Behavior of Lasso Quantile Regression with Small Sample Sizes

Dr. Elham Abdul-Razik Ismail
Assistant Professor of Statistics, Faculty of Commerce,
Al-Azhar University (Girls' Branch)

***Abstract*—Quantile regression is a statistical technique intended to estimate, and conduct inference about the conditional quantile functions. Just as the classical linear regression methods estimate models for conditional mean function, quantile regression offers a mechanism for estimating models for conditional median function, and the full range of other conditional quantile functions. In this paper describe, compare, and apply the two quantile regression (L_1-Lasso, L_2-Lasso) suggested approaches. The two quantile regression suggested approaches used to select the best subset of variables and estimate the parameters of the quantile regression equation when small sample sizes are used. The aim of this study is to study the behavior of L_1-Lasso and L_2-Lasso quantile regression method when small sample sizes are generated. Simulations show that the proposed approaches are very competitive in terms of variable selection, estimation accuracy and efficient when small sample sizes are used. All results showed superiority of L_1-Lasso compared with L_2-Lasso linear programming methods.**

Keywords—Quantile Regression – Small Sample size – Selection of Variables - estimated risk – relative estimated risk.

1-INTRODUCTION

Quantile regression (1) has gained increasing popularity as it provides richer information than the classic mean regression. Quantile regression is a statistical technique intended to estimate, and conduct inference about the conditional quantile functions. Just as the classical linear regression methods estimate models for conditional mean function, quantile regression offers a mechanism for estimating models for conditional median function, and the full range of other conditional quantile functions. An efficient algorithm proposed to compute the entire solution path of the lasso regularized quantile regression (2). (3) Focus on the variable selection aspect of penalized quantile regression. Under some mild conditions, the study demonstrates the oracle properties of the SCAD and adaptive-lasso penalized. (4) Consider quantile regression in high-dimensional sparse models. In such models, the overall number of regressors is very large, possibly much larger than the sample size. (5) Proposed the composite quantile regression estimator by averaging quantile regressions. The study showed that the composite quantile regression is selection consistent and can be more robust in various circumstances.

(6) Extended this work to analyzing a Tobit quantile regression model, a form of censored model in which $y_i = y_i^*$is observed if $y_i^*> 0$ and $y_i = 0$ is observed otherwise. A regression model then relates the unobserved y_i to the covariants x_i. (7)) used the AL likelihood and combine Markov Chain Monte Carlo (MCMC) with the expectation maximizing algorithm (EM) to carry out inference on quantile regression for longitudinal data. (8) used the AL likelihood combined with non-parametric regression modeling using piecewise polynomials to implement automatic curve fitting for quantile regression and (9) used the same approach but using natural cubic splines. (10) pointed out that the value of not only controls the quantile but also the skewness of the AL distribution resulting in limited explicitly. The residual distribution is symmetric when modeling the median. This motivated (11) and (10) considered a more exible residual distribution constructed using a Dirichlet process prior but still having the quantile equal to zero. The study of (10) included a general scale mixture of AL densities with skewness τ in their analysis, but conclude that in terms of ability to predict new observations, a general mixture of uniform distributions performs the best.

The study investigates the methodology and theory of non-convex, penalized quantile regression in ultra-high dimension. The proposed approach has two distinctive features: first it explore the entire conditional distribution of the response variable, given the ultra-high-dimensional covariates, and provides a more realistic picture of the scarcity pattern; two it requires substantially weaker conditions compared with alternative methods in the literature; thus, it greatly alleviates the difficulty of model checking in the ultra-high dimension. In theoretic development, it is challenging to deal with both the non smooth loss function and the non convex penalty function in ultra-high-dimensional parameter space. The study introduces a novel, sufficient optimality condition that relies on a convex differencing representation of the penalized loss function and the sub differential calculus (12).

A new non-parametric regression technique called local composite quantile regression smoothing to improve local polynomial regression further are proposed. Sampling properties of the estimation procedure proposed are studied. The study derives the asymptotic bias, variance and normality of the estimate proposed. The asymptotic relative efficiency of the estimate with respect to local polynomial regression is investigated. It is shown that the estimate can be much more efficient than the local polynomial regression estimate for various non-normal errors, while being almost as efficient as the local polynomial regression estimate for normal errors. Simulation is conducted to examine the performance of the estimates proposed. The simulation results are consistent with our theoretical findings. A real data example is used to illustrate the method proposed (13).

