Journal of the Korean Data & Information Science Society
2014, **25**(6), 1539–1547

http://dx.doi.org/10.7465/jkdi.2014.25.6.1539
한국데이터정보과학회지

Support vector quantile regression for autoregressive data[†]

Hyungtae Hwang[1]

[1]Department of Applied Statistics, Dankook University

Received 15 September 2014, revised 6 October 2014, accepted 16 October 2014

Abstract

In this paper we apply the autoregressive process to the nonlinear quantile regression in order to infer nonlinear quantile regression models for the autocorrelated data. We propose a kernel method for the autoregressive data which estimates the nonlinear quantile regression function by kernel machines. Artificial and real examples are provided to indicate the usefulness of the proposed method for the estimation of quantile regression function in the presence of autocorrelation between data.

Keywords: Autoregressive process, cross validation function, hyper-parameters, kernel function, support vector quantile regression.

1. Introduction

Since Koenker and Bassett (1978) introduced linear quantile regression, quantile regression has been a popular method for estimating the quantiles of a conditional distribution given input variables. Just as classical linear regression methods based on minimizing sum of squared residuals enable us to estimate a wide variety of models for conditional mean functions, quantile regression methods offer a mechanism for estimating models for the full range of conditional quantile functions, including the conditional median function. By supplementing the estimation of conditional mean functions with techniques for estimating an entire family of conditional quantile functions, quantile regression is capable of providing a better statistical analysis of the stochastic relationships among random variables. An introduction to, and look at current research areas of quantile regression can be found in Koenker and Hallock (2001), Yu *et al.* (2003), Koenker (2005) and Hwang (2010).

Most nonparametric regression methods focus on estimating the regression function for various data types (Kim *et al.*, 2008; Shim and Seok, 2008). The estimation of regression function from a data set is usually performed under the assumption that the error terms are iid (Juditsky *et al.*, 1995). This assumption is not satisfied when the correlation is present in the given data (e.g. time series data), which leads to severe problems on the estimation of a model under the iid assumption.

In this paper, we consider the autoregressive model, where $y_t - \mu(\boldsymbol{x}_t)$ follows AR(p) process and $\boldsymbol{x}_t$ is the input vector including a constant 1. We propose the support vector quantile

[†] The present research was conducted by the research fund of Dankook University in 2014.
[1] Professor, Department of Applied Statistics, Dankook University, Gyeonggi-do 448-701, Korea. E-mail: hthwang@dankook.ac.kr

regression to take the autocorrelation into account and estimate the quantile regression function under the AR model. The rest of this paper is organized as follows. In Section 2 the support vector quantile regression for non-autoregressive data is introduced. In Section 3 the support vector quantile regression (SVQR) for autoregressive data is introduced and the estimation method for AR(1) coefficient is presented. In Section 4 we perform the numerical studies through artificial and real examples. In Section 5 we give the conclusions.

2. Support vector quantile regression

Let the training data set denoted by $(\boldsymbol{x}_i, y_i)_{i=1}^n$, with each input $\boldsymbol{x}_i \in R^d$ including a constant 1 and the response $y_i \in R$, where the output variable y_i is related to the input vector $\boldsymbol{x}_i$. Here the feature mapping function $\phi(\cdot) : R^d \to R^{d_f}$ maps the input space to the higher dimensional feature space where the dimension d_f is defined in an implicit way. An inner product in feature space has an equivalent kernel in input space, $\phi(\boldsymbol{x}_i)'\phi(\boldsymbol{x}_j) = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ (Mercer, 1909). Several choices of the kernel $K(\cdot, \cdot)$ are possible. We consider the nonlinear regression case, in which the quantile regression function $q(\boldsymbol{x})$ of the response given $\boldsymbol{x}$ can be regarded as a nonlinear function of input vector $\boldsymbol{x}$.

