Journal of the Korean Data & Information Science Society
2010, **21**(6), 1319–1325

한국데이터정보과학회지

Restricted support vector quantile regression without crossing

Jooyong Shim[1] · Jangtaek Lee[2]

[1]Department of Applied Statistics, Catholic University of Daegu
[2]Department of Statistics, Dankook University

Received 18 September 2010, revised 12 November 2010, accepted 17 November 2010

Abstract

Quantile regression provides a more complete statistical analysis of the stochastic relationships among random variables. Sometimes quantile functions estimated at different orders can cross each other. We propose a new non-crossing quantile regression method applying support vector median regression to restricted regression quantile, restricted support vector quantile regression. The proposed method provides a satisfying solution to estimating non-crossing quantile functions when multiple quantiles for high dimensional data are needed. We also present the model selection method that employs cross validation techniques for choosing the parameters which affect the performance of the proposed method. One real example and a simulated example are provided to show the usefulness of the proposed method.

Keywords: Cross validation technique, location-scale model, quantile regression, restricted regression quantile, support vector quantile regression.

1. Introduction

Quantile regression introduced by Koenker and Bassett (1978) provides a more informative description of relationships among variables. It has been a popular method for estimating the quantiles of a conditional distribution on the values of covariates. 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 conditional median function, and the full range of other conditional quantile functions. Cole (1990) introduced a parametric method that has been termed the LMS method, which is based on three functions of the covariates: the Box-Cox power transformation (L), the mean or median function (M), and the coefficient of variation (S).

Koenker and Bassett (1978) introduced a nonparametric approach which is based on M-estimation similarly to least absolute deviation methods. This approach yields consistent

[1] Adjuct professor, Department of Applied Statistics, Catholic University of Daegu, Gyungbuk 702-701, Korea.
[2] Corresponding author: Department of Statistics, Dankook University, 126, Jukjeon-dong, Suji-gu, Yongin-si, Gyeonggi-do 448-701, Korea. E-mail: jtlee@dankook.ac.kr

estimates of the quantile regression under general conditions, without requiring that the form of the distribution of output variable be specified. However, a major drawback is that a separate specification and estimation are required for each quantile order of interest. Without special restriction, quantile functions estimated at different orders can cross each other, although true quantile functions are defined to be non-crossing. In an effort to modify the Koenker and Bassett algorithm to ensure that quantile regressions would not cross, He (1997) proposed the restricted regression quantile (RRQ), which is based on a location-scale model. It can be employed for a broad class of models including linear heteroscedastic models and nonlinear quantile regression models. Heagerty and Pepe (1999) proposed a semiparametric method where they model the location and scale as flexible regression spline functions and allow the distribution of error to vary as a function of covariates. Their method combines the strengths of both the parametric LMS method and the nonparametric methods of Koenker and Bassett (1978). In RRQ, He (1997) transformed the non-crossing constraint into positivity constraint. When conducting such transformation, we should impose some restrictions on the conditional moment structure of the problem, which is not desirable from nonparametric modeling view point. Takeuchi (2004) and Takeuchi *et al*. (2006) described non-crossing quantile regression method via support vector machine (SVM, Vapnik, 1995) which utilizes the non-crossing constraint as a simple linear constraint. This SVM approach shows good performance in prediction. For details of applications of SVM, refer to Hwang (2007), Hwang (2008), Shim and Seok (2008), Shim *et al*. (2009). However, non-crossing quantile regression method via SVM has disadvantage that every adjacent pair of conditional quantile functions should be computed when multiple quantiles are needed. Shim *et al*. (2009) proposed non-crossing quantile regression method using doubly penalized kernel machine which employs heteroscedastic location-scale model as basic model and then estimates both location and scale functions simultaneously by kernel machines.

In this paper, we propose a new non-crossing quantile regression method applying support vector median regression to RRQ, which is based on a location-scale model and uses a multi-step strategy. We also develop the model selection method that uses 10-fold cross validation technique and the generalized approximate cross validation (GACV) function for choosing the parameters which affect the performance of the proposed method. The rest of this paper is organized as follows. In Section 2 we propose the proposed method for non-crossing quantile regression. In Section 3 we present the model selection method using GACV function. In Section 4 we perform the numerical studies through two examples. In Section 5 we give the conclusions.