A new algorithm to directly solve the doubly regularized support vector machine without resorting to approximation as in hybrid huberized support vector machine is introduced. The suggested method is based on the alternating direction method of multipliers the alternating direction method of multipliers. The study demonstrate that the method is efficient even for large-scale problems with tens of thousands variables (14). A tuning parameter selection criterion based on variable selection stability is proposed. The key idea is that if multiple samples are available from the same distribution, a good variable selection method should yield similar sets of informative variables that do not vary much from one sample to another. The effectiveness of the proposed selection criterion is demonstrated in a variety of simulated examples and a real application. More importantly, its asymptotic selection consistency is established, showing that the variable selection method with the selected tuning parameter would recover the truly informative variable set with probability tending to one. The aim of this study is to study the behavior of L_1-Lasso and L_2-Lasso quantile regression method when small sample sizes are generated.

The organization of the study is as follows: In Section 2 the study described lasso and quantile Lasso methods which used in this study. Section 3 described the step which applied in this simulation study. Results and discussion are given in Section 4. Finally, concluding remarks are provided in Section 5.

2- LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR (Lasso)

The Lasso is a regression method proposed by R. Tibshirani in 1996. Similar to Ordinary Least Squares (OLS) regression, Lasso minimizes the Residual Sum of Squares (*RSS*) but poses a constraint to the sum of the absolute values of the coefficients being less than a constant. This additional constraint is moreover similar to that introduced in Ridge regression, where the constraint is to the sum of the squared values of the coefficients. This simple modification allows Lasso to perform also variable selection because the shrinkage of the coefficients is such that some coefficients can be shrunk exactly to zero.

The linkage between OLS, Lasso and Ridge is:

$$\text{OLS} = \min(\textstyle\sum_{i=1}^{n}(y_{i-}\ \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 \qquad (1)$$

$$\text{LASSO} = \min(\textstyle\sum_{i=1}^{n}(y_{i-}\ \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda\sum_{j=1}^{p}|\beta_j|) \qquad (2)$$

$$\text{Ridge} = \min(\textstyle\sum_{i=1}^{n}(y_{i-}\ \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda\sum_{j=1}^{p}\beta_j^2) \qquad (3)$$

where *n* is the number of objects, *p* the number of variables and λ a parameter.

It can be said that Lasso is an improvement over Ridge, in that Lasso has the beneficial aspects of Ridge, *i.e.* higher bias and lower variance (compared to OLS), but also allows to select variables, leading to an enhanced interpretability of the developed models. The λ parameter can be tuned in order to set the shrinkage level, the higher the λ is, and the more coefficients are shrunk to zero (16)(17).

The paper proposed and compares the ideas from Lasso by using the L_1 penalty and L_2 penalty. The study applied the two cases when the least squares loss replaced by the L_1 loss in quantile regression model. Applied L_1 penalty gain the following two advantages. First, it allows the researcher to penetrate the difficult problem of variable selection for the L_1 regression. Appealingly, the shrinkage property of the Lasso estimator continues to hold in L_1 regression. Second, the single criterion function with both components being of L_1-type reduces (numerically) the minimization to a strictly linear programming problem, making any resulting methodology extremely easy to implement. To be specific, our proposed estimator is a minimize of the following criterion function

$$\textstyle\sum_{i=1}^{n}\rho|Y_i - \beta' x_i| + \lambda\sum_{j=1}^{n} \parallel \beta_j \parallel_1 \qquad (4)$$

and

$$\textstyle\sum_{i=1}^{n}\rho|Y_i - \beta' x_i| + \lambda\sum_{j=1}^{n} \parallel \beta_j \parallel_2 \qquad (5)$$

