Comput Stat (2016) 31:1015–1030
DOI 10.1007/s00180-016-0647-5

ORIGINAL PAPER

Support vector quantile regression with varying coefficients

Jooyong Shim[1] · **Changha Hwang**[2] · **Kyungha Seok**[1]

Received: 19 August 2015 / Accepted: 31 January 2016 / Published online: 26 February 2016

Abstract Quantile regression has received a great deal of attention as an important tool for modeling statistical quantities of interest other than the conditional mean. Varying coefficient models are widely used to explore dynamic patterns among popular models available to avoid the curse of dimensionality. We propose a support vector quantile regression model with varying coefficients and its two estimation methods. One uses the quadratic programming, and the other uses the iteratively reweighted least squares procedure. The proposed method can be applied easily and effectively to estimating the nonlinear regression quantiles depending on the high-dimensional vector of smoothing variables. We also present the model selection method that employs generalized cross validation and generalized approximate cross validation techniques for choosing the hyperparameters, which affect the performance of the proposed model. Numerical studies are conducted to illustrate the performance of the proposed model.

Keywords Generalized approximate cross validation · Generalized cross validation · Iteratively reweighted least squares · Hyperparameter selection · Quadratic programming · Quantile regression · Support vector machine · Support vector quantile regression · Varying coefficient model

✉ Kyungha Seok
statskh@inje.ac.kr

Jooyong Shim
ds1631@hanmail.net

Changha Hwang
chwang@dankook.ac.kr

1 Department of Statistics and Institute of Statistical Information, Inje University, Kimhae, Obang-Dong 621-749, Korea

2 Department of Applied Statistics, Dankook University, Yongin, Gyeonggido 448-701, Korea

1 Introduction

Quantile regression (QR), introduced by Koenker and Bassett (1978), has been widely used as a way to estimate the conditional quantiles of a response variable distribution. Thus, QR in general provides a much more comprehensive picture of the conditional distribution of a response variable than the conditional mean function. Furthermore, QR is a useful and robust statistical method for estimating and conducting inferences about a model for conditional quantile functions (Yu et al. 2003). Applications of QR in many different areas, including medicine (Cole and Green 1992; Heagerty and Pepe 1999), survival analysis (Ying et al. 1995; Koenker and Geling 2001; Shim and Hwang 2009), econometrics (Hendricks and Koenker 1992; Koenker and Hallock 2001; Shim et al. 2011), and growth charts (Wei and He 2006), have been studied.

To address the curse of dimensionality problem in regression study, the additive model by Breiman and Friedman (1985) and the varying coefficient (VC) model by Hastie and Tibshirani (1993) have been proposed. It is well known that a general form of the VC model includes the additive model as a special case. VC models constitute an important class of nonparametric models. However, VC models have inherited the simplicity and easy interpretation of classical linear models. The introductions, various applications, and current research areas of VC models can be found in Hastie and Tibshirani (1993), Hoover et al. (1998), Fan and Zhang (2008), and Park et al. (2015). Recently, QR with VCs has been studied. Honda (2004) considered the estimation of conditional quantiles in VC models by estimating the coefficients by local L_1 regression. Kim (2007) also considered conditional quantiles with VCs and proposed a methodology for their estimation and assessment using polynomial splines. Cai and Xu (2008) considered QR with VCs for a time series model. They used local polynomial schemes to estimate the coefficients. In this paper, we propose a support vector quantile regression (SVQR) with VCs and its two estimation methods, which can be applied effectively to high-dimensional cases. This is the first article that deals with SVQR with VCs. By the way, we do not deal with the quantile crossing problems in this paper.

The support vector machine (SVM), first developed by Vapnik (1995) and his group at AT&T Bell Laboratories, has been successfully applied to a number of real world problems related to classification and regression problems. Takeuchi and Furuhashi (2004) first considered QR by SVM. Li et al. (2007) proposed a SVQR using quadratic programming (QP) and derived a simple formula for the effective dimension of the SVQR, which allows convenient selection of hyperparameters. Shim and Hwang (2009) considered a modified SVQR using an iterative reweighted least squares (IRWLS) procedure.

In this paper we present an SVQR with nonlinear coefficient functions and its two estimation methods. One uses QP and the other uses the IRWLS procedure. The IRWLS procedure uses a modified check function. This IRWLS procedure makes it possible to derive a generalized cross validation (GCV) method for choosing hyperparameters and to construct pointwise confidence intervals for coefficient functions. We also investigate the performance of the SVQR estimations through numerical studies. The rest of this paper is organized as follows. Section 2 introduces two versions of SVQR with VCs. Sections 3 and 4 present our numerical studies and conclusions, respectively.

2 SVQR with VCs

In this section we propose two versions of SVQR with VCs and their hyperparameter selection procedures.

2.1 SVQR with VCs using QP

We now illustrate SVQR with VCs using QP and its hyperparameter selection procedure. In this section we adapt a dimension reduction modeling method termed the VC modeling approach to explore dynamic patterns.

We assume the θth QR with VCs takes the form

$$q_\theta\left(\boldsymbol{x}_i, \boldsymbol{u}_i\right) = \sum_{k=0}^{d_x} x_{ik}\beta_{k,\theta}\left(\boldsymbol{u}_i\right) = \boldsymbol{\beta}_\theta^t\left(\boldsymbol{u}_i\right)\boldsymbol{x}_i, \tag{1}$$

where superscript t denote the transpose, $\boldsymbol{u}_i$ is called the smoothing variables, $\boldsymbol{x}_i = (x_{i0}, x_{i1}, \ldots, x_{id_x})^t$ with $x_{i0} \equiv 1$ is the input vector, $\{\beta_{k,\theta}(\cdot)\}$ are smooth coefficient functions, and $\boldsymbol{\beta}_\theta(\boldsymbol{u}_i) = (\beta_{0,\theta}(\boldsymbol{u}_i), \ldots, \beta_{d_x,\theta}(\boldsymbol{u}_i))^t$. Here all of the $\{\beta_{k,\theta}(\cdot)\}$ are allowed to depend on θ. For simplicity, we drop θ from $\{\beta_{k,\theta}(\cdot)\}$. The QR model (1) has been widely used to analyze conditional quantiles due to their flexibility and interpretability. In fact, this model constitutes an important class of nonparametric models.

