


Proposing Shrinkage Estimator of MCP and Elastic-Net penalties in Quantile Regression Model.

giddahz1402@uowasit.edu.iq

Jiddah Wally Zaher 1 

College of Administration and Economics
Wasit University, Iraq

ahameed@uowasit.edu.iq

Ali Hameed Yousif ²

College of Administration and Economics
, Wasit University, Iraq

Abstract— In some studies, there is a need to estimate the conditional distribution of the response variable at different points, and this is not available in linear regression. The alternative procedure to deal with these problems is quantile regression. In this research, a new estimator for estimating and selecting variables is proposed in the quantile regression model. A new estimator was combines two estimators Minimax Concave Penalty (MCP) and Elastic-Net called shrinkage estimator. It was compared with estimators (Minimax Concave Penalty (MCP) and Elastic-Net) by using simulation and based on Mean Square Error (MSE) and measures of sparsity False Positive Rate (FPR) and False negative rate (FNR). We concluded that the proposed method is the best in terms of estimation and selection of variables .

Keywords— MCP ,Elastic-Net, penalized Quantile Regression, Shrinkage Estimator

1 Introduction [1]

Regression analysis is considered the heart of statistics, which is concerned with the study and reliance on analyzing the relationship between a dependent variable and one or more explanatory variables. It is known that the method of Ordinary Least Squares (OLS) is used to estimate the factors of explanatory variables, but it is difficult to use it in some cases because one or all of its hypotheses are not available, just as random errors are abnormal, that is, they do not have a normal distribution. A suitable alternative to standard regression called quantile regression has been proposed by [2] . It is considered an extension of the standard linear regression and is complementary to the method of least squares (OLS), which estimates the conditional distribution of the response variable at different points. It is the most robust against outliers and anomalous values. In addition, it reduces the mean square error to the lowest possible. Quantile regression analyzes the relationship between one or more explanatory variables and a response variable conditional on divisions such as median, quar-

tiles, and percentiles, meaning that it works to find the amount of change that occurs in the response variable as a result of changes in the explanatory variables. Divisional regression is flexible when it comes to the error limit because it doesn't make any assumptions about how the error limit is distributed.[3].

Quantile regression has become of great use in many scientific applications in biomedicine, survival analysis, and in environmental[4] . The quantile function can be defined as the inverse of the cumulative distribution function.

$$Q_y(\tau) = F_y^{-1}(\tau) = \inf\{y: F(y|x) \geq \tau\} \quad 0 < \tau < 1 \quad (1)$$

In contrast, conditional quantile τ^{th} is conditional quantile $(Y|X)$ [5] Parameters of the quantile regression model can be estimated $\beta = (\beta_1, \beta_2, \dots, \beta_p) \in \mathcal{R}^p$. y represents the dependent variable and X It represents the matrix of explanatory variables .using the following formula:

$$\hat{\beta}_{(\tau)} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\tau}(y_i - \hat{X}_i \beta) \quad (2)$$

where $\rho_{\theta}(t)$ is called the quantile function or the loss function, which is expressed by the following formula:[6]

$$\rho_{\tau}(u) = \tau u I_{[0, \infty)} - (1 - \tau) u I_{(-\infty, 0)}(u) \quad (3)$$

We can also write it as follows:

$$\rho_{\tau}(u) = \begin{cases} \tau u & \text{if } u \geq 0 \\ -(1 - \tau)u & \text{if } u < 0 \end{cases} \quad (4)$$

Knowing that: $u = (y_i - \hat{X}_i \beta)$

The loss function is used to assign weight to τ with respect to positive errors $(y_i - \hat{X}_i \beta)$ and weight $(1 - \tau)$ with respect to negative errors $-(y_i - \hat{X}_i \beta)$, given that $0 < \tau < 1$, where $\tau = \frac{1}{2}$ makes the sum of absolute errors as small as possible and which corresponds to the median regression.

Non-penalized quantile regression is not appropriate in the case of an increase in the number of explanatory variables, as it is not possible to carry out the process of selecting variables, which has an important role in the process of building the model from a scientific point of view, which is one of the statistical problems that many statisticians are interested in. As we do not know any of the significant explanatory variables that could have a strong or weak effect on the response variable. The main goal is to obtain a simplified model with the least possible number of important explanatory variables, which is called the scattered model, where it is difficult to identify the significant variables that affect the response variable and are likely to lead to

the dropping of some important explanatory variables. These problems must be properly addressed in an ideal manner. Partial divisional regression is a useful tool to solve these problems. This is done by adding the penalty term to the divisional loss function. The quantile regression method is obtained, which is according to the following formula:[6]

$$Q_{\tau}(\beta) = \sum_{i=1}^n \rho_{\tau}(y_i - \hat{X}_i\beta) + \lambda \sum_{j=1}^p p_{\lambda}(|\beta_j|) \quad (5)$$

$p_{\lambda}(\cdot)$ It represents the penalty parameter. λ It represents the penalty parameter. $\lambda \geq 0$ It is called the tuning parameter.