It can be equivalently defined as a minimize of the objective function

$$\textstyle\sum_{i=1}^{n}\rho|Y_i - \beta' x_i| \qquad (6)$$

$$subject\ to \textstyle\sum_{j=1}^{n} \parallel \beta_j \parallel_1 \leq \lambda \qquad (7)$$

and

$$\textstyle\sum_{i=1}^{n}\rho|Y_i - \beta' x_i| \qquad (8)$$

$$subject\ to \textstyle\sum_{j=1}^{n} \parallel \beta_j \parallel_2 \leq \lambda \qquad (9)$$

where $\parallel \beta_j \parallel_1$ is the usual L_1 estimator and $\parallel \beta_j \parallel_2$ is the usual L_2 estimator. The tuning parameter λ there plays a crucial role of striking a balance between estimation of β_j and variable selection. Large values of β tend to remove variables and increase bias in the estimation aspect while small values tend to retain variables. Thus it would be ideal that a large β be used if a regression parameter is zero (to be removed) and small values be used if it not zero. To this end, it becomes clear that the researcher need a separate β for each parameter component β_j. Where *p* is the quantile values. The quantile notion generalizes specific terms like quartile, quintile, decile, and percentile. The p^{th} quantile denotes that value of the response below which the proportion of the population is p. Conditional-median regression is a special case of quantile regression in which the conditional 0.5^{th} quantile is modeled as a function of covariants. More generally, other quantiles can be used to describe noncentral positions of a distribution.

The L_1 penalty and L_2 penalty was used in the Lasso for variable selection. The least square (L_2) and least absolute deviation (L_1) regression are a useful method for robust regression, and the least absolute shrinkage and selection operator Lasso is a popular choice for shrinkage estimation and variable selection.

3-SIMULATION STUDY

This section discusses the numerical simulation of the two models under consideration, L_1-Lasso and L_2-Lasso for estimation and selection of variables of the quantile regression model if the error distribution has heavy tailed, skewed normal distribution, and long tails. The simulation study discusses steps which applied to evaluate the performance of the two approaches L_1-Lasso and L_2-Lasso, for selection of variables and estimation of parameters are as following:

-Quantile regression model was used as $y = x_i'\beta + \varepsilon_i$ or (y = 0.85x_1 + 0.85x_2 +0.85x_3 +0.85x_4 +0.85x_5 +0.85x_6 +0.85x_7 +0.85x_8), where the true value for the β's are set as (0.85,0.85,0.85,0.85,0.85,0.85,0.85,0.85) ′ . The previous values are dense scenario which chosen arbitrary from many studies like (2), (18), (19). The study generated x_i's from normal distribution (0, 1) during the simulation study.

-The simulation study applied $p \in (0.1; 0.90)$, depend on that the quantile nation generalizes specific terms like quartile, quintile, decile and percentile. Where p is quantile values which are arbitrary chosen from the previous specific terms.

-The study calculated the quantile regression models with intercept $\beta_\circ$ that for the intercept important in economic application. -The study when applied the (L_1-Lasso, L_2-Lasso) linear programming methods used a regularization or penalty parameter (¸ λ= 2) as a constant referred to (20).

-The study generated ε_i from different distributions which generated outliers with different parameter, so as to explain the influence of the change in the error distributions on the quantile regression equation, which is the basis for choosing through the L_1-Lasso and L_2-Lasso approaches.

-The Lognormal, Cauchy and skewed normal were employed to generate a long tailed distribution to estimate the parameters of the quantile regression model.

-Small random samples of size n=10, 15 and20 are generated using the GAMS 2.25. The study interested in the small sample sizes to evaluate the performance of the two approaches L_1-Lasso and L_2-Lasso, for selection of variables and estimation of parameters for quantile regression model.

-This study introduced a program by using GAMS 2.25 statistical package to calculate the L_1-Lasso and L_2-Lasso, estimators.

-For each error distribution with two different shape parameters and for each sample size, the estimates of β_j where j = 1,………………, 8 where calculated for L_1-Lasso and L_2-Lasso.

-Two suggested L_1-Lasso and L_2-Lasso method used to estimate quantile regression model with error distribution and follow each of the following four distributions and their parameters respectively,

Lognormal ~ log (0, 0.3) and (0, 2.5);

Cauchy ~ C (0, 0.5) and (0, 0.9);

Skewed Normal ~ N (0, 12) and (0, 20).

-These distributions have been generated using the above parameters that were chosen arbitrarily and taken from many previous studies to study the behavior and the performance for the methods under considerations when small sample sizes are applied.

-In this study the estimated risk and relative estimated risk are used as criteria to compare between the solutions of L_1-Lasso and L_2-Lasso.

