Research Reports on Computer Science
https://ojs.wiserpub.com/index.php/RRCS/

UNIVERSAL WISER PUBLISHER

Research Article

A *v*-Support Vector Quantile Regression Model with Automatic Accuracy Control

Pritam Anand[1*], Reshma Rastogi[2], Suresh Chandra[3]

[1]Dhirubhai Ambani Institute of Information and Communication Technology, Gujarat, India
[2]Department of Mathematics and Computer Science, South Asian University, New Delhi, India
[3]Department of Mathematics, Indian Institute of Technology, New Delhi, India
E-mail: ltpritamanand@gmail.com

Received: 4 July 2022; **Revised:** 24 September 2022; **Accepted:** 12 October 2022

Abstract: Quantile regression models have become popular among researchers these days. These models are being used frequently for obtaining the probabilistic forecast in different real-world applications. The Support Vector Quantile Regression (SVQR) model can obtain the conditional quantile estimate using kernel function in a non-parametric framework. The ϵ-SVQR model successfully incorporates the concept of the asymmetric ϵ-insensitive tube in the SVQR model and enables it to obtain a sparse and accurate solution. But, it requires a good choice of the user-defined parameter. A bad choice of ϵ value may result in poor predictions in the ϵ-SVQR model. In this paper, we propose a novel '*v*-Support Vector Quantile Regression' (*v*-SVQR) model for quantile estimation. It can efficiently obtain a suitable asymmetric ϵ-insensitive zone according to the variance present in the data. The proposed *v*-SVQR model uses the *v* fraction of training data points for the estimation of the quantiles. In the *v*-SVQR model, training points asymptotically appear above and below the asymmetric ϵ-insensitive tube in the ratio of $1-\tau$ and τ. Apart from these, there are other interesting properties of the proposed *v*-SVQR model, which we have briefly described in this paper. These properties have been empirically verified using simulated and real-world data sets also.

***Keywords*:** quantile regression, pinball loss function, support vector machine, ϵ-insensitive loss function

1. Introduction

Given the training set $T = \{(x_i, y_i) : x_i \in \mathbb{R}^n, y_i \in \mathbb{R}, i = 1, 2, ..., l\}$ and $\tau \in [0,1]$, the problem of quantile regression is to estimate a real-valued function $f_\tau(x)$ such that a proportion τ of *y/x* will be lying below of the estimate $f_\tau(x)$. For $\tau = 0.5$, the problem is equivalent to median estimation. The estimation of $f_\tau(x)$ is difficult, but more informative than estimation of only mean regression $f(x)$. The estimation of $f_\tau(x)$ for different values of τ can briefly describe the different characteristics of the conditional distribution of *y/x*. In many real-world problems, the estimation of mean regression $f(x)$ is not required or enough, rather they require the estimation of quantile $f_\tau(x)$.

The study of quantile regression problem has initially been started in 1978 by Koenkar and Bassett [1]. Later, it has been briefly discussed and described by Koenker in his book [2]. Koenkar and Bassett [1] proposed the pinball loss function for the estimation of the quantile function $f_\tau(x)$. For a given quantile $\tau \in (0,1)$, the pinball loss function was an asymmetric loss function suitable for quantile estimation. It was given by

DOI: https://doi.org/10.37256/rrcs.1220221662

$$P_\tau(u) = \begin{cases} \tau u & \text{if } u \geq 0, \\ (\tau - 1)u & \text{otherwise.} \end{cases} \quad (1)$$

Support Vector Regression (SVR) models [3-5] are one of the most popular class of regression models which can estimate the mean regression function $f(x)$ efficiently. SVR models commonly solve a Convex Program which guarantees the global optimal solution. These models have been widely used in solving real-world problems of diverse domains.

Takeuchi et al. [6] initiated the study of the quantile regression problem in non-parametric framework on the line of SVR models. They have proposed Support Vector Quantile Regression (SVQR) model in which they have minimized the pinball loss function in the SVR type optimization problem for estimation of the quantile function $f_\tau(x)$. The obtained solution of the SVQR model is not sparse as every training data point is allowed to contribute to the empirical risk which is measured by the asymmetric pinball loss function.

Anand et al. have proposed an asymmetric ϵ-insensitive pinball loss function in their work [7] which extends the concept of ϵ-insensitive zone in the pinball loss function in true sense. The asymmetric ϵ-insensitive pinball loss can obtain a suitable ϵ-insensitive zone of fixed width for every value of τ. The ϵ-insensitive zone was partitioned using the τ value in the asymmetric ϵ-insensitive pinball loss function. Using the asymmetric ϵ-insensitive pinball loss function, they have proposed the ϵ-SVQR model which can obtain better generalization ability than existing SVQR models and successfully brings the sparsity back into the SVQR model.

However, the ϵ-SVQR model [7] requires a good choice of the value of ϵ for obtaining a better prediction of quantiles. A bad choice of ϵ can distort the performance of the ϵ-SVQR model [7].

This paper proposes an efficient SVQR model which appropriately trade-offs the total width of the asymmetric ϵ-insensitive zone in its optimization problem via the user-defined parameter ν. The proposed model has been termed with the ν-Support Vector Quantile Regression (ν-SVQR) model. The ν-SVQR model can adjust the overall width of the asymmetric ϵ-insensitive zone such that at most ν fraction of training data points lie outside of it. This capability of ν-SVQR enables it to automatically adjust the width of the insensitive zone according to the variance present in the data without adjusting any parameter. In the ν-SVQR model, training points asymptotically appear above and below the asymmetric ϵ-insensitive tube in the ratio of $1 - \tau$ and τ. Further, there are other interesting asymptotic properties of the ν-SVQR model which we have briefly described in this paper. Several experiments on simulated as well as UCI data sets have been performed to empirically verify claims made in this paper.

We now describe the notations used in the rest of this paper. We have considered all vectors as column vectors unless we have specified them. For a given vector $x \in \mathbb{R}^n$, $\| x \|$ denotes the L_2 norm of x. We have considered the training set $T = \{(x_i, y_i) : x_i \in \mathbb{R}^n, y_i \in \mathbb{R}, i = 1, 2, ..., l\}$ for the quantile estimation in this paper. For a given quantile $\tau \in (0,1)$, the quantile estimate has been obtained using $f_\tau(x) = w^T \phi(x) + b$. Here $\phi : \mathbb{R}^n \to \mathcal{H}$ is a mapping from the input space to a higher dimensional feature space $\mathcal{H}$ such that for every pair of data points x_i and x_j, the $\phi(x_i)^T \phi(x_j)$ can be obtained using a positive semidefinite kernel $K(x_i, x_j)$ [8]. ξ_i and ξ_i^*, for i = 1, 2, …, l are slack variables and have been used for measuring the empirical risk. $\alpha_i, \beta_i, \gamma_i$ and η_i are Lagrangian multipliers. C and ν are user-defined positive parameters.

The rest of this paper is organized as follows. In Section 2, we have briefly reviewed the variants of quantile regression models available in the literature. Section 3 briefly describes the standard SVQR model [6] and ϵ-SVQR model [7]. In Section 4, we present our proposed ν-SVQR model. In Section 5, we have discovered different properties of our ν-SVQR model. Section 6 contains the numerical results obtained by different nature of experiments carried out on simulated as well as real-world data sets to empirically verify the properties of the proposed ν-SVQR model. Section 7 concludes this paper.

2. Related work

In this section, we shall briefly review different relevant quantile regression models. Koenkar and Bassett [1] have first introduced the notion of quantile regression. Thereafter, it has been briefly statistically studied and extended for several important real-world applications. Some of important effective variants of quantile regression models are

quantile smoothing splines [9], quantile autoregression [10], copula-based non-linear quantile autoregression [11] and interior point algorithm for quantile estimation [12].

Takeuchi et al. [6] have started the study of the SVQR model for obtaining the kernel quantile estimate in a non-parametric framework. Thereafter, researchers have attempted to extend the SVQR model on the line of the ϵ-SVR model for increasing its generalization ability as well as obtaining sparse solutions. For this, they have attempted to propose the ϵ-insensitive pinball loss functions to incorporate the concept of ϵ-insensitive zone in the asymmetric pinball loss function.

At first, Takeuchi and Furuhashi considered the ϵ-insensitive pinball loss function for estimation of the non-crossing quantile in their work [13]. Further, Hu et al. had also considered a similar kind of ϵ-insensitive pinball loss function in their work [14] for estimation of quantiles. However, the ϵ-insensitive zone in these pinball loss functions was symmetric. The use of the symmetric ϵ-insensitive zone in the asymmetric pinball loss function failed to perform well for the estimation of quantiles.

Soek et al. have first considered the asymmetric ϵ-insensitive zone in the pinball loss function in their proposed e-sensitive pinball loss function [15]. Later on, Park and Kim also proposed a similar kind of loss function in their work [16]. But the problem with these pinball loss functions was that they failed to provide a suitable ϵ-insensitive zone for every value of τ. Anand et al. have proposed an effective asymmetric ϵ-insensitive pinball loss function and extended the SVQR model on the line of the ϵ-SVR model efficiently [7].

Meinshausen has proposed quantile regression forest [17] by extending the idea of random forest regression in quantile estimation. Thereafter the idea of quantile regression forest was efficiently used in load forecasting [18, 19], heat waves [20], and digital soil mapping products [21].

In recent years, the quantile regression neural networks [22, 23] have gained popularity among researchers. Researchers have used them along with modern deep learning architectures for obtaining efficient probabilistic load forecasting [24-28], wind forecasting [29-31], and photovoltaic power forecasting [32, 33].

3. SVQR models

For the training set $T = \{(x_i, y_i) : x_i \in \mathbb{R}^n, y_i \in \mathbb{R}, i = 1, 2, ..., l\}$ and the quantile $\tau \in (0,1)$, the SVQR model estimates the function $f_\tau(x) = w^T \phi(x) + b$ in the feature space for the estimation of the τth quantile, where $\phi : \mathbb{R}^n \to \mathcal{H}$ is a mapping from the input space to a higher dimensional feature space $\mathcal{H}$.

3.1 *Standard SVQR model*

The standard SVQR model minimizes

$$\min_{w,b} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{l} L_\tau (y_i - (w^T x_i + b)), \tag{2}$$

where $L_\tau(v)$ is the asymmetric pinball loss function which is given by

$$L_\tau(v) = \begin{cases} \tau v & \text{if } v \geq 0, \\ (\tau - 1)v & \text{otherwise.} \end{cases} \tag{3}$$

Using the l-dimensional variables $\xi = (\xi_1, \xi_2, ..., \xi_l)$ and $\xi^* = (\xi_1^*, \xi_2^*, ..., \xi_l^*)$, the optimization problem (2) can be equivalently converted to the following Quadratic Programming Problem (QPP)

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\| w \|^2 + C\sum_{i=1}^{l}(\tau\xi_i + (1-\tau)\xi_i^*)$$

subject to,

$$y_i - (w^T\phi(x_i) + b) \le \xi_i,$$
$$(w^T\phi(x_i) + b) - y_i \le \xi_i^*,$$
$$\xi_i \ge 0, \quad \xi_i^* \ge 0, \quad i = 1, 2, ..., l. \tag{4}$$

Here $C \ge 0$ is a user-defined parameter that is used to find a good trade-off between empirical risk and model complexity of the estimator. The QPP (4) of the standard SVQR model can be easily solved by solving its corresponding Wolfe dual problem. More detail about the standard SVQR model can be found in Takeuchi et al. [6].