We first estimate the coefficients $\{\beta_k(\cdot)\}$ using the basic principle of SVQR based on the training data set $\mathcal{D} = \{(\boldsymbol{x}_i, \boldsymbol{u}_i, y_i)\}_{i=1}^n$. Then we estimate the conditional quantile $q_\theta(\cdot,\cdot)$ in VC model by estimating the coefficients. For the SVQR with VCs we assume that each coefficient function $\beta_k(\boldsymbol{u}_i)$ is nonlinearly related to the smoothing variables $\boldsymbol{u}_i$ such that $\beta_k(\boldsymbol{u}_i) = \boldsymbol{w}_k^t\boldsymbol{\phi}(\boldsymbol{u}_i) + b_k$ for $k = 0, \ldots, d_x$, where $\boldsymbol{w}_k$ is a corresponding weight vector of size $d_f \times 1$. Here the nonlinear feature mapping function $\boldsymbol{\phi} : R^{d_u} \to R^{d_f}$ 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, $\boldsymbol{\phi}(\boldsymbol{u}_i)^t\boldsymbol{\phi}(\boldsymbol{u}_j) = K(\boldsymbol{u}_i, \boldsymbol{u}_j)$, provided certain conditions hold Mercer (1909). Among several kernel functions we use in this paper Gaussian kernel, polynomial kernel and Epanechnikov kernel defined, respectively, as

$$\begin{aligned}
K\left(\boldsymbol{u}_i, \boldsymbol{u}_j\right) &= \exp\left(-\|\boldsymbol{u}_i - \boldsymbol{u}_j\|^2/2\sigma^2\right), \\
K\left(\boldsymbol{u}_i, \boldsymbol{u}_j\right) &= \left(1 + \boldsymbol{u}_i^t\boldsymbol{u}_j\right)^d, \quad i, j = 1, \ldots, n, \\
K\left(\boldsymbol{u}_i, \boldsymbol{u}_j\right) &= 0.75\left(1 - \|\frac{\boldsymbol{u}_i - \boldsymbol{u}_j}{h}\|^2\right) I\left(\|\frac{\boldsymbol{u}_i - \boldsymbol{u}_j}{h}\| < 1\right),
\end{aligned}$$

where σ, h and d are kernel parameters.

Then, using the basic principle of SVQR, the coefficient estimators $\{\hat{\beta}_k(\cdot)\}$ of SVQR with VCs can be obtained by minimizing the following equation,

$$L = \frac{1}{2}\sum_{k=0}^{d_x} \|\boldsymbol{w}_k\|^2 + C\sum_{i=1}^{n} \rho_\theta\left(y_i - \sum_{k=0}^{d_x} x_{ik}\left(\boldsymbol{w}_k^t \boldsymbol{\phi}\left(\boldsymbol{u}_i\right) + b_k\right)\right), \tag{2}$$

where $\rho_\theta(r) = \theta r I(r \geq 0) - (1-\theta) r I(r < 0)$ is the check function with the indicator function $I(\cdot)$, and $C > 0$ is a penalty parameter which controls the balance between the smoothness and fitness of the QR estimator.

We can express the optimization problem (2) by the formulation for SVQR as follows:

$$L = \frac{1}{2}\sum_{k=0}^{d_x} \|\boldsymbol{w}_k\|^2 + C\theta\sum_{i=1}^{n} \xi_i + C(1-\theta)\sum_{i=1}^{n} \xi_i^*$$

subject to

$$\begin{cases} y_i - \sum_{k=0}^{d_x} x_{ik}\left(\boldsymbol{w}_k^t \boldsymbol{\phi}(\boldsymbol{u}_i) + b_k\right) \leq \xi_i, \\ -y_i + \sum_{k=0}^{d_x} x_{ik}\left(\boldsymbol{w}_k^t \boldsymbol{\phi}(\boldsymbol{u}_i) + b_k\right) \leq \xi_i^*, \quad i = 1, \ldots, n. \end{cases}$$

We construct a Lagrange function as follows:

$$\begin{aligned} L = {} & \frac{1}{2}\sum_{k=0}^{d_x} \|\boldsymbol{w}_k\|^2 + C\theta\sum_{i=1}^{n} \xi_i + C(1-\theta)\sum_{i=1}^{n} \xi_i^* \\ & - \sum_{i=1}^{n} \alpha_i\left(\xi_i - y_i + \sum_{k=0}^{d_x} x_{ik}(\boldsymbol{w}_k^t \boldsymbol{\phi}(\boldsymbol{u}_i) + b_k)\right) \\ & - \sum_{i=1}^{n} \alpha_i^*\left(\xi_i^* + y_i - \sum_{k=0}^{d_x} x_{ik}(\boldsymbol{w}_k^t \boldsymbol{\phi}(\boldsymbol{u}_i) + b_k)\right) \\ & - \sum_{i=1}^{n} \eta_i \xi_i - \sum_{i=1}^{n} \eta_i^* \xi_i^*. \end{aligned} \tag{3}$$

We notice that the non-negative constraints with Lagrange multipliers $\alpha_i^{(*)}, \eta_i^{(*)} \geq 0$ should be satisfied. Taking partial derivatives of Eq. (3) with regard to the primal variables $(\boldsymbol{w}_k, \xi_i^{(*)}, b_k)$, we have

$$\begin{aligned} & \frac{\partial L}{\partial \boldsymbol{w}_k} = \boldsymbol{0} \Rightarrow \boldsymbol{w}_k = \sum_{i=1}^{n} x_{ik} \boldsymbol{\phi}(\boldsymbol{u}_i)(\alpha_i - \alpha_i^*), \quad k = 0, 1, \ldots, d_x, \\ & \frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C\theta = \alpha_i + \eta_i, \quad i = 1, \ldots, n, \end{aligned}$$

$$\frac{\partial L}{\partial \xi_i^*} = 0 \Rightarrow C(1-\theta) = \alpha_i^* + \eta_i^*, \quad i = 1, \ldots, n,$$
$$\frac{\partial L}{\partial b_k} = 0 \Rightarrow \sum_{i=1}^{n} x_{ik} \left(\alpha_i - \alpha_i^*\right) = 0, \quad k = 0, 1, \ldots, d_x.$$

Plugging the above results into the Eq. (3), we have the dual optimization problem to maximize

$$-\frac{1}{2} \sum_{i,j=1}^{n} \left(\alpha_i - \alpha_i^*\right) \left(\alpha_j - \alpha_j^*\right) \sum_{k=0}^{d_x} x_{ik} x_{jk} K(\boldsymbol{u}_i, \boldsymbol{u}_j) + \sum_{i=1}^{n} y_i \left(\alpha_i - \alpha_i^*\right) \tag{4}$$

subject to

$$\begin{cases} \sum_{i=1}^{n} x_{ik} \left(\alpha_i - \alpha_i^*\right) = 0, \quad k = 0, 1, \ldots, d_x, \\ 0 \le \alpha_i \le C\theta, \ 0 \le \alpha_i^* \le C(1-\theta), \quad i = 1, \ldots, n. \end{cases}$$

We notice that this SVQR with VCs works by solving a constrained QP problem.

