Atan Regularized in Quantile Regression for High Dimensional Data

Ali Hameed Yousif, Wafaa Jaafer Housain
Wasit University, College of Administration and Economics, Department of Statistics, Iraq
ahameed@uowasit.edu.iq

ABSTRACT. In this article, we proposed using a new version of penalty functions to estimate parameters and variable selection in quantile regression. This version of the penalty depends on arctangent function called the atan penalty. Ridge and elastic-net are the penalties that are used to compare with the proposed estimator in quantile regression. A simulation study and the real data application showed that the proposed estimator is the best compared to the other estimators.

1. Introduction
In the linear regression model, the relationship between the dependent variable and the other sets of explanatory variables is expressed by the conditional expectation function. Although the applications of linear regression are wide and common, it cannot measure the relationship between the response variable and the explanatory variables under different points. An alternative to the linear regression is the quantile regression, which was proposed by [1]. Quantile regression provides the possibility to measure these relationships through different points such as median and others. The response idea of this is summarized in the modeling of the conditional distribution function of the response variable by establishing the explanatory variables, which are symbolized by the $F(y|x)$ and is calculated as below:

$$F(y|x) = P(Y_i \leq y | x_i = x)$$

(1)

The $\tau$th continental quantile of the response variable $Y_i$ given $x_i = x$ is $\tau$, which is defined by the following function:

$$Q_{Y|X}(\tau) = \inf\{\tau : F_Y(y|x) \geq \tau\}$$

(2)

where $\tau$ is $0 < \tau < 1$. 
Suppose we have random samples \{(y_1, x_1) \ldots (y_n, x_n)\}. The quantile linear regression model of \(\text{th}\) is as follows:

\[ y_i = x_i' \beta + \epsilon_i, \quad i = 1, 2, \ldots, n \]  

(3)

Where \(y_i\) is the response variable, \(x_i\) is \(p\) dimensional covariates, \(\beta = (\beta_1, \ldots, \beta_p)\) is unknown parameters with \(p\) dimensional and \(\epsilon_i\) is the random error assumed distributed normally with mean zero and variance \(\sigma^2\). In linear quantile regression assumes that \(Q_{\tau}(x) = x' \beta\). The estimation of unknown parameters in quantile regression is given by the following equation:

\[ \min_{\beta} \sum_{i=1}^{n} \rho_{\tau}(y_i - x_i' \beta) \]  

(4)

where \(\rho_{\tau}(u) = u(|u - I(u < 0)|)\) is called the check function and \(\tau \in (0,1)\). In the quantile regression model and if \(\tau = 0.5\), then the quantile regression will be fit to the conditional median. The selection of variables is one of the most important problems in statistics. Many studies dealt with the problem of variables selection in regression such as [2], which proposed a bridge penalty that has the following formula \(\sum_{i=1}^{p} |\beta_i|^\gamma \cdot \gamma \geq 0\) [3] defined the Lasso penalty, which is a special case of bridge penalty, when \((\gamma = 1)\) and when \((\gamma = 2)\). Then we get to the ridge regression, which was proposed by[4],[5] introduced a non-concave penalty function called SCAD penalty which has Oracle properties which include unbiasedness, continuity, and sparsity. [6] used the Lasso penalty function in quantile regression based on longitudinal data. [7] proposed a penalty function called elastic net Penalty and the estimator is called elastic net. The elastic net estimator is similar to the Lasso estimator that selectively chooses the variable and continuously reduces and it can choose related categorical variables.[8] showed that that the Lasso estimator is inconsistent and did not achieve the property of Oracle and proposed an alternative penalty function called Adaptive Lasso, which used different penalties for all coefficient based on weights.[9] considered the L_1 (LASSO) regularized quantile regression which uses the sum of the absolute values of the coefficients as the penalty. [10] proposed a new version of penalized based on trigonometric functions is called atan estimator which is very close to L_0. The atan estimator consistently selects the correct mode.[11] proposed LAD-atan estimator for High Dimensional linear regression model. In this paper, we proposed a new version of penalty functions for quantile regression to the estimation of parameters and variable selection based on arctangent functions. The research is organized as follows. The second section includes the presentation of the regularized linear quantile regression based on the ridge, elastic-net and the atan estimators. The third section presents the choice of regularized parameter. The fourth and fifth sections deal with simulation results and the practical application of data. Finally, section six is devoted to recommendations.