3.2 *ϵ-SVQR model*

Anand et al. [7] have proposed an asymmetric ϵ-insensitive pinball loss function that can obtain a suitable asymmetric ϵ-insensitive zone for every value of τ. The asymmetric ϵ-insensitive pinball loss function is given by

$$L_\tau^\epsilon(u) = \max(-(1-\tau)(u+\tau\epsilon), 0, \tau(u-(1-\tau)\epsilon)). \tag{5}$$

It can be better understood in the following form

$$L_\tau^\epsilon(y_i, x_i, w, b) = \begin{cases} -(1-\tau)(y_i - (w^T x_i + b) + \tau\epsilon), & \text{if } y_i - (w^T x_i + b) < -\tau\epsilon. \\ 0, & \text{if } -\tau\epsilon < y_i - (w^T x_i + b) \le (1-\tau)\epsilon. \\ \tau(y_i - (w^T x_i + b) - (1-\tau)\epsilon), & \text{if } y_i - (w^T x_i + b) > (1-\tau)\epsilon. \end{cases} \tag{6}$$

Figure 1 shows that the asymmetric ϵ-insensitive pinball loss function can generate the suitable asymmetric ϵ-insensitive zone for different values of τ.

The ϵ-SVQR model minimizes

$$\min_{(w,b)} \frac{1}{2}\| w \|^2 + C\cdot\sum_{i=1}^{l} L_\tau^\epsilon(y_i, x_i, w, b)$$
$$= \min_{(w,b)} \frac{1}{2}\| w \|^2 + C\cdot\sum_{i=1}^{l} \max(-(1-\tau)(y_i - (w^T x_i + b) + \tau\epsilon), 0, \tau(y_i - (w^T x_i + b) - (1-\tau)\epsilon)) \tag{7}$$

which can be equivalently converted to the following QPP

$$\min_{(w,b,\xi,\xi^*)} \frac{1}{2}\| w \|^2 + C\cdot\sum_{i=1}^{l}(\tau\xi_i + (1-\tau)\xi_i^*)$$

subject to,

$$y_i - (w^T\phi(x_i) + b) \le (1-\tau)\epsilon + \xi_i,$$
$$(w^T\phi(x_i) + b) - y_i \le \tau\epsilon + \xi_i^*,$$
$$\xi_i \ge 0, \quad \xi_i^* \ge 0, \quad i = 1, 2, ..., l. \tag{8}$$

In the ϵ-SVQR model, $\epsilon \ge 0$ is the user-defined parameter and a good value of ϵ is required beforehand for the efficient estimate of quantiles. For the solution of the ϵ-SVQR primal problem (8), we obtain its corresponding Wolfe dual problem as follows

$$\min_{\alpha,\beta} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i-\beta_j)K(x_i,x_j)(\alpha_j-\beta_i)-\sum_{i=1}^{l}(\alpha_i-\beta_i)y_i+\sum_{i=1}^{l}((1-\tau)\epsilon\,\alpha_i+\tau\epsilon\,\beta_i)$$

subject to,

$$\sum_{i=1}^{l}(\alpha_i-\beta_i)=0,$$
$$0\le\alpha_i\le C\tau,\quad i=1,2,...,l,$$
$$0\le\beta_i\le C(1-\tau),\quad i=1,2,...,l. \tag{9}$$

After obtaining the solution of the dual problem (9), we can estimate $f_\tau(x)$, for any test data point $x\in\mathbb{R}^n$ using

$$f_\tau(x)=\sum_{i=1}^{l}(\alpha_i-\beta_i)K(x,x_i)+b. \tag{10}$$

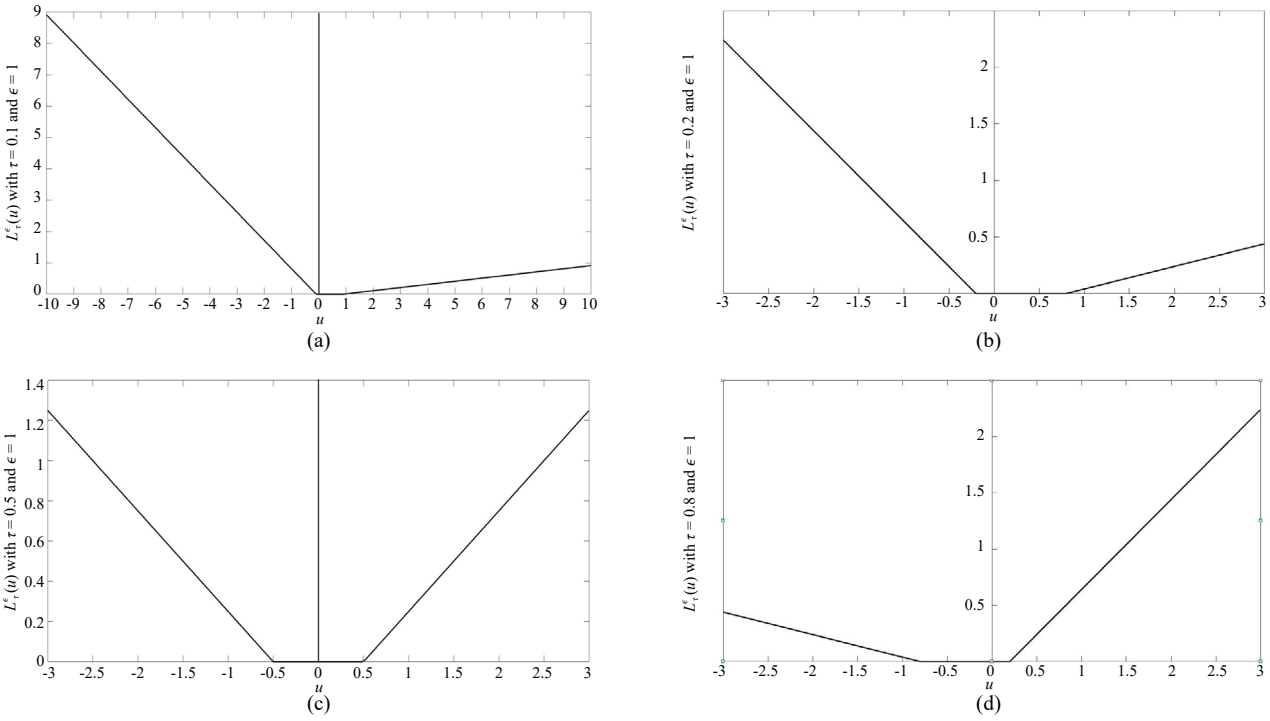


Figure 1. The asymmetric ϵ-pinball loss function for (a) $\tau = 0.1$ (b) $\tau = 0.2$ (c) $\tau = 0.5$ and (d) $\tau = 0.8$ with fixed $\epsilon = 1$

4. A *v*-SVQR model with automatic accuracy control

We can observe from the numerical results obtained from experiments with the proposed ϵ-SVQR model that it requires a good choice of the value of ϵ to be supplied beforehand into it. A bad choice of the value of ϵ value can distort its prediction.

Taking motivation from this, we propose an efficient SVQR model that appropriately trade-offs the total width of the asymmetric ϵ-insensitive zone in its optimization problem via the user-defined parameter v. The proposed model has been termed with v-SVQR model. The v-SVQR model can adjust the overall width of the asymmetric ϵ-insensitive zone such that at most v fraction of training data points lie outside of it. This capability of v-SVQR enables it to automatically adjust the width of the ϵ-insensitive zone according to the variance present in the data without adjusting any parameter.

In the v-SVQR model, training points asymptotically appear above and below the asymmetric ϵ-insensitive tube in the ratio of $1-\tau$ and τ. There are also other interesting asymptotic properties of v-SVQR model which make it more efficient and handful than ϵ-SVQR model.

4.1 *v-SVQR model*

The proposed v-SVQR model minimizes

$$\min_{(w,b,\epsilon,\xi,\xi^*)} \frac{1}{2}\|w\|^2 + C(v\tau(1-\tau)\epsilon + \frac{1}{l}\sum_{i=1}^{l}(\tau\xi_i + (1-\tau)\xi_i^*))$$

subject to,

$$\begin{aligned} &y_i - (w^T\phi(x_i)+b) \le (1-\tau)\epsilon + \xi_i, \\ &(w^T\phi(x_i)+b) - y_i \le \tau\epsilon + \xi_i^*, \\ &\xi_i \ge 0, \quad \xi_i^* \ge 0, \quad i = 1, 2, ..., l, \end{aligned} \tag{11}$$

where $C \ge 0$ and $v \ge 0$ are user-defined parameters.

For solving the primal problem (11) efficiently, we need to derive its Wolfe dual problem. The Lagrangian function for the primal problem (11) is obtained as

$$\begin{aligned} &L(w, b, \epsilon, \xi_i, \xi_i^*, \alpha_i, \beta_i, \gamma_i, \lambda_i, \eta) \\ &= \frac{1}{2}\|w\|^2 + C(v\tau(1-\tau)\epsilon + \frac{1}{l}\sum_{i=1}^{l}(\tau\xi_i + (1-\tau)\xi_i^*)) + \sum_{i=1}^{l}\alpha_i(y_i - (w^T\phi(x_i)+b) - (1-\tau)\epsilon - \xi_i) \\ &+ \sum_{i=1}^{l}\beta_i((w^T\phi(x_i)+b) - y_i - \tau\epsilon - \xi_i^*) - \sum_{i=1}^{l}\gamma_i\xi_i - \sum_{i=1}^{l}\lambda_i\xi_i^* - \eta\epsilon. \end{aligned} \tag{12}$$

We can now note the Karush-Kuhn-Tucker (KKT) conditions for (11) as follows

$$\frac{\partial L}{\partial w} = w + \sum_{i=1}^{l}(\beta_i - \alpha_i)\phi(x_i) = 0 \Rightarrow w = \sum_{i=1}^{l}(\alpha_i - \beta_i)\phi(x_i) \tag{13}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{l}(\beta_i - \alpha_i) = 0. \tag{14}$$

$$\frac{\partial L}{\partial \xi_i} = \frac{C}{l}\tau - \alpha_i - \gamma_i = 0, \quad i = 1, 2, ..., l. \tag{15}$$

$$\frac{\partial L}{\partial \xi_i} = \frac{C}{l}(1-\tau) - \beta_i - \lambda_i = 0, \quad i = 1, 2, ..., l. \tag{16}$$

$$\frac{\partial L}{\partial \epsilon} = Cv\tau(1-\tau) - (1-\tau)\sum_{i=1}^{l}\alpha_i - \tau\sum_{i=1}^{l}\beta_i - \eta = 0. \tag{17}$$

$$\alpha_i(y_i - (w^T\phi(x_i)+b) - (1-\tau)\epsilon - \xi_i) = 0, \quad i = 1, 2, ..., l. \tag{18}$$

$$\beta_i((w^T\phi(x_i)+b) - y_i - \tau\epsilon - \xi_i^*) = 0, \quad i = 1, 2, ..., l. \tag{19}$$

$$\gamma_i\xi_i = 0, \quad \lambda_i\xi_i^* = 0, \quad i = 1, 2, ..., l. \tag{20}$$

$$\eta\epsilon = 0. \tag{21}$$

$$y_i - (w^T\phi(x_i)+b) \le (1-\tau)\epsilon + \xi_i, \quad i = 1, 2, ..., l. \tag{22}$$

$$(w^T\phi(x_i)+b) - y_i \le \tau\epsilon + \xi_i^*, \quad i = 1, 2, ..., l. \tag{23}$$

$$\epsilon, \xi_i, \xi_i^*, \alpha_i, \beta_i \ge 0, \quad i = 1, 2, ..., l. \tag{24}$$