-The estimated risk of the estimators $\hat{\beta}_j$where j = 1,…, 8 are used to measure its performance, where the ER's of the estimators $\hat{\beta}_j$ for the parameters β_j the true value which suggested is defined by

$$ER(\hat{\beta}) = MSE = \sum_{r=1}^{R} \frac{(\hat{\beta}-\beta)^2}{R} \qquad (10)$$

Where R is the number of repeated samples and $\widehat{\beta}$ are the estimates calculated from the sample for β (true value) for each parameter (21).

-The study applied sampling runs (number of repeated) 500 replications for each distribution with the two different parameters and two different sample sizes to be sure of consistency of the results.

-For all sample sizes, for all approaches, and for all distribution parameters for the four distributions, the ER's for each parameter β_j were calculated using each method separately.

-The criteria to evaluate the performance for the two methods under considerations depend on the approach, which will produce a small ER for all parameters then it, would be considered more suitable when the objective is to select the variables and estimate the parameters.

4-RESULTS

This section concerned with the results related with simulation study for two methods under consideration; the two methods of linear programming (L_1 Lasso and L_2-Lasso). The study concerned with the behavior and the performance for the L_1 Lasso and L_2 -Lasso methods when small sample sizes are applied.

L_1 Lasso and L_2-Lasso methods used to select the best subset of variables and estimate the parameters of the quantile regression equation when three error distributions, with three different small sample sizes and two different parameters for each error distribution.

First: The two methods for variable selection in quantile regression.

The first aim of this section is to discuss the result of the comparison between L_1-Lasso and L_2-Lasso estimators when it used to select best subset of variables of the quantile regression models with small sample sizes. The two methods under consideration are used in selection of best subset of variables in quantile regression models. When the two methods are applied to select variables considering the true parameters β = (0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85) and fixed value of quantile p = (0.10, 0.90) and three different error distributions are considered (Cauchy, skewed normal and log normal) with two different parameters for each distribution.

The results for the first aim demonstrated that:

-two methods under consideration are tends to produce the same coefficient that are (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) zero and not exactly not zero coefficients . Table (1) shows the results on how frequently each variable was selected by L_1-Lasso and L_2-Lasso when ε ~ Lognormal distribution with (0, 2.5) parameter. Two suggested methods approximately deleting the same variables and selected the same values.

Table (1) Number of times each predictor variable was selected (out of 100 repetitions) by the L_1-Lasso and L_2-Lasso when ε ~ Lognormal distribution with (0, 2.5) parameter.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
L_1-Lasso	83	68	53	50	53	67	58	68
L_2-Lasso	55	75	58	50	53	62	65	50

- The estimators are calculated for each variable which selected and estimated risk for L_1-Lasso and L_2-Lasso programming estimators are calculated for the selection variables. Tables (2) to (7) collect the results of ER and named the variable which the two methods selected.

Second: The two methods for estimating the parameters

The second aim of this section is to discuss the results of the comparison between L_1-Lasso and L_2-Lasso estimators when the two methods under consideration are used to estimate the parameters for quantile regression models. Estimator which produced by L_1-Lasso and L_2-Lasso used to calculate the ER. The results for the second aim as follows:

-Estimated risk which produced by L_1-Lasso is less than ER which calculated by, L_2- Lasso.

Third: The two methods with the quantile regression values

The third aim of this section is to discuss the result of the comparison between L_1-Lasso and L_2-Lasso estimators when different value of quantiles are used to estimate the parameters and selection of variables of the quantile regression models. When the values of quantile are p = (0.1, 0.90). When three different error distributions are used, with three small sample sizes and each distribution have two different parameters. Table (2) to (7) appears the quantile regression values which used in this study. One, major advantage of quantile regression over classical mean regression is its exibility in assessing the effect of predictors on different locations of the response distribution.

Regression at multiple quantiles provides more comprehensive statistical views than analysis is at mean or at single quantile level. When the distribution is highly skewed, the mean can be challenging to interpret while the median remains highly informative.

The study showed ER with values of quantile regression (0.1) is less than ER with values (0.90) as shown in all tables. More generally, values of quantile regression (0.1 and 0.90) can be used to describe non-central positions of a distribution.

Fourth: The two methods with small different sample sizes

The fourth aim of this section is to discuss the results of suggested two methods when small different sample sizes are applied. The simulation study applied L_1-Lasso and L_2-Lasso when three small different sample sizes are generated when three different error distributions are used with two different parameters for quantile regression models. The simulation study generated three small sample sizes (n=10, n=15 and n=20) to explain if the small size of sample effect or not about the estimators and the behavior of the two methods with different samples sizes.

