IMPROVING ESTIMATIONS IN QUANTILE REGRESSION MODEL WITH AUTOREGRESSIVE ERRORS

by

***Bahadir YUZBASI*[a*], *Yasin ASAR*[b], *M. Samil SIK*[a], *and Ahmet DEMIRALP*[a]**

[a] Department of Econometrics, Inonu University, Malatya, Turkey
[b] Department of Mathematics-Computer Sciences, Necmettin Erbakan University, Konya, Turkey

Original scientific paper
https://doi.org/10.2298/TSCI170612275Y

An important issue is that the respiratory mortality may be a result of air pollution which can be measured by the following variables: temperature, relative humidity, carbon monoxide, sulfur dioxide, nitrogen dioxide, hydrocarbons, ozone, and particulates. The usual way is to fit a model using the ordinary least squares regression, which has some assumptions, also known as Gauss-Markov assumptions, on the error term showing white noise process of the regression model. However, in many applications, especially for this example, these assumptions are not satisfied. Therefore, in this study, a quantile regression approach is used to model the respiratory mortality using the mentioned explanatory variables. Moreover, improved estimation techniques such as preliminary testing and shrinkage strategies are also obtained when the errors are autoregressive. A Monte Carlo simulation experiment, including the quantile penalty estimators such as lasso, ridge, and elastic net, is designed to evaluate the performances of the proposed techniques. Finally, the theoretical risks of the listed estimators are given.

Key words: *preliminary estimation, Stein-type estimation, autocorrelation, quantile regression*

Introduction

Regression analysis is a statistical technique that is used to model the cumulative and linear relationship between covariates and response variables. The most common method used for this purpose is the ordinary least squares (OLS) method. The linear regression model can be written:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_{i,} \quad i = 1, 2, \ldots, n \,, \tag{1}$$

where y_i is the observations of the response variables, β_j – the unknown regression coefficients, x_{ij} – the known covariates, and ε_i – the unobservable random errors. When estimating the parameters using OLS method, the expectation of the dependent variable conditional on the independent variables is obtained. In other words, the relationship between the explanatory and explained variables in the co-ordinate plane is estimated with a mean regression line.

In order to use OLS estimator, there are three assumptions on the error terms showing white noise process of the regression model: the error terms have zero mean, the variance of the

* Corresponding author, e-mail: b.yzb@hotmail.com

error terms is constant, and the covariance between the errors is zero *i. e.*, there is no autocorrelation problem. In real life applications, most of the data does not provide the previous assumptions. Moreover, OLS provides a view of the relationship between covariates and response variable such that it models the expectation of the response conditional on the covariates without taking into account the outliers. To overcome these inadequacies of the classical regression, the quantile regression was proposed in [1] as an expansion of the classical regression model to a basic minimization problem which generates sample quantiles. For a random variable Z with distribution function $\mathcal{F}_Z(z) = P(Z \leq z) = \tau$ and $0 \leq \tau \leq 1$, the τ^{th} quantile function of Z, $\mathcal{Q}_\tau(z)$, is defined:

$$\mathcal{Q}_\tau(z) = \mathcal{F}_Z^{-1}(\tau) = \inf\{z : \mathcal{F}_Z(y) \geq \tau\}$$

which is the inverse function of $\mathcal{F}_Z(\tau)$ for the τ^{th} quantile. In other words, the τ^{th} quantile in a sample corresponds to the probability τ for a z value. Also an estimation of the full model (FM) $\hat{\boldsymbol{\beta}}_\tau$, τ^{th} quantile regression coefficients can be obtained by solving the following minimization problem:

$$\underset{\boldsymbol{\beta} \in \Re^p}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right) \tag{2}$$

where $\rho_\tau(u) = u[\tau - I(u < 0)]$ is the quantile loss function, $I(\cdot)$ is the indicator function, and $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$. Hence, it yields:

$$\hat{\boldsymbol{\beta}}_\tau = \underset{\boldsymbol{\beta} \in \Re^p}{\operatorname{argmin}} \left[\sum_{i \in \{i: y_i \geq \boldsymbol{x}_i'\boldsymbol{\beta}\}}^{n} \tau \left|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right| - \sum_{i \in \{i: y_i < \boldsymbol{x}_i'\boldsymbol{\beta}\}}^{n} (1-\tau)\left|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right| \right]$$

A quantile autoregression model which could be interpreted as a special case of the general random-coefficient autoregression model with strongly dependent coefficients was proposed in [2]. The authors studied statistical properties of the proposed model and associated estimators and derived the limiting distributions of the autoregression quantile process. The *quantreg* R package and it is implementations for linear, non-linear, and non-parametric quantile regression models were considered in [3]. Different computational intelligence methodologies based on artificial neural networks used for forecasting an air quality parameter were compared by [4]. Composite quantile regression for dependent data was suggested in [5]. The authors also showed the root-*n* consistency and asymptotic normality of the composite quantile estimator. Moreover, the authors apply their proposed method to NO_2 particle data in which air pollution on a road is modeled via traffic volume and meteorological variables. A penalized quantile estimator in semiparametric linear regression model and the authors also dealt with longitudinal data in [6]. The authors obtained the oracle properties of the estimator and selection consistency.

The books by Koenker [7] and Davino [8] are excellent sources for various properties of quantile regression as well as many computer algorithms. Moreover, an algorithm, called semismooth Newton co-ordinate descent was developed by [9], to obtain a better efficiency and scalability for computing the solution paths of penalized quantile regression. They also provide an R package called *hqreg*. Moreover, this package also obtains lasso of [10], ridge of [11] and elastic net of [12] estimators in the quantile regression models. The *hqreg* functions give the solution path while the *quantreg* package of [3] computes a single solution.