With a check function $\rho_\theta(\cdot)$, the estimator of the θ-th quantile regression function can be defined as any solution to the optimization problem,

$$\min \ell(q_\theta|\boldsymbol{x}) = \sum_{i=1}^n \rho_\theta(y_i - q(\boldsymbol{x}_i)) \tag{2.1}$$

where $\rho_\theta(r) = \theta r I_{(r\geq 0)} + (1-\theta) r I_{(r<0)}$. We can express the regression problem by formulation for SVM as follows:

$$\min L = \frac{1}{2}||\boldsymbol{\omega}||^2 + C\sum_{i=1}^n (\theta\xi_i + (1-\theta)\xi_i^*) \tag{2.2}$$
$$\text{subject to } y_i - \boldsymbol{\omega}'\phi(\boldsymbol{x}_i) \leq \xi_i,\ \boldsymbol{\omega}'\phi(\boldsymbol{x}_i) - y_i \leq \xi_i^*,\ \xi_i,\ \xi_i^* \geq 0,$$

where C is a penalty parameter penalizing the training errors. We construct a Lagrange function as follows:

$$\begin{aligned} L =& \frac{1}{2}||\boldsymbol{\omega}||^2 + C\sum_{i=1}^n (\theta\xi_i + (1-\theta)\xi_i^*) - \sum_{i=1}^n \alpha_i(\xi_i - y_i + \boldsymbol{\omega}'\phi(\boldsymbol{x}_i)) \\ & - \sum_{i=1}^n \alpha_i^*(\xi_i^* + y_i - \boldsymbol{\omega}'\phi(\boldsymbol{x}_i)) - \sum_{i=1}^n (\eta_i\xi_i + \eta_i^*\xi_i^*). \end{aligned} \tag{2.3}$$

We notice that the positivity constraints $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$ should be satisfied. After taking partial derivatives of equation (2.3) with regard to the primal variables $(\boldsymbol{\omega}, \xi_i, \xi_i^*)$ and plugging them into equation (2.3), we have the optimization problem below.

$$\max -\frac{1}{2}\sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sum_{i=1}^n (\alpha_i - \alpha_i^*)y_i - e\sum_i (\alpha_i + \alpha_i^*) \tag{2.4}$$

with constraints $0 \leq \alpha_i \leq \theta C, 0 \leq \alpha_i^* \leq (1-\theta)C$.

Solving the above equation with the constraints determines the optimal Lagrange multipliers, α_i, α_i^*, the estimator of the θ-th quantile regression function given the input vector $\boldsymbol{x}_t$ are obtained as follows:

$$\widehat{q}_\theta(\boldsymbol{x}_t) = \sum_{i=1}^{n} K(\boldsymbol{x}_t, \boldsymbol{x}_i)(\hat{\alpha}_i - \hat{\alpha}_i^*). \tag{2.5}$$

The functional structures of SVQR is characterized by the hyper-parameters, C and the kernel parameters. To select the optimal hyper-parameters of SVQR we consider the leave-one-out CV function as follows:

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n} \rho_\theta(y_i - \widehat{q}_\theta(\boldsymbol{x}_i)^{(-i)}), \tag{2.6}$$

where λ is the set of hyper-parameters and $\widehat{q}_\theta(\boldsymbol{x}_i)^{(-i)}$ is the quantile regression function estimated without ith observation. Since for each candidates of parameters, $\widehat{q}_\theta(\boldsymbol{x}_i)^{(-i)}$ for $i = 1, \cdots, n$, should be evaluated, selecting parameters using CV function is computationally formidable. Yuan (2006) proposed the generalized approximate cross validation (GACV) function to select the set of hyper-parameters λ for SVQR as follows:

$$GACV(\lambda) = \frac{\sum_{i=1}^{n} \rho_\theta(y_i - \hat{q}_\theta(\boldsymbol{x}_i))}{n - trace(H)}, \tag{2.7}$$

where H is the hat matrix such that $\widehat{q}(\theta|\boldsymbol{x}) = H\boldsymbol{y}$ with the (i, j)th element $h_{ij} = \frac{\partial \widehat{q}_\theta(\boldsymbol{x}_i)}{\partial y_j}$. From Li *et al.* (2007) we have that the trace of the hat matrix H equals to the size of set $I_s = \{i = 1, \cdots, n | \ 0 < \widehat{\alpha}_i < C\theta, 0 < \widehat{\alpha}_i^* < C(1-\theta)\}$.