The results for the fourth aim shown that the estimators values which used to calculate ER are improving with the increase in sample size when the (ε_i) generated fat-long tailed distribution. The results of L_1-Lasso better than L_2- Lasso when the three small different sample sizes are generated.

Fifth: Two methods with different distribution

The fifth aim of this section is to discuss the results of the comparison between L_1-Lasso and L_2-Lasso when different distribution with different parameters is used. The study considered with three distributions (fat-long tailed, skewed normal and chi- square distribution) for each distribution two different parameters. The simulation study applied with three different distributions and two different parameters for each to evaluate the performance for the two methods with existence of fat or long tailed distribution. If the estimate model effected by the existence of fat-long tailed distribution. The results for this aim are:

Cauchy distribution: From the results in Table (4) and (5) with three different sample sizes and two different parameters (0, 0.5), (0, 0.9) for Cauchy distribution. This study observed that the results of ER for L_1-Lasso, is less than L_2-Lasso. The variability for the estimation parameter which calculated by L_2-Lasso are greater than which calculated by L_1-Lasso when Cauchy distribution and three sample sizes.

Skewed normal distribution: Tables (6) and (7) showed the results when three different sample sizes, with two different parameter for skewed normal distribution, the results demonstrated that the ER for L_1-Lasso is less than ER L_2-Lasso. L_1-Lasso for estimation of parameters when skewed normal distribution is used much better than linear programming.

Log normal distribution: Tables (2) and (3) showed the results when three sample sizes with two different parameters (0, 0.3) and (0, 2.5) for Log normal distribution. The results demonstrated that ER L_1 -Lasso, *is* better than L_2 –Lasso for estimation of parameters when log normal distribution is used.

In this study, the two methods for estimating quantile regression parameter through Lasso linear programming are proposed. A simulation study has been made to evaluate the performance of the proposed estimators and the behavior of the methods with small sample sizes based on the estimated risk (ER) criterion. Quantile regression is an approach that allows us to examine the behavior of the response variable beyond its average of the Gaussian distribution, e.g., 10th percentile, and 90th percentile which applied in this study. Examining the different percentiles using quantile regression may be more beneficial for continuous improvement and cost savings.

Lasso quantile regression is a regularization technique for simultaneous estimation and variable selection where the classical variable selection methods are often highly time consuming and maybe suffer from instability. L_1 and L_2 penalized estimation methods shrink the estimates of the regression coefficients towards zero relative to the maximum likelihood estimates. The purpose of this shrinkage is to prevent over fit arising due to either collinearity of the covariates or high-dimensionality.

The study evaluates the performance for the two methods; Lasso linear (L_1-Lasso and L_2-Lasso). The two methods are used to select the best subset of variables and estimate the parameters of the quantile regression equation when three error distributions, with three sample sizes and two different parameters for each error distribution. Estimated risk which produced by L_1-Lasso is less than ER which produced L_2-Lasso methods.

5- Conclusion

1- The two methods are used to select the best subset of variables. Two suggested methods deleting the same variables that lead to the motivation for variable selection that the deleting variables from the model can improve the precision of parameter estimates.
2- The two methods are used to estimate the parameters of the quantile regression equation. Estimated risk is used to measure the performance of the methods. L_1-Lasso method is much better than, L_2-Lasso method.
3- Two methods are used to select the best subset of variables and to estimate the parameters with different quantile regression values. Different quantile regression may be more beneficial for continuous improvement and cost savings.
4- The performance for the L_1-Lasso when are used to select the best subset of variables, estimate the parameters with quantile regression is much better with small sample sizes.
5- The performance for the L_1-Lasso when it is used to select the best subset of variables, to estimate the parameters with two error distributions, with two different parameters for each error distribution much better with fat-long tailed distribution.

All results showed superiority of L_1 -Lasso compared with L_2-Lasso linear programming methods.