2. Regularized Quantile Regression

The quantile regression is not suitable in the cases of high-dimensional data and high correlations between the explanatory variables. To improve quantile regression a penalty function is added to the quantile loss function to get the method of the regularized quantile regression, which is as follows:

\[ Q_\tau(\beta) = \sum_{i=1}^{n} \rho_{\tau}(y_i - x_i' \beta) + \lambda \sum_{i=1}^{p} p_\delta(\|\beta_i\|) \]  

(5)
Where \( p \) is the dimension of the vector parameters \( \beta \), \( p, \lambda \) is the penalty function with a regularized parameter \( \lambda \geq 0 \). When the value of \( \lambda \) is equal to zero, then we get classical quantile regression\([9]\). There are many versions of penalty functions that have been used in estimation and variable selection. These versions will be described as follow:

### 2.1. Ridge Regularized

An alternative method to the ordinary least square (OLS) is a ridge regression which was proposed by [4]. A ridge regression method may deal with multicollinearity problems and high dimensional data. The idea of ridge regression adds small positive value to the information matrix \((\mathbf{x}^\prime \mathbf{x})\). A penalty function of ridge regression is as follow:

\[
\sum_{j=1}^{p} \beta_j^2 = 1
\]

The estimator of quantile regression based on a ridge penalty is obtained as follows:

\[
\hat{\beta}_{\text{ridge}} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} \rho_{\tau}(y_i - x_i^\prime \beta) + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}
\]

\[\text{(7)}\]

[12] proposed an algorithm for ridge regularized in quantile regression is called semi smooth Newton coordinate descent (SNCD).

### 2.2. Elastic-net Regularized

[7] proposed a new penalty called elastic net Penalty which is a combination of ridge and Lasso. The Elastic net estimators may deal with categorical variables and the formula of the elastic net penalty which is as follows:

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

Where \( 0 \leq \alpha \leq 1 \), when the value of \( \alpha = 0 \) then we get ridge regression, but if \( \alpha = 1 \), then we get Lasso regression. Regularized quantile regression based on elastic-net regularized is defined as:

\[
\hat{\beta}_{\text{Elastic-Net}} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} \rho_{\tau}(y_i - x_i^\prime \beta) + \lambda \sum_{j=1}^{p} \left[ \frac{1}{2} (1 - \alpha) \beta_j^2 + \alpha |\beta_j| \right] \right\}
\]

\[\text{(9)}\]

[12] also proposed an algorithm for elastic-net regularized in quantile regression is called semismooth Newton coordinate descent (SNCD).

### 2.3. Atan Regularized

[10] suggested using a new version of penalty functions based on arctangent functions (arctangent type) that are closed to \( l_0 \). This function is called atan penalty which is formulated defined below:
\[ P_{\lambda,Y}(\beta) = \lambda \left( y + \frac{2}{\pi} \right) \arctan \left( \frac{|\beta|}{\gamma} \right) \]  

(10)

With \( \lambda \geq 0 \) and \( y \geq 0 \). This type of function is concave to \( |\beta| \). It has the property of Oracle. The first derivative of the atan penalty is defined as follow:

\[ P_{\lambda,Y}'(\beta) = \lambda \frac{y+2/\pi}{y+\beta^2} \]  

(11)

In this research, we proposed the atan penalty function is added to the quantile function then we get on regularized quantile regression as follow:

\[ Q_{\gamma}(\beta) = \sum_{i=1}^{n} \rho_{\tau}(y_i - x_i'\beta) + \lambda \sum_{j=1}^{p} \left( y + \frac{2}{\pi} \right) \arctan \left( \frac{|\beta_j|}{\gamma} \right) \]  