3. SVQR for autoregressive data

Let the given data set be denoted by $\{\boldsymbol{x}_t, y_t\}_{t=1}^n$, with $\boldsymbol{x}_t \in \boldsymbol{R}^d$ including a constant 1 and $y_t \in \boldsymbol{R}$ and let the θ th quantile regression function of y_t given $\boldsymbol{x}_t$ by $q_\theta(y_t|\boldsymbol{x}_t)$. We consider the autoregressive model,

$$\Phi^p(B)(y_t - q_\theta(y_t)) = e_t, \ t = 1, 2, \cdots, n, \tag{3.1}$$

where $q_\theta(y_t) = q_\theta(y_t|\boldsymbol{x}_t)$ and $\Phi^p(B)$ is a polynomial in back-shift operator B with parameters $\lambda_i, i = 1, \cdots, p$, such that $\Phi^p(B)r_t = r_t - \lambda_1 r_{t-1} - \lambda_2 r_{t-2} - \cdots - \lambda_p r_{t-p}$. For clear explanation of the proposed model, we assume that y_t 's are known to follow AR(1) process such that $\Phi^0(B)(y_1 - q_\theta(y_1)) = e_1$, $\Phi(B)(y_t - q_\theta(y_t)) = e_t$, $t = 2, \cdots, n$. And we assume that e_t for $t > 1$ follow independently asymmetric Laplace distribution $AL_\theta(0, \sigma)$ so that the probability distribution function of y_t is given as follows:

$$p(y_1) \propto \exp\left(-\frac{1}{\sigma}\rho_\theta(y_1 - q_\theta(y_1))\right),$$

$$p(y_t|y_{t-1}) \propto \exp\left(-\frac{1}{\sigma}\rho_\theta(y_t - q_\theta(y_t|y_{t-1}))\right) \text{for } t > 1,$$

where $q_\theta(y_2|y_1) = q_\theta(y_2)+\lambda(y_1 - q_\theta(y_1))$ and $q_\theta(y_t|y_{t-1}) = q_\theta(y_t) + \lambda(y_{t-1}- q_\theta(y_{t-1}|y_{t-2}))$ for $t > 2$.

The probability distribution function of $(y_1, \cdots, y_n)$ is given as follows:

$$f(y_1, \cdots, y_n) \propto \quad \exp\left(-\frac{1}{\sigma}\rho_\theta(y_1 - q_\theta(y_1)) - \frac{1}{\sigma}\sum_{t=2}^{n}\rho_\theta(y_t - q_t(\theta) - \lambda(y_{t-1} - q_\theta(y_t))\right).$$

The negative log likelihood of the given data can be expressed as (constant terms are omitted),

$$L = \rho_\theta(y_1 - q_\theta(y_1)) + \sum_{t=2}^{n}\rho_\theta(y_t - q_\theta(y_t) - \lambda(y_{t-1} - q_\theta(y_{t-1}))). \tag{3.2}$$

For AR(1) the estimates of $\boldsymbol{\omega}$ is obtained by minimizing the penalized log likelihood,

$$L(\boldsymbol{\omega}) = \frac{1}{2}||\boldsymbol{\omega}||^2 + C\rho_\theta(y_1 - \boldsymbol{\omega}'\phi(\boldsymbol{x}_1)) + C\sum_{t=2}^{n}\rho_\theta(y_t - \boldsymbol{\omega}'\phi(\boldsymbol{x}_t) - \lambda(y_{t-1} - \boldsymbol{\omega}'\phi(\boldsymbol{x}_{t-1}))), \tag{3.3}$$

where C is a penalty parameter penalizing the training errors.

The likelihood (3.3) can be written as follows:

$$L(\boldsymbol{\omega}) = \sum_{t=1}^{n}\rho_\theta(y_t^* - \boldsymbol{\omega}'\phi^*(\boldsymbol{x}_t)), \tag{3.4}$$

where $\boldsymbol{y}^* = \begin{pmatrix} y_1 \\ y_2 - \lambda y_1 \\ \vdots \\ y_n - \lambda y_{n-1} \end{pmatrix}$ is $n \times 1$ vector, $\phi^*(\boldsymbol{x}) = \begin{pmatrix} \phi(\boldsymbol{x}_1) \\ \phi(\boldsymbol{x}_2) - \lambda\phi(\boldsymbol{x}_1) \\ \vdots \\ \phi(\boldsymbol{x}_n) - \lambda\phi(\boldsymbol{x}_{n-1}) \end{pmatrix}$ is $n \times d_f$ matrix and $\phi^*(\boldsymbol{x}_t)$ is a transpose of tth row of $\phi^*(\boldsymbol{x})$.