Making use of the above KKT conditions (13-24), the Wolfe dual problem of the primal problem (11) can be obtained as follows

$$\min_{(\alpha,\beta)} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i-\beta_j)K(x_i,x_j)(\alpha_j-\beta_i)-\sum_{i=1}^{l}(\alpha_i-\beta_i)y_i$$

subject to,

$$\sum_{i=1}^{l}(\alpha_i-\beta_i)=0,$$

$$(1-\tau)\sum_{i=1}^{l}\alpha_i+\tau\sum_{i=1}^{l}\beta_i \le Cv\tau(1-\tau),$$

$$0\le\alpha_i\le\frac{C}{l}\tau, \quad i=1,2,...,l,$$

$$0\le\beta_i\le\frac{C}{l}(1-\tau), \quad i=1,2,...,l. \tag{25}$$

5. Properties of *v*-SVQR model

The KKT conditions (13-24) will help us to discover the various properties of the proposed *v*-SVQR model. We list the properties of the proposed *v*-SVQR model below.

Proposition 5.1: If the *v*-SVQR model is applied to the training set $T=\{(x_i,y_i): x_i\in\mathbb{R}^n, y_i\in\mathbb{R}, i=1,2,...,l\}$, which results $\epsilon>0$, then $\alpha_i\beta_i=0$ and $\xi_i\xi_i^*=0, \forall i=1,2,...,l$.

Proof: If possible, let us suppose there exists an index *i* such that $\alpha_i\beta_i\neq 0$. It implies that $\alpha_i\neq 0$ and $\beta_i\neq 0$. Therefore, from the KKT condition (18) and (19) we can obtain

$$(y_i-(w^T\phi(x_i)+b)-(1-\tau)\epsilon-\xi_i)=0 \tag{26}$$

$$\text{and } ((w^T\phi(x_i)+b)-y_i-\tau\epsilon-\xi_i^*)=0. \tag{27}$$

Adding equations (26) and (27) gives $\xi_i^*+\xi_i=-\epsilon$ which is possible only when either $\xi_i<0$ or $\xi_i^*<0$. But, the KKT condition (24) requires $\xi_i\ge 0,\ \xi_i^*\ge 0,$ for *i* = 1, 2, ..., *l*, which contradicts our assumption. This proves $\alpha_i\beta_i=0\ \forall i=1,2,...,l$.

On the similar line, let us suppose that there exists an index *i* for which $\xi_i\xi_i^*\neq 0$. It means that $\xi_i\neq 0$ and $\xi_i^*\neq 0$ for which we can obtain $\gamma_i=0$ and $\lambda_i=0$ from KKT condition (20). For $\gamma_i=0$ and $\lambda_i=0,$ we will obtain $\alpha_i=\frac{c}{l}\tau$ and $\beta_i=\frac{c}{l}(1-\tau)$ from the KKT conditions (15) and (16) respectively, which is not possible as we have already proven that $\alpha_i\beta_i=0\ \forall i=1,2,...,l.$ This proves $\xi_i\xi_i^*=0\ \forall i=1,2,...,l.$

Proposition 5.2: For all those data points $(x_i,y_i),$ which lie inside or boundary of the asymmetric ϵ-insensitive tube, the corresponding ξ_i and ξ_i^* will take zero value.

Proof: The data point $(x_i,y_i),$ lying inside or boundary of the asymmetric ϵ-insensitive tube must satisfy

$$y_i-(w^T\phi(x_i)+b)-(1-\tau)\epsilon\le 0 \tag{28}$$

$$\text{and } (w^T\phi(x_i)+b)-y_i-\tau\epsilon\le 0. \tag{29}$$

If possible, let us suppose that $\xi_i\neq 0$ which means that $\xi_i>0$ (as the KKT condition (24) requires $\xi\ge 0$). Since we can obtain $\xi_i>0,$ we can obtain $\gamma_i=0$ and further $\alpha_i=0$ by using the KKT conditions (20) and (15) respectively. For $\alpha_i>0,$ the KKT condition (18) implies that

$$y_i - (w^T \phi(x_i) + b) - (1-\tau)\epsilon = \xi_i, \tag{30}$$

which is not possible as $\xi_i > 0$.

On a similar line, we can show that ξ_i^* also cannot take non-zero values.

It is also easy to prove that the data point, which lies outside of the asymmetric ϵ-insensitive tube, the corresponding ξ_i or ξ_i^* will take a positive value.

Proposition 5.3: For the data point (x_i, y_i), which lie inside of the asymmetric ϵ-insensitive tube, the corresponding α_i and β_i will take zero value.

Proof: The data point (x_i, y_i) lying inside of the ϵ-tube, the ξ_i and $\xi_i^* = 0$. Therefore, we can obtain the following for all data points (x_i, y_i) lying inside of the ϵ-tube

$$(y_i - (w^T \phi(x_i) + b) - (1-\tau)\epsilon - \xi_i) < 0 \tag{31}$$

$$\text{and } (w^T \phi(x_i) + b) - y_i - \tau\epsilon - \xi_i^* < 0. \tag{32}$$

Further, the use of the KKT condition (18) and (19) will let us obtain α_i and $\beta_i = 0$.

Proposition 5.4: For the data point (x_i, y_i), lying above of the ϵ-tube, $\alpha_i = \frac{C}{l}\tau$ and $\beta_i = 0$. For the data point (x_i, y_i), lying below of the ϵ-tube, $\alpha_i = 0$ and $\beta_i = \frac{C}{l}(1-\tau)$.

Proof: The data point (x_i, y_i) lying above of the ϵ-tube will hold

$$y_i - (w^T \phi(x_i) + b) - (1-\tau)\epsilon > 0, \tag{33}$$

for which the corresponding ξ_i will take a positive value for satisfying the KKT condition (22). For $\xi_i > 0$, we can get $\gamma_i = 0$ from (20) and further can obtain $\alpha_i = \frac{C}{l}\tau$ from the KKT condition (15). Further β_i will take zero value as $\alpha_i \beta_i = 0$.

On the similar line, we can prove that the data point (x_i, y_i), lying below of the ϵ-tube, $\alpha_i = 0$ and $\beta_i = \frac{C}{l}(1-\tau)$.

Proposition 5.5: For $0 < \alpha_i < \frac{C}{l}\tau$ $(0 < \beta_i < \frac{C}{l}(1-\tau))$, the corresponding data point (x_i, y_i) will be lying on the upper (lower) boundary of the asymmetric ϵ-insensitive tube.

Proof: For $0 < \alpha_i < \frac{C}{l}\tau$, the $\gamma_i > 0$ from KKT condition (15) which implies $\xi_i = 0$. Further, for $\alpha_i > 0$, we can obtain

$$y_i - (w^T \phi(x_i) + b) = (1-\tau)\epsilon, \tag{34}$$

which means that data point (x_i, y_i), will be lying on the upper boundary of the asymmetric ϵ-insensitive tube. On the similar line, we can obtain that for $0 < \beta_i < \frac{C}{l}(1-\tau)$, the corresponding data point (x_i, y_i) will be lying on the below boundary of the asymmetric ϵ-insensitive tube.

Remark 5.1: For $\epsilon > 0$, the η will take zero value and the inequalities constraint $(1-\tau)\sum_{i=1}^{l} \alpha_i + \sum_{i=1}^{l} \beta_i \leq Cv\tau(1-\tau)$ of the dual problem (25) will get converted to the equality constraint

$$(1-\tau)\sum_{i=1}^{l} \alpha_i + \tau \sum_{i=1}^{l} \beta_i = Cv\tau(1-\tau). \tag{35}$$

Now, we shall term the data points which are lying outside of the asymmetric ϵ-tube with 'errors'. The data points which are lying outside of the asymmetric ϵ-tube as well as the boundary of the tube are termed with 'support vectors'. These data points only contribute to the construction of the final regressor.

Proposition 5.6: Suppose the *v*-SVQR is applied to some data set and the resulting ϵ is non-zero then the following statements hold.

(a) v is the upper bound on the fraction of errors.

(b) v is the lower bound on the fraction of support vectors.

Proof: For the data point, lying above the asymmetric ϵ-tube, only α_i will take the value $\frac{C}{l}\tau$. For the data point, lying below the asymmetric ϵ-tube, only β_i will take the value $\frac{C}{l}(1-\tau)$. The data point (x_i, y_i) lying on the upper (lower) boundary of the asymmetric ϵ-insensitive tube will be taking $0 < \alpha_i < \frac{C}{l}\tau$ $(0 < \beta_i < \frac{C}{l}(1-\tau))$ values. Let us suppose that there are m_1 and m_2 data points which are lying above and below the asymmetric ϵ-tube respectively. Further, there are m_3 and m_4 data points which are lying on the upper and lower boundary of the asymmetric ϵ-insensitive tube.

For $\epsilon > 0$, we can obtain from using (17),

$$m_1 \frac{C}{l}(1-\tau)\tau + m_2 \frac{C}{l}(1-\tau)\tau \le Cv\tau(1-\tau)$$

which implies that $\frac{1}{l}(m_1 + m_2) \le v$.

Furthermore, there should exist at least m_1 and m_2 data points lying above and below the asymmetric ϵ-tube which would satisfy the equality (35). For these data points, we have

$$m_1 \frac{C}{l}(1-\tau)\tau + m_2 \frac{C}{l}(1-\tau)\tau = Cv\tau(1-\tau)$$

which implies that $\frac{1}{l}(m_1 + m_2) = v$. It further means that there should at least exist *v* fraction of the support vectors.

Remark 5.2: Asymptotically, the *v* equals the fraction of support vectors and errors. The probability of the data point lying on the boundary of the asymmetric ϵ-tube becomes zero asymptotically. This statement can be proved under a certain condition similar to the proof of Proposition 1 (iii) given in Scholkopf [34]. However here, we shall empirically verify that the *v* equals the fraction of support vectors and errors asymptotically.

Remark 5.3: Asymptotically, the data points appear above and below the asymmetric ϵ-tube in the ratio of $1 - \tau$ and τ respectively in the *v*-SVQR model. It means that for the large value of l, there would be $lv(1 - \tau)$ and $lv\tau$ data points lying above and below the asymmetric ϵ-tube respectively. It is because of the fact that α_i and β_i also have to satisfy the KKT condition (14).

Remark 5.4: If the proposed *v*-SVQR obtains the solution $(\overline{w}, \overline{b}, \overline{\epsilon})$ with parameter value C', then ϵ-SVQR model with parameters $\epsilon = \overline{\epsilon}$ and $C = C' * N$ will obtain the same solution $\overline{w}, \overline{b}$.