(12)

We can solve the above minimization problem in (12) by an algorithm proposed by [13]. Let \( \hat{\beta}_j^{(0)} = 0 \), \( i = 1,2,\ldots,p \) be initial values and the ith step \( (t > 1) \), then we get \( \hat{\beta}_j^{(t-1)} = \left( \hat{\beta}_1^{(t-1)}, \ldots, \hat{\beta}_p^{(t-1)} \right)' \). the update \( \hat{\beta}_j^{(t)} \) is obtained by minimizing

\[ \min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} \rho_{\tau}(y_i - x_i'\beta) + \sum_{j=1}^{p} w_j^{(t-1)} |\beta_j| \right\} \]  

(13)

Where \( w_j^{(t-1)} = P_{\lambda,Y}(\hat{\beta}_j^{(t-1)}) \) denotes to the weight and \( P_{\lambda,Y}(\cdot) \) denotes the derivative of \( P_{\lambda,Y}(\cdot) \) in (11). By adding slack variables \( \xi^+ \), \( \xi^- \) and \( \zeta_j \) the convex optimization problem in (13) can be equivalently written as

\[ \min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} c \xi^+_i + (1-c) \xi^-_i + \sum_{j=1}^{p} w_j^{(t-1)} \xi_j \right\} \]  

subject to

\[ \xi^+_i - \xi^-_i = y_i - x_i'\beta ; \quad i = 1,2,\ldots,n \]
\[ \xi^+_i \geq 0, \quad \xi^-_i \geq 0 ; \quad i = 1,2,\ldots,n \]
\[ \zeta_j \geq 0 ; \quad j = 1,2,\ldots,p \]

The linear programming problem in (14) can be solved by software packages.

3. Choice of Regularized Parameter

The choice of the regularized parameter plays a key role in regularized quantile regression,[14] used the regularized parameter by minimizing Schwartz Information Criterion (SIC) for quantile regression. The regularized parameter for the proposed estimator can be obtained by minimizing SIC as follow.

\[ \text{HBLIC}(\lambda) = \log(y_1 - x_1'\hat{\beta}_\lambda) + |S_0| \frac{\log \log n}{2n} C_n \]  

(15)

Where \( \hat{\beta}_\lambda = (\hat{\beta}_{\lambda,1}, \ldots, \hat{\beta}_{\lambda,p}) \) is the regularized quantile regression estimator which is obtained from the minimizing (12). \( S_0 = \{ j : \hat{\beta}_{\lambda,j} \neq 0, \ 0 \leq j \leq p \} \) represent support of estimated in the model. \( C_\lambda \) represents a sequence of a positive constant diverging to infinity as \( n \) increases such that \( C_n = O(\log(p)) \).
4. Simulation Experiments

In this section, we used the Monte Carlo simulation to compare the performance of Regularized quantile Regression estimators. The proposed atan regularized estimator was compared with each of the following estimators: ridge and elastic-net estimators. Simulation data are generated from the standard linear regression model:

\[ y_i = x_i^T \beta + \sigma \epsilon_i \]  \hspace{1cm} (16)

Where \( \beta = (3, 1.5, 0, 2, 0, ..., 0) \). The matrix \( x \) is generated from multivariate normal distribution i.e. \( x \sim \mathcal{N}(0, \Sigma) \), where the \( \Sigma_{ij} = \rho^{|i-j|} \) is covariance matrix and \( \rho \) is the correlation coefficient assume that equal to 0.5. The distribution of error assumes that standard normal distribution with mean zero and standard deviation 1. The sample size assumes that 30, 40, 100, number of covariates \( p=8.50 \) and three different values of quantile \( \tau = 0.25, 0.5 \) and 0.75. Two different values of \( \sigma \) are given by 1 and 3. The replicate of simulation is equal to 200.

