median Model Error across 500 iterations for the scenario where: p=100, n=50, r=0.95, and β values are the same as example 2.

	qLASSO	Gamma Divergence	qENet	qaLASSO	qSCAD	qMCP
N(0,1)	0.272	0.185	0.447	0.216	0.183	**0.169**
Laplace	0.316	0.303	0.584	0.288	0.211	**0.209**
t_3	0.351	0.285	0.541	0.270	0.242	**0.241**
G(3,1)	0.613	0.473	0.904	0.556	0.483	**0.462**
Normal. M	0.494	0.336	0.762	0.398	**0.369**	0.377
Chi(3)	1.099	0.807	1.465	1.020	**0.865**	0.914

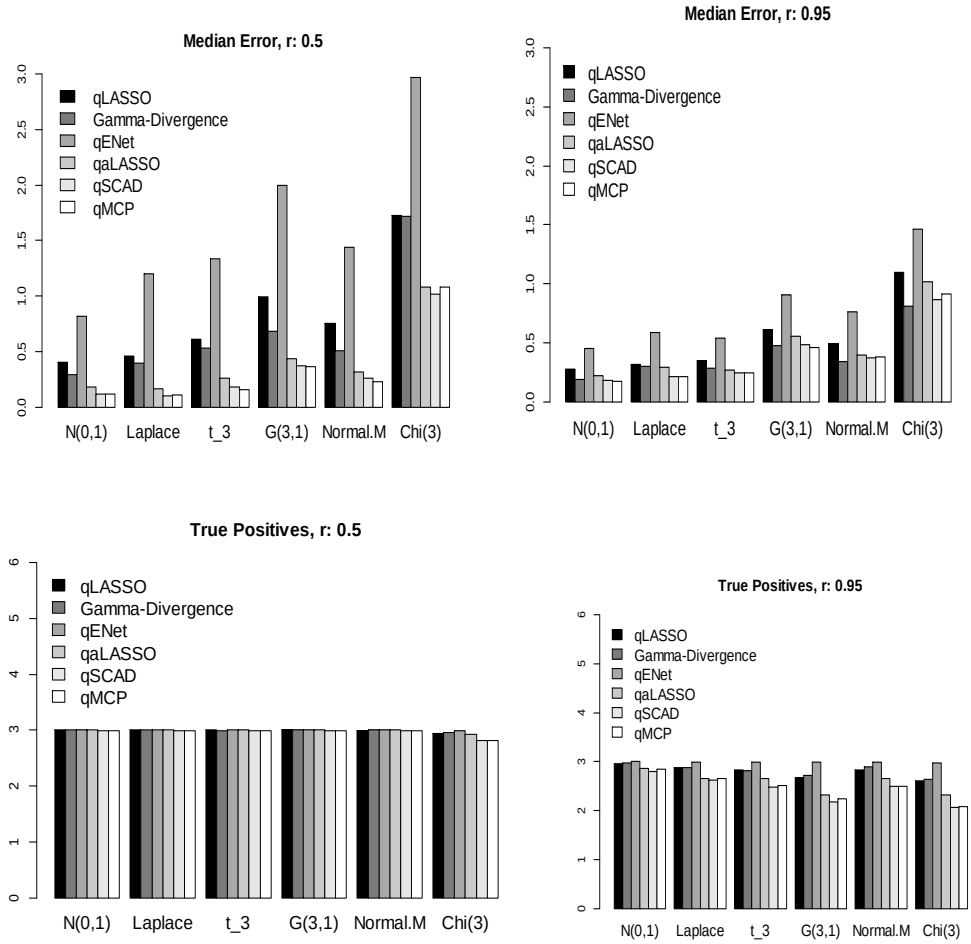

Figure 2: Quantile Regression results for different error distributions with small and large correlated predictors. The top panels show the median model error over 500 iterations for example 2, while the bottom panels depict the average true positives with 100 predictors (p = 100) and 50 samples (n = 50).

The research findings suggest that the effectiveness of the statistical methods is impacted by the degree of correlation among the predictors. Specifically, it was observed that the qENet method displayed subpar performance in scenarios where the predictors exhibited large levels of correlation. Conversely, the qMCP method showcased notably superior performance in comparison to other methods, especially in cases where deviations from normality were more prominent.

3.3 Example 3: Small- dimensional scenario with sparse parameters (Case 2)

In order to assess the effectiveness of different variable selection techniques, we established a new simulation scenario in which we encountered sparse situations, represented by β_j in case 2. Our findings, detailed in Table 3A, Table 3B, and Figure 3, present the median model error observed in 500 iterations for scenarios with $p = 50$ and $n = 100$.

