Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Shrinkage Estimator of SCAD and Adaptive Lasso penalties in Quantile Regression Model

JiddahWally Zaher[1], Ali Hameed Yousif [2]
[1,2]Wasit University , Iraq
giddahz1402@uowasit.edu.iqahameed@uowasit.edu.iq

Article Info
Page Number: 5945 - 5953
Publication Issue:
Vol 71 No. 4 (2022)

Article History
Article Received: *25 March 2022*
Revised: *30 April 2022*
Accepted: *15 June 2022*
Publication: *19 August 2022*

Abstract

Quantile regression is one of the most frequently used topics in data analysis. In this article, we proposed the shrinkage estimator for penalized quantile regression that combines SCAD (Smoothly Clipped Absolute Deviation) and Adaptive Lasso estimators. these estimators were compared by using simulationstudies based on statistical measures, mean squared error (MSE), false positive rate (FPR) and false negative rate (FNR).After applying theSimulation studies it was found that the proposed estimator is the best in estimation and selection of variable because it has the lowest mean squared error (MSE) and it has lowest False Positive Rate (FPR) and False Negative Rate (FNR).

Key words: SCAD, Adaptive lasso, Quantile, Penalized Quantile regression,Shrinkage Estimator.

I . INTRODUCTION[1-3]

One of the statistical analysis methods that are used to study and analyze the relationship between the studied variables and for all sciences is linear regression. It is a statistical method for studying the relationship between one or more explanatory variables and the dependent variable, and it expresses the relationship as an equation. Regression has many uses, including data description, parameter estimation, prediction, control, and others. Ordinary linear regression cannot often be used in many scientific studies and different phenomena because the regression conditions are not applicable. and the appropriate alternative for this is quantile regression, which was suggested by [4],It is a suitable alternative to linear regression, and interest in it has increased in recent years. Assuming a random sample$(\mathrm{Y}_1, \mathrm{X}_1), \dots, (\mathrm{Y}_n, \mathrm{X}_n)$,(0<θ<1),The linear regression model can be written as follows:[5]

$$\mathbf{Y_i} = \mathbf{\acute{X}_i B} + \boldsymbol{\epsilon}_\mathbf{i} \qquad \mathrm{i} = 1,2, \dots, \mathrm{n} \quad (1)$$

Where Yi represents the response variable, Xi represents the explanatory variables (independent variables).Brepresents the model parameters$\beta = (\beta_1, \beta_2, \dots, \beta_p)$,ϵ_i represents the random error that is normally distributed with zero mean and variance σ^2 ,$\epsilon_i \sim N(0, \sigma^2)$,The parameters of the Quantile Regression model $\beta = (\beta_1, \beta_2, \ldots, \beta_p) \in \mathcal{R}^P$ can be estimated using the following formula:

$$\widehat{\mathbf{B}}_{(\theta)} = \arg\min_{\mathbf{B}} \sum_{i=1}^{n} \rho_\theta(y_i - \acute{X}_i \mathbf{B}) \quad (2)$$

where the loss $\rho_\theta(t)$ is referred to as the check function and is expressed by the following formula:[4]

$$\rho_\theta(t) = \begin{cases} \theta t & \text{if} \quad t \geq 0 \\ -(1-\theta)t & \text{if} \quad t < 0 \end{cases} \quad (3)$$

Since each $(0 < \theta < 1)$ in the case where ($\theta = \frac{1}{2}$) makes the sum of the absolute errors as small as possible and which corresponds to the median regression. As the set of regression quantiles $\widehat{B}(\theta): \theta \in (0,1)\}$ are called quantile processes,The choice of the variable plays an important role in the process of building the model from a scientific point of view, and it is common for a large number of explanatory variables, to lead to an increase in dimensions, which causes the complexity of the model and it is difficult to interpret easilyby [6] ,studied the penalized quantile regression depending on the penalty estimator (SCAD) and (ADAPTIVE LASSO) and using the Difference Convex Algorithm (DCA). The results of real data and simulation also confirmed that these estimators used have the oracle properties [7]. studied the quantile regression in the case of high-dimensional data, and these data are often exposed to the problem of heterogeneity due to the increase in the number of explanatory variables. The penalty function (SCAD) was relied on to select the significant variables[8] .(QICD) algorithm, which is used to solve the problem of linear programming in quantile regression, which has complicated computational problems, and they showed that this algorithm is characterized by fast big arithmetic proposed by[9] . [5] studied the penalty function (SCAD) in quantile regression and used real data and simulation to show the efficiency of this method, and it was found that it has the oracle properties. [10]studied the quantile regression in the case of high-dimensional data (p > n) and the selection of significant variables using a set of penalty estimators (Ridge, Elastic-Net) that perform the process of estimation and selection of significant variables in the same process. They also proposed a new penalty function called Atan. for penalty divisional regression. After running the simulation, it was found that the proposed penalty estimator is the best when it comes to estimating and choosing variables. The rest of the research details were arranged as follows: In the second section, the quantile regression was addressed using the SCAD penalty estimator and the solution to the linear programming problem with the QICD algorithm. We also used the Adaptive Lasso penalty function and solved its linear programming problem using the DCA algorithm. We combined the SCAD penalty function with the adaptive lasso penalty functionto getestimator shrinkage . In the third section Applicationthesimulation study, and in the fourth section, the Conclusions.