On the other hand, the book [13] can be found the large literature and informations about shrinkage estimations in the context of linear and partially linear models (PLM). The preliminary and Stein-type estimations based on ridge regression were obtained by [14] for linear models and by [15, 16] for PLM. Yuzbasi and Arashi [17] proposed stein-type lasso estimator, and investigate its properties. Furthermore, the pretest and shrinkage estimation based on the quantile regression when the errors are both independent and identically distributed (i. i. d.) and non-identically independent distributed (non-i. i. d.), have been introduced, respectively, in [18, 19]. In these studies, asymptotic distributional bias, quadratic bias and risk functions are also obtained. The novelty of this study is the errors having the problem of autocorrelation which is very common in time series analysis.

Motivation example

In this section, we consider the possible effects of temperature and pollution on weekly mortality in Los Angeles (LA) Country as given in [20]. This data has weekly observations from 1970 to 1979. In tab. 1, we describe the variables of the cement data which is freely available in the *astsa* package with the function *lap* in R project. Note that we use raw data in this example.

The fig. 1 shows that the observations 152, 153, and 155 may be outliers. Applying *outlierTest* function in the *car* package in R, according to the results, we observe that the observations 152-155 and 260 are outliers. We also observe that the errors follow a heavy-tailed distribution.

Table 1. Descriptions of variables for the LA pollution-mortality data set

Variables	Descriptions
Response variable	
rmort	Respiratory mortality
Predictors	
tempr	Temperature
rh	Relative humidity
CO	Carbon monoxide
SO_2	Sulfur dioxide
NO_2	Nitrogen dioxide
HC	Hydrocarbons
O_3	Ozone
part	Particulates

According to fig. 1 and tab. 2, the residuals of this data have AR(5) process. Also, we consider the values of d_L and d_U as 1.686 and 1.852, respectively. Hence, there is a positive autocorrelation problem in this data.

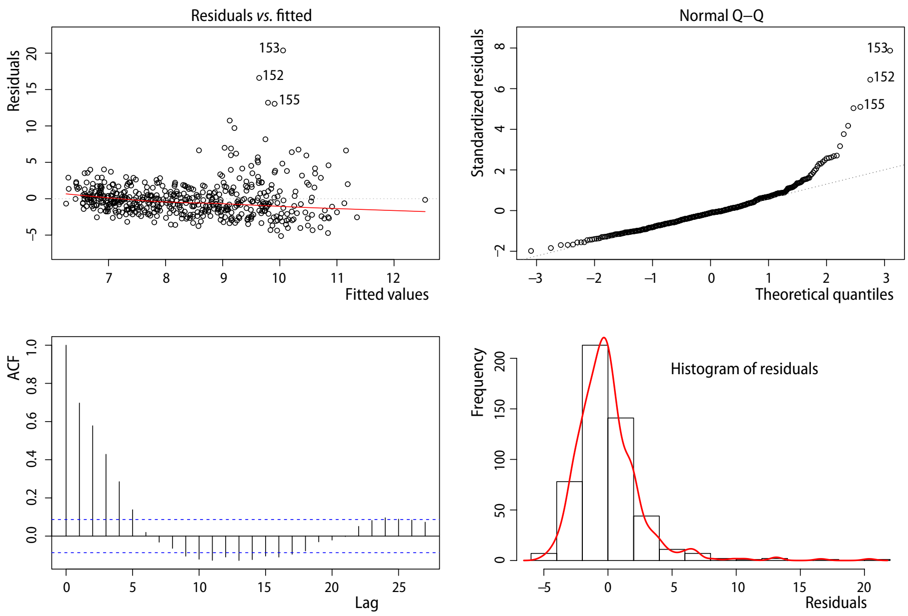

Figure 1. Residual diagnostics

Table 2. Durbin Watson test

lag	Autocorrelation	D-W statistic	p-value
1	0.697	0.604	0.000
2	0.578	0.841	0.000
3	0.428	1.140	0.000
4	0.285	1.427	0.000
5	0.138	1.719	0.006
6	0.019	1.955	0.714

The values of tab. 3 and the ratio of largest eigenvalue to smallest eigenvalue of design matrix in model eq. (1) is approximately 657.177 which show that there is a multicollinearity problem between independent variables. When we consider all results, this dataset suffers from the problems of multicollinearity, autocorrelation, heavy tailed errors and outliers simultaneously. Hence, we will use the quantile type estimation for this data.

Table 3. The VIF values

	tempr	rh	CO	SO_2	NO_2	HC	O_3	part
VIF	5.197	1.673	7.711	2.636	7.377	6.071	5.698	5.360

Statistical model

Linear regression model given in eq. (1) would be written in a partitioned form:

$$y_i = \boldsymbol{x}'_{1i}\boldsymbol{\beta}_1 + \boldsymbol{x}'_{2i}\boldsymbol{\beta}_2 + \varepsilon_i, \quad i = 1,2,\ldots,n$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2)'$ is partitioned so that the coefficient vector of $\boldsymbol{\beta}_1 = (\beta_1, \beta_2, \ldots, \beta_{p_1})'$, of order p_1, is our main interest and the coefficient vector of $\boldsymbol{\beta}_2 = (\beta_{p_1+1}, \beta_{p_2+2}, \ldots, \beta_p)'$ is the *irrelevant variables* with dimension p_2, where $p = p_1 + p_2$. Also, $\boldsymbol{x}_i = (\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i})$ and ε_i are errors with the same joint distribution function $\mathcal{F}$. The conditional quantile function of response variable y_i can be written:

$$\mathcal{Q}_\tau\left(y_i|\boldsymbol{x}_i\right) = \boldsymbol{x}'_{1i}\boldsymbol{\beta}_{1,\tau} + \boldsymbol{x}'_{2i}\boldsymbol{\beta}_{2,\tau}, \quad 0 < \tau < 1$$