Table3.A: Average Median Model Error across 500 iterations for the scenario where $p = 50, n = 100, r = 0.5$, and β values are the same as example 3.

	qLASSO	Gamma Divergence	qENet	qaLASSO	qSCAD	qMCP
N(0,1)	0.247	0.148	0.562	0.140	**0.099**	0.100
Laplace	0.280	0.305	0.770	0.141	**0.085**	0.087
t_3	0.324	0.327	0.803	0.206	**0.116**	**0.116**
G(3,1)	0.651	0.481	1.427	0.448	**0.297**	0.304
Normal. M	0.486	0.346	0.981	0.292	**0.222**	0.237
Chi(3)	1.066	**0.834**	2.214	1.083	0.915	0.919

Table3B: Average Median Model Error across 500 iterations for the scenario where $p = 50, n = 100, r = 0.95$, and β values are the same as example 3.

	qLASSO	Gamma Divergence	qENet	qaLASSO	qSCAD	qMCP
N(0,1)	0.194	**0.124**	0.442	0.184	0.164	0.169
Laplace	0.189	0.182	0.577	0.193	**0.169**	0.174
t_3	0.228	**0.189**	0.558	0.198	0.204	0.199
G(3,1)	0.440	**0.276**	0.871	0.423	0.428	0.435
Normal. M	0.309	**0.196**	0.674	0.295	0.300	0.293
Chi(3)	0.676	**0.463**	1.234	0.630	0.696	0.689

