Scholars Journal of Physics, Mathematics and Statistics OPEN ACCESS
Abbreviated Key Title: Sch J Phys Math Stat
ISSN 2393-8056 (Print) | ISSN 2393-8064 (Online)
Journal homepage: https://saspublishers.com

Investigating Performance of Composite Quantile Regression with and without Penalization

Nwakuya, M. T[1*], Nduka, E. C[2]

[1]Department of Mathematics and Statistics, University of Port Harcourt, VX49+P93, Rumuokwuta Rd, Mgbuoba 500272, Port Harcourt, Nigeria

DOI: 10.36347/sjpms.2022.v09i05.002 | **Received:** 19.05.2022 | **Accepted:** 15.06.2022 | **Published:** 22.06.2022

***Corresponding author:** Nwakuya, M. T
Department of Mathematics and Statistics, University of Port Harcourt, VX49+P93, Rumuokwuta Rd, Mgbuoba 500272, Port Harcourt, Nigeria

Abstract **Original Research Article**

The performance of composite quantile regression with and without penalization was investigated and results compared to Lasso Quantile Regression (LQR), Lasso Regression (LR) and Ridge Regression (RR). The mean square error, Akiake information criteria and mean absolute percentage error were used as the comparative criteria. The comparison was illustrated using real dataset and simulated data sets of sample sizes n=30,100,300,500 and 1000. Five consecutive quantiles; 0.19, 0.39, 0.59, 0.79 and 0.99 were used for the quantile regression methods while the lasso regression and ridge regression were based on the mean effect. Another set of quantiles; 0.25, 0.5, 0.75 and 0.95 were also accessed based on only CQR and CQR_AL. The results shows that the composite quantile regression without penalization (CQR) and composite quantile regression with penalization (CQR_AL) achieved same results with the lowest variance on estimated effects and was best fitted model for very large data sets (n=100, 300, 500 & 1000). For the real life data with sample size of 318, the CQR_AL showed the least MSE=421.7653 and AIC=1930.136, but its prediction accuracy was low at 19.03%. Considering sample size 30, the LQR at the 0.59th quantile had the lowest variance with MSE of 1.3940 and was best fitted with AIC of 340.20, also it was noticed that COR_AL achieved the next lowest value for MSE and AIC. These results led to the conclusion that CQR_AL and CQR can be used alternatively with large data sets. But when the sample size is small the LQR is most suited, but if a combined quantile effect is sort the CQR_AL should be opted for.

Keywords: Composite Quantile Regression, Composite quantile Regression with Adaptive Lasso, Lasso Quantile Regression, Lasso Linear Regression and Ridge Regression.