In this study, the main interest is to improve the performance of the important covariates under the following the null hypothesis:

$$H_0 : \boldsymbol{\beta}_{2,\tau} = \boldsymbol{0}_{p_2} \tag{3}$$

If the eq. (3) is true, then the sub-model (SM) quantile regression estimator of $\tilde{\boldsymbol{\beta}}_\tau$ is given by $\tilde{\boldsymbol{\beta}}_\tau = (\tilde{\boldsymbol{\beta}}_{1,\tau}, \boldsymbol{0}_{p_2})$, where

$$\tilde{\boldsymbol{\beta}}_{1,\tau} = \min_{\boldsymbol{\beta}_1 \in \Re^{p_1}} \sum_{i=1}^{n} \rho_\tau\left(y_i - \boldsymbol{x}'_{1i}\boldsymbol{\beta}_1\right)$$

The distribution function $\mathcal{F}_i$ is absolutely continuous, with continuous densities $f_i(\xi)$ uniformly bounded away from 0 and ∞ at the points $\xi_i(\tau), \quad i = 1,2,\ldots$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}'_i = \boldsymbol{D}, \quad \boldsymbol{D}_0 = \frac{1}{n}\boldsymbol{X}'\boldsymbol{X}$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}f_i\left[\boldsymbol{\xi_i}\left(\tau\right)\right]\boldsymbol{x}_i\boldsymbol{x}'_i = \boldsymbol{D}_1$$

$$\max_{1\le i\le n}\frac{\|\boldsymbol{x}_i\|}{\sqrt{n}} \to 0$$

where $\boldsymbol{D}_0$ and $\boldsymbol{D}_1$ are positive definite matrices.

Pretest and Stein-type estimations

The pretest was firstly applied by [21] for the validity of the unclear preliminary information by subjecting it to a preliminary test. First let us denote the FM estimator $\hat{\boldsymbol{\beta}}_\tau$ of $\boldsymbol{\beta}_\tau$ as a solution of eq. (2) which is defined by $\hat{\boldsymbol{\beta}}_\tau = (\hat{\boldsymbol{\beta}}'_{1,\tau}, \hat{\boldsymbol{\beta}}'_{2,\tau})'$. Hence, the pretest estimator $\hat{\boldsymbol{\beta}}^{PT}_{1,\tau}$ of $\boldsymbol{\beta}_{1,\tau}$ could be obtained by following equation:

$$\hat{\boldsymbol{\beta}}^{PT}_{1,\tau} = \hat{\boldsymbol{\beta}}_{1,\tau} - (\hat{\boldsymbol{\beta}}_{1,\tau} - \tilde{\boldsymbol{\beta}}_{1,\tau}) I(\mathcal{W} < c_{n,\alpha})$$

where $c_{n,\alpha}$ is the $100(1-\alpha)$ percentage point of the $\mathcal{W}$. In order to test eq. (3), under the previous assumptions, we consider the following Wald test statistics:

$$\mathcal{W} = nw^{-2} \hat{\boldsymbol{\beta}}'_{2,\tau} \boldsymbol{\Gamma}_{22.1} \hat{\boldsymbol{\beta}}_{2,\tau}$$

where

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}_{21} & \boldsymbol{\Gamma}_{22} \end{bmatrix} = \boldsymbol{D}^{-1} \boldsymbol{A} \boldsymbol{D}^{-1}, \quad \boldsymbol{A} = \lim_{n\to\infty} \frac{1}{n} \sum_i \sum_j \psi(e_i)\psi(e_j) \boldsymbol{x}_i \boldsymbol{x}'_j$$

the median $\psi(e_i) = \text{sgn}(e_i)$ and $\boldsymbol{\Gamma}_{22.1} = \boldsymbol{\Gamma}_{22} - \boldsymbol{\Gamma}_{21}\boldsymbol{\Gamma}_{11}^{-1}\boldsymbol{\Gamma}_{12}$. Under the null hypothesis, the distribution of $\mathcal{W}$ follows the chi-square distribution with p_2 degree of freedom.

The Stein-type shrinkage ($\hat{\boldsymbol{\beta}}^{S}_{1,\tau}$) estimator of $\boldsymbol{\beta}_{1,\tau}$ is a combination of the over-fitted model estimator $\hat{\boldsymbol{\beta}}_{1,\tau}$ with the under-fitted estimator $\tilde{\boldsymbol{\beta}}_{1,\tau}$, given:

$$\hat{\boldsymbol{\beta}}^{S}_{1,\tau} = \hat{\boldsymbol{\beta}}_{1,\tau} - d(\hat{\boldsymbol{\beta}}_{1,\tau} - \tilde{\boldsymbol{\beta}}_{1,\tau}) \mathcal{W}_n^{-1}, \quad d = (p_2 - 2) \geq 3$$

In an effort to avoid the over-shrinking problem inherited by $\hat{\boldsymbol{\beta}}^{S}_{1,\tau}$, we suggest using the positive part of the shrinkage ($\hat{\boldsymbol{\beta}}^{PS}_{1,\tau}$) estimator of $\boldsymbol{\beta}_{1,\tau}$ defined:

$$\hat{\boldsymbol{\beta}}^{PS}_{1,\tau} = \hat{\boldsymbol{\beta}}^{S}_{1,\tau} - \left(\hat{\boldsymbol{\beta}}_{1,\tau} - \tilde{\boldsymbol{\beta}}_{1,\tau}\right)\left(1 - d\mathcal{W}_n^{-1}\right) I\left(\mathcal{W}_n \leq d\right)$$