Solving the QP problem (4) with the constraints determines the optimal Lagrange multipliers $(\hat{\alpha}_i, \hat{\alpha}_i^*)$. Thus, for a given $(\boldsymbol{x}_t, \boldsymbol{u}_t)$ the SVQR with VCs using QP for coefficient function estimation takes the form:

$$\hat{\beta}_k(\boldsymbol{u}_t) = \sum_{i=1}^{n} x_{ik} K(\boldsymbol{u}_t, \boldsymbol{u}_i) \left(\hat{\alpha}_i - \hat{\alpha}_i^*\right) + \hat{b}_k,$$

and then for QR function estimator takes the form:

$$\hat{q}_\theta(\boldsymbol{x}_t, \boldsymbol{u}_t) = \sum_{i=1}^{n} \sum_{k=0}^{d_x} x_{tk} x_{ik} K(\boldsymbol{u}_t, \boldsymbol{u}_i) \left(\hat{\alpha}_i - \hat{\alpha}_i^*\right) + \sum_{k=0}^{d_x} x_{tk} \hat{b}_k.$$

We remark that $(\boldsymbol{x}_t, \boldsymbol{u}_t)$ could be an observation in the training data set or a new observation. Here $\hat{b}_k$ for $k = 0, 1, \ldots, d_x$ is obtained via Kuhn–Tucker conditions (Kuhn and Tucker 1951) such as

$$\begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \vdots \\ \hat{b}_{d_x} \end{pmatrix} = \left(\boldsymbol{X}_s^t \boldsymbol{X}_s\right)^{-1} \boldsymbol{X}_s^t \boldsymbol{y}_s, \tag{5}$$

where $\boldsymbol{X}_s$ is an $n_s \times (d_x + 1)$ matrix with ith row $\boldsymbol{x}_i^t$ for $i \in I_s = \{i = 1, \ldots, n | 0 < \alpha_i < C\theta, 0 < \alpha_i^* < C(1-\theta)\}$, $\boldsymbol{y}_s$ is an $n_s \times 1$ vector with ith element $\left(y_i - \sum_{j=1}^{n} \sum_{k=0}^{d_x} x_{ik} x_{jk} K(\boldsymbol{u}_i, \boldsymbol{u}_j) \left(\hat{\alpha}_j - \hat{\alpha}_j^*\right)\right)$ for $i \in I_s$ and n_s is the size of I_s.

We now consider the hyperparameter selection problem which determines the appropriate hyperparameters of the proposed SVQR with VCs using QP. The functional structure of the SVQR with VCs using QP is characterized by hyperparameters such as the regularization parameter C and the kernel parameter $\gamma \in \{\sigma, h, d\}$. To choose the values of hyperparameters of the SVQR with VCs using QP we first need to consider the cross validation (CV) function as follows:

$$CV(\boldsymbol{\lambda}) = \sum_{i=1}^{n} \rho_\theta \left(y_i - \hat{q}_\theta^{(-i)}(\boldsymbol{x}_i, \boldsymbol{u}_i) \right),$$

where $\boldsymbol{\lambda} = (C, \gamma)$ is the set of hyperparameters, and $\hat{q}_\theta^{(-i)}(\boldsymbol{x}_i, \boldsymbol{u}_i)$ is the θth QR function estimated without ith observation. Since for each candidate of hyperparameters, $\hat{q}_\theta^{(-i)}(\boldsymbol{x}_i, \boldsymbol{u}_i)$ for $i = 1, \ldots, n$, should be evaluated, selecting hyperparameters using CV function is computationally formidable. Applying Yuan (2006), a GACV function to select the set of hyperparameters $\boldsymbol{\lambda}$ for SVQR with VCs using QP is shown as follows:

$$GACV(\boldsymbol{\lambda}) = \frac{\sum_{i=1}^{n} \rho_\theta \left(y_i - \hat{q}_\theta \left(\boldsymbol{x}_i, \boldsymbol{u}_i \right) \right)}{n - df},$$

where df is a measure of the effective dimensionality of the fitted model. In this paper we used $df = n_s$ related with (5) from Li et al. (2007). Another common criterion is Schwarz information criterion (SIC) (Schwarz (1978), Koenker et al. (1994))

$$SIC(\boldsymbol{\lambda}) = \ln \left(\frac{1}{n} \sum_{i=1}^{n} \rho_\theta \left(y_i - \hat{q}_\theta \left(\boldsymbol{x}_i, \boldsymbol{u}_i \right) \right) \right) + \frac{\ln n}{2n} df.$$

2.2 SVQR with VCs using IRWLS

We now illustrate SVQR with VCs using IRWLS procedure and its hyperparameter selection procedure. This method enables us to derive GCV for selecting hyperparameters and obtain the variance of $\hat{\beta}_k(\boldsymbol{u}_t)$ so as to construct an approximate pointwise confidence interval for $\beta_k(\boldsymbol{u}_t)$.

The check function $\rho_\theta(\cdot)$ used in SVQR with VCs using QP can be seen as the weighted quadratic loss function such as

$$\rho_\theta(r) = \upsilon(\theta) r^2,$$

where $\upsilon(\theta) = (\theta I(r \geq 0) + (1 - \theta) I(r < 0))/|r|$. Now the optimization problem (2) becomes the problem of obtaining $(\boldsymbol{w}_k, b_k)$'s which minimize

$$L = \frac{1}{2} \sum_{k=0}^{d_x} \|\boldsymbol{w}_k\|^2 + \frac{C}{2} \sum_{i=1}^{n} \upsilon_i(\theta) \left(y_i - \sum_{k=0}^{d_x} x_{ik} \left(\boldsymbol{w}_k^t \boldsymbol{\phi}(\boldsymbol{u}_i) + b_k \right) \right)^2, \qquad (6)$$

where $\upsilon_i(\theta) = (\theta I(e_i \geq 0) + (1-\theta) I(e_i < 0))/|e_i|$ with $e_i = y_i - \sum_{k=0}^{d_x} x_{ik}(\boldsymbol{w}_k^t \boldsymbol{\phi}(\boldsymbol{u}_i) + b_k)$ and $C > 0$ is a penalty parameter.