[7] provided a comprehensive introduction to quantile regression. Many types of penalty functions that work to achieve variable selection and estimation have been introduced in the same process.[6] proposed an efficient algorithm that calculates the solution path for the penalty function (L1-Norm), which means the Lasso penalty function in partial quantile regression.[3] studied the selection of variables in quantitative regression. Using the (Smoothly Clipped Absolute Deviation (SCAD) penalty function and Adaptive Lasso.[8] proposed a new penalty function called MCP.[9] studied quantile regression in the case of high-dimensional data. [10] proposed a (Least absolute deviation) LAD-Atan penalty function that combines the idea of the least absolute difference LAD with the inverse tangent penalty function Atan. In quantile regression,[11] proposed a penalty function Atan.[22] Proposed two methods for estimating and selecting variables at the same time in quantile regression.

The aim from the research is to compare some penalty estimators with the proposed method to show its efficiency in terms of estimation and selection of variables.

The rest of the search is organized as follows. (2) the Penalized Quantile Regression. And in (2.1) Represent the function of MCP Penalized and in (2.2) the Represent function of the Elastic Net penalties, and in (2.3) the shrinkage.And In the (3) section, the simulation is applied and in the (4) section, the results and the (5) section the conclusions.

2. Penalized Quantile Regression

Quantile regression is not effective in dealing with high-dimensional data, as it is not possible to carry out the process of selecting variables, which is one of the important issues in statistics. In order to overcome these problems, we use the penalized quantile regression by using the quantile estimators in the quantile regression.

2.1 MCP Penalized

It is a concave penalty function, proposed by[8] . It has an acronym Minimax Concave Penalty (MCP). This method is characterized by the speed in estimating parame-

ters, and it is continuous, unbiased and accurate, in addition to that, it simplifies the model and improves its accuracy at the same time. It is superior to the Lasso method in terms of accuracy, speed of estimation and reduction. The penalty function (MCP) can be expressed in the following form:

$$p_{\lambda,\gamma}(|\beta|) = \begin{cases} \lambda|\beta| - \frac{\beta^2}{2\gamma} & |\beta| \leq \gamma\lambda \\ \frac{1}{2}\gamma\lambda^2 & |\beta| > \gamma\lambda \end{cases} \quad (6)$$

knowing that $\gamma > 1$

As for the objective function of the penalty quantile regression of the (MCP) penalty, it can be written in the following form

$$\hat{\beta}_{MCP} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \rho_{\tau}(y_i - \hat{X}_i\beta) + \lambda \sum_{j=1}^p p_{\lambda\gamma}(\beta_j) \right\} \quad (7)$$

We used the Quantile Iterative Coordinate Descent algorithm (QICD) to solve the linear programming problem of the MCP penalty function [12].

2.2 Elastic-Net

In order to bypass the defects of the Lasso penalty function, such as the presence of a set of variables with a high correlation or when the number of explanatory variables is greater than the sample size ($p > n$) in which case the Lasso estimator does not work well, an Elastic-Net penalty function is proposed, which is used to select groups of interrelated variables. which is a combination between the lasso penalty function, which encourages obtaining scattered solutions for the parameters, and the Ridge penalty function, which improves the prediction process. The penalty function (Elastic-Net) can be expressed by the following formula [13]

$$P_{\lambda\alpha}(\beta) = \lambda \sum_{j=1}^p \left[\frac{1}{2} (1 - \alpha) \beta_j^2 + \alpha |\beta_j| \right], \quad 0 < \alpha < 1 \quad (8)$$

$p_{\lambda\alpha}(\beta)$ means Penalty function, λ Means Penalty parameter, note that $\lambda > 0$

And $0 < \alpha < 1$ if it was $\alpha = 0$ The Elastic Net penalty function is reduced to (Ridge Regression), and if it was $\alpha = 1$ reduced to lasso estimator.

When adding the penalty function (Elastic Net) to the divisional loss function, we get the penalty estimator for quantile regression (QR) as in the following formula [14]

$$\hat{\beta}_{Elastic-Net} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \rho_{\tau}(y_i - \hat{X}_i\beta) + \lambda \sum_{j=1}^p \left[\frac{1}{2} (1 - \alpha) \beta_j^2 + \alpha |\beta_j| \right] \right\} \quad (9)$$

$i=1,2,\dots,n$, X it represents the matrix of explanatory variables. Y represents the dependent variable .