Table (2) Estimated risks of the $\boldsymbol{L_1}$-Lasso and $\boldsymbol{L_2}$-Lasso Estimates for Lognormal (0, 2.5)

			N=10		N=15		N=20	
			L_1	L_2	L_1	L_2-	L_1	L_2
$\rho = 0.1$	ER	β_0	0.072	0.072	0.072	0.072	0.072	0.072
		β_1	0.428	0.466	0.073	0.482	0.071	0.212
		β_2	0.420	0.631	0.142	0.489	0.098	0.228
		β_3	0.558	0.327	0.152	0.408	0.111	0.210
		β_4	0.582	0.450	0.157	0.746	0.127	0.198
		β_5	0.385	0.237	0.223	0.522	0.105	0.265
		β_6	1.520	0.377	0.13	0.464	0.096	0.200
		β_7	0.532	0.440	0.166	0.542	0.092	0.285
		β_8	0.524	0.332	0.144	0.427	0.114	0.268
$\rho = 0.9$	ER	β_0	0.072	0.072	0.072	0.072	0.072	0.072
		β_1	2.867	3.75	0.099	0.086	0.070	0.068
		β_2	1.491	2.02	0.220	0.067	0.076	0.068
		β_3	2.605	2.77	0.237	0.074	0.080	0.076
		β_4	3.008	3.85	0.196	0.079	0.100	0.071
		β_5	1.212	2.80	0.192	0.074	0.082	0.071
		β_6	1.960	3.90	0.185	0.084	0.087	0.068
		β_7	1.096	1.88	0.246	0.087	0.081	0.070
		β_8	1.274	1.32	0.172	0.078	0.098	0.070

Table (3) Estimated risks of the L_1-Lasso and L_2-Lasso Estimates for lognormal (0, 0.3)

			N=10		N=15		N=20	
			L_1	L_2	L_1	L_2	L_1	L_2
$\rho=0.1$	ER	β_0	0.72	0.072	0.072	0.072	0.072	0.072
		β_1	0.075	0.157	0.072	0.072	0.066	0.068
		β_2	0.118	0.93	0.065	0.071	0.065	0.067
		β_3	0.071	0.259	0.072	0.069	0.066	0.072
		β_4	0.098	0.111	0.068	0.071	0.066	0.067
		β_5	0.122	0.123	0.064	0.071	0.066	0.069
		β_6	0.075	0.143	0.067	0.080	0.066	0.068
		β_7	0.069	0.92	0.063	0.070	0.065	0.069
		β_8	0.088	0.178	0.064	0.073	0.065	0.066
$\rho=0.9$	ER	β_0	0.072	0.072	0.072	0.072	0.072	0.072
		β_1	0.109	0.160	0.080	0.073	0.070	0.070
		β_2	0.279	0.103	0.107	0.067	0.075	0.070
		β_3	0.364	0.147	0.087	0.066	0.079	0.068
		β_4	0.282	0.149	0.084	0.068	0.098	0.073
		β_5	0.245	0.115	0.096	0.074	0.082	0.071
		β_6	0.546	0.135	0.095	0.069	0.089	0.069
		β_7	0.357	0.119	0.096	0.075	0.081	0.069
		β_8	0.239	0.120	0.119	0.069	0.098	0.68

Table (4) Estimated risks of the L_1-Lasso and L_2-Lasso Estimates for Cauchy (0, 0.9)

			N=10		N=15		N=20	
			L_1	L_2	L_1	L_2	L_1	L_2
$\rho=0.9$	ER	β_0	0.072	0.072	0.072	0.072	0.072	0.072
		β_1	3.875	61064.623	0.896	17022.547	0.271	4431.811
		β_2	7.226	36554.527	0.835	15232.509	0.358	6091.063
		β_3	7.228	35935.199	0.905	13030.700	0.225	5550.617
		β_4	8.208	34853.870	1.121	14294.785	0.258	4480.318
		β_5	18.596	27180.332	1.288	13437.302	0.675	5037.966
		β_6	9.604	47656.138	1.065	17228.064	0.38	5478.853
		β_7	3.387	24040.182	1.252	11652.591	0.342	4382.749
		β_8	6.429	33259.976	0.679	13330.215	0.286	5100.825
$\rho=0.1$	ER	β_0	0.072	0.072	0.072	0.072	0.072	0.072
		β_1	27.774	30587.310	4.847	10.724	0.433	0.068
		β_2	75.468	79361.83	6.652	2.280	0.820	0.075
		β_3	271.99	28971.374	1.107	16.649	0.459	0.074
		β_4	815.509	28874.940	49.391	0.129	2.29	0.072
		β_5	21.229	32789.991	3.274	2.638	0.584	0.074
		β_6	237.121	43268.701	141.47	10.871	0.746	0.070
		β_7	131.379	35764.517	6.81	7.026	1.609	0.074
		β_8	124.347	26498.139	27.021	2.621	0.753	0.079