We can express the minimization of the likelihood (2.4) in the formulation of SVQR as follows:

$$\min \frac{1}{2}||\boldsymbol{\omega}||^2 + C\sum_{t=1}^{n}(\theta\xi_i + (1-\theta)\xi_i^*) \tag{3.5}$$

$$\text{subject to } y_t^* - \boldsymbol{\omega}'\phi^*(\boldsymbol{x}_t) \le \xi_i,\ t = 1, \cdots, n,\ \boldsymbol{\omega}'\phi^*(\boldsymbol{x}_t) - y_t \le \xi_t^*,\ t = 1, \cdots, n,\ \xi_i, \xi_i^* \ge 0,$$

where C is a penalty parameter which controls the tradeoff between the goodness-of-fit on the data and $||\boldsymbol{\omega}||^2$.

We construct a Lagrange function as follows:

$$\begin{aligned} L =& \frac{1}{2}||\boldsymbol{\omega}||^2 + C\sum_{t=1}^{n}(\theta\xi_i + (1-\theta)\xi_i^*) - \sum_{t=1}^{n}\alpha_i(\xi_i - y_t^* + \boldsymbol{\omega}'\phi^*(\boldsymbol{x}_t)) \\ & - \sum_{t=1}^{n}\alpha_t^*(\xi_t^* + y_t - \boldsymbol{\omega}'\phi^*(\boldsymbol{x}_t)) - \sum_{t=1}^{n}(\eta_t\xi_t + \eta_t^*\xi_t^*). \end{aligned} \tag{3.6}$$

We notice that the positivity constraints $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$ should be satisfied. After taking partial derivatives of equation (3.6) with regard to the primal variables $(\boldsymbol{\omega}, \xi_i, \xi_i^*)$ and plugging them into equation (3.6), we have the optimization problem below,

$$\max -\frac{1}{2}\sum_{t,s=1}^{n}(\alpha_t-\alpha_t^*)(\alpha_s-\alpha_s^*)K^*(\boldsymbol{x}_t,\boldsymbol{x}_s)+\sum_{t=1}^{n}(\alpha_t-\alpha_t^*)y_t^* \tag{3.7}$$
$$\text{with constraints } 0 \leq \alpha_i \leq \theta C, 0 \leq \alpha_i^* \leq (1-\theta)C,$$

where

$$K^*(\boldsymbol{x}_i,\boldsymbol{x}_j)=\begin{cases} K_{11} & \text{if } i{=}j{=}1 \\ K_{1j}-\lambda K_{1,j-1} & \text{if } i{=}1,\ j>1 \\ K_{i1}-\lambda K_{i-1,1} & \text{if } i>1,\ j{=}1 \\ K_{ij}+\lambda^2 K_{i-1,j-1}-\lambda K_{i,j-1}-\lambda K_{i-1,j} & \text{if } i>1,\ j>1, \end{cases}$$

where $K_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

Solving the above equation with the constraints determines the optimal Lagrange multipliers, α_i, α_i^*. The estimated quantile regression function of y_t given $y_t - 1$ and $\boldsymbol{x}_t$ is obtained as follows:

$$\hat{q}_\theta(y_t|y_{t-1},\boldsymbol{x}_t)=K^*(\boldsymbol{x}_t,\boldsymbol{x})(\widehat{\boldsymbol{\alpha}}-\widehat{\boldsymbol{\alpha}}^*). \tag{3.8}$$

Using that

$$\widehat{q}_\theta(y_t|\boldsymbol{x}_t)=\phi(\boldsymbol{x}_t)'\widehat{\boldsymbol{\omega}}=\phi(\boldsymbol{x}_t)'\phi^*(\boldsymbol{x})(\widehat{\boldsymbol{\alpha}}-\widehat{\boldsymbol{\alpha}}^*)$$

and

$$\phi(\boldsymbol{x}_t)'\phi^*(\boldsymbol{x})=\big(K_{t1},K_{t2}-\lambda K_{t1},\cdots,K_{tn}-\lambda K_{t,n-1}\big)$$

the estimator of the θ-th quantile regression function of y_t given $\boldsymbol{x}_t$ is obtained as follows:

$$\widehat{q}_\theta(y_t|\boldsymbol{x}_t)=\tilde{K}(\boldsymbol{x}_t,\boldsymbol{x})(\widehat{\boldsymbol{\alpha}}-\widehat{\boldsymbol{\alpha}}^*), \tag{3.9}$$

where $\tilde{K}(\boldsymbol{x}_t, \boldsymbol{x}_j) = K(\boldsymbol{x}_t, \boldsymbol{x}_1)$ if $j = 1$ and $\tilde{K}(\boldsymbol{x}_t, \boldsymbol{x}_j) = K(\boldsymbol{x}_t, \boldsymbol{x}_j) - \lambda K(\boldsymbol{x}_t, \boldsymbol{x}_{j-1})$ if $j > 1$.