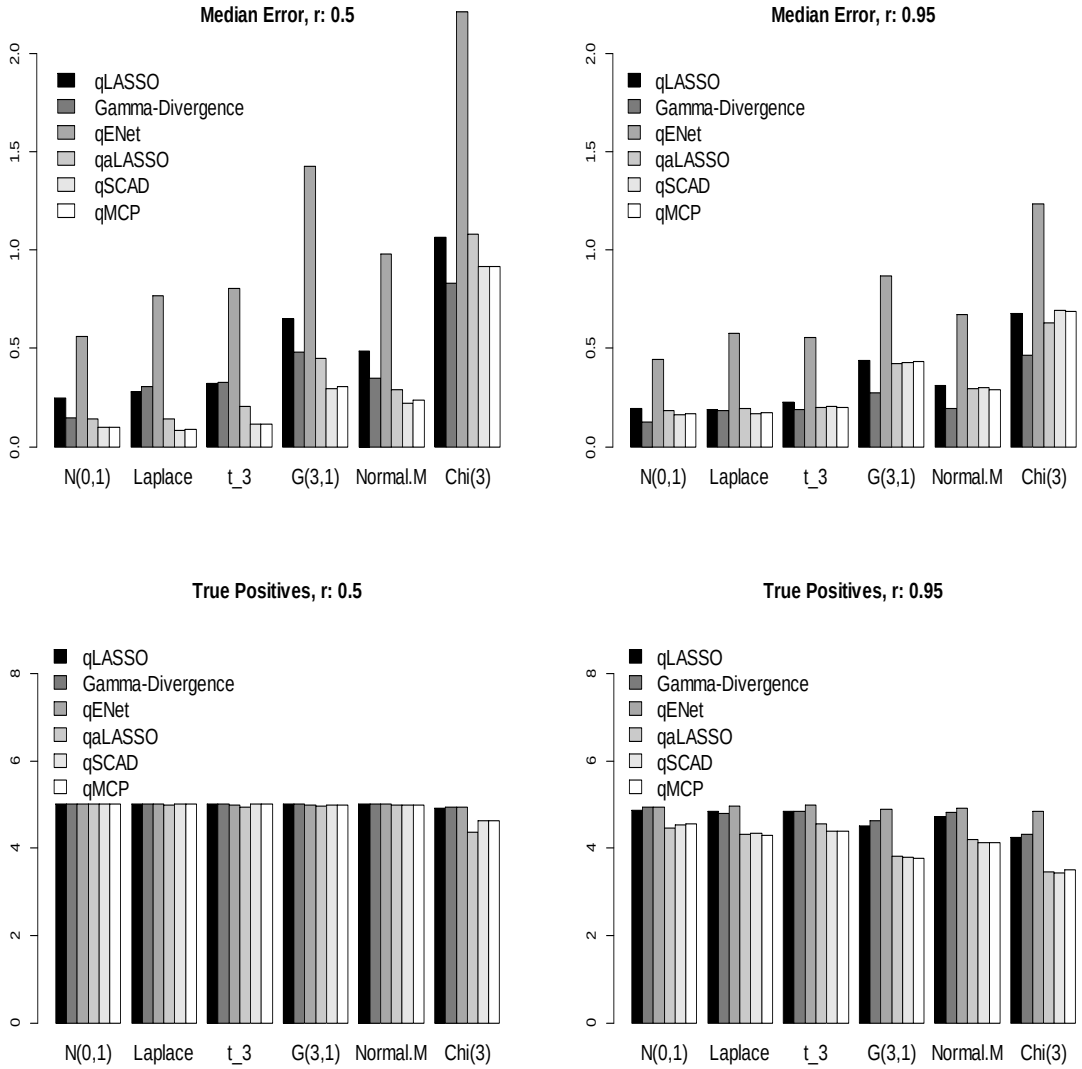

Figure 3: Quantile Regression results for different error distributions with small and large correlated predictors. The top panels show the median model error over 500 iterations for example 3, while the bottom panels depict the average true positives with 50 predictors (p = 50) and 100 samples (n = 100).

The results of our simulation study, as shown in Table 3A, Table 3B, and Figure 3, indicate that the gamma-divergence method outperforms all other methods as the departure from normality increases. This trend is particularly noticeable when dealing with highly correlated predictors.

3.4 Example 4: Large-dimensional scenario with sparse parameters (Case 2)

To analyze how different variable selection methods perform, we designed a new simulation to simulate a sparse situation similar to case 2, where we have β_j . The results are reported in Table 4A, Table 4B, and Figure 4, showing the median model error over 500 iterations for cases where $p = 100$ and $n = 50$.

Table 4A: Average Median Model Error across 500 iterations for the scenario where: $p = 100, n = 50, r = 0.5$, and β values are the same as example 4.

	qLASSO	Gamma Divergence	qENet	qaLASSO	qSCAD	qMCP
N(0,1)	0.608	0.379	1.051	0.660	0.278	**0.276**
Laplace	0.833	0.825	1.831	0.716	**0.532**	0.537
t_3	1.025	0.887	1.999	0.847	**0.855**	0.945
G(3,1)	1.702	**1.290**	2.718	1.695	1.784	1.837
Normal. M	1.123	0.872	1.949	**0.687**	0.792	0.803
$Chi(3)$	2.205	1.916	4.160	**1.867**	2.101	2.122

Table 4B: Average Median Model Error across 500 iterations for the scenario where$p = 100, n = 50, r = 0.95$, and β values are the same as example 4.

	qLASSO	Gamma Divergence	qENet	qaLASSO	qSCAD	qMCP
N(0,1)	0.361	**0.232**	0.733	0.324	0.365	0.359
Laplace	0.424	**0.377**	0.970	0.391	0.418	0.404
t_3	0.494	**0.396**	1.146	0.431	0.517	0.487
G(3,1)	0.831	**0.610**	1.575	0.779	0.802	0.807
Normal. M	0.616	**0.405**	1.346	0.530	0.579	0.556
Chi(3)	1.466	1.100	2.228	1.168	**1.045**	1.049