Table (5) Estimated risks of the L_1-Lasso and L_2-Lasso Estimates for Cauchy (0, 0.5)

			N=10		N=15		N=20	
			L_1	L_2	L_1	L_2	L_1	L_2
$\rho=0.1$	ER	β_0	0.072	0.072	0.072	0.072	0.072	0.072
		β_1	4.650	13327.26	3.792	5245.262	0.193	1383.480
		β_2	14.201	25209.96	1.440	4841.157	0.171	1851.774
		β_3	47.505	10127.5	2.122	3926.302	0.106	1756.187
		β_4	12.589	8785.645	6.284	4346.578	0.322	1389.320
		β_5	3.380	11750.23	1.141	3877.916	0.110	1593.225
		β_6	53.704	12944.63	6.708	5246.261	0.163	1749.799
		β_7	39.847	11450.78	1.085	3584.331	0.230	1456.853
		β_8	33.689	10056.92	4.834	4093.229	0.091	1483.577
$\rho=0.9$	ER	β_0	0.072	0.072	0.072	0.072	0.072	0.072
		β_1	3.307	19174.57	0.330	3.275	0.070	0.076
		β_2	0.701	11498.83	0.216	0.735	0.073	0.078
		β_3	0.830	10965.4	0.113	5.151	0.078	0.071
		β_4	5.158	10738.5	0.280	0.094	0.118	0.173
		β_5	1.719	8121.073	0.168	0.847	0.098	0.170
		β_6	4.811	14885.38	0.177	3.351	0.076	0.076
		β_7	1.623	7534.33	0.177	2.166	0.072	0.072
		β_8	5.242	10080.526	0.295	0.819	0.105	0.172

Table (6) Estimated risks of the L_1-Lasso and L_2-Lasso Estimates for Skewed normal (0, 12)

			N=10		N=15		N=20	
			L_1	L_2	L_1	L_2	L_1	L_2
$\rho=0.1$	ER	β_0	0.072	0.072	0.072	0.072	0.072	0.072
		β_1	0.653	6.64	0.758	2.564	0.063	1.245
		β_2	0.749	5.408	0.336	2.711	0.058	1.172
		β_3	0.298	8.449	0.176	2.416	0.057	1.127
		β_4	0.258	5.603	0.292	2.468	0.060	1.042
		β_5	0.634	5.668	0.129	2.785	0.060	1.158
		β_6	0.667	6.908	0.109	2.031	0.061	1.136
		β_7	0.490	5.581	0.521	2.651	0.059	1.083
		β_8	0.072	7.299	0.124	2.474	0.071	1.147
$\rho=0.9$	ER	β_0	0.072	0.072	0.072	0.072	0.072	0.072
		β_1	4.609	6.64	1.134	2.564	0.501	1.37
		β_2	4.932	5.408	1.290	2.711	0.942	1.212
		β_3	7.417	8.449	1.494	2.416	0.821	1.241
		β_4	4.276	5.603	1.614	2.468	1.090	0.97
		β_5	5.764	5.668	1.695	2.785	1.107	1.187
		β_6	5.481	6.908	1.785	2.031	0.772	1.078
		β_7	5.304	5.581	1.874	2.651	0.920	1.269
		β_8	4.629	7.299	1.867	2.474	0.751	1.488

Table (7) Estimated risks of the L_1-Lasso and L_2-Lasso Estimates for Skewed normal (0, 20)