We can express the optimization problem (6) by formulation for weighted least squares SVM as follows:

$$L = \frac{1}{2} \sum_{k=0}^{d_x} \|\boldsymbol{w}_k\|^2 + \frac{C}{2} \sum_{i=1}^{n} \upsilon_i(\theta) e_i^2$$

subject to

$$y_i - \sum_{k=0}^{d_x} x_{ik} \left(\boldsymbol{w}_k^t \boldsymbol{\phi}(\boldsymbol{u}_i) + b_k \right) = e_i, \quad i = 1, \ldots, n.$$

We construct a Lagrange function as follows:

$$L = \frac{1}{2} \sum_{k=1}^{d_x} \|\boldsymbol{w}_k\|^2 + \frac{C}{2} \sum_{i=1}^{n} \upsilon_i(\theta) e_i^2 - \sum_{i=1}^{n} \alpha_i \left(e_i - y_i + \sum_{k=0}^{d_x} x_{ik} \left(\boldsymbol{w}_k^t \boldsymbol{\phi}(\boldsymbol{u}_i) + b_k \right) \right), \tag{7}$$

where α_i's are Lagrange multipliers. Taking partial derivatives of Eq. (7) with regard to $(\boldsymbol{w}_k, b_k, e_i, \alpha_i)$ we have,

$$\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{w}_k} &= \boldsymbol{0} \Rightarrow \boldsymbol{w}_k = \sum_{i=1}^{n} x_{ik} \boldsymbol{\phi}(\boldsymbol{u}_i) \alpha_i, \quad k = 0, \ldots, d_x, \\
\frac{\partial L}{\partial b_k} &= 0 \Rightarrow \sum_{i=1}^{n} x_{ik} \alpha_i = 0, \quad k = 0, \ldots, d_x, \\
\frac{\partial L}{\partial e_i} &= 0 \Rightarrow C \upsilon_i(\theta) e_i - \alpha_i = 0, \quad i = 1, \ldots, n, \\
\frac{\partial L}{\partial \alpha_i} &= 0 \Rightarrow e_i - y_i + \sum_{k=0}^{d_x} x_{ik} (\boldsymbol{w}_k^t \boldsymbol{\phi}(\boldsymbol{u}_i) + b_k) = 0, \quad i = 1, \ldots, n.
\end{aligned}$$

After eliminating e_i's and $\boldsymbol{w}_k$'s, we have the optimal values of α_i's and b_k's from the linear equation as follows:

$$\begin{pmatrix} \boldsymbol{X}\boldsymbol{X}^t \odot \boldsymbol{K} + \frac{1}{C} \boldsymbol{V}(\theta)^{-1} & \boldsymbol{X} \\ \boldsymbol{X}^t & \boldsymbol{0}_{(d_x+1)\times(d_x+1)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{b} \end{pmatrix} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{0}_{(d_x+1)\times 1} \end{pmatrix} \tag{8}$$

where $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^t$, $\boldsymbol{K}$ is an $n \times n$ kernel matrix with (i, j)th element $K(\boldsymbol{u}_i, \boldsymbol{u}_j)$, $\boldsymbol{V}(\theta)$ is an $n \times n$ diagonal matrix of $\upsilon_i(\theta)$, $\boldsymbol{0}_{p\times q}$ is a $p \times q$ zero matrix, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^t$, $\boldsymbol{b} = (b_0, \ldots, b_{d_x})^t$ and $\odot$ denotes a componentwise product. We notice that the solution to (8) cannot be obtained in a single step since $\boldsymbol{V}(\theta)$ contains

$(\boldsymbol{\alpha}, \boldsymbol{b})$, which leads to apply the IRWLS procedure which starts with initialized values of $(\boldsymbol{\alpha}, \boldsymbol{b})$.

Solving the linear Eq. (8) determines the optimal Lagrange multipliers $\hat{\alpha}_i$'s and bias terms $\hat{b}_k$'s. Thus, for a given $(\boldsymbol{x}_t, \boldsymbol{u}_t)$ the SVQR with VCs using IRWLS for coefficient function estimation takes the form:

$$\hat{\beta}_k(\boldsymbol{u}_t) = \sum_{i=1}^{n} x_{ik} K(\boldsymbol{u}_t, \boldsymbol{u}_i)\hat{\alpha}_i + \hat{b}_k, \tag{9}$$

and then for QR function estimation takes the form:

$$\hat{q}_\theta(\boldsymbol{x}_t, \boldsymbol{u}_t) = \sum_{i=1}^{n} \sum_{k=0}^{d_x} x_{tk} x_{ik} K(\boldsymbol{u}_t, \boldsymbol{u}_i)\hat{\alpha}_i + \sum_{k=0}^{d_x} x_{tk}\hat{b}_k. \tag{10}$$

For the purpose of utilizing in constructing confidence intervals for $\beta_k(\boldsymbol{u}_t)$ and $q_\theta(\boldsymbol{x}_t, \boldsymbol{u}_t)$ we are going to express $\hat{\beta}_k(\boldsymbol{u}_t)$ and $\hat{q}_\theta(\boldsymbol{x}_t, \boldsymbol{u}_t)$ as the linear combination of $\boldsymbol{y}$ in what follows. From (8) we can express $\hat{\beta}_k(\boldsymbol{u}_t)$ as follows:

$$\begin{aligned}\hat{\beta}_k(\boldsymbol{u}_t) &= \left(\boldsymbol{x}_{(k)}^t \odot \boldsymbol{k}_t, \boldsymbol{v}_{d_x+1}^t(k)\right) \boldsymbol{M}\boldsymbol{y} \\ &= \boldsymbol{s}_k(\boldsymbol{u}_t)\boldsymbol{y},\end{aligned} \tag{11}$$