1. INTRODUCTION

Linear regression is routinely used in most research works because it is a technique well established in modelling and prediction of the response variable conditional on a set of predictor variables or covariates but it is usually restricted by its dependency on some specific assumptions and can only be used to predict the mean effect of the covariates. Quantile regression which was presented by Koenker (2005) as an alternative to linear regression on the other hand is liberated from the dependency of the linear regression's limiting assumptions and further more quantile regression can be used to predict any desired quantile of the response variable, conditional on a set of predictor variables or covariates thereby portraying a holistic covariate effects.Quantile regression has gained so many footings in many areas of research, including survival analysis, Koenker and Geling(2001), economics, Hendricks and Koenker(1992) and many more. Quantile regression estimates the effect of the covariates on different independent quantiles of response variable per time, thus having say "*a*" independent quantile regression models. Zou and Yuan (2008) extended the quantile regression to accommodate a process of analyzing not just a single independent quantile per time but simultaneously estimating the effects on a sequence of "a" multiple conditional quantiles of Y given X at consecutive quantiles such that, $0 < \tau_1 < \tau_2 < \cdots < \tau_a < 1$, this they termed the Composite Quantile Regression (CQR). This CQR model assumes the same covariate effects across the *"a"* quantile levels. The major difference between the quantile regression and the composite quantile regression is that the CQR accommodates additional quantile levels, produces same covariate effect at all quantiles. Jiang and Sun (2021) stated that CQR is a robust and efficient estimation method. In their paper the proposed a penalized CQR estimator for single index models, where they combined CQR with a method that removes bias and produces asymptotical normal estimator that produces valid confidence intervals. The composite quantile regression has shown to achieve a lower variance on estimated effects relative to quantile regression, Pietrosanu *et al*., (2020). Yang and Yang (2020), developed a CQR estimator for linear regression with errors in variable based on instrumental variables. Ma and Wang (2022) proposed a robust and variable selection method for linear quantile regression models based on CQR and Empirical likelihood that accommodates both within-subject correlations and non-ignorable dropouts. Yuan et al (2022), investigated optimal sub-sampling for CQR with massive dataset, where the established that CQR estimator is consistent and asymptotically normal. In other to reduce variance and instability of the regression estimates the coefficients may be regularized by applying some form of penalization, such as the ridge penalization and the Lasso penalization. Regularization is basically placing a penalty on the best fitted model in order to reduce variance and shrink the regression coefficients. The two major linear regression penalization methods are the Lasso Regression (LR) and the Ridge Regression (RR), which have been applied by many researchers including Seber and Lee (2003) that applied RR to suppress the effects of collinearity and reduce the apparent magnitude of the correlation among the predictors in other to obtain more stable estimates of the coefficients in comparison to the Ordinary Least Square estimates and they also observed that it improved the accuracy of the prediction. Many penalized quantile regression methods have also been well developed to inquire into covariate effects at a single or multiple pre-specified quantile levels; this includes the works of Zou and Yuan (2008) and Wang et al (2012). The lasso variable selection is recently the most widely applied penalty for quantile regression. Zou (2006) showed the inconsistency of the lasso variable selection in certain scenarios and then presented adaptive lasso regularization as a solution to the inconsistency. This work intends to investigate the performance of composite quantile regression with and without penalization and compare the results to that of Lasso linear regression, Ridge regression and Lasso Quantile regression. The comparison was illustrated through the use of simulated data sets and real life data set. The analysis was done in R. The rest of the paper is structured in the following headings; Regularized Regression, Quantile regression, Composite Regression, Methodologies, Results & Discussion.

2. REGULARIZED REGRESSION

There are two main regularization techniques in statistical analysis namely Ridge Regression known as the L2 regularization and Lasso Regression known as the L1 regularization. The aim of regularization is stop overfitting of the data and it is achieved through penalization, which is placing a penalty on the best fitted model in order to reduce variance and shrink the regression coefficients. The difference between the two methods is in the way the penalty is assigned. L1 regularization places a constraint that is equivalent to the absolute value of the size of the coefficient while L2 regularization places a constraint on the sum of the squared values of the coefficients. LASSO was introduced by Tibshirani (1996) as a modification of Ridge regression. The L1 penalty not only shrinks the fitted coefficients toward zero but also causes some of the fitted coefficients to be exactly zero when making λ sufficiently large, Li and Zhu (2008). LASSO minimizes the Residual Sum of Squares (RSS) with a selection operator. The selection operator enables it to perform variable selection by shrinking the regression coefficients sometimes to zero. This helps it to select important explanatory variables and estimate regression parameters simultaneously. These penalized methods are linear regression methods with a lasso and ridge penalty, hence the estimate the mean effect of the covariate at the response. Although the penalized linear regression approach is useful, it only focuses on the central tendency of the conditional distribution. Given the linear model thus;

$y_i = X_i'\beta + e_i; i = 1, \dots, N$ …………………….. (1)

Given that;

$e_i = y_i - x_i'\beta$ …………………….. (2)

Where y_i is the univariate response, x_i is a vector of p-dimensional covariates, β is the unknown regression parameter, ε_i is independent and identically distributed random error. The theory of Ridge regression was proposed by Hoerl and Kennard in (1970) and the LASSO coefficient estimation as presented by Tibshirani (1996) is obtained by minimizing the following;

$\sum_{i=1}^{n} |y_i - x_i'\beta| + \lambda_{lasso}|\beta_j|$ …………………….. (3)

Where $\lambda > 0$ is regularization parameter that strikes the balance between estimation of β_j and the variable selection, basically it is the amount of shrinkage that controls the strength of the penalty. The Ridge Regression estimator is presented thus;