Obtaining the value of ϵ and *b:* At first, we can obtain the value of the ϵ which is the effective width of the asymmetric tube. For this, we find out the data points which are lying on the upper and lower boundary of the ϵ-tube using Proposition 5.5. For $0 < \alpha_i < \frac{C}{l}\tau$, we can obtain the upper width of the asymmetric tube using $y_i - (w^T \phi(x_i) + b)$. For $0 < \beta_j < \frac{C}{l}(1-\tau)$, we can obtain the lower width of the asymmetric tube using $(w^T \phi(x_j) + b) - y_j$. But, the computation of the final width of the asymmetric ϵ-tube does not require the value of *b* and can be obtained by

$$\epsilon = (y_i - (\sum_{k=1}^{l} (\alpha_k - \beta_k) K(x_k, x_i))) + (\sum_{k=1}^{l} (\alpha_k - \beta_k) K(x_k, x_j)) - y_j), \tag{36}$$

where $0 < \alpha_i < \frac{C}{l}\tau$ and $0 < \beta_j < \frac{C}{l}(1-\tau)$.

After obtaining the value of ϵ, we can obtain the value of *b*. For $0 < \alpha_i < \frac{C}{l}\tau$, we can obtain

$$b = y_i - \sum_{k=1}^{l} (\alpha_k - \beta_k) K(x_k, x_i) - (1-\tau)\epsilon. \tag{37}$$

For $0 < \beta_j < \frac{C}{l}(1-\tau)$, we can also obtain

$$b = y_j - \sum_{k=1}^{l}(\alpha_k - \beta_k)K(x_k, x_j) + \tau\epsilon. \tag{38}$$

In practice, we compute values of b from equations (37) and (38) and use their average value as the final value of b. After computing the values of decision variables α, β and b the quantile regression is estimated by

$$f_\tau(x) = \sum_{i=1}^{l}(\alpha_i - \beta_i)K(x, x_i) + b. \tag{39}$$

Further, like the ϵ-SVR model described in Gunn [5], if the kernel contains a bias term then, the v-SVQR dual problem (25) can be solved without equality constraint and the quantile regression function is simply estimated by

$$f_\tau(x) = \sum_{i=1}^{l}(\alpha_i - \beta_i)K(x, x_i). \tag{40}$$

6. Experimental section

In this section, we shall empirically verify the claims made in this paper. For this, we first describe our experimental setup. We have performed all experiments with MATLAB 17.0 environment (http://in.mathworks.com/) on an Intel i7 processor with 8.0 GB of RAM. The QPPs of the proposed v-SVQR and ϵ-SVQR have been solved by the quadprog function with the interior-point convex algorithm available in the MATLAB 16.0 environment. For all of the experiments, we have used the radial basis function (RBF) kernel $\exp(\frac{-\|x-y\|^2}{q})$, where q is the kernel parameter and quantile regression function is estimated by (39). The proposed v-SVQR model requires three parameters to be tuned namely RBF kernel parameters q, C and v whereas the ϵ-SVQR model requires the tuning of parameters q, C and ϵ. All these parameters have been tuned using the exhaustive search method (Hsu and Lin, [35]). The parameter q and C has been searched in the set $\{2^i : i = -15, -9, ...9, 15\}$.

6.1 *Performance criteria*

For the evaluation of the efficacy of SVQR models, we have used some evaluation criteria which is also mentioned in Xu et al. [36]. Given the training set $T = \{(x_i, y_i) : x_i \in \mathbb{R}^n, y_i \in \mathbb{R}, i = 1, 2, ..., l\}$ and true τth conditional quantile function $Q_\tau(y/x)$, we list the evaluation criteria as follows.

(i) RMSE: It is Root Mean Square of Error.
It is given by $\sqrt{\frac{1}{l}\sum_{i=1}^{l}(Q_\tau(y_i/x_i) - f_\tau(x_i))^2}$.

(ii) MAE: It is Mean of the Absolute Error.
It is given by $\frac{1}{l}\sum_{i=1}^{l}\left|(Q_\tau(y_i/x_i) - f_\tau(x_i))\right|$.

(iii) Error E_τ: It is the measure which is used when the true quantile function is unknown. It is given by $E_\tau = \left|p_\tau - \tau\right|$, where $p_\tau = P(y_i \le f_\tau(x_i))$ is the coverage probability. For the real-world UCI data sets experiments, we would be using this measure. We shall compute the coverage probability p_τ by obtaining the estimated τ value in 100 random trials.

(iv) Sparsity (u) $= \frac{\#(u=0)}{\#(u)}$, where $\#(r)$ determines the number of the component of the vector r.

Simulated data sets

We shall show different properties of the proposed v-SVQR model and its advantages over the ϵ-SVQR model empirically. The best way to do this is to generate data sets as actual true quantiles can be easily computed for these data sets and unbiased comparisons can be made. We have generated the training set T where x_i is drawn from the univariate uniform distribution with [−4, 4]. The response variable y_i is obtained from polluting a nonlinear function of x_i with different natures of noises in simulated data sets as follows.

$$\text{AD1: } y_i = (1 - x_i + 2x_i^2)e^{-0.5x_i^2} + \xi_i, \text{ where } \xi_i \text{ is from N}(0, \sigma).$$

$$\text{AD2: } y_i = (1 - x_i + 2x_i^2)e^{-0.5x_i^2} + \xi_i, \text{ where } \xi_i \text{ is from U(a, b).}$$

The true quantile function $Q_\tau(y/x)$ in these simulated data sets can be obtained as

$$y_i = (1 - x_i + 2x_i^2)e^{-0.5x_i^2} + F_\tau^{-1}(\xi_i),$$

where $F_\tau^{-1}(\xi_i)$ is the τth quantile of random error ξ_i. We have evaluated the SVQR models by generating 1,000 testing points in each trial.

Experiment 1

Our first experiment will empirically verify the Proposition 5.6 of this paper. We shall verify that the user-defined positive parameter v in our v-SVQR model is the upper bound on the fraction of errors and the lower bound on the fraction of support vectors.

For this, we have generated 200 training data points of AD1 simulated data set and obtained the numerical results for 10 random simulations. Table 1 shows the performance of the proposed v-SVQR model on several values of v for τ = 0.2, 0.5, 0.7, and 0.8. Figure 2 shows the proposed v-SVQR model on several values of v for τ = 0.1, 0.3, 0.6 and 0.9. The following inferences can be easily drawn from numerical results listed in Table 1 and plots in Figure 2.

(a) Irrespective of τ values, as the value v increases, the total width ϵ of asymmetric ϵ-insensitive zone decreases.
(b) Irrespective of τ values, v is the upper bound on the fraction of errors.
(c) Irrespective of τ values, v is the lower bound on the fraction of support vectors.
(d) RMSE and MAE obtained by the proposed ϵ-SVQR model also vary with the parameter v.

Experiment 2

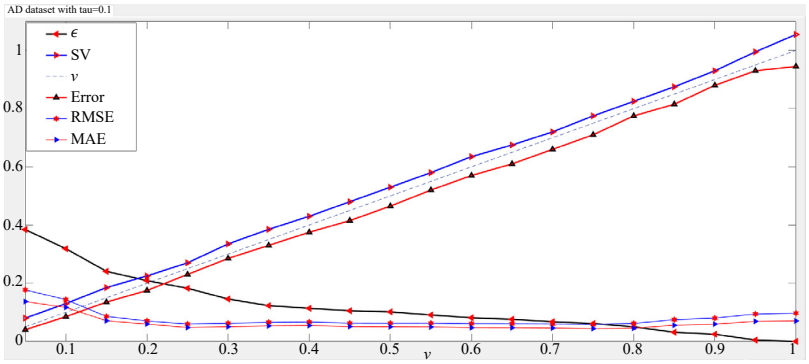
The objective of this experiment is to study the asymptotic behavior of the proposed v-SVQR model. We shall also verify the claims made in Remarks 5.2 and 5.3. We perform this experiment with the varying number of training points of the AD1 data set and fix the v = 0.8 in the proposed v-SVQR model.

Table 2 results the numerical results obtained by the proposed v-SVQR model on AD1 data set with different sizes of training set. In this table, 'ratio' is the ratio of training data points lying above and below the asymmetric ϵ-insensitive tube. The following facts can be easily observed from the numerical results listed in Table 2.

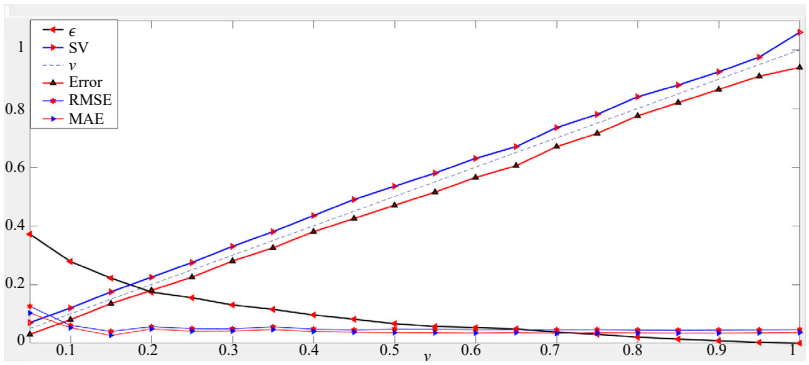
(a) Irrespective of τ values, the fraction of support vectors and errors converge to the v value in the proposed v-SVQR model. It has also been well illustrated by the plot in Figures 3(a) and 3(b). It is only because of the fact that the probability of a training data point lying on boundaries of asymmetric ϵ-insensitive tube vanishes, as the number of training point increases.
(b) Irrespective of τ values, the ratio of training data point lying above and below of the asymmetric ϵ-tube converges to $\frac{1-\tau}{\tau}$. It has also been well illustrated by the plot in Figure 3(c) and 3(d). It means that the asymmetric ϵ-insensitive zone used in the proposed v-SVQR model is very suitable for handling quantile estimation problem.
(c) The resulting overall width of ϵ-insensitive zone converges to a constant value in the proposed v-SVQR model.
(d) As the number of training points increases, there are more information available to the proposed v-SVQR model. It results in decrease in RMSE values obtained by the proposed v-SVQR model.

