International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 8 Issue 3 March - 2024, Pages: 49-53

Comparative study of Lasso penalty function and Laplace error penalty function in the model of Quantile Regression

Asaad Naser Hussein Mzedawee

Asaad.nasir@qu.edu.iq

University of Al-Qadisiyah – College of Administration & Economics, Iraq

Abstract. *In this paper dealt with the study the method of choosing the best model of quantile regression through the use of penalty methods within the model of linear quantile regression . Where the researcher used two penalty methods, and these two methods are Lasso penalty function and Laplace error penalty function) This research dealt with the study of the method of choosing the best model of quantile regression through the use of penalty methods within the model of linear quantile regression. Where the researcher used two penalty methods, and these two methods are Lasso penalty function and Laplace error penalty function). We conducted one Monte Carlo simulation experiment with the assumption of the existence of a real vector for the parameters to be estimated, and this experiment was conducted with the assumption of generating different sample sizes in order to improve each method's level of accuracy. The results of the Monte Carlo simulation indicate that the Laplace penalty function method gave the lowest value to the average square of errors and the positive and negative parameters(FNR,FPR), and thus is better than the Lasso method .*

Keywords: Lasso, Quantile Regression, Laplace penalty function , The problem of linear correlation

1- Introduction

Quantile regression is a statistical model used to estimate conditional subdivisions of a response variable, given a set of predictors. Unlike traditional regression models that estimate the conditional mean, quantile regression provides insight into the distribution of the response variable over different quantities. The quantile regression model can be used to analyze the relationship between variables when there is heterogeneity in the data or when the answer variable's distribution is asymmetric. It estimates the impact of the expected variables on specific quantities of the response variable, capturing information about the upper or lower tails of the distribution. To fit the model of quantile regression, you can use various algorithms such as iterative weighted least squares or optimization methods. The model allows you to estimate multiple quantities at once, providing a comprehensive understanding of the relationship between the predictors and the response variable at different points in its distribution. Applications for quantile regression exist in many domains, including economics, finance, healthcare, and Social Sciences. It is especially useful when analyzing variables with outliers or when the focus is on specific percentages of the distribution. Quantile Regression was proposed by Koenker and Basset in 1978 [4] . Quantitative regression weighs the distances between the values that the regression line predicts and the differentially observed values, and then tries to reduce the weighted distances(1998)[1] .

Suppose we have a random variable with the value the cumulative distribution function

$$F_Y(y) = p(Y \le y) \qquad (1)$$

The £ is given by the estimated Y

$$Q_Y(£) = F_Y^{-1} = \inf\{y : F_Y \ge £\} \qquad (2)$$

Since the : $0 \le £ \le 1$

The quantile regression model is written as follows :

$$\mathrm{Y_i} = X_i^{'}\beta + e_i \qquad (3)$$

Where

$\mathrm{Y_i}$: Variable response vector

$X_i^{'}$: Vector explanatory variables

β : Vector parameters

e_i : Vector of errors

The vector of parameters β of the Model estimates the quantile regression from the following formula :

$$\hat{\beta}_{£} = \arg\min_{\beta\in R} \sum_{i=1}^{n} \rho_{£}(y_i - X_i^{'}\beta) \qquad (4)$$

$\rho_{£}$ is called an adjustment function and its formula is as follows :