$\sum_{i=1}^{n} |y_i - x_i'\beta| + \lambda_{ridge} \sum_{j=1}^{m} \beta_j^2$ ………………….. (4)

3. QUANTILE REGRESSION

The quantile regression (QR) framework provides a pragmatic approach in understanding the differential impacts of covariates along the distribution of an outcome. In a quantile regression we are concerned with the linear model as given in equation (1). When the effect on a single quantile $\tau \epsilon$ 0,1 on distribution of the response variable is of interest, the quantile regression estimation of β as proposed by Koenker (2005) is achieved by solving the linear programming problem below;

$$\left(\hat{\beta}_{\tau}^{QR}\right) = \underset{\beta}{augmin} \sum_{i=1}^{n} \rho_{\tau}(y_i - x_i'\beta_{\tau}) \text{ ……………. (5)}$$

Given the residual/error notation as $e_i = y_i - x_i'\beta_{\tau}$, the loss function $\rho_{\tau}$is defined as;

$$\rho_{\tau}(.) = \begin{cases} y_i - x_i'\beta_{\tau}(\tau - 1) \; if \; y_i - x_i'\beta_{\tau} < 0 \\ y_i - x_i'\beta_{\tau}(\tau) \; if \; y_i - x_i'\beta_{\tau} \geq 0 \end{cases} \text{ …….. (6)}$$

Where τ represents the quantile. The estimated coefficients ($\hat{\beta}_{\tau}'s$) may be interpreted as marginal effect on conditional quantile of interest, Koenker and Hallock (2001). Going with the regularization condition, quantile regression is more efficient than the least square and its relative efficiency with respect to the least square is very small, Jiang and Yu (2018). A useful alternative approach to the penalized linear regression is to use penalized quantile regression. The quantile regression can be regularized by imposing the lasso penalty to the quantile model giving rise to the quantile regression with lasso penalty with the linear programming problem given below;

$$\left(\hat{\beta}_{\tau}^{QR}\right) = \underset{\beta}{augmin} \sum_{i=1}^{n} \rho_{\tau}\left(y_i - \sum_{j}^{m} x_i'\beta_{\tau}\right) + \lambda_{lasso}|\beta_{\tau j}| \text{ …………………….. (7)}$$

This simple modification allows Lasso to also perform variable selection because the shrinkage of the coefficients is such that some coefficients can be shrunk exactly to zero. Lasso can effectively select important explanatory variables and estimate regression parameters simultaneously. The combination of the quantile regression and Lasso penalty is computationally easy to implement via the standard linear programming procedure.

4. Composite Quantile Regression

In the quest to improve on quantile regression, Zou and Yuan (2008) proposed the composite quantile regression (CQR) for estimating the regression coefficients in the classical linear regression model. They revealed that in comparison to the least squares estimator the relative efficiency of the CQR estimator is greater than 70% notwithstanding the distribution of the error term. This composite quantile regression is a method designed to simultaneously analyze a multiple quantile regression model to yield the same regression coefficients across different quantiles, it simultaneously estimates a sequence of q conditional quantiles of y given X at levels 0< τ_1< τ_2<…<τ_q where q=. Zou and Yuan (2008) presented evidence to show the strength of the combination of multiple quantile regression models. The composite quantile regression has been showed to be selection consistent and can be more robust in various circumstances, Yuan *et al.*, (2022). The CQR method is basically a sum of different quantile regressions with equal weights.

The CQR method to estimate β_0 as follows;

$$\left(\hat{b}_1, \dots, \hat{b}_q, \hat{\beta}^{CQR}\right) = \text{argmin}_{(\hat{b}_1,\dots,\hat{b}_q,\beta)} \sum_{q=1}^{Q} \sum_{i=1}^{N} \rho_{\tau_q}\{y_i - b_q - X_i'\beta\} \text{ … (8)}$$

Where

$\rho_{\tau_q}(.), q = 1, \dots, Q, is\ as\ defined\ by\ equation\ (3), given\ that; 0 < \tau_1 < \tau_2 < \dots < \tau_Q < 1,$ and $\hat{b}_q$ is the estimator of $b_{\tau_q}$and b_{τ_q} is the τ_q quantile of e_i.