Table 1. Performance of the proposed v-SVQR model on AD1 data set for different τ values

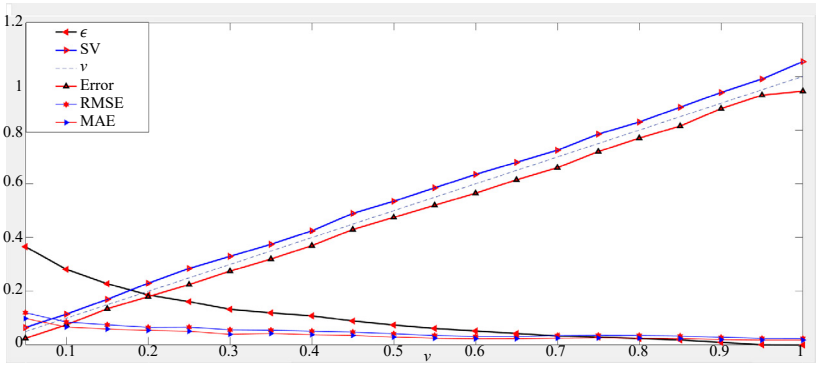
τ																					
0.2	v	0.050	0.100	0.150	0.200	0.250	0.300	0.350	0.400	0.450	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	1.000
	ϵ	0.384	0.288	0.225	0.190	0.156	0.143	0.120	0.108	0.091	0.076	0.068	0.055	0.047	0.043	0.037	0.027	0.019	0.013	0.003	0.000
	SV	0.080	0.125	0.175	0.230	0.290	0.335	0.380	0.425	0.480	0.525	0.580	0.630	0.685	0.725	0.770	0.840	0.885	0.925	0.980	1.060
	Error	0.030	0.075	0.130	0.185	0.230	0.275	0.330	0.370	0.425	0.470	0.520	0.570	0.620	0.660	0.710	0.775	0.820	0.860	0.920	0.940
	RMSE	0.160	0.072	0.050	0.038	0.044	0.052	0.063	0.063	0.060	0.058	0.059	0.058	0.060	0.061	0.063	0.064	0.065	0.066	0.067	0.067
	MAE	0.127	0.059	0.037	0.032	0.037	0.044	0.051	0.053	0.048	0.043	0.046	0.045	0.046	0.047	0.047	0.047	0.050	0.050	0.051	0.050
0.5	v	0.050	0.100	0.150	0.200	0.250	0.300	0.350	0.400	0.450	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	1.000
	ϵ	0.366	0.293	0.220	0.174	0.151	0.132	0.119	0.103	0.084	0.072	0.058	0.048	0.042	0.032	0.027	0.020	0.017	0.011	0.002	0.000
	SV	0.065	0.110	0.175	0.230	0.275	0.325	0.375	0.435	0.490	0.535	0.580	0.625	0.680	0.735	0.775	0.840	0.890	0.940	0.990	1.055
	Error	0.030	0.070	0.135	0.185	0.220	0.270	0.320	0.375	0.430	0.470	0.525	0.565	0.620	0.675	0.710	0.780	0.820	0.885	0.925	0.945
	RMSE	0.104	0.079	0.058	0.060	0.049	0.042	0.045	0.044	0.041	0.038	0.032	0.033	0.033	0.029	0.025	0.025	0.024	0.025	0.027	0.029
	MAE	0.090	0.063	0.049	0.046	0.035	0.032	0.036	0.035	0.029	0.028	0.023	0.026	0.025	0.023	0.020	0.020	0.019	0.020	0.022	0.024
0.7	v	0.050	0.100	0.150	0.200	0.250	0.300	0.350	0.400	0.450	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	1.000
	ϵ	0.363	0.281	0.243	0.196	0.164	0.147	0.126	0.107	0.090	0.079	0.067	0.052	0.045	0.036	0.029	0.025	0.016	0.012	0.005	0.000
	SV	0.080	0.120	0.175	0.225	0.285	0.330	0.380	0.430	0.485	0.530	0.590	0.630	0.680	0.730	0.790	0.835	0.890	0.925	0.985	1.045
	Error	0.035	0.085	0.130	0.185	0.235	0.275	0.320	0.375	0.420	0.470	0.525	0.570	0.615	0.665	0.720	0.765	0.830	0.860	0.920	0.955
	RMSE	0.128	0.090	0.091	0.090	0.075	0.063	0.057	0.047	0.044	0.042	0.043	0.047	0.046	0.045	0.043	0.043	0.044	0.046	0.047	0.046
	MAE	0.103	0.070	0.070	0.069	0.053	0.048	0.043	0.037	0.033	0.032	0.033	0.033	0.031	0.031	0.029	0.030	0.030	0.031	0.033	0.033
0.8	v	0.050	0.100	0.150	0.200	0.250	0.300	0.350	0.400	0.450	0.500	0.550	0.600	0.650	0.700	0.750	0.800	0.850	0.900	0.950	1.000
	ϵ	0.367	0.309	0.244	0.214	0.190	0.161	0.145	0.119	0.096	0.084	0.075	0.056	0.049	0.037	0.031	0.026	0.014	0.012	0.004	0.000
	SV	0.080	0.130	0.175	0.220	0.280	0.335	0.380	0.435	0.485	0.530	0.570	0.635	0.695	0.730	0.780	0.835	0.890	0.935	0.990	1.060
	Error	0.035	0.090	0.135	0.175	0.230	0.280	0.325	0.385	0.430	0.470	0.515	0.575	0.625	0.670	0.715	0.775	0.820	0.870	0.920	0.940
	RMSE	0.143	0.131	0.082	0.087	0.083	0.074	0.071	0.064	0.061	0.061	0.061	0.051	0.055	0.054	0.053	0.054	0.057	0.058	0.058	0.058
	MAE	0.112	0.102	0.060	0.062	0.060	0.058	0.053	0.046	0.043	0.044	0.044	0.039	0.042	0.041	0.040	0.042	0.044	0.044	0.044	0.044



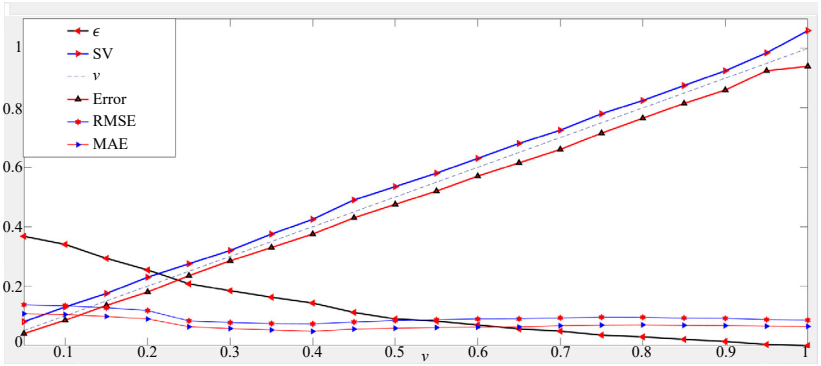
(a)



(b)



(c)



(d)

Figure 2. Performance of the proposed v-SVQR model on AD1 data set for (a) $\tau = 0.1$ (b) $\tau = 0.3$ (c) $\tau = 0.6$ (d) $\tau = 0.9$

Table 2. Performance of the proposed v-SVQR with different sizes of training set

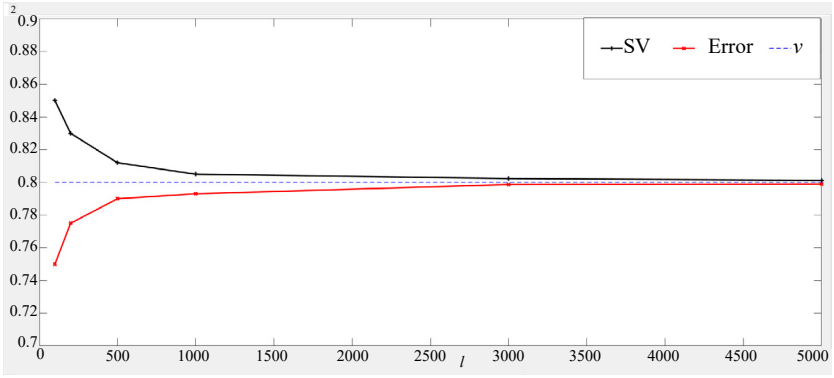
τ	l	100	200	500	1000	3000	5000
0.1	SV	0.85	0.83	0.81	0.81	0.80	0.80
	Error	0.75	0.78	0.79	0.79	0.80	0.80
	Ratio	14.00	13.09	9.97	9.72	9.20	9.17
	ϵ	0.04	0.05	0.05	0.04	0.05	0.05
	RMSE	0.11	0.05	0.06	0.05	0.01	0.01
0.3	l	100	200	500	1000	3000	5000
	SV	0.85	0.82	0.82	0.81	0.80	0.80
	Error	0.73	0.77	0.79	0.80	0.80	0.80
	Ratio	2.65	2.56	2.41	2.38	2.36	2.35
	ϵ	0.04	0.03	0.03	0.03	0.03	0.03
	RMSE	0.10	0.05	0.03	0.04	0.01	0.01
0.7	l	100	200	500	1000	3000	5000
	SV	0.85	0.82	0.82	0.81	0.80	0.80
	Error	0.73	0.76	0.79	0.79	0.80	0.80
	Ratio	0.35	0.38	0.42	0.43	0.43	0.43
	ϵ	0.02	0.03	0.03	0.03	0.03	0.03
	RMSE	0.07	0.05	0.04	0.02	0.01	0.01
0.9	l	100	200	500	1000	3000	5000
	SV	0.87	0.82	0.81	0.81	0.80	0.80
	Error	0.76	0.77	0.79	0.79	0.80	0.80
	Ratio	0.09	0.08	0.10	0.11	0.11	0.11
	ϵ	0.03	0.04	0.04	0.05	0.05	0.05
	RMSE	0.13	0.06	0.07	0.03	0.02	0.02

Experiment 3

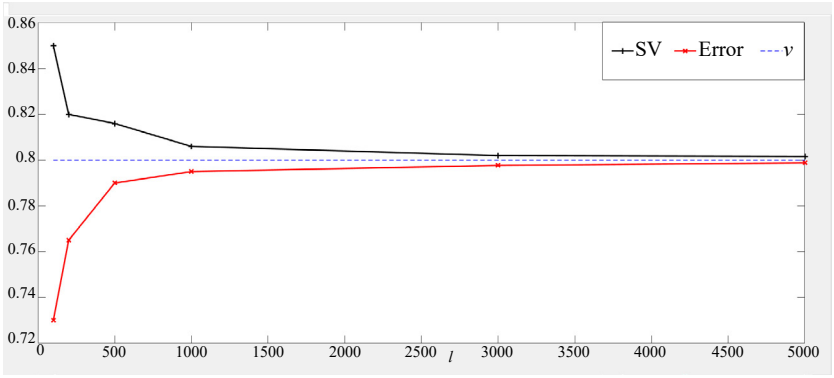
This experiment has been performed to show the capability of the proposed v-SVQR model to automate the control over accuracy. The proposed v-SVQR model has capability to automatically adjust the width of the asymmetric ϵ-insensitive zone for efficient prediction. For fixed values of parameters with $v = 0.3$, we have simulated the proposed v-SVQR model on AD1 data set with noise variance $\sigma = 0.2$ and $\sigma = 1$. Figure 4 shows the estimates obtained by the proposed v-SVQR along with the ϵ-insensitive zone for $\tau = 0.1$ and 0.9 at a fixed value of $v = 0.3$. It can be observed that the proposed v-SVQR model can automatically adjust the width of the asymmetric ϵ-insensitive zone according to the variance present in data for obtaining efficient estimates of quantiles.

Further, we have checked the performance of the proposed v-SVQR model on AD1 data set with different noise variance σ. For this, we have fixed the number of training data points to 500. The v parameter in the proposed v-SVQR was fixed to 0.4. Other parameters were also fixed. Table 3 lists numerical results obtained by the proposed v-SVQR model on AD1 data set with different noise variance σ for several τ values. Figure 5 illustrates the numerical results listed in Table 3 well for some τ values. The following inferences can be easily drawn from numerical results presented in Table 3 and Figure 5.