II.PENALIZED QUANTILE REGRESSION

When the number of explanatory variables is increased at the expense of the sample size, the process of estimating the model parameters is difficult to implement. It is also difficult to choose the significant variables, which is one of the important issues in statistical modeling. In this case, quantile regression cannot be applied, and therefore, the appropriate alternative is to use an appropriate regression method to deal with high-dimensional data regression. It's called penalized quantile regression [10],The penalty function is added to the quantile loss function, so we get the goal function, as in the following formula; [11]

$$\mathbf{Q_{\theta}(B;\lambda)=\sum_{i=1}^{n}\rho_{\theta}\left(y_i-\acute{X}_iB\right)+\lambda\sum_{j=1}^{p}p_{\lambda}\left(\left|B_j\right|\right)}\quad(\mathbf{4})$$

$p_{\lambda}(.)$it expresses the penalty function.

λ: It expresses the penalty parameter, whose value is called the tuning parameter.

A.SCAD

It is one of the penalty functions that are used to estimate the parameters of the linear regression model simultaneously. It is expressed by the term (Smoothly clipped Absolute Deviation) that was proposed by[12] and it was shown that this method achieves the properties of oracle. This method keeps the good parts of both penalty group selection and ridge regression, but it also makes scattered solutions that make sure the selected models stay the same (model selection stability) and give unbiased estimates for large parameters. [6] The penalty function (SCAD) is according to the following formula:

$$p_{\lambda}\left(\left|B_j\right|\right)=\begin{cases}\lambda\left|B_j\right| & \text{if } 0\le\left|B_j\right|<\lambda\\ \frac{(a^2-1)\lambda^2-(\left|B_j\right|-a\lambda)^2}{2(a-1)} & \text{if } \lambda\le\left|B_j\right|<a\lambda\\ \frac{(a+1)^2\lambda^2}{2} & \text{if } \left|B_j\right|>a\lambda\end{cases}\quad(\mathbf{5})$$

$a>2$, It is also called the tuning parameter, and its value was suggested by two scientists [6], $\lambda>0$ it's called the tuning parameter.Therefore, the penalty least squares estimator with a penalty function (SCAD) in quantile regression can be expressed using the following formula: [5]

$$\mathbf{\hat{B}_{SCAD}=argmin_B\left\{\sum_{i=1}^{n}\rho_{\theta}(y_i-\acute{X}_iB)+\lambda\sum_{j=1}^{p}p_{\lambda}\left(\left|B_j\right|\right)\right\}}\quad(\mathbf{6})$$

One of the important issues in penalized regression is the selection of the penalty parameter, which is also called the tuning parameter and is symbolized by the symbol λ. Although penalized estimators have oracle properties, choosing them remains important because it controls the amount of reduction in the parameters of the final model. [10]This parameter is selected according to the (HBIC) criterion that was proposed in the quantile regression by[13] As in the following formula:

$$\mathbf{HBIC(\lambda) = \log\left(\sum_{i=1}^{n} \rho_{\theta}\left(Y_i - \acute{X}_i\beta_{\lambda}\right)\right) + |s_{\lambda}|\frac{\log(\log n)}{n}C_n}, \qquad (7)$$

$|s_{\lambda}|$ represents the origin of the group$S_{\lambda} \equiv \{j: \beta_{\lambda,j} \neq 0, 1 \leq j \leq p\}$, As for C_n Represents a series of constants that diverge infinitely with each increment (n). The penalty parameter λ is chosen, which makes HBIC (λ) at its lower limit. It is based on the (Iterative Coordinate Descent) algorithm, which is symbolized by the symbol (QICD), which was proposed by [9], and which is used to solve the quantile penalized regression problem in the case of high-dimensional data p > n. It is used to solve an estimator (SCAD).