The CQR with adaptive lasso estimates the regression parameters by solving the linear programming problem given by;

$$\left(\hat{b}_1, \dots, \hat{b}_q, \hat{\beta}^{CQR}\right) = \text{argmin}_{(\hat{b}_1,\dots,\hat{b}_q,\beta)} \sum_{q=1}^{Q} \sum_{i=1}^{N} \rho_{\tau_q}\{y_i - b_q - X_i'\beta\} + \lambda \sum_{j=1}^{m} \frac{|\beta_j|}{|\hat{\beta}^{CQR}|^2} \text{ ……………… (9)}$$

Where λ > 0 is a regularization parameter and$\hat{\beta}^{CQR}$is the solution (without intercepts) to the non-regularized composite quantile regression problem. The minimization problem in equation (8) and equation (9) were done using the one-at-a- time coordinate descent algorithm.

5. METHODOLOGY

The methods applied in this work include the lasso linear regression (LLR), Ridge regression (RR), Lasso Quantile regression (LQR), Composite Quantile Regression (CQR) and Composite Quantile Regression with adaptive lasso (CQRAL).Three data sets; a real life data of 318 individuals with measurements on their age, body mass index, pulse rate and systolic blood pressure was used as well as two simulated datasets with sample size of 30, 100, 300, 500 and 1000. The five regression methods (LLR, RR, LQR, CQR and CQRAL) were all applied to these three data sets. The LQR, CQR and CQRAL were analyzed at five consecutive quantiles; 0.19, 0.39, 0.59, 0.79 and 0.99 while only CQR and CQRL were also analyzed at another set of quantiles 0.25, 0.50, 0.75 and 0.95. In the analysis the LQR estimated the covariate effect at all the five quantiles independently while CQR and CQRAL estimated a combined quantile effect that is producing the same covariate effect at all the different sets of quantiles. We investigated the results of these methods using the following performance criteria; the mean square error (MSE) which measures the amount of error in a statistical model, mean absolute percentage error (MAPE) which measures in percentages the accuracy of a method in predicting the response from the observed and Akaike information criterion (AIC) which is used in model selection. The lasso Linear regression is presented in equation (3), The Ridge regression is presented in equation (4), Quantile regression with Lasso penalty is given in equation(7), Composite quantile regression is given in equation (8) and the Composite quantile regression with adaptive lasso is given in equation (9). The comparison criteria are defined as follows;

MSE= $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

MAPE= $\frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$

AIC= $2 * ln(L) + 2k$

Where y_i is the observed response, $\hat{y}_i$ is the predicted response, n is the sample size, L is the likelihood and k is the number of estimated parameters.

6. RESULTS AND DISCUSSIONS

Table (1) to Table (6) presents results for quantiles 0.19, 0.39, 0.59, 0.79 and 0.99.

Table 1: Results from real life data

Methods	Tau	Intercept	age	PR	BMI	Comparison criteria		
						MSE	AIC	MAPE
QR_Lasso	0.19	63.3225	-0.2952	-0.2016	0.2801	785.1009	2127.728	20.6085
	0.39	60.4034	-0.1473	-0.3268	0.4726	500.4407	1984.526	17.6882
	0.59	67.4297	-0.6288	-0.5998	0.7753	439.0811	1942.930	19.7902
	0.79	56.0467	-0.0537	-0.1131	0.8132	704.2579	2093.172	28.9144
	0.99	40.0392	-0.7173	3.3272	0.6044	1428.7231	2317.989	33.1888
LR_Lasso	-	53.3744	-	-	0.5505	425.5982	1933.012	19.3089
Ridge R	-	56.9252	-0.1729	-0.1529	0.6017	422.5469	1930.722	19.1070
CQR	-	54.9729	-0.2294	-0.2004	0.6519	422.4469	1930.644	18.7588
CQR_AL	-	55.3339	-0.2267	-0.1681	0.6467	**421.7653**	**1930.135**	19.0331

Table 1 above shows the results from the real life data considered in this work. The results reveal that the composite quantile regression with lasso penalty is the best method compared to the other methods with the lowest MSE of 421.7653, AIC of 1930.135 and prediction accuracy of 19.033, and its results is approximately same with CQR. We also notice that the LLR method shrunk the coefficients of Age and PR to zero, thereby selecting only BMI, indicating that both variables (Age and PR) do not contribute significantly to the response but the BMI does.