where $\boldsymbol{s}_k(\boldsymbol{u}_t) = (\boldsymbol{x}_{(k)}^t \odot \boldsymbol{k}_t, \boldsymbol{v}_{d_x+1}^t(k))\boldsymbol{M}$, $\boldsymbol{x}_{(k)}$ is the $(k+1)$th column of $\boldsymbol{X}$, $\boldsymbol{k}_t = (K(\boldsymbol{u}_t, \boldsymbol{u}_1), \ldots, K(\boldsymbol{u}_t, \boldsymbol{u}_n))$, $\boldsymbol{v}_{d_x+1}(k)$ is a vector of size $(d_x + 1)$ with 0's but 1 in $(k+1)$th, and $\boldsymbol{M}$ is the $(n + d_x + 1) \times n$ submatrix of the inverse of the leftmost matrix in (8). For a point $(\boldsymbol{u}_t, \boldsymbol{x}_t)$ we can also express $\hat{q}_\theta(\boldsymbol{x}_t, \boldsymbol{u}_t)$ as follows:

$$\hat{q}_\theta(\boldsymbol{x}_t, \boldsymbol{u}_t) = \boldsymbol{h}_t(\theta)\boldsymbol{y},$$

where $\boldsymbol{h}_t(\theta) = \left((\boldsymbol{x}_t^t \boldsymbol{X}^t) \odot \boldsymbol{k}_t, \boldsymbol{x}_t^t\right)\boldsymbol{M}$. From (11) we can obtain the estimator of $Var(\hat{\beta}_k(\boldsymbol{u}_t))$ for $k = 0, 1, \ldots, d_x$ as follows:

$$\widehat{Var}\left(\hat{\beta}_k(\boldsymbol{u}_t)\right) = \boldsymbol{s}_k(\boldsymbol{u}_t)\hat{\boldsymbol{\Sigma}}\boldsymbol{s}_k^t(\boldsymbol{u}_t), \tag{12}$$

where $\hat{\boldsymbol{\Sigma}}$ is an estimator of $Var(\boldsymbol{y})$.

For nonparametric inference the confidence interval is really useful. There are two types of confidence intervals. One is the pointwise confidence interval. The other is the simultaneous confidence interval. Our interest here is in estimating coefficient functions rather than the QR function itself. Thus we illustrate the pointwise confidence intervals only for the coefficient functions in the SVQR with VCs using IRWLS. But the pointwise confidence interval for the QR function can be derived in the same way. The estimated variance (12) can be used to construct pointwise confidence intervals. Under certain regularity conditions (Shiryaev 1996), the central limit theorem for linear smoothers is valid and we can show asymptotically

$$\frac{\hat{\beta}_k(\boldsymbol{u}_t) - E\left(\hat{\beta}_k(\boldsymbol{u}_t)\right)}{\widehat{Var}\left(\hat{\beta}_k(\boldsymbol{u}_t)\right)} \rightarrow^D N(0,1), \quad k = 0, 1, \ldots, d_x,$$

where $\rightarrow^D$ denotes convergence in distribution. If the estimator is conditionally unbiased, i.e., $E(\hat{\beta}_k(\boldsymbol{u}_t)) = \beta_k(\boldsymbol{u}_t)$ for $k = 0, 1, \ldots, d_x$, approximate $100(1-\alpha)\%$ pointwise confidence interval takes the form

$$\left(\hat{\beta}_k(\boldsymbol{u}_t) \pm z_{1-\frac{\alpha}{2}}\sqrt{\widehat{Var}\left(\hat{\beta}_k(\boldsymbol{u}_t)\right)}\right), \quad k = 0, 1, \ldots, d_x, \tag{13}$$

where $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$th quantile of the standard normal distribution. In fact, the interval (13) is a confidence interval for $E(\hat{\beta}_k(\boldsymbol{u}_t))$. It is a confidence interval for $\beta_k(\boldsymbol{u}_t)$ under the assumption $E(\hat{\beta}_k(\boldsymbol{u}_t)) = \beta_k(\boldsymbol{u}_t)$. Thus it is actually the bias-ignored approximate $100(1-\alpha)\%$ pointwise confidence interval.

We now consider the hyperparameter selection problem which determines the appropriate hyperparameters of the proposed SVQR with VCs using IRWLS. To determine the values of hyperparameters of the SVQR with VCs using IRWLS we first need to consider the CV function as follows:

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n} \upsilon_i(\theta)\left(y_i - \hat{q}_\theta^{(-i)}(\boldsymbol{x}_i, \boldsymbol{u}_i)\right)^2$$

By leaving-out-one Lemma of Craven and Wahba (1979),

$$\begin{aligned}\left(y_i - \hat{q}_\theta^{(-i)}(\boldsymbol{x}_i, \boldsymbol{u}_i)\right) - \left(y_i - \hat{q}_\theta(\boldsymbol{x}_i, \boldsymbol{u}_i)\right) &= \hat{q}_\theta(\boldsymbol{x}_i, \boldsymbol{u}_i) - \hat{q}_\theta^{(-i)}(\boldsymbol{x}_i, \boldsymbol{u}_i) \\ &\simeq \frac{\partial \hat{q}_\theta(\boldsymbol{x}_i, \boldsymbol{u}_i)}{\partial y_i}\left(y_i - \hat{q}_\theta^{(-i)}(\boldsymbol{x}_i, \boldsymbol{u}_i)\right)\end{aligned}$$

we have

$$\left(y_i - \hat{q}_\theta^{(-i)}(\boldsymbol{x}_i, \boldsymbol{u}_i)\right) \simeq \frac{y_i - \hat{q}_\theta(\boldsymbol{x}_i, \boldsymbol{u}_i)}{1 - \frac{\partial \hat{q}_\theta(\boldsymbol{x}_i, \boldsymbol{u}_i)}{\partial y_i}}.$$

Then the ordinary cross validation (OCV) function can be obtained as

$$\begin{aligned}OCV(\boldsymbol{\lambda}) &= \frac{1}{n}\sum_{i=1}^{n}\upsilon_i(\theta)\left(\frac{y_i - \hat{q}_\theta(\boldsymbol{x}_i, \boldsymbol{u}_i)}{1 - \frac{\partial \hat{q}_\theta(\boldsymbol{x}_i, \boldsymbol{u}_i)}{\partial y_i}}\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\upsilon_i(\theta)\left(\frac{y_i - \hat{q}_\theta(\boldsymbol{x}_i, \boldsymbol{u}_i)}{1 - h_{ii}}\right)^2,\end{aligned}$$

where $h_{ij} = \partial \hat{q}_\theta(\boldsymbol{x}_i, \boldsymbol{u}_i)/\partial y_j$ is an (i, j)th element of the hat matrix $\boldsymbol{H}$. Replacing h_{ii} by their average $tr(\boldsymbol{H})/n$, the GCV function can be obtained as

$$GCV(\boldsymbol{\lambda}) = \frac{n \sum_{i=1}^n \upsilon_i(\theta) \left(y_i - \hat{q}_\theta\left(\boldsymbol{x}_i, \boldsymbol{u}_i\right)\right)^2}{(n - tr(\boldsymbol{H}))^2}.$$