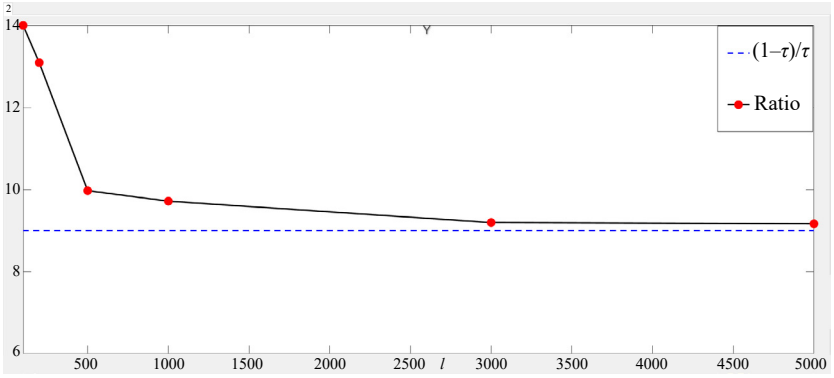
(a) Irrespective of values of τ, as the noise variance σ increases, the v-SVQR model accordingly increases the width of the asymmetric ϵ-insensitive zone.
(b) Irrespective of values of σ, v is the upper bound on fraction of errors and the lower bound on fraction of support vectors in the proposed v-SVQR model.
(c) As the noise variance σ increases, the RMSE obtained by the v-SVQR model increases.



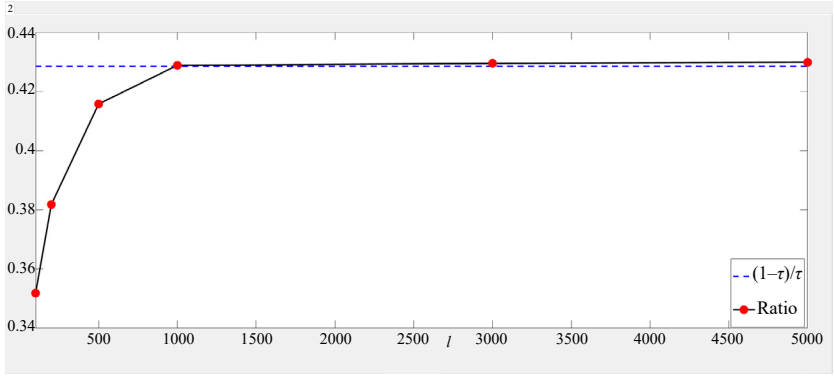
(a)



(b)

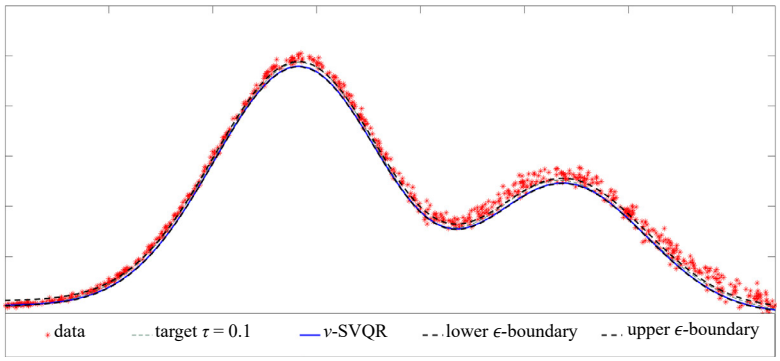


(c)

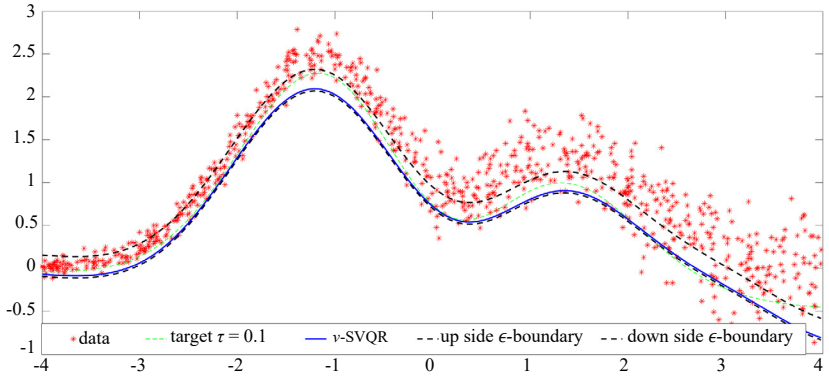


(d)

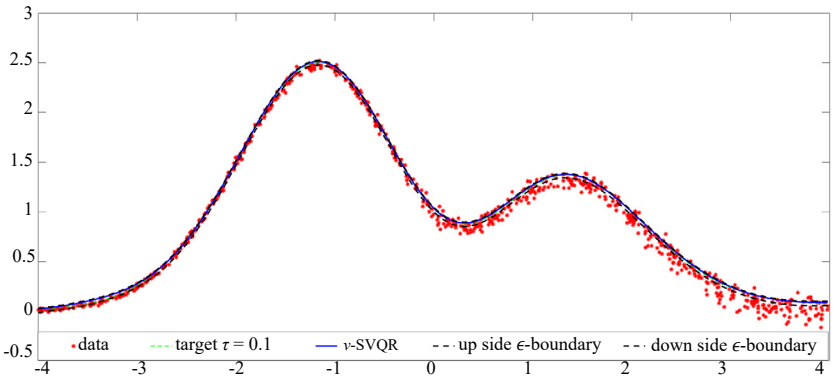
Figure 3. Asymptotic behavior of ν-SVQR model for (a) $\tau = 0.1$ (b) $\tau = 0.3$ (c) $\tau = 0.1$ and (d) $\tau = 0.7$



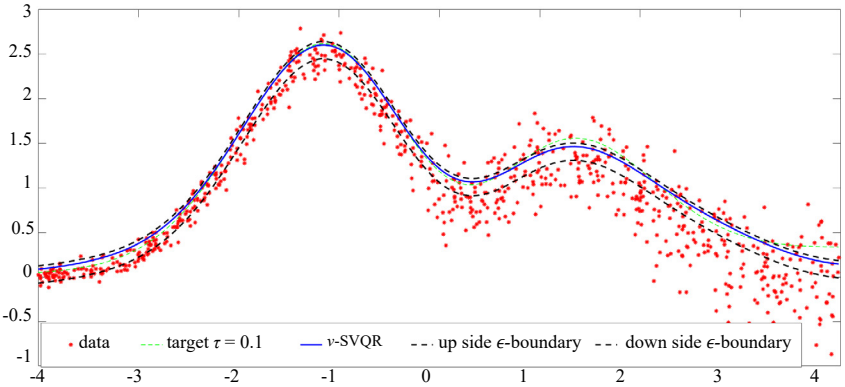
(a)



(b)



(c)



(d)

Figure 4. Automatic adjustment of the width of the ϵ-insensitive zone in the proposed v-SVQR model

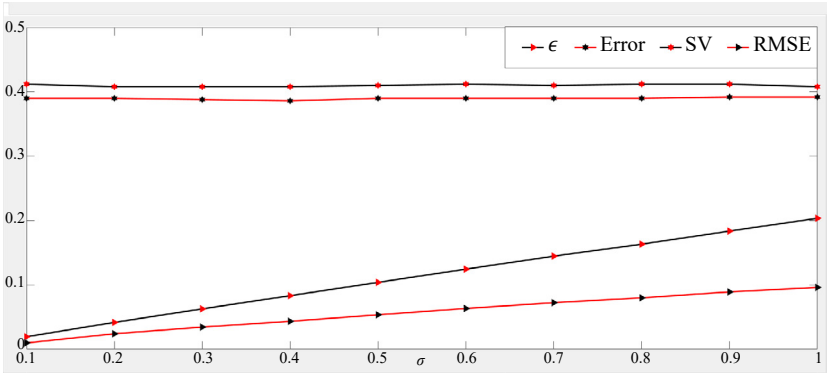
Table 3. Performance of the proposed v-SVQR with different values of noise variance σ

τ	σ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.9	ϵ	0.02	0.04	0.05	0.07	0.09	0.11	0.12	0.14	0.16	0.18
	Error	0.39	0.39	0.40	0.39	0.39	0.39	0.39	0.39	0.39	0.39
	SV	0.41	0.41	0.42	0.41	0.41	0.41	0.41	0.41	0.41	0.42
	RMSE	0.01	0.02	0.02	0.03	0.04	0.05	0.05	0.06	0.07	0.08
0.7	σ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
	ϵ	0.01	0.03	0.04	0.06	0.07	0.09	0.10	0.12	0.13	0.15
	Error	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39
	SV	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41
	RMSE	0.00	0.01	0.02	0.02	0.03	0.03	0.04	0.04	0.05	0.05
0.5	σ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
	ϵ	0.01	0.03	0.05	0.06	0.07	0.09	0.10	0.12	0.13	0.15
	Error	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39
	SV	0.41	0.42	0.42	0.41	0.41	0.41	0.41	0.41	0.41	0.41
	RMSE	0.00	0.01	0.01	0.02	0.02	0.03	0.03	0.03	0.04	0.04
0.3	σ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
	ϵ	0.01	0.03	0.05	0.07	0.08	0.10	0.11	0.13	0.15	0.16
	Error	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39
	SV	0.42	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41
	RMSE	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.04	0.05	0.06
0.1	σ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
	ϵ	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20
	Error	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39
	SV	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41
	RMSE	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10

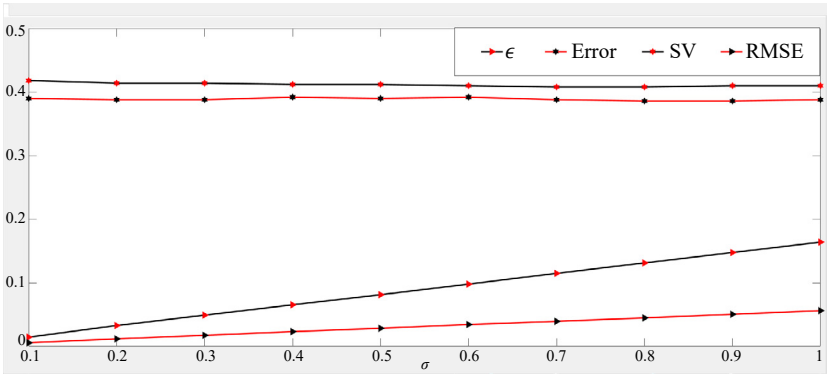
Experiment 4

As stated in Remark 5.4, the proposed v-SVQR model is similar to the ϵ-SVQR model in the sense that any solution $(\overline{w}, \overline{b})$ obtained by the proposed v-SVQR can also be obtained by the ϵ-SVQR. But, the proposed v-SVQR model has the capability of adjusting the ϵ-insensitive zone according to the variance present in the data. For realizing this direct benefit of the proposed v-SVQR model over the ϵ-SVQR model, we perform the following experiment.