2. Restricted support vector quantile regression

2.1. Support vector median regression

Let the training data set D be denoted by $(\boldsymbol{x}_i, y_i)_{i=1}^n$, with each input vector $\boldsymbol{x}_i \in R^d$ and the output $y_i \in R$ which is linearly or nonlinearly 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 case, in which the median function given $\boldsymbol{x}$, $m(\boldsymbol{x})$, can be regarded as a nonlinear function of input vector $\boldsymbol{x}$. With an absolute loss function, the median function can be defined as a function of any solution to the optimization problem,

$$\min \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{i=1}^{n}|y_i - m(\boldsymbol{x}_i)|. \tag{2.1}$$

We can express the median regression problem by formulation for SVM as follows.

$$\min \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{2.2}$$

subject to

$$y_i - \boldsymbol{w}'\phi(\boldsymbol{x}_i) - b \le \xi_i,\ \boldsymbol{w}'\phi(\boldsymbol{x}_i) + b - y_i \le \xi_i^*,\ \xi_i, \xi_i^* \ge 0,$$

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

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

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

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

subject to $\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$, $\alpha_i\alpha_i^* = 0$, $0 \le \alpha_i \le C$ and $0 \le \alpha_i^* \le C$.

Solving the above problem with the constraints determines the optimal Lagrange multipliers $\widehat{\alpha}_i$ and $\widehat{\alpha}_i^*$. Thus, the estimated median function given the input vector $\boldsymbol{x}_o$ is obtained as

$$\widehat{m}(\boldsymbol{x}_o) = K(\boldsymbol{x}_o, \boldsymbol{x})(\widehat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}^*) + \widehat{b}, \tag{2.5}$$

where $\widehat{b}$ is obtained via Kuhn-Tucker conditions (Kuhn and Tucker, 1951) such as,

$$\widehat{b} = \frac{1}{n_s}\sum_{i \in I_s}(y_i - K_i(\widehat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}^*)), \tag{2.6}$$

where n_s is the size of the set $I_s = \{i = 1, \cdots, n | -C < \widehat{\alpha}_i - \widehat{\alpha}_i^* < C\}$.

2.2. Restricted support vector quantile regression

Here we consider a nonlinear heteroscedastic model

$$y_i = m(\boldsymbol{x}_i) + s(\boldsymbol{x}_i)\epsilon_i \tag{2.7}$$

where $s(\boldsymbol{x}_i)$ is assumed to be positive, ϵ_i is assumed to have a median 0 and $|\epsilon_i|$ is assumed to have a median 1.

For non-crossing quantile regression, we employ the restricted regression quantile (He, 1997) with support vector median regression as follows:

1) Apply support vector median regression on $(\boldsymbol{x}_i, y_i)_{i=1}^n$ to obtain the median function $\widehat{m}(\boldsymbol{x}_i)$ of y given $\boldsymbol{x}_i$ and residuals $\hat{r}_i = y_i - \widehat{m}(\boldsymbol{x}_i)$.

2) Apply the support vector median regression on $(\boldsymbol{x}_i, |\hat{r}_i|)_{i=1}^n$ to obtain the estimated median function of $|\hat{r}_i|$ given $\boldsymbol{x}_i$, $\hat{s}_i$, which is the estimate of $s(\boldsymbol{x}_i)$ since the median of $|\epsilon_i|$ is assumed 1.

3) Find the θ th quantile of $\hat{r}_i$, $\widehat{\beta}\hat{s}_i$, by minimizing $\sum_{i=1}^n h_\theta(\hat{r}_i - \beta\hat{s}_i)$, where h_θ is a check function such that $h_\theta(e) = \theta I_{(e\geq 0)} + (1-\theta)I_{(e<0)}$. Since $r_i = y_i - m(\boldsymbol{x}_i) = s(\boldsymbol{x}_i)\epsilon_i$, quantiles of r_i depend on $\boldsymbol{x}_i$ through $s(\boldsymbol{x}_i)$ whose estimate is $\hat{s}_i$.

Then the θ th quantile function of y given $\boldsymbol{x}_i$ is obtained as

$$\widehat{q}_\theta(\boldsymbol{x}_i) = \widehat{m}(\boldsymbol{x}_i) + \widehat{\beta}\hat{s}_i \tag{2.8}$$

Here quantiles of $\hat{r}_i$ are non-crossed since the linear quantile regression is performed, which leads quantile functions of y given $\boldsymbol{x}_i$ non-crossed.

3. Model selection

In the step 1 and 2 of estimation procedure of non-crossing quantile regression, we apply support vector median regression. The functional structures of support vector median regression is characterized by C and the kernel parameters. To select the hyper-parameters of support vector median regression we consider the cross validation (CV) function as follows:

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^n |y_i - \widehat{m}(\boldsymbol{x}_i)^{(-i)}|, \tag{3.1}$$

where λ is the set of parameters and $\widehat{m}(\boldsymbol{x}_i)^{(-i)}$ is the median function estimated without i th observation. Since for each candidates of parameters, $\widehat{m}(\boldsymbol{x}_i)^{(-i)}$ 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 as follows,

$$GACV(\boldsymbol{\lambda}) = \sum_{i=1}^n \frac{h_\theta(y_i - \hat{q}_\theta(\boldsymbol{x}_i))}{n - d_f}, \tag{3.2}$$

where $\widehat{q}_\theta(\boldsymbol{x}_i)$ is the estimated θ th quantile function given $\boldsymbol{x}_i$ and d_f is a measure of the effective dimensionality of the fitted model. Yuan (2006) used $d_f = \sum_{i=1}^n \partial\hat{q}_\theta(\boldsymbol{x}_i)/\partial y_i$ with a differentiable modified check function. Li *et al.* (2007) showed that d_f is equal to the number of interpolated y_i's, which is, d_f is the size of the set $\{i = 1, \cdots, n | \widehat{q}_\theta(\boldsymbol{x}_i) = y_i\}$. In support vector median regression, $\widehat{m}(\boldsymbol{x}_i) = y_i$ for $i \in I_s$, which leads to use the size of set I_s in the equation (2.6) as d_f in GACV function (3.2) for the model selection. Thus we have GACV function as follows;

$$GACV(\boldsymbol{\lambda}) = \sum_{i=1}^{n} \frac{|y_i - \widehat{m}(\boldsymbol{x}_i)|}{n - n_s}. \tag{3.3}$$

4. Numerical studies

In this section, we illustrate the performance of the proposed method for non-crossing quantile regression through a well-known motorcycle data set (Table 1 on pp. 302 of Härdle, 1989) and simulated data sets. We use the Gaussian kernel for the examples, which is,

$$K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \exp\left(-\frac{||\boldsymbol{x}_1 - \boldsymbol{x}_2||^2}{\sigma^2}\right).$$

Example 1. In this example we consider the motorcycle data, which have been widely used to demonstrate the performance of nonparametric quantile regression methods. The data were collected performing crash tests with dummies sitting on motorcycles. The head acceleration (y) of the dummies (in g) was recorded a certain time (measured in milliseconds (x)) after they had hit a wall. The estimated quantile functions for θ=0.1, 0.25, 0.5, 0.75, 0.9 are superimposed on the scatter plot in Figure 4.1. The values of (C_1, C_2, σ_1^2, σ_2^2) are chosen as (200, 10, 0.5, 0.5) by the proposed model selection method, where C_1 and σ_1^2 are the regularization parameter and kernel parameter, respectively, for the step 1 and C_2 and σ_2^2 are for the step 2.

As seen from Figure 4.1, as x increases the variance of y increases when $x < 33$ and decreases when $x > 33$. The conditional quantile functions estimated by the proposed method do reasonably well even in the region beyond 50 milliseconds where the data is so sparse that all the quantile functions want to coalesce. As a whole, the proposed method seems to give a good estimation of non-crossing quantile functions.

The conditional quantile functions estimated by the proposed method do reasonably well even in the region beyond 50 milliseconds where the data is so sparse that all the quantile functions want to coalesce. As a whole, the proposed method seems to give a good estimation of non-crossing quantile functions.

Example 2. For this example we generate 100 data sets of size 150 in a similar manner to Cawley *et al.* (2004). The univariate input observations x are drawn from a uniform distribution on the interval $(0, \pi)$, the corresponding responses y are drawn from a univariate normal distribution with mean and variance that vary smoothly with x

$$x \sim U(0, \pi), \quad y \sim N(\mu(x), V(x))$$

where $\mu(x) = \sin(3x/2)\sin(5x/2)$ and $V(x) = 1/100 + (1 - \sin(5x/2))^2/4$.

The values of $(C_1, C_2, \sigma_1^2, \sigma_2^2)$ are chosen by the proposed model selection method for every single data set whenever a new data set is generated. As seen from Figure 4.2, five estimated

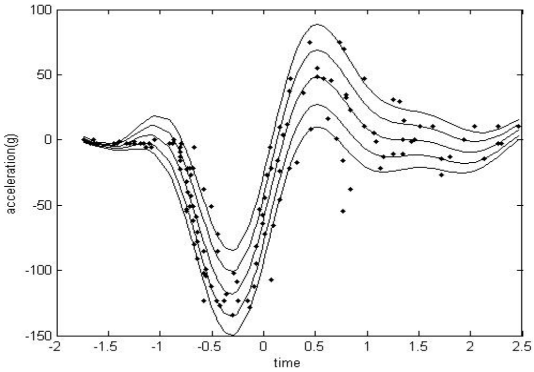

Figure 4.1 An illustration of the proposed quantile regression analysis for the motorcycle data. Estimated quantile functions for θ =0.1, 0.25, 0.5, 0.75, 0.9 are superimposed on the scatter plot.

quantile functions reflect well the heteroscedastic structure of the error term. They have their (local) minima and (local) maxima at different x values. For example, the 0.1th, 0.5th and 0.9th quantile functions have maxima at x =0.75, 0.75 and 2.15, respectively, and minima at 1.75, 0.25 and 0.25, respectively. There are no quantile crossing even in the region where the data is sparse.For 100 data sets of size 150, we now compare the proposed method with other quantile function estimation methods - the RRQ method by He (1997) and the SVM method by Takeuchi (2004). For comparison we calculate the mean and standard deviation of 100 mean absolute errors (MAEs) for each estimated quantile function as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|q_\theta(x_i) - \hat{q}_\theta(x_i)|$$

In Table 4.1 standard deviations are given in parenthesis. As seen from Table 4.1, the proposed method yields the smallest mean of MAEs. We can see that the proposed method works better than the RRQ and the SVM on simulated data sets.