In simulation experiments, comparison between estimators are based on root mean squared prediction error (RMSPE) as followed:

\[ \text{RMSPE}(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^{n} (y_i^* - x_i^T \hat{\beta})^2 \]  \hspace{1cm} (17)

Where \( y_i^* \) is based on test data. To measure the sparsity sparsity, we used false positive rate (FPR) and false negative rate (FNR) which was proposed by [15].

| \( \tau \) | \( \sigma \) | estimators | RMSPE | FPR | FNR |
|-----|-----|-----------|--------|-----|-----|
| 0.25 | 1   | Ridge     | 2.0729 | 1   | 0   |
|      |     | Elastic-net| 1.1966 | 0.759| 0   |
|      |     | Atan      | 1.1630 | 0.449| 0   |
|      |     | Ridge     | 2.9180 | 1   | 0   |
|      | 3   | Elastic-net| 2.3226 | 0.732| 0   |
|      |     | Atan      | 2.3021 | 0.456| 0   |
|      |     | Ridge     | 2.0499 | 1   | 0   |
|      | 0.5 | Elastic-net| 1.8834 | 0.706| 0   |
|      |     | Atan      | 1.1373 | 0.406| 0   |
|      |     | Ridge     | 2.8917 | 1   | 0   |
|      | 0.75| Elastic-net| 2.3133 | 0.701| 0   |
|      |     | Atan      | 2.2757 | 0.417| 0   |
|      |     | Ridge     | 2.0779 | 1   | 0   |
|      |     | Elastic-net| 1.2021 | 0.727| 0   |
|      |     | Atan      | 1.1497 | 0.423| 0   |
|      |     | Ridge     | 2.9392 | 1   | 0   |
|      | 3.0 | Elastic-net| 2.3592 | 0.702| 0.001|
|      |     | Atan      | 2.2994 | 0.432| 0   |
### Table 2. Simulation results for an experiment when n=100, p=8

| τ  | σ  | estimators | RMSPE | FPR | FNR |
|----|----|------------|-------|-----|-----|
| 0.25 |    | Ridge      | 1.9354 | 1    | 0   |
|     | 1  | Elastic-net | 1.0709 | 0.681 | 0   |
|     |    | Atan       | 1.0551 | 0.53  | 0   |
|     | 3  | Ridge      | 2.8068 | 1    | 0   |
|     |    | Elastic-net | 2.1245 | 0.681 | 0   |
|     |    | Atan       | 2.1114 | 0.456 | 0   |
| 0.50 |    | Ridge      | 1.0476 | 0.45  | 0   |
|     | 1  | Elastic-net | 1.0632 | 0.70  | 0   |
|     |    | Atan       | 1.0476 | 0.45  | 0   |
|     | 3  | Elastic-net | 2.1122 | 0.71  | 0   |
|     |    | Atan       | 2.0956 | 0.463 | 0   |
| 0.75 |    | Ridge      | 1.0709 | 0.70  | 0   |
|     | 1  | Elastic-net | 1.0709 | 0.70  | 0   |
|     |    | Atan       | 1.0508 | 0.476 | 0   |
|     | 3  | Elastic-net | 2.1281 | 0.709 | 0   |
|     |    | Atan       | 2.1029 | 0.489 | 0   |