The functional structures of autoregressive SVQR is characterized by the autoregressive parameter λ and hyper-parameters, C and the kernel parameters. Given λ we consider the leave-one-out CV function to select the optimal hyper-parameters as follows:

$$CV(v)=\frac{1}{n}\sum_{t=1}^{n}\rho_\theta(y_t-\widehat{q}_\theta(\boldsymbol{x}_t)^{(-t)}), \tag{3.10}$$

where v is the set of hyper-parameters and $\widehat{q}_\theta(\boldsymbol{x}_t)^{(-t)}$ is the quantile regression function estimated without ith observation. We use k-fold CV function as follows:

$$kCV(v)=\sum_{j=1}^{k}\frac{1}{n^j}\sum_{i\in F_k}\rho_\theta(y_i-\widehat{q}_\theta(\boldsymbol{x}_i)^{(-j)}), \tag{3.11}$$

where $\widehat{q}_\theta(\boldsymbol{x}_t)^{(-j)}$ is the quantile regression function estimated without jth subset, F_j is the j th subset such that $\bigcup_{j=1}^k F_j = \{1, 2, \cdots, n\}$ and $F_i \cap F_j = \{\}$ for $i \neq j$, n_j is the size of F_j, and $n = \sum_{j=1}^k n_j$.

Under the assumption that the optimal hyper-parameters and corresponding $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\alpha}}^*)$ are given, the estimate of λ is obtained by minimizing the log likelihood (3.3) with respect to λ, which is,

$$\min\ L(\lambda) = \sum_{t=2}^{n} \rho_\theta(\tilde{y}_t - \lambda y_{t-1} + \lambda K_t^{(1)}(\widehat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}^*) - \lambda^2 K_t^{(2)}(\widehat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}^*)), \qquad (3.12)$$

where $\widetilde{y}_t = y_t - K_t(\widehat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}^*)$, K_t is the tth row of $K(\boldsymbol{x}, \boldsymbol{x})$, $K_t^{(1)} = (K_{t-1,1}, K_{t-1,2} + K_{t1}, \cdots, K_{t-1,n} + K_{t,n-1})$ and $K_t^{(2)} = (0, K_{t-1,1}, \cdots, K_{t-1,n-1})$.

Thus the optimal hyper-parameters and the estimate of λ are obtained iteratively as follows:

(i) Set the initial value of λ.
(ii) Obtain the optimal hyper-parameters from kCV function (3.11).
(iii) Obtain the estimate of λ from (3.12).
(iv) Reiterate (ii)-(iii) until convergence.

4. Numerical studies

We compare the performance of the proposed SVQR for autoregressive data (SVQR_AR) with SVQR through a simulated data set and a real data set. The Gaussian kernel function such that $K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \exp(-||\boldsymbol{x}_1 - \boldsymbol{x}_2||^2/\gamma)$, $\gamma > 0$, is utilized for the quantile regression function estimation in numerical studies. The optimal values of (C, γ) of SVQR are obtained by GACV function (2.8), the optimal values of (C, γ) and estimated λ of SVQR_AR are obtained by the iterative procedure using 10-fold CV function (3.11) and (3.12).

Example 4.1: We consider the autoregressive model:

$$y_1 = q_\theta(x_1) + e_1, y_2 - q_\theta(x_2) = 0.2(y_1 - q_\theta(x_1)) + e_2, t = 2, \cdots, 100$$

where $x_t = t/100$, $q_\theta(x_t) = 1 + \sin(2\pi x_t)$, e_t follows an asymmetric Laplacian distribution $AL_\theta(0, 0.1)$. For each quantile level $\theta = 0.1, 0.5, 0.9$ we generate 100 data sets of size 100. Figure 4.1 shows true quantile regression functions (solid lines) and estimated quantile regression functions (left, dotted lines) by SVQR which assumes iid errors, and estimated quantile regression functions (right, dotted lines) by SVQR_AR, imposed on the scatter plots of 100 data points of y_t 's in a data set. In Figure 4.1 (left) we can see that the proposed method seems to represent the behavior of quantile regression function of given data better than SVQR. We repeated the above procedure 100 times (we generated 100 data sets) to have the root mean squared error (RMSE)s for the true mean functions as follows,

$$RMSE_\theta = \sqrt{\frac{1}{100}\sum_{t=1}^{100}(q_\theta(y_t|x_t) - \widehat{q}_\theta(y_t|x_t))^2}.$$