3 Numerical studies

In this section, we illustrate the performance of the SVQR with VCs using QP and IRWLS with synthetic data and the wage data in Wooldridge (2003). For our numerical studies, we compare the proposed methods with SVQR in the study by Li et al. (2007) and local polynomial quantile regression with VCs (LPQRVC) in the study by Cai and Xu (2008). Throughout this paper, we use the Epanechnikov kernel for the LPQRVC and the Gaussian kernel function for the SVQR, SVQR with VCs using QP and IRWLS. For hyperparameter selection we use the CV function for the LPQRVC method, the GCV function for the SVQR with VCs using IRWLS, and the GACV function for the SVQR and SVQR with VCs using QP. To obtain the best of each method, we use different kernels and criteria. The hyerparameters are selected to minimize each objective function with the grid search method. The candidates sets of the regularization parameter C and kernel parameter σ in SVQR with VCs using QP and IRWLS, and SVQR are $\{10, 20, 40, 70, 100, 200, 400, 600, 800, 1000, 1200\}$ and $\{0.5, 1, 2, \ldots, 8\}$, respectively. The parameter h in Epanechnikov kernel function is selected from the set $\{0.1, 0.2, \ldots, 1\}$. We use $\mathbf{0}$'s as the initial values of $\boldsymbol{\alpha}$ and $\boldsymbol{b}$ for the IRWLS procedure associated with the SVQR with VCs using IRWLS.

3.1 Synthetic data example

For the synthetic data example we generate $\{(\boldsymbol{x}_i, u_i, y_i)\}_{i=1}^n$ from the location-scale model,

$$y_i = \beta_1(u_i)x_{1i} + \beta_2(u_i)x_{2i} + \sigma(u_i)e_i, \quad i = 1, \ldots, n,$$

where $\beta_1(u_i) = \sin(\sqrt{2}\pi u_i)$, $\beta_2(u_i) = \cos(\sqrt{2}\pi u_i)$, $\sigma(u_i) = \exp(\sin(0.5\pi u_i))$, $u_i \sim\ i.i.d.\ U(0, 3)$, $x_{1i}, x_{2i} \sim\ i.i.d.\ N(1, 1)$, and $e_i \sim\ i.i.d.\ N(0, 1)$ or Student's t with three degrees of freedom. The θth QR is

$$q_\theta(u_i, x_{1i}, x_{2i}) = \beta_0(u_i) + \beta_1(u_i)x_{1i} + \beta_2(u_i)x_{2i}, \tag{14}$$

where $\beta_0(u_i) = \sigma(u_i)\Phi^{-1}(\theta)$ and $\Phi^{-1}(\theta)$ is the θth quantile of the standard normal.

The performance of the estimators $\hat{q}_\theta$'s and $\hat{\beta}_k$'s is assessed by the mean integrated squared errors (MISE) and by the standard deviation of ISEs, respectively, defined as

Table 1 Comparison of the MISE and SDISE values for the case that $e_i \sim$ *i.i.d.* $N(0, 1)$

θ	f	SVQR	LPQRVC	SVQRVCQP	SVQRVCLS
0.1	$q_\theta(u, \boldsymbol{x})$	6.4934	1.7888	1.6547	**1.3862**
		(0.1769)	(0.0670)	(0.0678)	(0.0589)
	$\beta_0(u)$		2.1280	1.6473	**1.2641**
			(0.1572)	(0.1245)	(0.1082)
	$\beta_1(u)$		0.8982	0.6590	**0.5490**
			(0.0629)	(0.0487)	(0.0395)
	$\beta_2(u)$		0.8400	0.6905	**0.5398**
			(0.0554)	(0.0541)	(0.0433)
0.5	$q_\theta(u, \boldsymbol{x})$	2.6081	0.9794	0.7577	**0.7576**
		(0.0758)	(0.0339)	(0.0324)	(0.0323)
	$\beta_0(u)$		1.1896	0.9176	**0.5910**
			(0.0851)	(0.0822)	(0.0477)
	$\beta_1(u)$		0.4653	0.3105	**0.3100**
			(0.0319)	(0.0252)	(0.02422)
	$\beta_2(u)$		0.4944	0.3433	**0.3129**
			(0.0276)	(0.0253)	(0.01924)
0.9	$q_\theta(u, \boldsymbol{x})$	5.8312	1.5767	1.4763	**1.2343**
		(0.1679)	(0.0555)	(0.0581)	(0.0509)
	$\beta_0(u)$		1.7497	1.6832	**1.1605**
			(0.1329)	(0.1580)	(0.1067)
	$\beta_1(u)$		0.7082	0.6284	**0.5002**
			(0.0486)	(0.0492)	(0.0371)
	$\beta_2(u)$		0.7074	0.6256	**0.5276**
			(0.0376)	(0.0437)	(0.03697)

The SDISE values are given in parentheses
SVQRVCQP SVQR with VCs using QP, *SVQRVCLS* SVQR with VCs using IRWLS
Boldfaced values indicate the best performance for the given quantity

$$MISE = \frac{1}{N}\sum_{j=1}^{N} ISE_j,$$

$$SDISE = \left(\frac{1}{N}\sum_{j=1}^{N}\left(ISE_j - MISE\right)^2\right)^{1/2},$$

where $ISE_j = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_i - f_i)^2$, $f_i = q_\theta(u_i, \boldsymbol{x}_i)$ or $\beta_k(u_i)$, $k = 0, 1, 2$, for the jth data set, and n and N are the numbers of observations and data sets, respectively. For our experiment, we repeat $N = 100$ times with each sample size $n = 100$ for each $\theta = 0.1, 0.5$ and 0.9.