$$\rho_{£} = u\{£ - \mathrm{I}(\mathrm{u} < 0)\} \qquad (5)$$

It is difficult to determine the effect of explanatory variables on the response variable in the scatter model (Sparse) because of the large number of explanatory variables in that model, causing the problem of P >n and when this situation occurs, the model suffers from the problem of linear correlation, which caused the collapse of the least squares method OLS, and therefore it is necessary to find some treatments for this problem, one of the prevailing treatments in modern research is to use penalty methods with regression models in order to reduce the number of variables and to obtain the Parsimonious model, where these methods reduce the values of the parameters of the model with zeroing the parameters of explanatory variables that are not important or relevant The effect is very minimal. Quantile regression also contributes to addressing that problem [2]. The partition regression with adaptive lasso SCAD was studied by the researchers Wu and Liu in 2009 and they came to the conclusion from the results of real and simulated data that the model's estimators have Oracle properties [10]. Quantitative regression allows comprehensive and flexible assessments of the changing effects on the results of the stay of interest while providing simple physical explanations on a time scale . In view of this, many quantitative regression methods have easy and stable calculations [5]. In the year 2023, the researchers (Vettam and John) presented a scientific paper entitled some theoretical LEP results sequencing linear regression estimators with Ordinary Least Squares. Theoretical outcomes demonstrated how new penalty function-based solutions differed from those based on convex and non-convex penalty functions. The study's findings demonstrated that the orcal property is satisfied by LEP and arctan penalties[3]. The problem of the paper is the inability of linear regression to estimate the conditional distribution the dependent variable at various times, so use the sectional regression, and use the sectional model with penalty functions due to the large number of variables compared to the sample size .

The paper aims to compare the LEP and Lasso estimators of the partition regression model using simulation, and the FNR pseudo-positive criterion and the FPR pseudo-negative criterion are used as criteria for selecting variables, on the other hand, the LEP and lasso methods are compared through the MSE mean square error criterion .

The paper is included as follows in the second section in which the LEP and Lasso estimator is explained . In the third section, we explain Lasso, and in the fourth section, we explain the simulation procedure . The fifth section contains the most important conclusions and recommendations .Finally, the sixth section includes scientific reference .

2- Laplace error penalty function (LEP)[6],[9]

A LEP two-tuning error penalty function or contraction parameters was proposed by the scientists (wen and wang) in the year (2015). In contrast to other penalty functions, this one was naturally created as a defined series function instead of multi-defined lines, since the smooth function with boundaries is continuously curved and distinguishable without making sudden changes at breakpoints, providing a smoother representation of the data . By contrast, multi-profile lines can exhibit oscillations and vibrations at breakpoints, leading to undesirable results (2022) . The LEP error penalty function is smoother to deal with compared to some penalty functions such as the SCAD function, which encourages us to use it with the quantile regression model , and its mathematical formula is as follows :

$$\hat{\beta}_{£} = \arg\min_{\beta\in R} \sum_{i=1}^{n} \rho_{£}(y_i - X_i^{'}\beta) + \lambda \sum_{j=1}^{p} (1 - e^{\frac{|\beta|}{k}}) \qquad \ldots \quad (6)$$

Since that $p_{\lambda,k}$ is LEP $p_{\lambda,k} = \lambda(1-e^{\frac{\vartheta}{k}})$, β: model parameters , λ, k they are two non-negative tuning or contraction parameters that regulate the size of the penalty and control the degree of approximation to the penalty L_0 , respectively . The penalty function is called **LEP** because the function has the form of a density **LEP** $e^{\frac{\vartheta}{k}}$.(Trzasko & Manuka) (2009) used the $1-e^{\frac{\beta}{k}}$ function, in the reconstruction of the MRI image .

3- Penalty Method (Lasso)[7][8]

The Lasso method is one of the most common methods for excluding variables that have no or little effect on the response variable or on the time model proposed by researcher Tibshirani 1996 . The Lasso method of organization is considered both as a method of selecting variables that have an impact on the response variable and as a method of estimating parameters, and this makes the estimated model highly interpretable .Its formula is formed with the divisional regression model as in the following figure :

$$\hat{\beta}_{£} = \arg\min_{\beta \in R} \sum_{i=1}^{n} \rho_{£}(y_i - X_i^{'}\beta) + \lambda \sum_{j=1}^{p} |\beta_j| \quad \dots \quad (7)$$

λ is a contraction parameter whose value is $\lambda \geq 0$ and is chosen by the Cross_ Validation method or C.V the Generalized method . This was noted by the scientist Tibshirani 1996 . The Lasso method is considered non-derivable, being absolute .