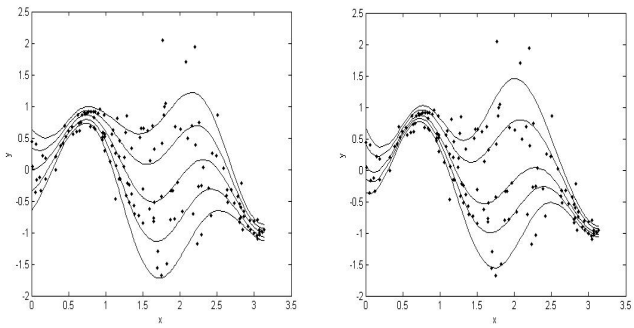

Figure 4.2 True quantile functions (Left) and estimated quantile functions estimated by the proposed method (Right) for one of 100 data sets. The quantile functions for θ =0.1, 0.25, 0.5, 0.75, 0.9 are superimposed on the scatter plot.

Table 4.1 Comparison of MAEs for the simulated data sets.

	θ				
	0.1	0.25	0.5	0.75	0.9
RRQ	0.1632 (0.0344)	0.1523 (0.0382)	0.1827 (0.0546)	0.2521 (0.0751)	0.3604 (0.1031)
SVM	0.1828 (0.0994)	0.1626 (0.0573)	0.1936 (0.0707)	0.3051 (0.0955)	0.4601 (0.1393)
proposed	0.1362 (0.0446)	0.1030 (0.0306)	0.0891 (0.0278)	0.1054 (0.0342)	0.1352 (0.0452)

5. Conclusions

In this paper we have proposed a new non-crossing quantile regression method which uses support vector median regression and heteroscedastic location-scale model as basic model. To show the effectiveness of the proposed method we have used a well-known motorcycle data set and simulated data sets. Through numerical experiment, we found that the proposed method derives a satisfying solution to estimating non-crossing quantile functions when multiple quantiles are needed and captures well the characteristics of data. The functional characteristics are obtained through the selection of free parameters of the proposed method in case of the Gaussian kernel. These parameters have been tuned using GACV function.

References

Cawley, G. C., Talbot, N. L. C., Foxall, R. J., Dorling, S. R. and Mandic, D. P. (2004). Heteroscedastic kernel ridge regression. *Neurocomputing*, **57**, 105-124.

Cole, T. J. (1990). The LMS method for constructing normalized growth standards. *European Journal of Clinical Nutrition*, **44**, 45-60.

Härdle, W. (1989). *Applied nonparametric regression*, Cambridge, Cambridge University Press.

He, X. (1997). Quantile curves without crossing. *The American Statistician,* **51**, 86-192.

Heagerty, P. J. and Pepe, M. S. (1999). Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. *Applied Statistics,* **48**, Part 4, 533-551.

Hwang, C. (2007). Kernel machine for Poisson regression. *Journal of Korean Data & Information Science Society*, **18**, 767-772.

Hwang, C. (2008). Mixed effects kernel binomial regression. *Journal of Korean Data & Information Science Society*, **19**, 1327-1334.

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

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. *Proceedings of 2nd Berkeley Symposium. Berkeley*, University of California Press, 481-492.

Li, Y., Liu, Y. and Ji, Z. (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 and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society* A, 415-44.

Shim, J. and Seok, K. H. (2008). Kernel poisson regression for longitudinal data. *Journal of Korean Data & Information Science Society*, **19**, 1353-1360.

Shim, J., Park, H. and Hwang, C. (2009). A kernel machine for estimation of mean and volatility functions. *Journal of Korean Data & Information Science Society*, **20**, 905-912.

Shim, J., Seok, K. H. and Hwang, C. (2009). Non-crossing quantile regression via doubly penalized kernel machine. *Computational Statistics*, **24**, 83-94.

Takeuchi, I. (2004). Non-crossing quantile regression curves by support vector and its efficient implementation. *Proceedings of 2004 IEEE IJCNN*, **1**, 401-406.

Takeuchi, I., Le, Q. V., Sears, T. D. and Smola, A. J. (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research*, **7**, 1231-1264.

Vapnik, V. (1995). *The nature of statistical learning theory*, Springer-Verlag, New York.

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