Tables 1 and 2 show the results for the MISE and SDISE values for q_θ's and β_k's for $\theta = 0.1, 0.5, 0.9$ when the distribution of error term is the standard normal $N(0, 1)$

Table 2 Comparison of the MISE and SDISE values for the case that $e_i \sim i.i.d.\ t_3$

θ	f	SVQR	LPQRVC	SVQRVCQP	SVQRVCLS
0.1	$q_\theta(u, \boldsymbol{x})$	7.8250	2.4951	2.6691	**2.1686**
		(0.4512)	(0.1733)	(0.2672)	(0.1477)
	$\beta_0(u)$		3.6666	2.8780	**1.7324**
			(0.6591)	(0.4208)	(0.2580)
	$\beta_1(u)$		1.2462	1.1476	**0.8406**
			(0.1313)	(0.1503)	(0.0751)
	$\beta_2(u)$		1.3206	1.0730	**0.8092**
			(0.2169)	(0.1196)	(0.0797)
0.5	$q_\theta(u, \boldsymbol{x})$	5.9087	1.3078	**1.0066**	1.3507
		(0.6735)	(0.0529)	(0.0464)	(0.0694)
	$\beta_0(u)$		1.6457	1.0665	**0.6408**
			(0.1453)	(0.1289)	(0.0614)
	$\beta_1(u)$		0.7641	**0.4654**	0.5503
			(0.0756)	(0.0428)	(0.0396)
	$\beta_2(u)$		0.6531	**0.4304**	0.4808
			(0.0495)	(0.0405)	(0.0340)
0.9	$q_\theta(u, \boldsymbol{x})$	12.4839	5.5506	6.1202	**4.9098**
		(1.0247)	(1.1426)	(1.1660)	(1.1526)
	$\beta_0(u)$		5.6208	4.7308	**3.7451**
			(0.8061)	(0.6877)	(0.6860)
	$\beta_1(u)$		2.2973	2.1786	**1.6913**
			(0.3880)	(0.4007)	(0.39023)
	$\beta_2(u)$		2.3749	2.6247	**1.9944**
			(0.4778)	(0.5056)	(0.5004)

The SDISE values are given in parentheses
SVQRVCQP SVQR with VCs using QP, *SVQRVCLS* SVQR with VCs using IRWLS
Boldfaced values indicate the best performance for the given quantity

and Student's t with three degrees of freedom, respectively. The SDISE values are in parentheses. Boldfaced values indicate the best performance for the given quantity. We know from Table 1 that the proposed SVQR with VCs using QP and IRWLS outperform SVQR and LPQRVC in estimating all q_θ's and LPQRVC in estimating all β_k's for the standard normal error distribution. In particular, the SVQR with VCs using IRWLS has the smallest values of MISE and SDISE for all θ's. We know from Table 2 that the SVQR with VCs using IRWLS outperforms SVQR and LPQRVC in estimating q_θ's and LPQRVC in estimating β_k's except $\theta = 0.5$ for the t_3 error distribution. However, the SVQR with VCs using QP performs best in estimating q_θ, β_1 and β_2 except β_0 for $\theta = 0.5$ for the t_3 error distribution.

Table 3 Estimated coefficients for linear QR for $\theta = 0.1, 0.5$ and 0.9 for the wage data set

θ	Bias	u	x_1	x_2	x_3
0.1	0.3678	0.0435	0.1948	0.2192	−0.0009
0.5	0.1008	0.0876	0.3759	0.1240	0.0065
0.9	0.4685	0.1026	0.3399	0.0284	0.0134

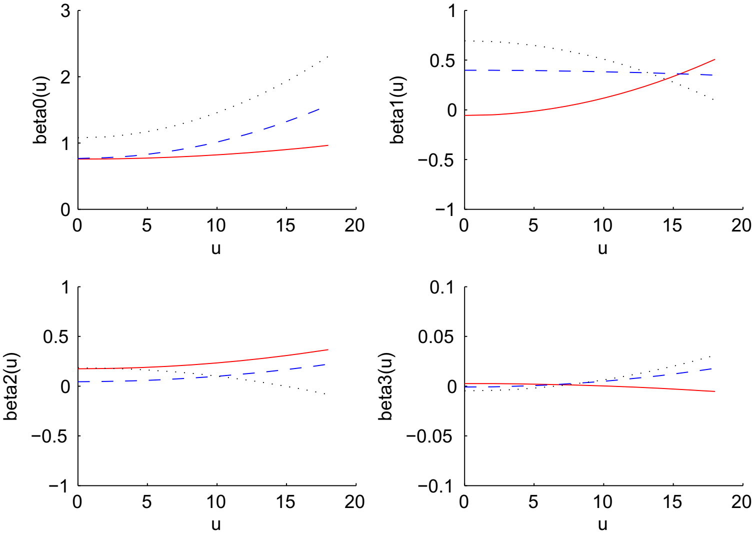

Fig. 1 Plots of the estimated coefficient functions by SVQR with VCs using IRWLS (SVQRVCLS) for three quantiles, $\theta = 0.1$ (*solid line*), $\theta = 0.5$ (*dashed line*) and $\theta = 0.9$ (*dotted line*). *Top left* $\beta_0(u)$ versus u, *top right* $\beta_1(u)$ versus u, *bottom left* $\beta_2(u)$ versus u, and *bottom right* $\beta_3(u)$ versus u

3.2 Real data example

For a real example we consider a subset of the wage data set studied in Wooldridge (2003), which consists of three variables collected regarding each of 526 working individuals for the year 1976. The dependent variable y is the logarithm of wages in dollars per hour. Among major independent variables possibly affecting wages, we use years of education (u), indicator of gender (x_1), marital status (x_2), and years of potential labor force experience (x_3). Two variables, x_1 and x_2, are binary in nature and serve to indicate qualitative features of the individual. We define x_1 to be a binary variable taking on the value one for males and the value zero for females. We also define x_2 to be one if the person is married and zero if the person is not married. For a complete description of all 24 variables, refer to http://fmwww.bc.edu/ec-p/data/wooldridge/wage1.des.

Simple correlation analysis shows that all variables u, x_1, x_2 and x_3 have positive correlation coefficient values with y, which are 0.4311, 0.3737, 0.2707, and 0.1114, respectively. From the coefficients we might interpret that a married man with higher education and longer experience will have a higher chance of getting higher wages.