The averages of 100 RMSEs from SVQR and SVQR_AR are shown in Table 4.1. From Table 4.1 we can see that SVQR_AR shows smaller RMSE's than SVQR, which implies that SVQR_AR provides better estimation performance.

Table 4.1 Average of RMSEs of predicting the quantile model using SVQR (left) and SVQR_AR (right) (standard error is in parenthesis)

θ	SVQR	SVQR_AR
0.1	0.2119(0.0056)	0.1225(0.0045)
0.5	0.0642(0.0021)	0.0619(0.0021)
0.9	0.2289(0.0071)	0.1321(0.0066)

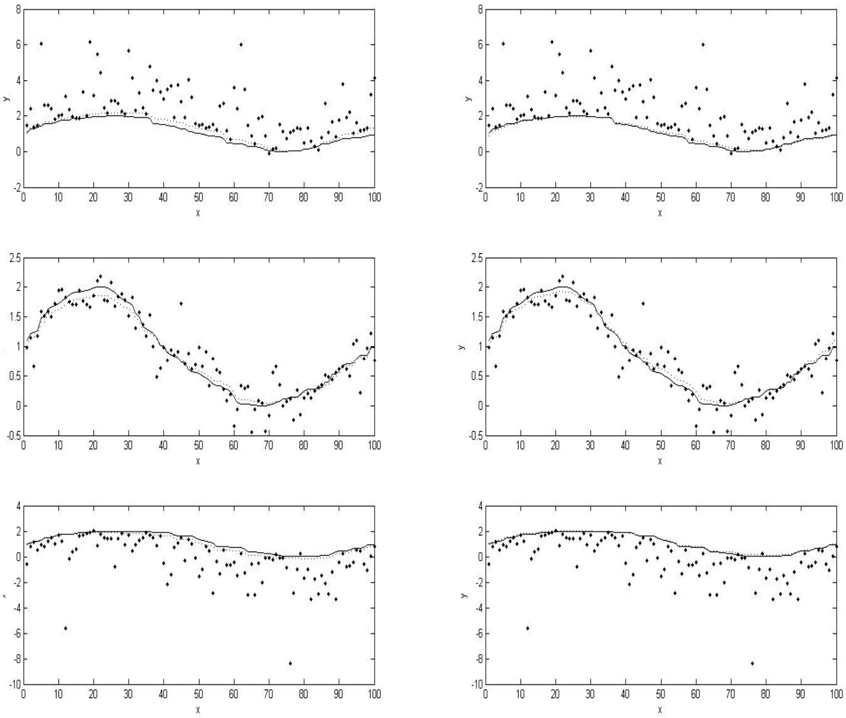

Figure 4.1 True quantile regression functions (solid lines) and estimated quantile regression functions (dotted lines) (left: SVQR, right: SVQR_AR), superimposed on the scatter plots of 100 data points (upper: $\theta = 0.1$, middle: $\theta = 0.5$, lower: $\theta = 0.9$)

Example 4.2: For the second example, we use the car sales data set of one brand name which is available at www.autotimes.co.kr. The data set consists of the number of Sonatas sold in each month from January 2006 to April 2009. We assumed AR(1) model for this data set:

$$y_1 = q_{0.5}(\boldsymbol{x}_1) + e_1, y_t - q_{0.5}(\boldsymbol{x}_t) = \lambda(y_{t-1} - q_{0.5}(\boldsymbol{x}_{t-1})) + e_t, t = 2, \cdots, 40$$

where e_t is assumed to follow an asymmetric Laplacian distribution $AL_\theta(0, \sigma)$, y_t is the logarithm of car sales, x_{t1} and x_{t2} are logarithms of prices at tth month of Sonata and