Quantile penalized estimation

We briefly mention about the penalized estimators, given by [9] in quantile regression in a general form:

$$\hat{\boldsymbol{\beta}}^{\text{penalized}}_\tau = \underset{\boldsymbol{\beta}}{\text{argmin}} \sum_i \rho_\tau \left(y_i - \boldsymbol{x}'_i \boldsymbol{\beta}\right) + \lambda P(\boldsymbol{\beta})$$

where ρ is a quantile loss function, P – a penalty function, and λ – a tuning parameter. Also,

$$P(\boldsymbol{\beta}) \equiv P_\alpha(\boldsymbol{\beta}) = \alpha \|\boldsymbol{\beta}\|_1 + \frac{1-\alpha}{2} \|\boldsymbol{\beta}\|_2^2$$

which is the lasso penalty for $\alpha = 1$ [10], the ridge penalty for $\alpha = 0$ [11] and the elastic-net penalty for $0 \leq \alpha \leq 1$ [12].

Motivation example continues

In order to apply the proposed methods, we use a two step approach:

- Step 1: A set of covariates are selected based on a suitable model selection technique since the prior information is not available here.
- Step 2: The FM and SM estimates are combined in such a way that minimizes the quadratic risk.

For Step 1, one may use the model selection criterion such as AIC, BIC or best subset selection. We, however, use BIC. In tab. 4, we show the FM and candidate SM.

Table 4. The FM and candidate SM

Models	Formulas
Full model	$\text{rmort} = \beta_0 + \beta_1\text{tempr} + \beta_2\text{rh} + \beta_3\text{CO} + \beta_4\text{SO}_2 + \beta_5\text{NO}_2 + \beta_6\text{HC} + \beta_7\text{O}_3 + \beta_8\text{part}$
Sub-model	$\text{rmort} = \beta_0 + \beta_1\text{tempr} + \beta_3\text{CO}$

Figure 2 presents a summary of the OLS and the FM quantile regression results. Here, we have eight covariates and the intercept. For each of the nine coefficients, we plot the 19 distinct quantile regression estimates for τ ranging from 0.05 to 0.95 as the solid curve filled with dots. For each covariate, these point estimates may be interpreted as the impact of a one-unit change of the covariate on the response variable respiratory mortality other covariates fixed. Thus, each of the plots has a horizontal quantile, or τ , scale, and the vertical axes indicates the covariate effect. The solid line in each figure shows the OLS estimate of the conditional mean effect. The two dotted lines represent conventional 90% confidence intervals for the OLS estimate. The shaded gray area depicts a 90% point-wise confidence band for the quantile regression estimates.

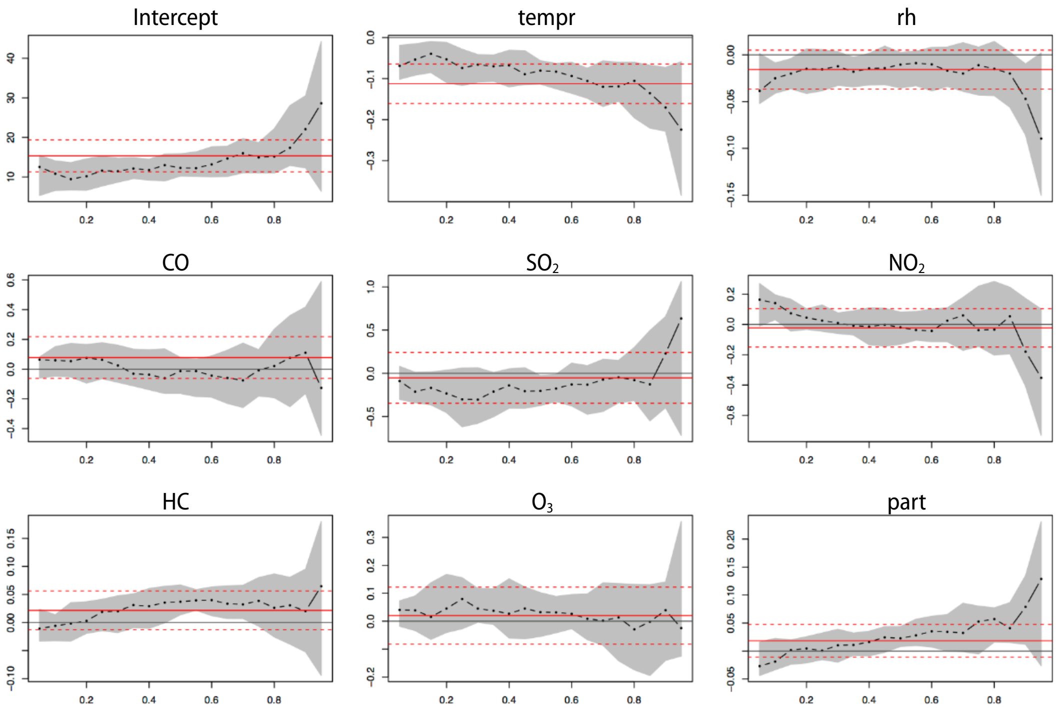

Figure 2. The OLS and FM quantile regression estimates for LA pollution-mortality data set

We will confine our discussion: The intercept estimates seem more dependent on the particular quantile. For example, up to the third quantile, quantile estimates are lower than the OLS while it is larger than the OLS for the upper quantile. With the exception of the coefficients

CO, HC, and O_3, the quantile regression estimates lie at some point outside the confidence intervals for the OLS regression, suggesting that the effects of these covariates may change across the conditional distribution of the independent variable.