Table 2: Results from Simulated data with sample size 30

Methods	Tau	Intercept	Age	PR	BMI	Comparison criteria		
						MSE	AIC	MAPE
QR_Lasso	0.19	0.1985	-0.2243	-0.1925	0.6912	1.9248	662.281	1.2952
	0.39	-0.4695	-0.1741	-0.1822	0.6647	1.4526	382.5873	5.545
	0.59	0.3902	-0.1945	-0.1856	0.6619	**1.3940**	**340.201**	8.0115
	0.79	0.4824	-0.2055	-0.1922	0.6963	1.4517	380.7435	11.4368
	0.99	-0.2748	-0.0547	-0.1319	0.4807	4.6499	1544.856	26.000
LR_Lasso	-	1.0665	-	-0.1064	-	10.6933	2377.618	2.3039
Ridge R	-	-0.6077	-0.0796	-0.0307	0.3741	3.4772	1254.213	10.0883
CQR	-	0.1594	-0.1937	-0.1865	0.6852	1.4188	357.846	11.3332
CQR_AL	-	.	-0.1918	-01879	0.6819	1.4093	351.1215	7.6049

Table 2 shows the results from simulated data of sample size 30. The results show that the quantile regression with lasso penalty produced the best results at the 0.59 quantile with the lowest MSE and AIC with a low prediction accuracy compared to that of 0.79 quantile. Based on these results we can say that quantile regression with lasso penalty is better suited when the sample size is not very large in comparison to the other methods considered. The results also show a noticeable difference between the MSE and AIC of CQR and CQR_AL. The LLR was seen to select Pr for the model while age and BMI coefficients were shrunk to zero.

Table 3: Results from Simulated data with sample size 100

Methods	Tau	Intercept	Age	PR	BMI	Comparison criteria		
						MSE	AIC	MAPE
QR_Lasso	0.19	2.2527	-0.0427	-0.0163	-0.0256	10.7373	245.3721	131.1697
	0.39	1.5206	-0.2038	-0.1939	0.6379	0.1795	-163.783	140.898
	0.59	2.0417	-0.2141	-0.1946	0.6330	0.1689	-169.836	104.5314
	0.79	1.9221	-0.18367	-0.1924	0.6201	0.6166	-40.3525	139.1497
	0.99	3.6472	-0.1915	-0.1851	0.5515	1.0628	14.0906	231.8258
LR_Lasso	-	2.9902	-	-0.0547	-	8.2789	219.372	146.7483
Ridge R	-	0.8828	-0.1263	-0.1232	-0.4021	33.6300	359.5419	518.194
CQR	-	1.5999	-0.2067	-0.1914	0.6352	**0.1631**	**-173.317**	120.059
CQR_AL	-	1.58711	-0.2055	-0.1917	0.6358	**0.1632**	**-173.263**	118.6314

Table 3 results show that the CQR and CQR_AL produced similar and best results with the lowest MSE and AIC with prediction accuracy of 120. Based on these results we can say that either CQR or CQRL can be used when the sample size is large. The LLR was seen to select PR for the model while Age and BMI coefficients were shrunk to zero.

Table 4: Results from Simulated data with sample size 300

Methods	Tau	Intercept	Age	PR	BMI	Comparison criteria		
						MSE	AIC	MAPE
QR_Lasso	0.19	1.6783	-0.1896	-0.1898	0.6380	1.8580	187.013	83.8157
	0.39	1.1093	-0.2014	-0.1900	0.6399	1.8124	186.3928	173.2947
	0.59	1.6329	-0.2049	-0.1909	0.6321	2.6523	300.6225	213.6329
	0.79	1.2810	-0.1999	-0.1920	0.6595	1.0629	26.2890	126.2709
	0.99	1.7860	-0.1857	-0.1943	0.6640	1.981	199.8312	93.4510
LR_Lasso	-	3.0856	-	-0.0542	-	18.7816	887.863	441.85
Ridge R	-	1.0088	-0.1296	-0.1222	0.3975	2.9729	334.8684	157.3103
CQR	-	1.2534	-0.1971	-0.1918	0.6414	**1.8050**	**185.1692**	171.531
CQR_AL	-	1.2533	-0.1971	-0.1918	0.6414	**1.8031**	**184.8504**	171.425