### Table 3. Simulation results for an experiment when n=30, p=50

| τ  | σ  | estimators | RMSPE | FPR | FNR |
|----|----|------------|-------|-----|-----|
| 0.25 |    | Ridge      | 2.0734 | 1    | 0   |
|     | 1  | Elastic-net | 1.8449 | 0.494 | 0   |
|     |    | Atan       | 1.6882 | 0.381 | 0   |
|     |    | Ridge      | 4.6854 | 1    | 0   |
|     | 3  | Elastic-net | 4.2937 | 0.376 | 0.078 |
|     |    | Atan       | 4.0579 | 0.321 | 0.075 |
|     |    | Ridge      | 2.9153 | 1    | 0   |
| 0.50 |    | Elastic-net | 1.8041 | 0.520 | 0   |
|     | 1  | Atan       | 1.7807 | 0.50  | 0   |
|     |    | Ridge      | 4.5976 | 1    | 0   |
|     | 3  | Elastic-net | 4.2979 | 0.435 | 0   |
|     |    | Atan       | 5.3351 | 0.413 | 0   |
|     |    | Ridge      | 3.0468 | 1    | 0   |
| 0.75 |    | Elastic-net | 1.8370 | 0.4858 | 0 |
|     | 1  | Atan       | 1.7416 | 0.40  | 0   |
|     |    | Ridge      | 4.6374 | 1    | 0   |
|     | 3  | Elastic-net | 4.2834 | 0.389 | 0.078 |
|     |    | Atan       | 4.1664 | 0.373 | 0.067 |
Table 1 and 2 summarize the results of simulation when \( p=8 \) and sample size 40 and 100 respectively while Table 3 summarizes the results of simulation when \( p=50 \) and sample size 30. It is obvious that the atan estimator has the best performance because it has the smallest values of RMSPE, FPR and FNR respectively.

5. Application
In this section, the performance of the proposed estimator is presented with a comparison with other estimators and the application of the resulting method to the semen data of people with delayed reproduction for 35 patients. The response variable is sperms and there are 49 predictors. These predictors are:

Random Blood Sugar (RBS), Age, Haemoglobin A1C (HBA1C), Follicle Stimulating Hormone (FSH), Brolactin, Testosterone, Triiodothyronine (T3), Tetraiodothyronine-Thyroxine (T4), Thyroid Stimulating Hormone (TSH), Weight, Luteinizing Hormone (LH), The Venereal Disease Research Laboratory (VDRL), Prostate Specific Antigen (PSA), Treponema Pallidum Hemagglutinations (TPHA), Alanine aminotransferase (ALT), Free Thyroid Stimulating Hormone (FTSH), C.Pepitide, Vitamin D3(D3), Vitamin B12(B12), Erythrocyte Sedimentation Rate (ESR), White Blood Cells (WBC), Haemoglobin (HB), Blood viscosity test (PCV), Calcium (Ca), Sodium (Na), Potassium (K), Magnesium (Mg), Chlorine, phosphate, Total Serum Bilirubin (TSB), Glutamic oxaloacetic transaminase (Got), Glutamate pyruvic transaminase (Gpt), Widal, Cholesterol, Triglyceride (TG), C-reactive protein (CRP), Adrenocorticotropin hormone (ACTH), Antisperm antibodies (AA), Low Density Lipoproteins (LDL), Very Low Density Lipoproteins (VLDL), High-Density Lipoprotein (HDL), Blood Urine Analysis (B.U.A.), Serum Creatinine (S.Creatinine), Serum Albumin (S.Albumin), Serum Cortisol (S.Cortisol), General Urine Analysis (G.U.A), Zinc (Zn), Lactate Dehydrogenase (LDH), Rose Bengal (RB), Triglycerides (TG), Cholesterol. The response variable represents the number of sperms and the number of observation is 40 and this data set appears in [16]. Ridge, elastic net and atan are used in the application part. The application results are presented in the following table.

Table 4. Represent nonzero coefficients for the model relating \( \log(\text{sperms}) \) and explanatory covariates in the sperm data set.