In order to analyze this example, we bootstrap the data using 1000 resamplings. After that, we split the data into two which are training and test data sets. Furthermore, we center the co-variates of training and test data set based on the training data set independently. Finally, we computed the predictive mean absolute deviation (PMAD) criterion which is defined:

$$\text{PMAD}\left(\hat{\boldsymbol{\beta}}_\tau^*\right) = \frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}}\left|\boldsymbol{y}_{\text{test}} - \boldsymbol{X}_{\text{test}}\hat{\boldsymbol{\beta}}_\tau^*\right|$$

We evaluate the performance of the estimators by averaged cross validation (CV) error using a 5-fold CV. In tab. 5, we report the performances of the estimators in the sense of PMAD for the real data application. As expected, the SM estimator has the lowest PMAD value for all τ values. The PS performs better than the lasso, elastic-net, FM, and OLS, especially in the first and second quantile (median), while the ridge outperforms all others except for SM since the data has highly the problem of multicollinearity. Also, the performance of PT is also well in median.

Table 5. The PMAD values of the listed estimations

τ	FM	SM	PT	PS	Ridge	Lasso	Enet
0.25	2.612	2.209	2.612	2.515	2.337	2.531	2.535
0.50	2.391	1.881	2.359	2.249	2.220	2.372	2.372
0.75	3.082	2.041	3.063	2.802	2.275	2.557	2.469
	OLS						
Mean	2.803						

Simulation

We conduct Monte-Carlo simulation experiments to study the performances of the proposed estimators under various practical settings. In order to generate the response variable, we use:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad i = 1,\ldots,n$$

where $\boldsymbol{x}_i$ is standard normal. The correlation between the j^{th} and k^{th} components of $\boldsymbol{X}$ equals to $0.5^{|j-k|}$ and also ε_i is dependent.

We consider $\boldsymbol{\beta}' = (3, 1.5, 0, 0, 2, 0, 0, 0)$. Also, we simulate data which contains a training dataset, validation set and an independent test set. Note that the co-variates are scaled to have mean zero and unit variance. We fitted the models only using the training data and the tuning parameters were selected using the validation data.

We also use the notation $\cdot/\cdot/\cdot$ to describe the number of observations in the training, validation and test datasets, respectively. Hence, we consider that the each data set consists of 50/50/200 observations and $\boldsymbol{X} \sim N\left(0,\boldsymbol{\Sigma}\right)$, where $\sum_{ij} = 0.5^{|i-j|}$. Furthermore, the errors follow AR(1) process, that is:

$$\varepsilon_i = \rho\varepsilon_{i-1} + \omega_t$$

where $|\rho| < 1$ is called the *autocorrelation parameter* and the ω_t term is a new error term that follows the usual regression assumptions: $\omega_t \sim_{iid} \mathcal{N}(0,1)$.

Table 6 presents a summary for the different illustrative models used in the case of autoregressive errors where $\rho = \pm 5$ characterized by heavier tails while $\rho = \pm 2$ corresponds to the median. First, we note that the OLS fails against to quantile-type estimations. As expected, the SM has the lowest PMAD value since the data is generated from an empirical distribution

Table 6. Simulated PMAD values of estimators, and the values in parenthesis present the standard errors of each estimation

τ		$\rho = -0.2$	$\rho = 0.2$	$\rho = -0.5$	$\rho = 0.5$
0.25	FM	0.190(0.004)	0.188(0.004)	0.206(0.005)	0.220(0.004)
	SM	0.060(0.002)	0.057(0.002)	0.069(0.002)	0.063(0.002)
	PT	0.078(0.005)	0.076(0.006)	0.090(0.007)	0.083(0.006)
	PS	0.120(0.004)	0.134(0.005)	0.143(0.005)	0.149(0.005)
	Ridge	0.135(0.003)	0.139(0.004)	0.154(0.003)	0.158(0.003)
	Lasso	0.081(0.002)	0.076(0.003)	0.087(0.003)	0.089(0.002)
	Enet	0.078(0.002)	0.074(0.003)	0.085(0.003)	0.087(0.002)
0.50	FM	0.173(0.004)	0.165(0.003)	0.199(0.005)	0.191(0.004)
	SM	0.055(0.002)	0.053(0.002)	0.058(0.002)	0.057(0.002)
	PT	0.059(0.005)	0.061(0.004)	0.072(0.006)	0.069(0.006)
	PS	0.103(0.004)	0.098(0.004)	0.126(0.005)	0.122(0.005)
	Ridge	0.133(0.003)	0.124(0.003)	0.150(0.003)	0.149(0.003)
	Lasso	0.073(0.002)	0.073(0.002)	0.078(0.003)	0.077(0.003)
	Enet	0.072(0.002)	0.072(0.002)	0.078(0.003)	0.075(0.003)
0.75	FM	0.183(0.004)	0.184(0.004)	0.217(0.005)	0.210(0.005)
	SM	0.060(0.002)	0.059(0.002)	0.062(0.002)	0.066(0.002)
	PT	0.082(0.005)	0.072(0.005)	0.080(0.006)	0.090(0.006)
	PS	0.121(0.004)	0.115(0.004)	0.146(0.005)	0.144(0.005)
	Ridge	0.140(0.003)	0.139(0.003)	0.161(0.004)	0.159(0.004)
	Lasso	0.080(0.002)	0.074(0.002)	0.087(0.003)	0.089(0.003)
	Enet	0.078(0.002)	0.073(0.002)	0.085(0.003)	0.083(0.003)
Mean	OLS	0.137(0.009)	0.136(0.009)	0.154(0.010)	0.154(0.010)

where the candidate SM is nearly true. Furthermore, the pretest and positive shrinkage estimators are superior to the FM estimator. On the other hand, the results indicate that the PT mostly performs better than penalty estimators while positive shrinkage does not have a good performance due to the small value of p_1.