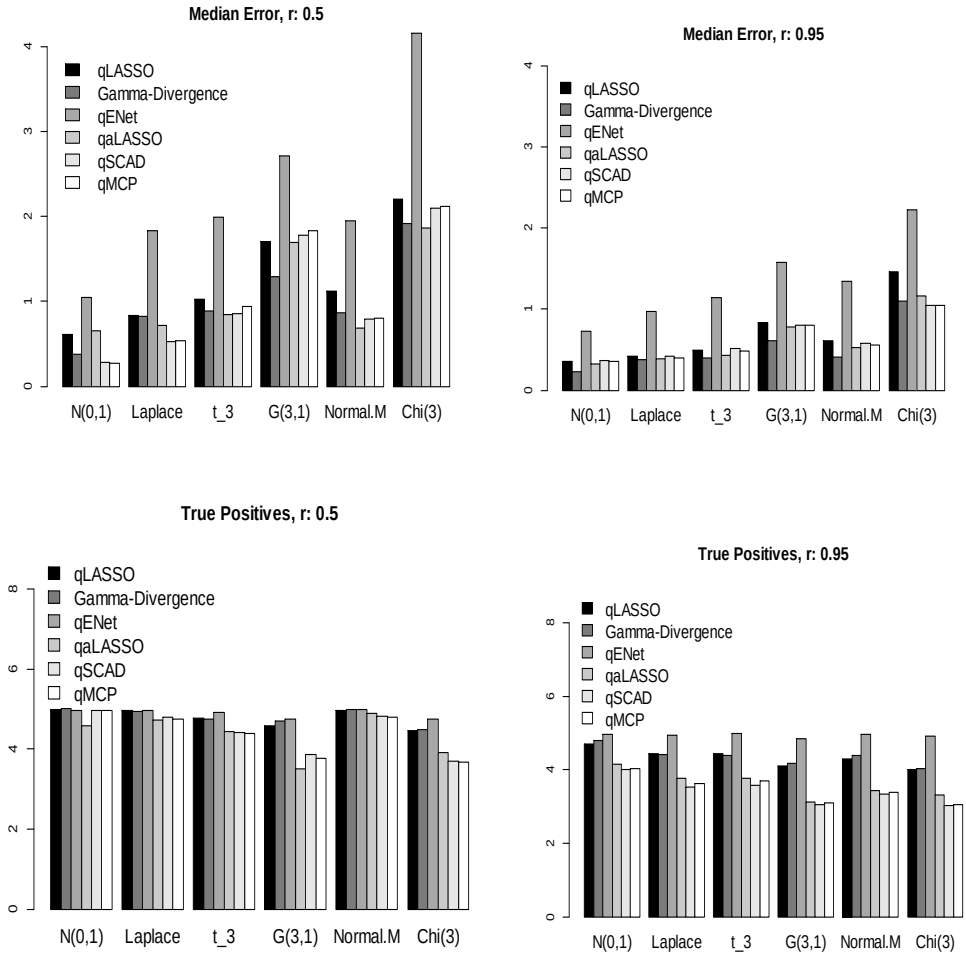

Figure 4: Quantile Regression results for different error distributions with small and large correlated predictors. The top panels show the median model error over 500 iterations for example 4, while the bottom panels depict the average true positives with 100 predictors (p = 100) and 50 samples (n = 50).

In our simulation study, we found that when comparing the results in Table 4A and Table 4B, as well as Figure 4, the performance of the gamma-divergence method outperformed all other methods for most error distributions. Additionally, the results indicated that the performance of qENet was the worst, particularly in scenarios where there was a departure from normality, especially when the predictors were largely correlated.

3.5 Example 5: Small-dimensional scenario with non-sparse parameters (Case 3)

To evaluate the effectiveness of different variable selection methods, we established a new simulation in which β_j is situated as in case 3, representing a non-sparse scenario. The results are presented in Table 5A, Table 5B, and Figure 5, show casing the median model error across 500 iterations for scenarios where $p = 50$ and $n = 100$.

Table5.A: Average Median Model Error across 500 iterations for the scenario where $p = 50, n = 100, r = 0.5$, and β values are the same as example 5.

	qLASSO	Gamma Divergence	qENet	qaLASSO	qSCAD	qMCP
N(0,1)	0.422	**0.339**	0.398	0.530	0.529	0.525
Laplace	0.500	0.460	**0.454**	0.647	0.670	0.652
t_3	0.520	0.484	**0.478**	0.672	0.686	0.673
G(3,1)	0.790	**0.656**	0.704	0.988	1.054	1.031
Normal. M	0.606	**0.486**	0.559	0.757	0.799	0.779
Chi(3)	1.023	**0.862**	0.917	1.266	1.409	1.387

Table5B: Average Median Model Error across 500 iterations for the scenario where $p = 50, n = 100, r = 0.95$, and β values are the same as example 5.

	qLASSO	Gamma Divergence	qENet	qaLASSO	qSCAD	qMCP
N(0,1)	0.230	0.161	**0.186**	0.262	0.257	0.250
Laplace	0.234	0.225	**0.174**	0.290	0.266	0.262
t_3	0.267	0.229	**0.209**	0.318	0.310	0.299
G(3,1)	0.432	**0.307**	0.348	0.506	0.487	0.481
Normal. M	0.347	**0.245**	0.286	0.395	0.390	0.385
Chi(3)	0.611	**0.485**	0.506	0.743	0.731	0.709