Grandeur, respectively, and x_{t3} is logarithm of KOSPI200 (Korea composite stock price index 200) at tth month which is also assumed to affect the car sales. In SVQR the optimal values of (C, γ) are obtained as (100,0.5). In SVQR_AR the optimal values of (C, γ) and $\widehat{\lambda}$ are obtained as (100,2) and -0.2, respectively, which implies that the previous car sales provides a small negative effect on the present car sales in SVQR_AR. Figure 4.2 shows the estimated median regression functions by SVQR (left, solid lines) and SVQR_AR (right, solid lines) superimposed on the scatter plots of 40 data points in a data set. From the figure we can see that SVQR_AR fits data points more smoothly. We obtained the root mean squared error (RMSE) such as

$$RMSE_{0.5} = \sqrt{\frac{1}{40}\sum_{t=1}^{40}(y_t - \widehat{q}_{0.5}(\boldsymbol{x}_t))^2}.$$

$RMSE_{0.5}$ are obtained as 0.2141 and 0.2103 for SVQR and SVQR_AR, respectively. The smaller value of $RMSE_{0.5}$ indicates that SVQR_AR works better than SVQR on the median regression function estimation in this example.

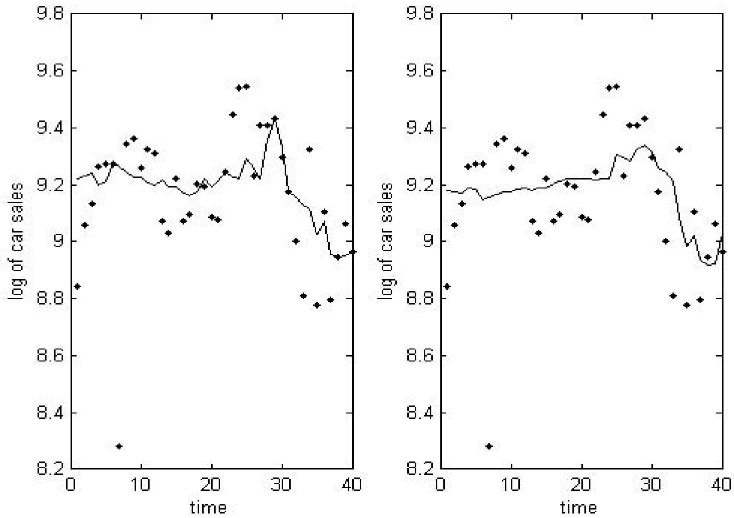

Figure 4.2 Estimated median regression function by SVQR (left) and SVQR_AR (right lines) superimposed on the scatter plots of 40 data points

5. Conclusions

In this paper, we dealt with estimating the quantile regression functions for autoregressive data. Through the examples we showed that the proposed method yields the satisfying results, better estimation performance and estimation ability of autoregressive parameter. We also found that the proposed method has an advantage of easy extension to the heteroscedastic autoregressive model by incorporating a doubly penalizing method.

References

Hwang, C. (2010). M-quantile regression using kernel machine technique. *Journal of the Korean Data & Information Science Society*, **21**, 973-981.

Juditsky, A., Hjalmarsson, H., Benveniste, A., Deylon, B., Ljung, L., Sjöberg, J. and Zhang, Q. (1995). Nonlinear black-box modelling in system identification : Mathematical foundations. *Automatica*, **31**, 1725-1750.

Kim, M. S., Park, H., Hwang, C. and Shim, J. (2013). GACV for claims reserving via kernel machine. *Journal of the Korean Data & Information Science Society*, **19**, 1419-1427.

Koenker, R. (2005). *Quantile regression,* Cambridge University Press, Cambridge.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, **40**, 122-142.

Koenker, R. and Bassett. G. (1978). Regression quantile. *Econometrica*, **46**, 33-50.

Li, Y., Liu, Y. and Zhu, J. (2007). Quantile regression in reproducing kernel Hilbert spaces. *Journal of the American Statistical Association*, **102**, 255-268.

Mercer, J. (1909). Functions of positive and negative type and their connection with theory of integral equations. *Philosophical Transactions of Royal Society A*, 415-446.

Shim, J and Seok, K. H. (2013). GACV for partially linear SVR. *Journal of the Korean Data & Information Science Society*, **24**, 391-399.

Yu, K., Lu, Z. and Stander, J. (2003). Quantile regression: Applications and current research area. *The Statistician*, **52**, 331-350.

Yuan, M. (2006). GACV for quantile smoothing splines. *Computational Statistics & Data Analysis*, **50**, 813-829.