The table 4 shows that the best fitted model was CQR and CAR_AL with a prediction accuracy of 171. It was also observed that the CQR and CQR_AL both produced similar results agreeing with the fact that when sample size is large the lasso penalty is of no effect hence either of the methods can be applied.

Table 5: Results from Simulated data with sample size 500

Methods	Tau	Intercept	Age	PR	BMI	Comparison criteria		
						MSE	AIC	MAPE
QR_Lasso	0.19	0.9956	-0.2078	-0.1908	0.6432	12.1559	268.7095	102.147
	0.39	0.9548	-0.2032	-0.1891	0.6476	13.14623	265.614	99.9503
	0.59	1.1373	-0.2014	-0.1896	0.6499	13.1450	265.605	98.658
	0.79	1.2843	-0.1983	-0.1903	0.6561	13.1480	265.6271	98.725
	0.99	2.2199	-0.2056	-0.1917	0.6514	13.1951	265.9844	108.338
LR_Lasso	-	3.2964	-	-0.0588	-	13.7486	270.0939	105.0024
Ridge R	-	0.6228	-0.1359	-0.1228	0.4201	13.2611	266.45837	106.000
CQR	-	1.0443	-0.1999	-0.1906	0.6510	**13.1434**	**265.592**	98.889
CQR_AL	-	1.0442	-0.1999	-0.1906	0.6509	**13.1434**	**265.592**	98.889

The table shows that CQR and CQR_AL produced same results and the also produced the least MSE and AIC, making them the best suited model with a prediction accuracy of 98.89.

Table 6: Results from Simulated data with sample size 1000

Methods	Tau	Intercept	Age	PR	BMI	Comparison criteria		
						MSE	AIC	MAPE
QR_Lasso	0.19	0.5952	-0.1993	0.1914	0.6537	0.3484	-1046.3	66.3049
	0.39	0.7374	-0.1989	-0.1898	0.6540	0.2134	-1536.7	48.258
	0.59	0.6825	-0.1909	-0.1905	0.6620	0.2060	-1571.8	49.626
	0.79	0.9817	-0.1921	-0.1906	0.6612	0.3168	-1141.4	64.184
	0.99	2.0637	-0.2049	-01935	0.6646	1.3767	327.68	162.311
LR_Lasso	-	-	-0.0537	-	0.0018	17.2204	691.118	484.56
Ridge R	-	-	-0.1315	-0.1227	0.4232	1.9860	2854.09	101.37
CQR	-	0.8047	-0.1959	-0.1909	0.6578	**0.1957**	**-1623.4**	47.574
CQR_AL	-	0.8046	-0.1959	-0.1909	0.6578	**0.1957**	**-1623.4**	47.575

The table 6 above has shown that the composite quantile regression and composite quantile regression with lasso penalty produced the same results, though their prediction accuracy is seen to be very low compared to the other methods. Based on these results we can say that when the sample size is very large, the lasso penalty has no effect in a composite quantile regression, therefore either of them can be used when the sample size is very large.

Table 7: Combined Coefficients results for quantiles 0.25, 0.50, 0.75 and 0.95

Sample sizes	Methods	Intercept	Age	PR	BMI
30	CQR_AL	3.3377	-0.2171	-0.1998	0.5963
	CQR	3.4969	-0.2078	-0.1967	0.6049
100	CQR_AL	1.5964	-0.2061	-0.1915	0.6349
	CQR	1.6071	-0.2062	-0.1915	0.6354
300	CQR_AL	1.2533	-0.1970	-0.1919	0.6415
	CQR	1.2533	-0.1970	-0.1919	0.6415
500	CQR_AL	1.6441	-0.1978	-0.1906	0.6510
	CQR	1.6444	-0.1998	-0.1906	0.6510
1000	CQR_AL	0.8043	-0.1955	-0.1908	0.1908
	CQR	0.8044	-0.1958	-0.1909	0.6578

The table shows that both methods produced similar results at large sample sizes with a little difference at n=30. When these results are compared to the results from quantiles 0.19, 0.39, 0.59, 0.79 and 0.99 it is observed that the coefficients values are equally similar. This can be attributed to the fact that both sets of quantiles practically cover the whole distribution of the response variable, hence giving the same combined results.