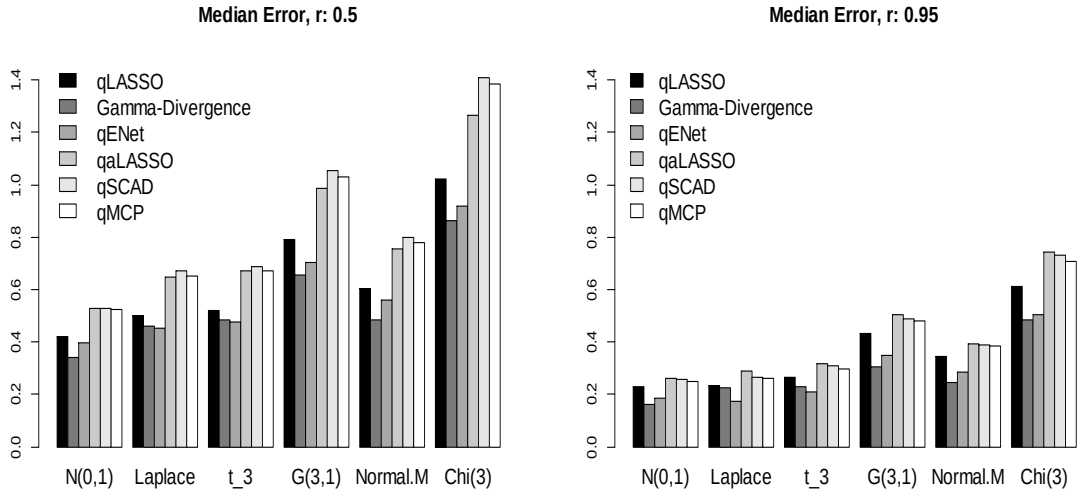

Figure 5: Quantile Regression results for different error distributions with small and large correlated predictors. The median model error over 500 iterations for example 5 with 50 predictors (p = 50) and 100 samples (n = 100).

Our simulation study shows that, according to the findings shown in Figure 5, Table 5A, and Table 5B, the gamma-divergence and qENet approaches are superior to all other methods as the deviation from normality becomes more significant. This advantage is particularly evident when the predictors demonstrate high correlation.

3.6 Example 6: Large-dimensional scenario with non-sparse parameters (Case 3)

In order to evaluate the efficacy of methods for selecting variables, we developed a new simulation where we treat β_j as described in situation 3, representing a non-sparse situation. The outcomes, displayed in Table 6A, Table 6B, and Figure 6, demonstrate the median model error over 500 iterations for cases where $p = 100$ and $n = 50$.

Table 6.A: Average Median Model Error across 500 iterations for the scenario where $p = 100, n = 50, r = 0.5$, and β values are the same as example 6.

	qLASSO	Gamma Divergence	qENet	qaLASSO	qSCAD	qMCP
N(0,1)	0.945	1.574	**0.893**	1.234	1.023	1.022
Laplace	1.331	1.614	**1.222**	1.627	1.573	1.586
t_3	1.370	1.705	**1.269**	1.803	1.720	1.688
G(3,1)	2.044	2.217	**1.875**	2.433	2.498	2.444
Normal. M	1.428	1.919	**1.316**	1.624	1.686	1.625
Chi(3)	2.421	2.487	**2.218**	3.196	3.294	3.153

Table 6.B: Average Median Model Error across 500 iterations for the scenario where $p = 100, n = 50, r = 0.95$, and β values are the same as example 6.

	qLASSO	Gamma Divergence	qENet	qaLASSO	qSCAD	qMCP
N(0,1)	0.597	0.523	**0.497**	0.734	0.671	0.669
Laplace	0.806	0.777	**0.636**	0.933	0.931	0.914
t_3	0.874	0.807	**0.676**	1.016	1.014	0.988
G(3,1)	1.188	1.062	**0.982**	1.335	1.390	1.370
Normal. M	0.943	0.796	**0.778**	1.026	1.093	1.100
Chi(3)	1.744	1.584	**1.411**	1.926	2.146	2.050