			N=10		N=15		N=20	
			L_1	L_2	L_1	L_2	L_1	L_2
$\rho=0.1$	ER	β_0	0.072	0.072	0.072	0.072	0.072	0.072
		β_1	0.653	17.741	0.758	6.773	0.063	4.123
		β_2	0.749	20.415	0.336	7.242	0.058	3.577
		β_3	0.298	20.301	0.176	8.479	0.057	3.971
		β_4	0.258	16.370	0.292	12.020	0.060	3.220
		β_5	0.634	21.106	0.129	6.899	0.060	3.607
		β_6	0.674	17.984	0.109	7.639	0.061	3.512
		β_7	0.667	15.155	0.521	7.569	0.059	3.577
		β_8	0.490	21.573	0.124	5.751	0.071	4.445
$\rho=0.9$	ER	β_0	0.072	0.072	0.072	0.072	0.072	0.072
		β_1	4.609	17.157	1.134	6.054	0.501	3.726
		β_2	4.932	17.918	1.29	6.808	0.942	3.629
		β_3	7.417	22.314	1.494	8.057	0.821	3.299
		β_4	4.276	20.737	1.614	7.532	1.090	3.233
		β_5	5.764	14.802	1.695	7.364	1.107	3.608
		β_6	5.481	19.507	1.785	5.816	0.772	3.603
		β_7	5.304	15.642	1.874	8.103	0.92	3.401
		β_8	4.629	17.854	1.867	8.240	0.751	3.140

6-References

1- R. Koenker and G. Bassett. Regression Quantiles. Econometrica, 1978. 46, 33-50.

2- Y. L and J. Zhu, L_1-Normal Quantile Regression. Journal of Computational and Graphical Statistics, 2008.17, 163-185.

3- Y. Wu and Y. Liu, Variable Selection in Quantile Regression. Statistica Sinica, 2009. 19, 801-817.

4- A. Belloni and V. Chernozhukov, ℓ_1- Penalized Quantile Regression in High- Dimensional Sparse Models. The Annals of Statistics. 2011. Vol.39. No.1. 82-130.

5- H. Zou and M. Yuan Regularized Simultaneous Model Selection in Multiple Quantiles Regression. Computational Statistics and Data Analysis. 2008. 52. 5296-5304.

6- K. Yu and J. Stander Bayesian Analysis of a Tobit Quantile Regression Model. Journal of Econometrics. 2007.137, 260-276.

7- M. Geraci and M. Bottai Quantile regression for longitudinal data using the asymmetric Laplace distribution. Biostatistics. 2007. 8,140-154.

8- C. Chen and K. Yu Automatic Bayesian quantile regression curve fitting. Statistics and Computing 2009. 19(3), 271-281.

9- P. Thompson, Y. Cai, R. Moyeed, D. Reeve and J. Stander Bayesian nonparametric quantile regression using splines. Computational Statistics and Data Analysis. 2010. 54(4), 1138-1150.

10- A. Kottas and M. Krnjajifc Bayesian Semi parametric Modeling in Quantile Regression. Scandinavian Journal of Statistics. 2009 .36, 297-319.

11- A. Kottas. and A. E. Gelfand, Bayesian Semi parametric Median Regression Model. Journal of the American Statistical Association. 2001 96, 1458-1468.

12- L. Wang, Y. Wu and R. Li Quantile Regression for Analyzing Heterogeneity in Ultra – High Dimension. Journal of the American statistical association. 2012 Vol.107, No.497.

13- B. Kai, R. Li and H. Zou Local Composite Quantile Regression Smoothing: An Efficient and Safe Alternative to Local Polynomial Regression. Journal of Royal Statistical Society. 2010 72. 49- 69.

14- G. Ye, Y. Chen and X. Xie Efficient Variable Selection In Support Vector Machines Via The Alternating Direction Method Of Multipliers. 14th International Conference on Artificial Intelligence and Statistics (AISTATS) Vo. 15. 2011

15- W. Sun, J. Wang and Y. Fang Consistent Selection Of Tuning Parameters Via Variable Selection Stability. Journal of Machine Learning Research. 2013.14. 3419-3440.

16- R. Tibshirani, Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society, Series B, 1996 58, 267-288.

17- R. Tibshirani. Regression Shrinkage and Selection via the Lasso: a Retrospective. Journal of the Royal Statistical Society. 2011. 73, 273-282.

18- Q. Li, R. Xi and N. Lin Bayesian regularized quantile regression. Bayesian Analysis, 2010. 5, 1-24.

19- R. Alhamzawi, K. Yu and T. Benoit (2012). Bayesian adaptive lasso quantile regression. Statistical Modeling, 2012 12(3), 279-297.

20- J. Fan and J. Lv A selective overview of variable selection in high dimensional feature space. Statistica Sinica, 2010. 20, 101-148.

21- E. Ismail Curve fitting for data with outliers using fuzzy linear programming. Unpublished Ph.D. thesis, Faculty of Commerce, Al-Azhar University-Girls' Branch.2003.