4- Simulation

Simulation experiments are carried out in the Monte Carlo method based on the R program in order to compare the LEP and Lasso penalty capabilities

4-1 Generating independent variables

We use the multivariate normal distribution with the covariance matrix, covariance η and mean(0) to obtain the independent variables p

$x \sim \mathrm{MN}(0,\eta)$ where $\eta_{ij} = \rho^{|i-j|}$, $\rho = 0.5$

- The procedure for generating random errors is according to the normal distribution with an average of 0 and variance $e_i \sim \mathrm{N}(0,\sigma^2)$, $\mathrm{i} = 1,2,.......,\mathrm{n}$.
- Calculation of the response variable using the x Matrix generated in Paragraph (1) plus the error limit generated in Paragraph (2) .
- We make fixed assumptions for the model . σ: Takes the two values 0.6 and 1

$(p = 10, n = 27)$, $\beta = (2,1,6,0,4,0,.....,0)$, $£ = (0.2,0.8)$

4-2 Comparison criteria

The pseudo-positive criterion FNR and the pseudo-negative criterion FPR are used as the criteria for selecting variables, on the other hand, the LEP and lasso methods are compared by the mean square error criterion MSE .

Table (1) when $(p = 10, n = 27)$

£	σ	Estimators	MSE	FPR	FNR

0.2	0.5	LASSO	0.0425	0.141	0
		LEP	0.0327	0.006	0
	1	LASSO	0.6795	0.310	0.116
		LEP	0.5270	0.068	0.35
0.8	0.5	LASSO	0.0504	0.007	0
		LEP	0.0502	0	0
	1	LASSO	0.0932	0.183	0.067
		LEP	0.0499	0.060	0

Table (2) when $(p=10, n=100)$

£	σ	Estimators	MSE	FPR	FNR
0.2	0.5	LASSO	0.0077	0.268	0
		LEP	0.0026	0.250	0
	1	LASSO	0.0609	0.260	0.097
		LEP	0.0478	0.237	0
0.8	0.5	LASSO	0.0950	0.263	0
		LEP	0.0781	0.237	0
	1	LASSO	0.0723	0.276	0.03
		LEP	0.0419	0.247	0.077

It is clear from the above simulation results in Tables (1) and (2) that the LEP method is the best in terms of estimation because it gives the lowest values of the mean square of errors in various sample sizes ,the number of independent variables and all default values(quantities), on the other hand, it is the best in choosing variables because it gave the lowest values at the positive and negative parameters (FNR, FPR).

5- Conclusions and recommendations

1- The LEP penalty method is better than the Lasso penalty method with the divisional regression model in terms of estimation ,selection of independent variables .
2- The LEP penalty method is easy to apply and therefore we recommend using it in other regression models .
3- The penalty methods (LEP and Lasso) are affected by an increase in sample sizes and give better results and give better results when increased, as well as affected by the amounts of quantities that are counterproductive whenever the amount of default values increases, the average error boxes increase with it, so we recommend reducing the amounts of default values

6-Reference

[1] *Buchinsky, M. (1998). Recent advances in quantile regression models: a practical guideline for empirical research. Journal of human resources, 88-126.*

[2] *Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. 96, 1348-1360.*

[3] *John, M., & Vettam, S. (2023). A few theoretical results for Laplace and arctan penalized ordinary least squares linear regression estimators. Communications in Statistics-Theory and Methods, 1-22.*

[4] *Koenker, R., and Bassett, G. (1978), .Regression Quantiles. Econometrica, 46, 33–50.*

[5] *Peng, L. (2021). Quantile regression for survival data. Annual review of statistics and its application, 8, 413-437.*

[6]*Song, Y., Li, Z., & Fang, M. (2022). Robust variable selection based on penalized composite quantile regression for high-dimensional single-index models. Mathematics, 10(12), 2000.*

[7] *Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267-288.*

[8] *Tibshirani, R. J., & Taylor, J. (2012). Degrees of freedom in lasso problems.*

[9] *Wen, C., Wang, X., & Wang, S. (2015). Laplace Error Penalty-based Variable Selection in High Dimension. Scandinavian Journal of Statistics, 42(3), 685-700.*

[10] *Wu, Y., & Liu, Y. (2009). Variable selection in quantile regression. Statistica Sinica, 801-817.*