Theoretical results

In this section, we demonstrate the asymptotic risk properties of suggested estimators. So, we consider the following theorem.

Theorem 1. The distribution of quantile regression model with AR(1) process is given:

$$\sqrt{n}\,(\hat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_\tau) \xrightarrow{D} \mathcal{N}(0, \omega^2 \boldsymbol{\Gamma})$$

where $\xrightarrow{D}$ denotes convergence in distribution.

Proof. The proof can be obtained from [8].

Let a sequence of local alternatives $\{K_n\}$ given:

$$K_n : \boldsymbol{\beta}_{2,\tau} = \frac{\boldsymbol{\gamma}}{\sqrt{n}}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_{p_2})' \in \Re^{p_2}$ is a fixed vector. If $\boldsymbol{\gamma} = \mathbf{0}_{p_2}$, then the null hypothesis is true. Furthermore, we consider the following proposition to establish the asymptotic properties of the estimators.

Proposition. Let $\vartheta_1 = n^{1/2}(\hat{\boldsymbol{\beta}}_{1,\tau} - \boldsymbol{\beta}_{1,\tau})$, $\vartheta_2 = n^{1/2}(\tilde{\boldsymbol{\beta}}_{1,\tau} - \boldsymbol{\beta}_{1,\tau})$ and $\vartheta_3 = n^{1/2}(\hat{\boldsymbol{\beta}}_{1,\tau} - \tilde{\boldsymbol{\beta}}_{1,\tau})$. Under the regularity assumptions of eqs. (1)-(3), *Theorem 1* and the local alternatives $\{K_n\}$, as $n \to \infty$ we have the following joint distributions:

$$\begin{pmatrix} \vartheta_1 \\ \vartheta_3 \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0_{p_1} \\ -\boldsymbol{\delta} \end{pmatrix}, \begin{pmatrix} \omega^2 \boldsymbol{\Gamma}_{11.2}^{-1} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Phi} \end{pmatrix}\right]$$

$$\begin{pmatrix} \vartheta_3 \\ \vartheta_2 \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} -\boldsymbol{\delta} \\ \boldsymbol{\delta} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Phi} & \boldsymbol{\Sigma}^* \\ \boldsymbol{\Sigma}^* & \omega^2 \boldsymbol{\Gamma}_{11}^{-1} \end{pmatrix}\right]$$

where $\boldsymbol{\delta} = \boldsymbol{\Gamma}_{11}^{-1}\boldsymbol{\Gamma}_{12}\boldsymbol{\gamma}$, $\boldsymbol{\Gamma}_{11.2} = \boldsymbol{\Gamma}_{11} - \boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{22}^{-1}\boldsymbol{\Gamma}_{21}$, $\boldsymbol{\Phi} = \omega^2 \boldsymbol{\Gamma}_{11}^{-1}\boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{22.1}^{-1}\boldsymbol{\Gamma}_{21}\boldsymbol{\Gamma}_{11}^{-1}$, $\boldsymbol{\Sigma}_{12} = -\omega^2 \boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{21}\boldsymbol{\Gamma}_{11}^{-1}$ and $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_{21} + \omega^2 \boldsymbol{\Gamma}_{11.2}^{-1}$.

Now, we are ready to obtain the asymptotic distributional risks of estimators which are given the following subsection.

The performance of risk

The asymptotic distributional risk of an estimator $\hat{\boldsymbol{\beta}}_{1,\tau}^*$ is defined:

$$\mathcal{R}\left(\hat{\boldsymbol{\beta}}_{1,\tau}^*\right) = \mathrm{tr}\left(\mathbf{W}\boldsymbol{\Gamma}\right)$$

where $\mathbf{W}$ is a positive definite matrix of weights with dimensions of $p_1 \times p_1$, and $\boldsymbol{\Gamma}$ is the asymptotic covariance matrix of an estimator $\hat{\boldsymbol{\beta}}_{1,\tau}^*$ is defined:

$$\boldsymbol{\Gamma}\left(\hat{\boldsymbol{\beta}}_{1,\tau}^*\right) = \mathbb{E}\left\{\lim_{n\to\infty} n\left(\hat{\boldsymbol{\beta}}_{1,\tau}^* - \boldsymbol{\beta}_{1,\tau}\right)\left(\hat{\boldsymbol{\beta}}_{1,\tau}^* - \boldsymbol{\beta}_{1,\tau}\right)'\right\}$$

Theorem 2. Under the assumed regularity conditions given in eqs. (1) and (3), the *Proposition*, the *Theorem 1* and the local alternatives $\{K_n\}$, the expressions of asymptotic risks of listed estimators are:

$$\mathcal{R}\left(\hat{\boldsymbol{\beta}}_{1,\tau}\right) = \omega^2 \mathrm{tr}\left(\mathbf{W}\boldsymbol{\Gamma}_{11.2}^{-1}\right)$$

$$\mathcal{R}\left(\tilde{\boldsymbol{\beta}}_{1,\tau}\right) = \omega^2 \mathrm{tr}\left(\mathbf{W}\boldsymbol{\Gamma}_{11}^{-1}\right) + \mathrm{tr}\left(\mathbf{W}\boldsymbol{M}\right) \quad \text{where} \quad \left(\boldsymbol{M} = \boldsymbol{\Gamma}_{11}^{-1}\boldsymbol{\Gamma}_{12}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{\Gamma}_{21}\boldsymbol{\Gamma}_{11}^{-1}\right) = \boldsymbol{\delta}\boldsymbol{\delta}'$$