B.ADAPTIVE LASSO

mentioned that the Lasso estimator has many defects. It also does not achieve the properties of oracle, which are unbiasedness, sparsity and continuity [6]. [14] Proposed a new penalty function, which is a suitable alternative to the penalty function lasso called (Adaptive Lasso). The idea of this method is to add different adaptive weights $w_j$for penalty coefficients in the penalty function lasso, which represents the reciprocal of the least squares estimates of non-penalized quantile regression raised to some power as weights. Where it leads to an increase in the penalty for the parameters that are close to zero, the bias in estimating the function is reduced. This results in an improvement in the accuracy of the selection variable.The penalty estimator in quantile regression (QR) depending on the penalty function (Adaptive Lasso) is calculated according to the following formula: [7]

$$\mathbf{\widehat{B}_{ALASSO} = argmin_B\left\{\sum_{i=1}^{n} \rho_{\theta}(y_i - \acute{X}_i B) + \lambda \sum_{j=1}^{p} w_j\left|B_j\right|\right\}} \qquad (8)$$

$w_j = \frac{1}{|\widehat{B}_{\tau,j}|^{\gamma}}$, J=1,2,…,p

γ represents the value of the reducing parameter, $\gamma > 0$, It is a fixed value $\gamma = 1$, $\widehat{B}_{\tau}$ is estimator $\sqrt{n}$ consistent for B_{τ}

$$\widehat{B}_{\theta} = \operatorname{argmin}_{B\tau} \sum_{i=1}^{n} \rho_{\theta}\left(yi - \acute{X}_i B_{\tau}\right) \qquad (9)$$

The researcher (Zou) showed that if the weights were chosen efficiently and dependent on the data, then the (adaptive lasso) estimator could achieve the (efficient) property. To solve the estimator

(Adaptive Lasso) we use the (difference convex algorithm) (DCA) that was proposed in [15] it is used to solve the linear programming problem. The DCA algorithm is as follows:

$$\mathbf{min}\sum_{i=1}^{n}(\boldsymbol{\theta\xi_i}+(\mathbf{1}-\boldsymbol{\theta})\boldsymbol{\zeta_i})+\mathbf{n}\boldsymbol{\lambda_n}\sum_{j=1}^{p}\mathbf{w_j}\boldsymbol{\eta_j}\quad(\mathbf{10})$$

$\xi_i \geq 0, \zeta_i \geq 0, \xi_i - \zeta_i = yi - x_i^T B_j, i = 1, \dots, n$

$\eta_j \geq B_j, \eta_j \geq -B_j, j = 1, \dots, p$, Where w_j weights are chosen appropriately.

The Schwarz information criterion for selecting the penalty parameter was used in quantile regression by[11] .

$$\mathbf{SIC}(\boldsymbol{\lambda}) = \mathbf{log}\left(\frac{\mathbf{1}}{\mathbf{n}}\sum_{i=1}^{n}\boldsymbol{\rho_\theta}\left(\mathbf{y_i}-\acute{\mathbf{X}}_{\mathbf{i}}\mathbf{B}\right)\right)+\frac{\mathbf{log\,n}}{\mathbf{2d}}\mathbf{df}, \qquad (11)$$

Where df represents the measurement of the effective dimensions of the model installation. The penalty parameter λ is chosen, which makes SIC(λ) at its lower limit.

C.SHRINKAGE METHOD

In this section, the SCAD estimator has been combined with the (Adaptive Lasso) estimator to get a new estimator, which called the shrinkage estimator. The shrinkage estimator used to estimate the parameters of the quantile regression model and select variables. Therefore, the shrinkage estimator can be expressed using the following formula:

$$\widehat{\boldsymbol{\beta}}_{\mathbf{Shrinkage}} = \boldsymbol{\alpha}\widehat{\boldsymbol{\beta}}_{\mathbf{Scad}} + (\mathbf{1}-\boldsymbol{\alpha})\widehat{\boldsymbol{\beta}}_{\mathbf{Adabtive\ lasso}} \quad (12)$$

Where that $0 < \alpha$

Ⅲ.SIMULATION

In this section, we used Monte Carlo simulation in order to compare the SCAD, Adaptive Lasso, and Shrinkage estimators based on the R program with replicate (500). The simulation data based on the following model.

$$\mathbf{y} = \acute{\mathbf{x}}\boldsymbol{\beta} + \mathbf{u} \qquad (\mathbf{13})$$

Note that y represents the vector of the dependent variables, x represents the matrix of explanatory variables, β is the vector of parameters, and u represents the vector of random errors. The x matrix was generated using a multivariate normal distribution ,i.e.$x \sim MN(0, \Sigma)$,with $(\Sigma_x)_{ij} = 0.5^{|i-j|}$.A random variable u is generated based on standard normal distribution with mean zero and variance one. Thetrue parameters $\beta = (1,1,{}^{1}/_{2},{}^{1}/_{3},{}^{1}/_{4},{}^{1}/_{5},0,\ldots,0)$,p=10,n=30 and 100.The default values of $\theta = (0.25,0.50,0.75)$.