Another analysis through linear QR for $\theta = 0.1, 0.5$, and 0.9 has been done, and the coefficient estimators are shown in Table 3. From Table 3 we know that marital status and gender are more important factors in predicting wages compared to the education length for the low and median wage group ($\theta = 0.1, 0.5$). For the high wage group ($\theta = 0.9$), the effect of years of education is greater than that of marital status, and gender is still a major factor. We can see that gender has the largest coefficient values of 0.3759 and 0.3399 for $\theta = 0.5$ and 0.9, respectively. The coefficient of x_3 is negligibly small for all θ's. It is even a negative value for $\theta = 0.1$

We now analyze the wage data with the SVQR with VCs using IRWLS only. Figure 1 depicts the estimated coefficient functions for three quantiles, $\theta = 0.1$ (solid line), $\theta = 0.5$ (dashed line), and $\theta = 0.9$ (dotted line), $\beta_0(u)$ vs. u in the top left, $\beta_1(u)$ vs. u in the top right, $\beta_2(u)$ vs. u in the bottom left, and $\beta_3(u)$ vs. u in the bottom right. As seen in Fig. 1, wages increase as the years of education increase for the high and median wage groups and remains almost unchanged for the low wage group. The positive effect of gender on wages for the high wage group is strong for subjects with low education status, but the effect gradually disappears as the years of education increase. On the contrary, gender barely affects wages for subjects with low education status, but it has a strong effect on subjects with high education status in the low wage group. However, the positive effect of gender remains almost unchanged regardless of the years of education for the median wage group.

Figure 1 also shows that the effect of marriage slightly increases for the low and median wage groups as the years of education increase, and it remains almost unchanged for the high wage group. The experience length barely affects wages for subjects with low education status in all wage groups. A slight positive effect of experience length on wages is seen for subjects with high education status in the high and median wage groups. However, experience length does not help to increase wages regardless of education status for the low wage group. Thus, we notice that the smoothing variable, u, and the independent variables have different effects on the different quantiles of the conditional distribution of wages.

According to the linear QR analysis, the coefficients of x_1 are 0.1948, 0.3759, and 0.3399 for $\theta = 0.1, 0.5$ and 0.9, respectively. Figure 1 shows that the order of magnitude of these coefficient values is maintained only in the vicinity of $u = 13$. Also, the strong positive effect of gender on wages for the case of the low wage and high education status has not been revealed by the linear QR. Thus, the SVQR with VCs using IRWLS method reveals what we can not observe through linear QR.

4 Conclusion

In this paper, we considered the estimation of conditional quantiles in VC models by estimating the coefficient functions. We proposed the SVQR with VCs using QP and IRWLS for estimating quantiles. The coefficient functions are estimated by using the kernel trick of SVM. The proposed estimators are easy to compute via standard SVQR algorithms. Through two examples, we observed that the proposed methods derive satisfying results and overall give more accurate and stable estimators than the SVQR and LPQRVC. Thus, our methods appear to be useful in estimating QR function and nonlinear coefficient functions. In particular, SVQR with VCs using IRWLS is preferred since this method makes it possible to construct confidence intervals for coefficient functions and save computing time. The SVQR with VCs using QP and IRWLS also make hyperparameter selection easier and faster than a leave-one-out CV or k-fold CV. Thus, the SVQR with VCs using QP and IRWLS methods can be easily and effectively applied to nonlinear regression coefficients depending on the high-dimensional vector of smoothing variables. We conclude that SVQR with VCs using IRWLS is a promising nonparametric estimation method of QR function.

Acknowledgments The authors wish to thank two anonymous reviewers for their valuable and constructive comments on an earlier version of this article. The research of J. Shim was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology with Grant No. (NRF-2015R1D1A1A01056582). The research of C. Hwang was was supported by the Human Resources Program in Energy Technology of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (No. 20154030200830), and the research of K. Seok was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology with Grant No. (2011-0009705).

References

Breiman L, Friedman JH (1985) Estimating optimal transformations for multiple regression and correlation (with discussion). J Am Stat Assoc 80:580–619

Cai Z, Xu X (2008) Nonparametric quantile estimations for dynamic smooth coefficient models. J Am Stat Assoc 103:1595–1608

Cole T, Green P (1992) Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 11:1305–1319

Craven P, Wahba G (1979) Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. Numer Math 31:377–403

Fan J, Zhang W (2008) Statistical methods with varying coefficient models. Stat Interface 1:179–195

Hastie T, Tibshirani R (1993) Varying-coefficient models. J R Stat Soc B 55:757–796

Heagerty P, Pepe M (1999) Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. J R Stat Soc C 48:533–551

Hendricks W, Koenker R (1992) Hierarchical spline models for conditional quantiles and the demand for electricity. J Am Stat Assoc 87:58–68

Honda T (2004) Quantile regression in varying coefficient models. J Stat Plan Inference 121:113–125

Hoover DR, Rice JA, Wu CO, Yang LP (1998) Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika 85:809–822

Kim MO (2007) Quantile regression with varying coefficients. Ann Stat 35:92–108

Koenker R, Ng P, Portnoy S (1994) Quantile smoothing splines. Biometrika 81:673–680

Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46:33–50

Koenker R, Geling R (2001) Reappraising medfly longevity: a quantile regression survival analysis. J Am Stat Assoc 96:458–468

Koenker R, Hallock KF (2001) Quantile regression. J Econ Perspect 15:143–156

Kuhn H, Tucker A (1951) Nonlinear programming. In: Proceedings of 2nd Berkeley symposium. University of California Press, Berkeley

Li Y, Kiu Y, Zhu J (2007) Quantile regression in reproducing kernel Hilbert space. J Am Stat Assoc 103:255–268

Mercer J (1909) Functions of positive and negative type and their connection with theory of integral equations. Philos Trans R Soc A 209:415–446

Park BU, Mammen E, Lee YK, Lee ER (2015) Varying coefficient regression models: a review and new developments. Int Stat Rev 83:36–64

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Shim J, Kim Y, Lee J, Hwang C (2011) Estimating value at risk with semiparametric support vector quantile regression. Comput Stat 27:685–700
Shim J, Hwang C (2009) Support vector censored quantile regression under random censoring. Comput Stat Data Anal 53:912–919
Shiryaev AN (1996) Probability. Springer, New York
Takeuchi I, Furuhashi T (2004) Non-crossing quantile regressions by SVM. In: Proceedings of 2004 IEEE international joint conference on neural networks, pp 401–406
Vapnik VN (1995) The nature of statistical learning theory. Springer, New York
Wei Y, He X (2006) Conditional growth charts (with discussions). Ann Stat 34:2069–2097
Wooldridge JM (2003) Introductory econometrics. Thompson South-Western, Mason
Ying Z, Jung SH, Wei LJ (1995) Survival analysis with median regression models. J Am Stat Assoc 90:178–184
Yu K, Lu Z, Stander J (2003) Quantile regression: applications and current research area. Statistician 52:331–350
Yuan M (2006) GACV for quantile smoothing splines. Comput Stat Data Anal 5:813–829