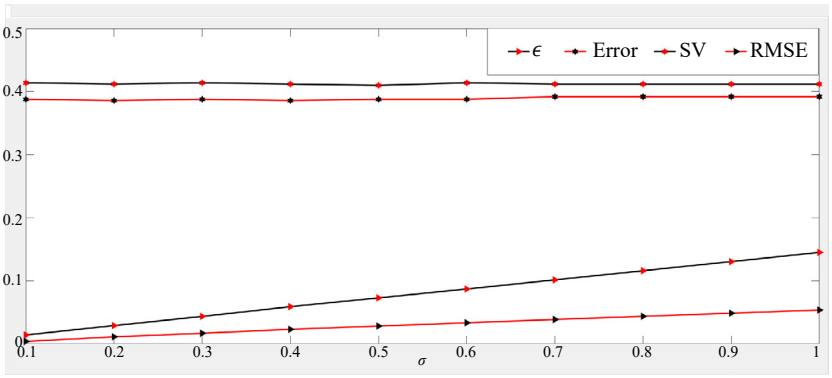
We generate 500 training data points of the AD2 data set where response points were polluted with noise from $U(-0.1, 0.1)$. For predicting the $\tau = 0.3$ quantile, we have tuned the parameters of ϵ-QSVR as well as the proposed v-SVQR model. We have found that at $\epsilon = 0.1$ and $C = 2^0$, the ϵ-QSVR model obtains the minimum RMSE of 0.0057. The proposed v-SVQR model obtains the RMSE value 0.0056 with parameters $v = 0.5$ and $C = 2^0 * 500$. The v-SVQR model obtains the asymmetric ϵ tube of width 0.0074. Now, with the same parameters setting in both ϵ-QSVR and v-SVQR models, we increase the variance present in the noise of AD2 data set to $U(-5, 5)$. The ϵ-SVQR model which has a fixed ϵ value could obtain the RMSE 0.2168. But, the v-SVQR model automatically adjusts the width of asymmetric ϵ-tube to 0.3734 and can obtain the RMSE value 0.1840. The estimate obtained by the ϵ-SVQR and the proposed v-SVQR model on AD2 data set with different noise variance has been well illustrated in Figure 6. It verifies our claim that the proposed v-SVQR model has the capability of automatically adjusting the asymmetric ϵ-insensitive zone.



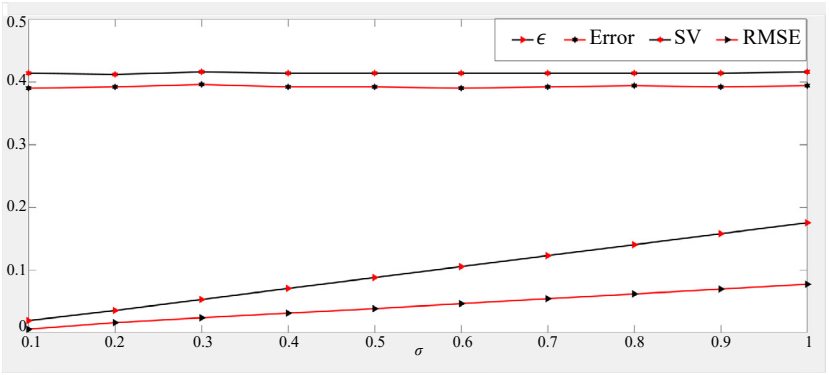
(a)



(b)



(c)



(d)

Figure 5. The proposed ν-SVQR model with different noise variance σ for (a) $\tau = 0.1$ (b) $\tau = 0.3$ (c) $\tau = 0.7$ and (d) $\tau = 0.9$

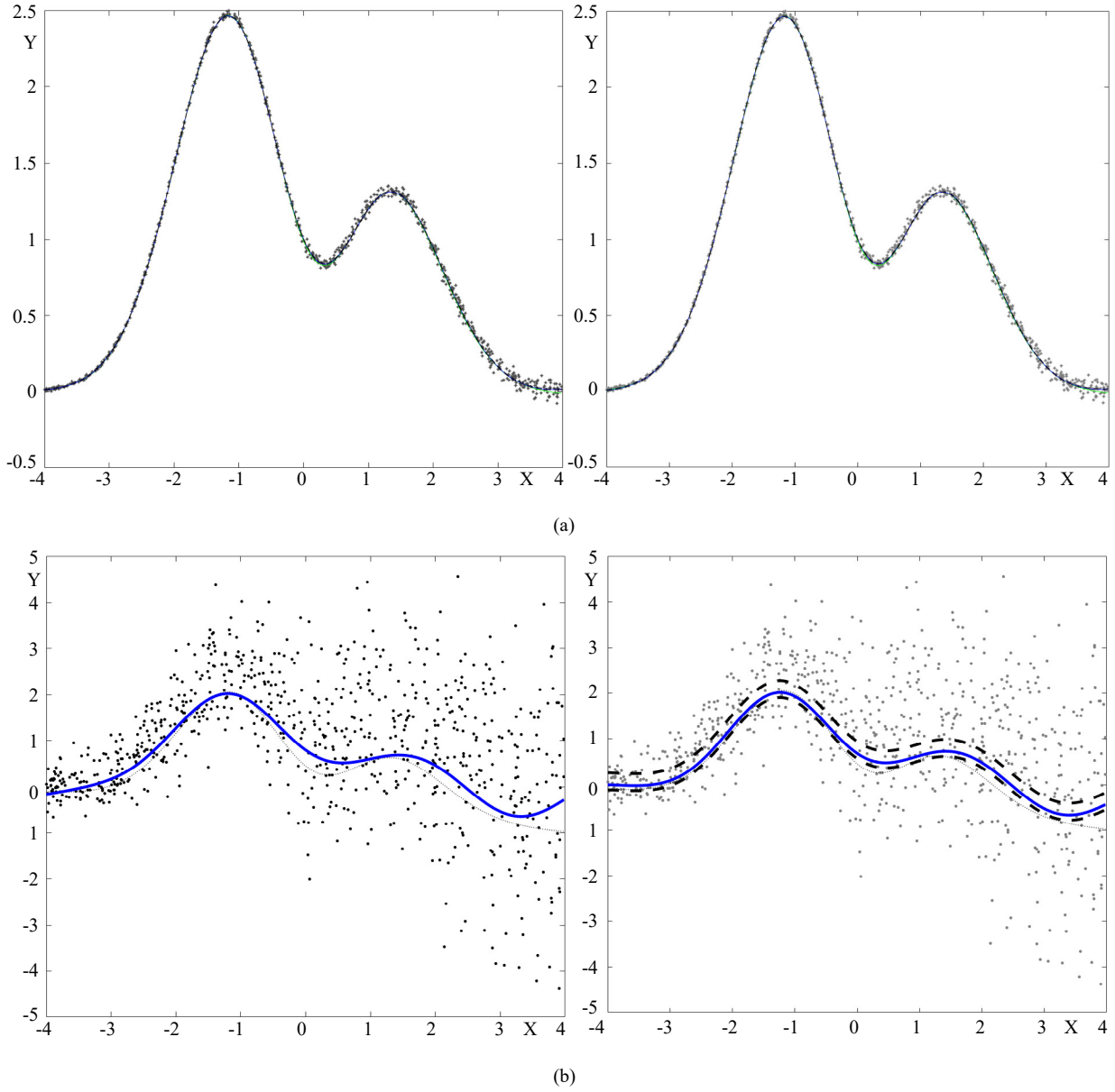


Figure 6. The ϵ-SVQR model (left) fails to adjust the total width of the asymmetric ϵ-insensitive tube with the increase in noise variance. The proposed ϵ-SVQR model (right) adjusts the total width of the asymmetric ϵ-insensitive tube according to the increase in noise variance and hence can obtain a better estimate

Experiment 5

The above experiments are enough to empirically verify the claims made in this paper. But, we still want to check the performance of the v-SVQR model on real-world data sets. For this, we have performed the experiments with the Servo (167×5) data set which is taken from the UCI repository [37]. We have used 80% of this data set for training the proposed v-SVQR model and the rest of the data points were used for the testing. The 100 random trials have been used to obtain the error E_τ and sparsity. Table 4 shows the error obtained by the proposed v-SVQR for different values of τ with different values of v. Table 5 shows the sparsity obtained by the proposed v-SVQR model for different values of v with different τ values. It can be observed that irrespective of values of τ, the sparsity decreases with an increase in the v value.

Table 4. Error E_τ obtained by the proposed v-SVQR model with different values of v for different τ values

v/τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.174	0.445	0.464	0.401	0.316	0.222	0.129	0.056	0.054
0.15	0.169	0.403	0.452	0.399	0.316	0.221	0.129	0.056	0.055
0.20	0.040	0.248	0.395	0.387	0.315	0.220	0.129	0.056	0.057
0.25	0.037	0.062	0.187	0.313	0.300	0.217	0.129	0.056	0.059
0.30	0.040	0.073	0.087	0.171	0.251	0.210	0.129	0.056	0.059
0.35	0.043	0.072	0.066	0.092	0.160	0.195	0.127	0.056	0.060
0.40	0.043	0.069	0.069	0.069	0.094	0.161	0.123	0.056	0.061
0.45	0.044	0.059	0.067	0.067	0.076	0.112	0.119	0.056	0.062
0.50	0.045	0.055	0.062	0.069	0.071	0.076	0.116	0.057	0.062
0.55	0.043	0.062	0.062	0.067	0.081	0.085	0.074	0.061	0.067
0.60	0.038	0.066	0.062	0.068	0.083	0.070	0.072	0.060	0.067
0.65	0.038	0.058	0.059	0.067	0.081	0.085	0.074	0.061	0.067
0.70	0.038	0.057	0.060	0.069	0.082	0.092	0.079	0.065	0.070
0.75	0.038	0.054	0.061	0.070	0.082	0.094	0.081	0.072	0.070
0.80	0.038	0.054	0.061	0.074	0.082	0.092	0.082	0.072	0.073
0.85	0.038	0.055	0.064	0.071	0.081	0.091	0.081	0.067	0.075
0.90	0.041	0.054	0.064	0.068	0.084	0.085	0.078	0.068	0.075
0.95	0.040	0.055	0.065	0.071	0.082	0.083	0.072	0.068	0.074
1.00	0.040	0.055	0.067	0.071	0.078	0.084	0.073	0.069	0.073

Table 5. Sparsity obtained by the proposed v-SVQR model with different values of v for different τ values

v/τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	89.47	88.72	89.47	89.47	89.47	89.47	88.72	88.72	89.47
0.15	83.46	82.71	84.21	83.46	83.46	83.46	84.21	84.21	84.21
0.2	78.95	78.95	79.70	78.20	78.95	78.95	79.70	78.95	78.95
0.25	74.44	74.44	73.68	74.44	73.68	73.68	72.93	73.68	74.44
0.3	69.17	68.42	69.17	69.17	69.17	69.17	69.17	69.17	69.17
0.35	63.91	63.91	63.91	63.91	63.91	63.91	63.16	63.91	64.66
0.4	57.89	57.89	57.89	57.89	58.65	58.65	59.40	59.40	58.65
0.45	54.14	52.63	53.63	53.38	52.63	54.14	54.14	54.14	54.14
0.5	48.87	48.12	48.87	48.12	47.37	48.12	49.62	48.87	48.12
0.55	42.86	42.86	43.61	43.61	43.61	44.36	42.86	43.61	43.61
0.6	38.35	38.35	37.59	36.84	38.35	38.35	38.35	39.10	39.10
0.65	33.08	33.08	33.83	32.33	33.08	32.33	33.08	34.59	34.59
0.7	27.82	29.32	27.07	27.82	27.07	27.82	27.82	29.32	29.32
0.75	22.56	22.56	21.80	22.56	23.31	22.56	23.31	23.31	24.06
0.8	17.29	17.29	17.29	18.80	17.29	16.54	18.05	18.80	18.80
0.85	13.53	12.03	12.03	11.28	13.53	12.78	12.03	13.53	14.29
0.9	8.27	9.02	8.27	8.27	7.52	7.52	7.52	8.27	9.02
0.95	3.01	2.26	3.76	3.01	3.01	3.01	3.01	3.76	4.51
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