7. CONCLUSION

Motivated by the lack of awareness of the composite quantile regression, this paper tried to show its advantages over the penalized quantile regression, lasso linear regression and the ridge regression. The paper considered both the penalized and un-penalized composite quantile regression (ie CQR and CQR_AL) and did a comparative study between them and quantile regression, lasso regression and ridge regression. The analysis was done on two sets of quantiles, for the quantile regression methods while the lasso regression and ridge regression are based on the mean effect. The combined results at quantiles 0.19, 0.39, 0.59, 0.79 and 0.99 from CQR and CQR_AL were compared to the combined results at quantiles 0.25, 0.50, 0.75 and 0.95, and it was seen that the results from both sets of consecutive quantiles was the same. This can be attributed to the fact that both sets of quantiles practically cover the whole distribution of the response variable, hence giving the same combined results. The results also show that CQR and CQRAL achieves same results with the lowest variances on estimated effects for very large data sets (100, 300, 500 and 1000) in comparison to the other methods considered in this paper. But for n=30 the results show that QRL achieved the least mean square error and AIC hence the best fitted model. We can say that even though the QR produced the least variance when the sample size 30 (ie, not very large), it cannot combine the individual quantile effects to produce a unique effect on all the quantiles. In conclusion we can say that the CQR and CQRAl can be used interchangeably for large data sets as the both produce same result with large data sets. But for not very large data sets the CQRAl is better suited compared to CQR.

REFERENCES

- Hendricks, W., & Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American statistical Association*, *87*(417), 58-68.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67.
- Jiang, R., Hu, X., Yu, K., & Qian, W. (2018). Composite quantile regression for massive datasets. *Statistics*, *52*(5), 980-1004. DOI:10.1080/02331888.2018.1500579
- Jiang, R., & Sun, M. (2021). Single Index Composite Quantile Regression for Ultra High Dimensional Data. https:// doi.org/10.1007/s11749-021-00785-9
- Koenker, R. (2005). Quantile Regression. New York: Cambridge University Press.
- Koenker, R., & Hallock, K. (2001). Quantile Regression; Journal of Economic Perspectives, vol 15, pp 143–156.
- Li, Y., & Zhu, J. (2008). L 1-norm quantile regression. *Journal of Computational and Graphical Statistics*, *17*(1), 163-185.
- Ma, W., & Wang, L. (2021). Improved composite quantile regression and variable selection with nonignorable dropouts. *Random Matrices: Theory and Applications*, 2250010.
- Pietrosanu, M., Gao, J., Kong, L., Jiang, B., & Niun, D. (2020). Advanced Algorithms for Penalized Quantile and Composite Quantile Regression; Computational Statistics. https://doi.org/10.1007/s00180-020-01010-1
- Seber, G. A. F., & Lee, A. J. (2003). Linear Regression Analysis. John Wiley & Sons, New Jersey, 2 edition, pp 326-329.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.
- Wang, L., Wu, Y., & Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, *107*(497), 214-222.
- Yang, W., & Yang, Y. (2020). Composite quantile regression estimation of linear error-in-variable

models using instrumental variables. *Metrika*, *83*(1), 1-16.

- Yuan, X., Li, Y., Dong, X., & Liu, T. (2022). Optimal subsampling for composite quantile regression in big data. *Statistical Papers*, 1-28. https://doi.org/10.1007/s00362-022-01292-1
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418-1429.Zou, H., & Yuan, M. (2008). Regularized simultaneous model selection in multiple quantiles regression. *Computational Statistics & Data Analysis*, *52*(12), 5296-5304.