$$\begin{aligned}\mathcal{R}\left(\hat{\boldsymbol{\beta}}_{1,\tau}^{\mathrm{PT}}\right) = {} & \mathcal{R}\left(\hat{\boldsymbol{\beta}}_{1,\tau}\right) + \omega^2 \mathrm{tr}\left(\mathbf{W}\boldsymbol{\Gamma}_{11}^{-1}\boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_{22.1}^{-1}\boldsymbol{\Gamma}_{21}\boldsymbol{\Gamma}_{11}^{-1}\right) + \mathrm{tr}\left(\boldsymbol{\delta}\mathbf{W}\boldsymbol{\delta}'\right)\mathbb{H}_{d+4}\left(\chi_{d+4}^2;\Delta\right) + \\ & + \mathbf{W}\boldsymbol{\Phi}\mathbb{H}_{d+4}\left[\chi_{d+2,\alpha}^2\left(\Delta\right)\right] + \mathrm{tr}\left(\boldsymbol{\delta}\mathbf{W}\boldsymbol{\delta}'\right)\mathbb{H}_{d+6}\left[\chi_{d+2,\alpha}^2\left(\Delta\right)\right]\end{aligned}$$

$$\begin{aligned}\mathcal{R}\left(\hat{\boldsymbol{\beta}}_{1,\tau}^{S}\right) = {} & \mathcal{R}\left(\hat{\boldsymbol{\beta}}_{1,\tau}\right) - 2d\mathbb{E}\left[\chi_{d+4}^{-2}\left(\Delta\right)\right]\mathrm{tr}\left(\mathbf{W}\boldsymbol{\Sigma}_{21}\right) - \\ & - 2d\mathbb{E}\left[\chi_{d+6}^{-2}\left(\Delta\right)\right]\mathrm{tr}\left(\mathbf{W}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}\right) + 2d\mathbb{E}\left[\chi_{d+4}^{-2}\left(\Delta\right)\right]\mathrm{tr}\left(\mathbf{W}\boldsymbol{\delta}\boldsymbol{\delta}'\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}\right) + \\ & + d^2\mathbb{E}\left[\chi_{d+4}^{-4}\left(\Delta\right)\right]\mathrm{tr}\left(\mathbf{W}\boldsymbol{\Sigma}^{*-1}\right) + d^2\mathbb{E}\left[\chi_{d+6}^{-2}\left(\Delta\right)\right]\mathrm{tr}\left(\mathbf{W}\boldsymbol{\delta}\boldsymbol{\delta}'\right)\end{aligned}$$

$$\begin{aligned}\mathcal{R}\left(\hat{\boldsymbol{\beta}}_{1,\tau}^{\mathrm{PS}}\right) = {} & \mathcal{R}\left(\hat{\boldsymbol{\beta}}_{1,\tau}^{S}\right) - 2\mathbb{E}\left[1 - d\chi_{d+4}^{-2}\left(\Delta\right)\right]I\left[\chi_{d+4}^2\left(\Delta\right) < d\right]\mathrm{tr}\left(\mathbf{W}\boldsymbol{\Sigma}_{21}\right) - \\ & - 2\mathbb{E}\left[1 - d\chi_{d+6}^{-2}\left(\Delta\right)\right]I\left[\chi_{d+6}^2\left(\Delta\right) < d\right]\mathrm{tr}\left(\mathbf{W}\boldsymbol{\delta}'\boldsymbol{\delta}\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Sigma}_{21}\right) + \\ & + 2\mathbb{E}\left[1 - d\chi_{d+4}^{-2}\left(\Delta\right)\right]I\left[\chi_{d+6}^2\left(\Delta\right) < d\right]\mathrm{tr}\left(\mathbf{W}\boldsymbol{\delta}\boldsymbol{\delta}'\right) +\end{aligned}$$

$$+\mathbb{E}\left[1-d\chi_{d+4}^{-2}\left(\Delta\right)\right]^2 I\left[\chi_{d+4}^{2}\left(\Delta\right)<d\right]\mathrm{tr}\left(\mathbf{W}\boldsymbol{\Sigma}^*\right)+$$
$$+\mathbb{E}\left[1-d\chi_{d+6}^{-2}\left(\Delta\right)\right]^2 I\left[\chi_{d+6}^{2}\left(\Delta\right)<d\right]\mathrm{tr}\left(\mathbf{W}\boldsymbol{\delta\delta}'\right)$$

Noting that if $\boldsymbol{\Gamma}_{21}=0$, then all the risks reduce to common value $\omega^2\mathrm{tr}(\mathbf{W}\boldsymbol{\Gamma}_{11}^{-1})$ for all $\mathbf{W}$. For $\boldsymbol{\Gamma}_{12}\neq 0$, the risk of $\hat{\boldsymbol{\beta}}_{1,\tau}$ remains constant while the risk of $\tilde{\boldsymbol{\beta}}_{1,\tau}$ is an bounded function of Δ since $\Delta\in[0,\infty]$. The risk of $\hat{\boldsymbol{\beta}}_{1,\tau}^{\mathrm{PT}}$ increases as Δ moves away from zero, achieves its maximum and then decreases towards the risk of the FM estimator. Thus, it is a bounded function of Δ. The risk of $\hat{\boldsymbol{\beta}}_{1,\tau}$ is smaller than the risk of $\hat{\boldsymbol{\beta}}_{1,\tau}^{\mathrm{PT}}$ for some small values of Δ and opposite conclusions hold for rest of the parameter space. It can be seen that:

$$\mathcal{R}\left(\hat{\boldsymbol{\beta}}_{1,\tau}^{\mathrm{PS}}\right)\leq\mathcal{R}\left(\hat{\boldsymbol{\beta}}_{1,\tau}^{S}\right)\leq\mathcal{R}\left(\hat{\boldsymbol{\beta}}_{1,\tau}\right)$$

strict inequality holds for small values of Δ. Thus, positive shrinkage is superior to the shrinkage estimator. However, both shrinkage estimators outperform the full model estimator in the entire parameter space induced by Δ. On the other hand, the pretest estimator performs better than the shrinkage estimators when Δ takes small values and outside this interval the opposite conclusion holds.