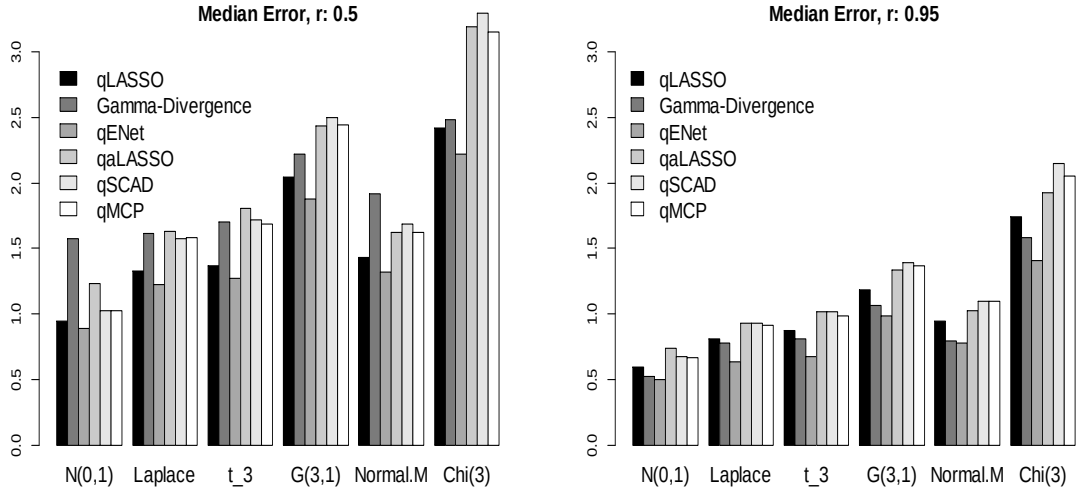

Figure 6: Quantile Regression results for different error distributions with low and high correlated predictors. The median model error over 500 iterations for simulation 6 with 100 predictors ($p = 100$) and 50 samples ($n = 50$).

After conducting our simulation study, the results from Table 6A, Table 6B, and Figure 6 support the superiority of the qENet method over all other methods as the departures from normality become more pronounced. Furthermore, our findings suggest that the performance of the qSCAD LASSO method is subpar when dealing with departures from normality, especially in cases where the predictors are highly correlated.

4. CONCLUSION

In the field of statistics, numerous methods rely on the assumption of normality. However, these approaches may not be appropriate for datasets that exhibit significant deviations from the normal distribution, often observed in the presence

of outliers or other forms of contamination. In our research, we have explored several recently developed variable selection techniques, including the quantile LASSO, gamma-divergence, quantile elastic net, quantile adaptive LASSO, quantile SCAD, and quantile MCP. We specifically focused on scenarios where the number of variables is equal to or exceeds the number of observations ($p \geq n$).

Through a comprehensive simulation study, we have demonstrated the efficacy of the gamma-divergence and quantile elastic net (qENet) methods, indicating their superiority over alternative approaches, particularly in cases where there is a substantial departure from normality within the dataset.

REFERENCES

[1] J. Fan and J. Lv , "A selective overview of variable selection in high dimensional feature space", Statist. Sin., 20, 101–148,(2010).

[2] R. Tibshirani, "Regression shrinkage and selection via the LASSO", Journal of the Royal Statistical Society, Series B 58, 267–288, (1996).

[3] H., Zou, and T., Hastie, "Regularization and variable selection via the elastic net'', Journal of the Royal Statistical Society B, 67, 301–320, (2005).

[4] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties", Journal of the American Statistical Association 96(456), 1348–1360, (2001).

[5] Y. Li and J. Zhu, "L1-norm quantile regression" , Journal of Computational and Graphical Statistics, 17(1), 163–185, (2008).

[6] Y. Wu, and Y. Liu, "Variable Selection in Quantile Regression, Statistica Sinica", 19, 801-817, (2009).

[7] R. Koenker, and G. Bassett ,"Regression quantiles", Econometrica, 46(1), 33–50, (1978).

[8] A. Yan, and F. Song, "Adaptive elastic net-penalized Quantile Regression for variable selection". Communications in Statistics- Theory and Methods, (2019).

[9] R. Alhamzawi , K. Yu, and D. F. Benoit, "Bayesian adaptive LASSO quantile regression", Statistical Modelling, 12(3), 279–297, (2012).

[10]S. M. Ajeel and H. A. Hashem, "Comparison Some Robust Regularization Methods in Linear Regression via Simulation Study", Academic Journal of Nawroz University, 9(2), 244–252, (2020).

[11] A.E., Hoerl, and R.W. ,Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems''. Technometrics , 12 (1), 55–67, (1970).

[12] H. Zou, "The adaptive LASSO and its oracle properties," Journal of the American Statistical Association 101, 1418–1429, (2006).

[13] R. Koenker , "Quantile Regression for longitudinal data", Journal of Multivariate Analysis, 91(1), 74–89, (2004).

[14] H. Fujisawa and S. Eguchi, "Robust parameter estimation with a small bias against heavy contamination", Journal of Multivariate Analysis, 99(9), 2053-2081, (2008).

[15]T. Kawashima and H. Fujisawa, "Robust and sparse regression via _divergence. Entropy", 19(11), 608, (2017).

[16] C. H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," Annals of Statistics 38, 894–942, (2010).