7. Conclusion

In this paper, we propose a novel v-SVQR model which improves the existing ϵ-SVQR model [7]. The ϵ-SVQR model can obtain an efficient prediction of the conditional quantile by considering an asymmetric ϵ-insensitive zone around the data points. But for this, it requires a good choice of the value of ϵ. A bad choice of the value of ϵ can lead to poor prediction in ϵ-SVQR. Our proposed v-SVQR model can automatically obtain a suitable asymmetric ϵ-insensitive zone around data points according to the variance present in data for conditional quantile estimation. For this, the proposed v-SVQR model trades off the size of the ϵ-insensitive zone against the model complexity and empirical risk via the user-defined parameter v. We have briefly derived the different properties of the v-SVQR model using the KKT condition of the optimization problem of the v-SVQR model. In the v-SVQR model, the user-defined parameter v is the upper bound of the fraction of errors and the lower bound of the fraction of Support Vectors. Further, we have carried out extensive numerical experiments on several simulated and real-world data sets and verified our claims regarding the properties of the v-SVQR model.

The v-SVQR model is not suitable for large-scale data set processing, as it requires the solution of a QPP, which involves $2l$ constraints. In the future, we would like to solve the QPP of the v-SVQR model using efficient variants of the stochastic gradient descent method which may enable the proposed model to deal with online and large-scale data efficiently. We have planned to use the variants of the metaheuristic algorithms to tune the parameters of the v-SVQR model in the future. We have also planned to check the performance of the v-SVQR model in wind hybrid models for obtaining the probabilistic forecast of wind energy.

Acknowledgment

We would like to acknowledge Ministry of Electronics and Information Technology, Government of India, as this work has been funded by them under Visvesvaraya PhD Scheme for Electronics and IT, Order No. Phd-MLA/4(42)/201516.

Conflict of interest

We authors hereby declare that we do not have any conflict of interest with the content of this manuscript.

References

[1] Koenker R, Bassett Jr G. Regression quantiles. *Econometrica*. 1978; 46(1): 33-50. https://doi.org/10.2307/1913643

[2] Koenker R, Hallock KF. Quantile regression. *Journal of Economic Perspectives*. 2001; 15(4): 143-156. https://doi.org/10.1257/jep.15.4.143

[3] Vapnik V, Golowich S, Smola A. Support vector method for function approximation, regression estimation and signal processing. In: Mozer MC, Jordan M, Petsche T. (eds.) *Proceedings of the 9th International Conference on Neural Information Processing Systems*. Cambridge, MA, United States: MIT Press; 1996. p.281-287.

[4] Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V. Support vector regression machines. In: Mozer MC, Jordan M, Petsche T. (eds.) *Proceedings of the 9th International Conference on Neural Information Processing Systems*. Cambridge, MA, United States: MIT Press; 1996. p.155-161.

[5] Gunn SR. Support vector machines for classification and regression. *ISIS Technical Report*. 1998; 14(1): 5-16.

[6] Takeuchi I, Le QV, Sears TD, Smola AJ. Nonparametric quantile estimation. *Journal of Machine Learning Research*. 2006; 7(45): 1231-1264.

[7] Anand P, Rastogi R, Chandra S. A new asymmetric ϵ-insensitive pinball loss function based support vector quantile regression model. *Applied Soft Computing*. 2020; 94: 106473. https://doi.org/10.1016/j.asoc.2020.106473

[8] Mercer J, Andrew RF. XVI. Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character*. 1909; 209(441-458): 415-446. https://doi.org/10.1098/rsta.1909.0016

[9] Koenker R, Ng P, Portnoy S. Quantile smoothing splines. *Biometrika*. 1994; 81(4): 673-680. https://doi.org/10.1093/biomet/81.4.673
[10] Koenker R, Xiao Z. Quantile autoregression. *Journal of the American Statistical Association*. 2006; 101(475): 980-990. https://doi.org/10.1198/016214506000000672
[11] Chen X, Koenker R, Xiao Z. Copula-based nonlinear quantile autoregression. *The Econometrics Journal*. 2009; 12: S50-S67. https://doi.org/10.1111/j.1368-423X.2008.00274.x
[12] Koenker R, Park BJ. An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*. 1996; 71(1-2): 265-283. https://doi.org/10.1016/0304-4076(96)84507-6
[13] Takeuchi I, Furuhashi T. Non-crossing quantile regressions by SVM. In: *Proceedings of the 2004 IEEE International Joint Conference on Neural Networks*. IEEE; 2004. p.401-406. https://doi.org/10.1109/ijcnn.2004.1379939
[14] Hu T, Xiang DH, Zhou DX. Online learning for quantile regression and support vector regression. *Journal of Statistical Planning and Inference*. 2012; 142(12): 3107-3122. https://doi.org/10.1016/j.jspi.2012.06.010
[15] Seok KH, Cho D, Hwang C, Shim J. Support vector quantile regression using asymmetric e-insensitive loss function. In: Mahadevan V, Tomar GS. (eds.) *Proceedings of the 2010 2nd International Conference on Education Technology and Computer*. IEEE; 2010. p.V1-438-V1-439. https://doi.org/10.1109/icetc.2010.5529214
[16] Park J, Kim J. Quantile regression with an epsilon-insensitive loss in a reproducing kernel Hilbert space. *Statistics & Probability Letters*. 2011; 81(1): 62-70. https://doi.org/10.1016/j.spl.2010.09.019
[17] Meinshausen N. Quantile regression forests. *Journal of Machine Learning Research*. 2006; 7(35): 983-999. https://dl.acm.org/doi/10.5555/1248547.1248582
[18] Aprillia H, Yang HT, Huang CM. Statistical load forecasting using optimal quantile regression random forest and risk assessment index. *IEEE Transactions on Smart Grid*. 2021; 12(2): 1467-1480. https://doi.org/10.1109/tsg.2020.3034194
[19] Dang S, Peng L, Zhao J, Li J, Kong Z. A quantile regression random forest-based short-term load probabilistic forecasting method. *Energies*. 2022; 15(2): 663. https://doi.org/10.3390/en15020663
[20] Khan N, Shahid S, Juneng L, Ahmed K, Ismail T, Nawaz N. Prediction of heat waves in Pakistan using quantile regression forests. *Atmospheric Research*. 2019; 221: 1-11. https://doi.org/10.1016/j.atmosres.2019.01.024
[21] Vaysse K, Lagacherie P. Using quantile regression forest to estimate uncertainty of digital soil mapping products. *Geoderma*. 2017; 291: 55-64. https://doi.org/10.1016/j.geoderma.2016.12.017
[22] Taylor JW. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *Journal of Forecasting*. 2000; 19(4): 299-311. https://doi.org/10.1002/1099-131x(200007)19:4<299::aid-for775>3.3.co;2-m
[23] Xu Q, Deng K, Jiang C, Sun F, Huang X. Composite quantile regression neural network with applications. *Expert Systems with Applications*. 2017; 76: 129-139. https://doi.org/10.1016/j.eswa.2017.01.054
[24] Zhang W, Quan H, Srinivasan D. An improved quantile regression neural network for probabilistic load forecasting. *IEEE Transactions on Smart Grid*. 2018; 10(4): 4425-4434. https://doi.org/10.1109/tsg.2018.2859749
[25] Gan D, Wang Y, Yang S, Kang C. Embedding based quantile regression neural network for probabilistic load forecasting. *Journal of Modern Power Systems and Clean Energy*. 2018; 6(2): 244-254. https://doi.org/10.1007/s40565-018-0380-x
[26] Zhang W, Quan H, Gandhi O, Rajagopal R, Tan CW, Srinivasan D. Improving probabilistic load forecasting using quantile regression NN with skip connections. *IEEE Transactions on Smart Grid*. 2020; 11(6): 5442-5450. https://doi.org/10.1109/tsg.2020.2995777
[27] Lu S, Xu Q, Jiang C, Liu Y, Kusiak A. Probabilistic load forecasting with a non-crossing sparse-group Lasso-quantile regression deep neural network. *Energy*. 2022; 242: 122955. https://doi.org/10.1016/j.energy.2021.122955
[28] Wang Y, Gan D, Sun M, Zhang N, Lu Z, Kang C. Probabilistic individual load forecasting using pinball loss guided LSTM. *Applied Energy*. 2019; 235: 10-20. https://doi.org/10.1016/j.apenergy.2018.10.078
[29] Yu Y, Han X, Yang M, Yang J. Probabilistic prediction of regional wind power based on spatiotemporal quantile regression. *IEEE Transactions on Industry Applications*. 2020; 56(6): 6117-6127. https://doi.org/10.1109/tia.2020.2992945
[30] Zhang Z, Qin H, Liu Y, Yao L, Yu X, Lu J, et al. Wind speed forecasting based on quantile regression minimal gated memory network and kernel density estimation. *Energy Conversion and Management*. 2019; 196(9): 1395-1409. https://doi.org/10.1016/j.enconman.2019.06.024
[31] Zhou B, Ma X, Luo Y, Yang D. Wind power prediction based on LSTM networks and nonparametric kernel density estimation. *IEEE Access*. 2019; 7: 165279-165292. https://doi.org/10.1109/access.2019.2952555
[32] Yu Y, Wang M, Yan F, Yang M, Yang J. Improved convolutional neural network-based quantile regression for

regional photovoltaic generation probabilistic forecast. *IET Renewable Power Generation*. 2020; 14(14): 2712-2719. https://doi.org/10.1049/iet-rpg.2019.0949

[33] Huang Q, Wei S. Improved quantile convolutional neural network with two-stage training for daily-ahead probabilistic forecasting of photovoltaic power. *Energy Conversion and Management*. 2020; 220: 113085. https://doi.org/10.1016/j.enconman.2020.113085

[34] Schölkopf B, Smola AJ, Williamson RC, Bartlett PL. New support vector algorithms. *Neural Computation*. 2000; 12(5): 1207-1245. https://doi.org/10.1162/089976600300015565

[35] Hsu CW, Lin CJ. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*. 2002; 13(2): 415-425. https://doi.org/10.1109/72.991427

[36] Xu Q, Zhang J, Jiang C, Huang X, He Y. Weighted quantile regression via support vector machine. *Expert Systems with Applications*. 2015; 42(13): 5441-5451. https://doi.org/10.1016/j.eswa.2015.03.003

[37] Dua D, Graff C. *UCI Machine Learning Repository*. [Data set] 2017. http://archive.ics.uci.edu/ml [Accessed 16th October 2022].