Conclusion

In this paper, we obtained pretest and stein-type shrinkage estimations based on quantile regression when the distribution of errors have the problem of autocorrelation. Also, we investigated the performance of the listed estimators in a real world example using the data analyzed by [20] such that the effects of air pollution and temperature on weekly mortality in LA are considered. The results showed that the quantile type estimators outperform the OLS. Not surprisingly, the SM estimator has the lowest PMAD since the candidate SM is assumed as true. Furthermore, the PT and PS perform better than the FM. Also, the performance of the proposed estimators are mostly superior to penalty estimators, moreover ridge has a better performance since the data has the multicollinearity problem. On the other hand, we conducted a Monte Carlo simulation study in order to investigate the performance of the suggested estimators. The results of simulation study coincide with the results of real data example. Finally, we demonstrated the asymptotic distributional risk performance of the listed estimators. Our asymptotic theory is well supported by numerical analysis.

Acknowledgment

We would like to thank two anonymous reviewers for constructive comments which significantly improved the presentation of paper and let to put many details. This work is supported by Research Fund of the Inonu University with project number: SBA-2017-946.

References

[1] Koenker, R., Bassett, Jr., G., Regression Quantiles, *Econometrica: Journal of the Econometric Society*, (1978), Jan., pp. 33-50
[2] Koenker, R., Xiao, Z., Quantile Autoregression, *Journal of the American Statistical Association*, *101* (2006), 475, pp. 980-990
[3] Koenker, R. Quantreg: Quantile Regression. R Package Version 5.05, R Foundation for Statistical Computing: Vienna, Available at: http://CRAN. R-project. org/package= quantreg, 2013
[4] Ćirić, I. T., *et al.*, Air Quality Estimation by Computational Intelligence Methodologies, *Thermal Science*, *16* (2012), Suppl. 2, pp. S493-S504

[5] Tang, Y., *et al*., Variable Selection via Composite Quantile Regression with Dependent Errors, *Statistica Neerlandica*, *69* (2015), 1, pp. 1-20
[6] Wang, K., Lin, L., Variable Selection in Semiparametric Quantile Modeling for Longitudinal Data, *Communications in Statistics-Theory and Methods*, *44* (2015), 11, pp. 2243-2266
[7] Koenker, R., *Quantile Regression,* No: 38, Cambridge University Press, New York, USA, 2005
[8] Davino, C., *et al*., *Quantile Regression: Theory and Applications*, John Wiley & Sons, Ltd, New York, USA, 2014
[9] Yi, C., Huang, J., Semismooth Newton Coordinate Descent Algorithm for Elastic-Net Penalized Huber Loss Regression and Quantile Regression, *Journal of Computational and Graphical Statistics*, *26* (2016), 3, pp. 547-557
[10] Tibshirani, R., Regression Shrinkage and Selection Via the Lasso, *Journal of the Royal Statistical Society. Series B (Methodological)*, *58* (1996) 1, pp. 267-288
[11] Hoerl, A. E., Kennard, R. W., Ridge Regression: Biased Estimation for Non-Orthogonal Problems, *Technometrics, 12* (1970), 1, pp. 69-82
[12] Zou, H. Hastie, T., Regularization and Variable Selection via the Elastic Net, *Journal of the Royal Statistical Society:Series B (Statistical Methodology)*, *67* (2005), 2, pp. 301-320
[13] Ahmed, S. E., *Penalty, Shrinkage and Pretest Strategies: Variable Selection and Estimation*, Springer, New York, USA, 2014
[14] Yuzbası, B., *et al.,* Improved Penalty Strategies in Linear Regression Models, *REVSTAT-Statistical Journal, 15* (2017), 2, pp. 251-276
[15] Yuzbası, B., Ejaz Ahmed, S., Shrinkage and Penalized Estimation in Semi-Parametric Models with Multicollinear Data, *Journal of Statistical Computation and Simulation, 86* (2016), 17, pp. 3543-3561
[16] Yuzbası, B., *et al*., Ridge-Type Pretest and Shrinkage Estimations in Partially Linear Models, *Statistical Papers*, On-line first, https://doi.org/10.1007/s00362-017-0967-8, 2017
[17] Yuzbası, B., Arashi, M. Double Shrunken Sselection Operator, *Communications in Statistics-Simulation and Computation*, On-line first, https://doi.org/10.1080/03610918.2017.1395040, 2017
[18] Yuzbası, B., *et al.*, Pretest and Stein-Type Estimations in Quantile Regression Model, On-line first, *arXiv preprint arXiv:1707.03820*, 2017
[19] Yuzbası, B., *et al*., Improved Quantile Regression Estimators when the Errors are Independently and Non-Identically Distributed, On-line first, *arXiv preprint arXiv:1709.02244*, 2017
[20] Shumway, R. H., *et al*., Modeling Mortality Fluctuations in Los Angeles as Functions of Pollution and Weather Effects, *Environmental Research*, *45* (1988), 2, pp. 224-241
[21] Bancroft, T. A. On Biases in Estimation Due to the Use of Preliminary Tests of Significance, *The Annals of Mathematical Statistics*, *15* (1944), 2, pp. 190-204

Paper submitted: June 12, 2017
Paper revised: November 15, 2017
Paper accepted: November 24, 2017