The Semi smooth Newton Coordinate Descent (SNCD) algorithm was used to solve the problem of linear programming of a penalty function Elastic Net.[19]

2.3 Shrinkage Estimator

The (MCP) estimator has been combined with the (Elastic-Net) estimator to get a new estimator, which we called the shrinkage estimator, which is used to estimate the parameters of the quantile regression model and to select variables. Therefore, the estimator (shrinkage) can be expressed using the following formula.

$$\hat{\beta}_{\text{Shrinkage}} = \alpha \hat{\beta}_{\text{mcp}} + (1 - \alpha) \hat{\beta}_{\text{Elastic}} \quad (10)$$

Whereas $0 < \alpha < 1$, It represents the tuning parameter.

The penalty parameter was selected based on the (HBIC) high-dimensionality standard [15] .

$$HBIC(\lambda) = \log \left(\sum_{i=1}^n \rho_{\tau}(Y_i - \hat{X}_i \beta_{\lambda}) \right) + |s_{\lambda}| \frac{\log(\log n)}{n} C_n, \quad (11)$$

Where $|s_{\lambda}|$ is the set $S_{\lambda} \equiv \{j: \beta_{\lambda,j} \neq 0, 1 \leq j \leq p\}$ as for C_n it represents a series of constants that diverge infinitely with increment (n). $\beta_{\lambda} = (\beta_{\lambda,1}, \dots, \beta_{\lambda,p})'$. The penalty parameter λ is chosen which makes $HBIC(\lambda)$ at its lower limit.

3. Simulation

In this section, we try a Monte Carlo simulation study to compare the estimators MCP, Elastic-Net, and Shrinkage based on the R program with replicate (200). We generated the data using a regression model.

$$y = x\beta + \varepsilon_i \quad (12)$$

Note that y represents the dependent variable vector, x represents the matrix of the explanatory variable, β is the parameter vector, and ε_i represents the random error vector. A random variable ε_i is generated based on

Standard normal distribution with mean zero and variance one. The true parameters

$$\beta = (3, 1.5, 2, 0, 0, 0, \dots, 0)$$

,
(P=15,n=50,and n=150)

The default values τ are

$\tau=(0.20,0.40,0.80)$

The x matrix is generated by a multivariate normal distribution $x \sim MN(0, \Sigma_x)$ with $(\Sigma_x)_{ij} = 0.5^{|i-j|}$ Whereas, $1 \leq i, j \leq p_n$, the penalty estimators are compared based on the mean square error (MSE), and the selection of variables is based on the two measures of false positive rate (FPR) and false negative rate (FNR). Whereas, the method that has the lowest value for these two criteria is the best in terms of selecting variables.

TABLE (I)

the simulation results for the first experiment when (n = 50, p = 15)

τ	Estimators	MSE	FPR	FNR
0 .20	MCP	0.0563	0.044	0.4
	Elastin-Net	0.0314	0.344	0.066
	Shrinkage	0.0214	0.334	0.056
0 .40	MCP	0.0763	0.0333	0.45
	Elastin-Net	0.0530	0.444	0.116
	shrinkage	0.043	0.434	0.106
0 .80	MCP	0.0663	0.0111	0.4
	Elastin-Net	0.0315	0.4111	0.133
	shrinkage	0.0215	0.4011	0.123

TABLE (II)

the simulation results for the first experiment when ($n=150$, $p=15$)

τ	Estimators	MSE	FPR	FNR
0.20	MCP	0.0396	0	0.433
	Elastin-Net	0.0108	0.355	0.016
	Shrinkage	0.008	0.345	0.006
0.40	MCP	0.0326	0	0.4333
	Elastin-Net	0.0102	0.3	0
	Shrinkage	0.0226	0.01	0.4233
0.80	MCP	0.0201	0	0.5
	Elastin-Net	0.0068	0.333	0
	Shrinkage	0.0032	0.323	0.01

4. Results .

Table (I) and Table (II) summarize the results of the simulation. When the sample size is $n=50$, $n=150$ and the number of explanatory variables $p=15$, The results show that the shrinkage method has the best performance from the MCP and Elastic-Net in case of estimator and selecting variables because it has the smallest values of MSE and FPR, FNR.

next in preference as shown in the table (I) and table (II) when the $n=50$, $n=150$ Elastin-Net has the best from the MCP hand to estimate because it has the smallest values of MSE. As for the selection of variables, the estimator of the MCP penalty function has the best because it gives the lowest value for FPR.