| Covariates | Ridge  | Elastic-net | Atan |
|------------|--------|-------------|------|
| RBS        | -0.0673| -0.0460     | 0    |
| Age        | -0.0622| -0.0360     | -0.0107|
| HBA1C      | -0.0733| -0.0255     | -0.0187|
| FSH        | -0.1128| -0.2153     | -0.0464|
| Prolactin  | -0.0608| -0.0764     | 0    |
| Testosterone| 0.1318 | 0.1730     | 0.0643|
| T3         | -0.0263| 0           | 0    |
| T4         | -0.0354| -0.0376     | 0    |
| TSH        | -0.0415| -0.0256     | -0.0374|
| Weight     | 0.0118 | 0           | 0.0126|
| LH         | -0.0165| 0           | -0.0258|
| VDRL       | 0.0150 | 0           | 0    |
| PSA        | -0.0389| -0.0384     | -0.0173|
| TPHA       | -0.0159| 0           | 0    |
From the above table, we note that a ridge does not make a variable selection, while elastic-net contains 22 nonzero coefficients, the resulting of Atan contains 16 nonzero coefficients, the atan has the ability of making shrinkage coefficients to be zero.
6. CONCLUSION
In this article, we proposed using atan penalty for Quantile regression. The performance, atan estimator for quantile regression showed that from the simulation and real data application experiments are the best from other estimators in estimation and variable selection.

7. References
[1] R. Koenker and G. J. Bassett, “Regression quantiles., Econometrica, 46 (1): 33–50,” Math. Rev. MR0474644 Zentralblatt MATH, vol. 373, 1978.
[2] L. E. Frank and J. H. Friedman, “A statistical view of some chemometrics regression tools,” Technometrics, vol. 35, no. 2, pp. 109–135, 1993.
[3] R. Tibshirani, “Regression shrinkage and selection via the lasso,” J. R. Stat. Soc. Ser. B, vol. 58, no. 1, pp. 267–288, 1996.
[4] A. E. Hoerl and R. W. Kennard, “Ridge regression: Biased estimation for nonorthogonal problems,” Technometrics, vol. 12, no. 1, pp. 55–67, 1970.
[5] J. Fan and R. Li, “Variable selection via nonconcave penalized likelihood and its oracle properties,” J. Am. Stat. Assoc., vol. 96, no. 456, pp. 1348–1360, 2001.
[6] R. Koenker, “Quantile regression for longitudinal data,” J. Multivar. Anal., vol. 91, no. 1, pp. 74–89, 2004.
[7] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” J. R. Stat. Soc. Ser. B (statistical Methodol.), vol. 67, no. 2, pp. 301–320, 2005.
[8] H. Zou, “The adaptive lasso and its oracle properties,” J. Am. Stat. Assoc., vol. 101, no. 476, pp. 1418–1429, 2006.
[9] Y. Li and J. Zhu, “L 1-norm quantile regression,” J. Comput. Graph. Stat., vol. 17, no. 1, pp. 163–185, 2008.
[10] Y. Wang and L. Zhu, “Variable selection and parameter estimation with the Atan regularization method,” J. Probab. Stat., vol. 2016, 2016.
[11] A. H. Yousif and O. A. Ali, “Proposing robust lad-atan penalty of regression model estimation for high dimensional data,” Baghdad Sci. J., vol. 17, no. 2, 2020, doi: 10.21123/bsj.2020.17.2.0550.
[12] C. Yi and J. Huang, “Semismooth newton coordinate descent algorithm for elastic-net penalized huber loss regression and quantile regression,” J. Comput. Graph. Stat., vol. 26, no. 3, pp. 547–557, 2017.
[13] L. Wang, Y. Wu, and R. Li, “Quantile regression for analyzing heterogeneity in ultra-high dimension,” J. Am. Stat. Assoc., vol. 107, no. 497, pp. 214–222, 2012.
[14] E. R. Lee, H. Noh, and B. U. Park, “Model selection via Bayesian information criterion for quantile regression models,” J. Am. Stat. Assoc., vol. 109, no. 505, pp. 216–229, 2014.
[15] A. Alfons, C. Croux, S. Gelper, and others, “Sparse least trimmed squares regression for analyzing high-dimensional large data sets,” Ann. Appl. Stat., vol. 7, no. 1, pp. 226–248, 2013.
[16] A. H. Yousif, “The use of robust penalized estimators of linear regression model in the presence of problems of dimensions and outliers with application,” University of Baghdad, 2017.