TABLE (I)

shows the simulation results for the first experiment when n = 30, p = 10

θ	**Estimators**	**MSE**	**FPR**	**FNR**
0.25	**SCAD**	**0.0697**	**0.125**	**0.35**
	Adaptive lasso	**0.0613**	**0.025**	**0.35**
	Shrinkage	**0.0387**	**0.075**	**0.25**
0.50	**SCAD**	**0.0609**	**0.2**	**0.3**
	Adaptive lasso	**0.0846**	**0.1**	**0.333**
	shrinkage	**0.0391**	**0.1**	**0.2**
0.75	**SCAD**	**0.0497**	**0.3**	**0.283**
	Adaptive lasso	**0.0510**	**0.125**	**0.266**
	shrinkage	**0.0503**	**0.2**	**0.183**

TABLE (II)

shows the simulation results for the first experiment when n = 100, p = 10

θ	**Estimators**	**MSE**	**FPR**	**FNR**
0.25	**SCAD**	**0.0350**	**0.025**	**0.3**
	Adaptive lasso	**0.0716**	**0**	**0.3**
	Shrinkage	**0.025**	**0.015**	**0.29**

0.50	**SCAD**	**0.0297**	**0.025**	**0.233**
	Adaptive lasso	**0.0575**	**0**	**0.283**
	shrinkage	**0.0197**	**0.015**	**0.223**
0.75	**SCAD**	**0.0473**	**0.05**	**0.333**
	Adaptive lasso	**0.0424**	**0**	**0.2**
	shrinkage	**0.0324**	**0.01**	**0.19**

Ⅳ. RESULTS

- From Tables (Ⅰ) shown, the shrinkage estimator has the best because it gives the lowest value of MSE,FPR and FNR in estimation and the selection of variables. And when (n=30) the estimator of the (Adaptive lasso) penalty function has the best because it gives the lowest value for the FPR rate in case $\theta = 0.25$. Either in case $\theta = 0.50{,}0.75$ The SCAD has the best .
- From Tables (Ⅱ)showns, the Shrinkage estimator has the best, When (n=100) the estimator of a penalty function SCAD is the best.

Ⅴ.CONCLUSIONS

In this article, we proposed the (shrinkage) method and it was compared with the SCAD (Smoothly Clipped Absolute Deviation) and Adaptive Lasso penalty functions. After running the simulations, it was found that the proposed method is the best way to estimate and choose the variables. Becauseit has the lowest mean squared error (MSE) and it has lowestFalse Positive Rate(FPR) and False Negative Rate(FNR).

Ⅵ.REFERENCES

[1] 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.

[2] Yousif, A.H., Proposing robust LAD-shrink set estimator for high dimensional regression model. Int. J. Agric. Stat. Sci, 2021. **17**: p. 1229-1234.

[3] 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.

4] Koenker, R. and G. Bassett Jr, Regression quantiles. Econometrica: journal of the Econometric Society, 1978: p. 33-50.

[5] Amin, M., et al., SCAD-penalized quantile regression for high-dimensional data analysis and variable selection. Statistica Neerlandica, 2015. **69**(3): p. 212-235.

[6] Fan, J. and R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association, 2001. **96**(456): p. 1348-1360.

[7] Wu, Y. and Y. Liu, Variable selection in quantile regression. Statistica Sinica, 2009: p. 801-817.

[8] 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.

[9] 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.

[10] 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.

[11] Li, Y. and J. Zhu, L 1-norm quantile regression. Journal of Computational and Graphical Statistics, 2008. **17**(1): p. 163-185.

[12] Bühlmann, P. and S. Van De Geer, Statistics for high-dimensional data: methods, theory and applications. 2011: Springer Science & Business Media.

[13] 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.

[14] Zou, H., The adaptive lasso and its oracle properties. Journal of the American statistical association, 2006. **101**(476): p. 1418-1429.

[15] Le An, T.H. and P. Dinh Tao, Solving a class of linearly constrained indefinite quadratic problems by DC algorithms. Journal of global optimization, 1997. **11**(3): p. 253-285.

[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, 2020.