5. Conclusions.

In this section, we conclude with the following: After using the simulation and conducting two types of experiments shown in the tables of the previous paragraph, the best way to estimate the parameters is (shrinkage) as they contain the lowest value of the mean square errors (MSE) shown in the results of the experiments in the previous paragraph above. Also, the best way to choose the variables is shrinkage because they contain the lowest value of the false positive rate (FPR) and the false negative rate (FNR). We also conclude that when the sample size increases, the mean squares of errors decrease.

6. References.

- [1] Yousif, A.H., *Proposing robust LAD-shrink set estimator for high dimensional regression model*. Int. J. Agric. Stat. Sci, 2021. **17**: p. 1229-1234.
- [2] Koenker, R. and G. Bassett Jr, *Regression quantiles*. Econometrica: journal of the Econometric Society, 1978: p. 33-50.
- [3] Wu, Y. and Y. Liu, *Variable selection in quantile regression*. Statistica Sinica, 2009: p. 801-817.
- [4] Chen, C. and Y. Wei, *Computational issues for quantile regression*. Sankhyā: The Indian Journal of Statistics, 2005: p. 399-417.
- [5] Kuan, C.-M., *An introduction to quantile regression*. Access from avesbiodiv. mncn. csic. es, 2007.
- [6] Li, Y. and J. Zhu, *L 1-norm quantile regression*. Journal of Computational and Graphical Statistics, 2008. **17**(1): p. 163-185.
- [7] Buhai, S., *Quantile regression: overview and selected applications*. Ad Astra, 2005. **4**(4): p. 1-17.
- [8] Zhang, C.-H., *Nearly unbiased variable selection under minimax concave penalty*. The Annals of statistics, 2010. **38**(2): p. 894-942.
- [9] Wang, L., Y. Wu, and R. Li, *Quantile regression for analyzing heterogeneity in ultra-high dimension*. Journal of the American Statistical Association, 2012. **107**(497): p. 214-222.
- [10] Yousif, A.H. and O.A. Ali, *Proposing robust lad-atan penalty of regression model estimation for high dimensional data*. Baghdad Science Journal, 2020. **17**(2): p. 550-555.
- [11] Yousif, A.H. and W.J. Housain. *Atan regularized in quantile regression for high dimensional data*. in *Journal of Physics: Conference Series*. 2021. IOP Publishing.
- [12] Peng, B. and L. Wang, *An iterative coordinate descent algorithm for high-dimensional nonconvex penalized quantile regression*. Journal of Computational and Graphical Statistics, 2015. **24**(3): p. 676-694.
- [13] Zou, H. and T. Hastie, *Regularization and variable selection via the elastic net*. Journal of the royal statistical society: series B (statistical methodology), 2005. **67**(2): p. 301-320.

- [14] Yousif, A.H. and A.H. Gatea, *Atan regularized for the high dimensional Poisson regression model*. International Journal of Nonlinear Analysis and Applications, 2021. **12**(Special Issue): p. 2197-2202.
- [15] Lee, E.R., H. Noh, and B.U. Park, *Model selection via Bayesian information criterion for quantile regression models*. Journal of the American Statistical Association, 2014. **109**(505): p. 216-229.
- [16] Mendez-Civieta, A., Aguilera-Morillo, M. C., & Lillo, R. E. (2020). Adaptive sparse group LASSO in quantile regression. *Advances in Data Analysis and Classification*, 1-27.
- [17] Rahman, E., & Alkenani, A. (2020, July). Sparse minimum average variance estimation via the adaptive elastic net when the predictors correlated. In *Journal of Physics: Conference Series* (Vol. 1591, No. 1, p. 012041). IOP Publishing.
- [18] De Oliveira, S. H., Gonçalves, A. R., & Von Zuben, F. (2019, August). Group lasso with asymmetric structure estimation for multi-task learning. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence* (pp. 3202-3208).
- [19] Yi, C., & Huang, J. (2017). Semismooth Newton coordinate descent algorithm for elastic-net penalized Huber loss regression and quantile regression. *Journal of Computational and Graphical Statistics*, 26(3), 547-557.
- [20] Staffa, S. J., Kohane, D. S., & Zurakowski, D. (2019). Quantile regression and its applications: a primer for anesthesiologists. *Anesthesia & Analgesia*, 128(4), 820-830.
- [21] Midi, H., Alshaybawee, T., & Alguraibawi, M. (2020). Modified Least Trimmed Quantile Regression to Overcome Effects of Leverage Points. *Mathematical Problems in Engineering*.
- [22] Jiang, L., Bondell, H. D., & Wang, H. J. (2014). Interquantile shrinkage and variable selection in quantile regression. *Computational statistics & data analysis*, 69, 208-219.