

Expected shortfall estimation using kernel machines[†]

Jooyong Shim¹ · Changha Hwang²

¹Department of Data Science, Inje University

²Department of Statistics, Dankook University

Received 1 February 2013, revised 22 February 2013, accepted 1 April 2013

Abstract

In this paper we study four kernel machines for estimating expected shortfall, which are constructed through combinations of support vector quantile regression (SVQR), restricted SVQR (RSVQR), least squares support vector machine (LS-SVM) and support vector expectile regression (SVER). These kernel machines have obvious advantages such that they achieve nonlinear model but they do not require the explicit form of nonlinear mapping function. Moreover they need no assumption about the underlying probability distribution of errors. Through numerical studies on two artificial and two real data sets we show their effectiveness on the estimation performance at various confidence levels.

Keywords: Expected shortfall, expectile, least squares support vector regression, restricted support vector quantile regression, support vector quantile regression, value at risk.

1. Introduction

Value at risk (VaR) and expected shortfall (ES) are two closely related and widely used risk measures (Jorion, 2007). VaR is defined as the minimum potential return of a portfolio with a given confidence level over a certain time horizon. ES is defined as the conditional expectation of the return given that it is less than the VaR (Artzner *et al.*, 1999; Taylor, 2008a). In the insurance literature, ES is called conditional tail expectation. ES is a coherent risk measure whereas VaR is not, because VaR does not satisfy the subadditivity condition (Artzner *et al.*, 1999). The issue that is considered in this paper is the efficient computation of accurate estimate of ES. Although several methods have been presented, this issue is still a very challenging statistical problem.

Taylor (2008a) proposed to estimate VaR and ES via linear asymmetric least squares regression. Efron (1991) showed that the expectile could be used to estimate VaR and ES because there existed a one-to-one mapping from expectiles to quantiles. Cai and Wang (2008), Chen (2008) and Taylor (2008b) considered nonparametric econometric tools for ES

[†] This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012S1A3A2033330).

¹ Adjunct professor, Institute of Statistical Information, Department of Data Science, Inje University, Kyungnam 621-749, Korea

² Corresponding author: Professor, Department of Statistics, Dankook University, Gyeonggido 448-701, Korea. E-mail: chwang@dankook.ac.kr

computation. Zhu and Galbraith (2011) estimate the ES with asymmetric t and exponential power distributions. Leorato *et al.* (2012) proposed a class of ES estimators based on representing the estimator as an integral of quantile regression function (IQRF). Kato (2012) introduced a weighted Nadaraya-Watson estimation of conditional ES.

Closer inspection of this literature, however, suggests that the estimation methods of ES can be further improved. In this paper, we consider four kernel machines for estimating ES, which are based on IQRF, SVQR by Takeuchi *et al.* (2006), RSVQR by Shim and Lee (2010), LS-SVM by Suykens and Vanderwalle (1999) and SVER by Wang *et al.* (2011). The support vector machine (SVM), first developed by Vapnik (1995) and his group at AT&T Bell Laboratories, solves the weak point of neural network such as the existence of local minima in the area of statistical learning theory and structural risk minimization. SVM has been successfully applied to a number of real world problems related to classification and regression problems. One of its prominent advantages is the idea of using kernels to realize the nonlinear transformations without knowing the detailed transformations. In LS-SVM concerning classification problems, we have regression interpretations and direct links to work in classical statistics and the solution is given by a linear system instead of a quadratic programming. Takeuchi *et al.* (2006) first considered quantile regression by SVM formulation. Li *et al.* (2007) derived a simple formula for the effective dimension of the SVQR model, which allows convenient selection of the hyperparameters.

The rest of this paper is organized as follows. Section 2 introduces the definitions of VaR and ES. Section 3 outlines SVQR, RSVQR and LS-SVM. Section 4 introduces four methods for estimating ES based on the combinations of kernel methods such as SVQR, RSVQR, LS-SVM and SVER. Sections 5 and 6 present the numerical studies and conclusion, respectively.

2. VaR and expected shortfall

Let y be a real-valued random variable with a continuous and non-decreasing distribution function F . In market risk measurement and management, y is usually the return on a portfolio over a certain holding period. The quantile regression function on y given the input vector \mathbf{x} is defined as follows,

$$q_y(\theta|\mathbf{x}) = \inf\{y : F(y|\mathbf{x}) \geq \theta\}, \theta \in (0, 1). \quad (2.1)$$

Alternatively, it can be defined as the minimizer of the objective function,

$$(1 - \theta) \int_{-\infty}^{q_y(\theta|\mathbf{x})} |y - q_y(\theta|\mathbf{x})| f(y|\mathbf{x}) dy + \theta \int_{q_y(\theta|\mathbf{x})}^{\infty} |y - q_y(\theta|\mathbf{x})| f(y|\mathbf{x}) dy. \quad (2.2)$$

VaR is defined as the minimum potential return in value of a portfolio with a given confidence level $(1 - \theta)$ over a certain holding period, which is θ th quantile of the probability distribution of changes in the value of a portfolio. The VaR of the random return y given \mathbf{x} with confidence level $(1 - \theta)$ is defined as

$$VaR(\theta|\mathbf{x}) = q_y(\theta|\mathbf{x}). \quad (2.3)$$

ES is defined as the conditional expectation of the random variable y given that it is below the VaR, which leads the ES with confidence level $(1 - \theta)$ is defined as

$$ES(\theta|\mathbf{x}) = E(y|y \leq VaR(\theta)) = E(y|y \leq q_y(\theta|\mathbf{x})), \quad (2.4)$$

which can be rewritten as

$$ES(\theta|\mathbf{x}) = E(y|y \leq q_y(\theta|\mathbf{x})) = \frac{1}{\theta} \int_0^\theta q_y(p|\mathbf{x}) dp = q_y(\theta|\mathbf{x}) - \frac{1}{\theta} \int_{-\infty}^{q_y(\theta|\mathbf{x})} F(y|\mathbf{x}) dy. \quad (2.5)$$

This shows that the ES is larger, in absolute value, than the VaR.

The expectile regression function is defined as the minimizer of the objective function,

$$(1 - \theta) \int_{-\infty}^{e_y(\theta|\mathbf{x})} (y - e_y(\theta|\mathbf{x}))^2 f(y|\mathbf{x}) dy + \theta \int_{e_y(\theta|\mathbf{x})}^{\infty} (y - e_y(\theta|\mathbf{x}))^2 f(y|\mathbf{x}) dy. \quad (2.6)$$

From Efron (1991) the expectile regression function can be used to estimate VaR and ES because there exists a one-to-one mapping from expectile regression functions to quantile regression functions. The estimation of ES from expectile regression function can be obtained by solving the following equation (Taylor, 2008a),

$$ES(\theta|\mathbf{x}) = \left(1 + \frac{1}{\theta} \frac{\tau}{1 - 2\tau}\right) q_y(\theta|\mathbf{x}) - \frac{1}{\theta} \frac{\tau}{1 - 2\tau} E(y|\mathbf{x}), \quad (2.7)$$

where τ is selected such that $q_y(\theta|\mathbf{x}) = e_y(\tau|\mathbf{x})$ and $E(y|\mathbf{x})$ is the mean function. If y is defined to be a zero mean residual term, this becomes:

$$ES(\theta|\mathbf{x}) = \left(1 + \frac{1}{\theta} \frac{\tau}{1 - 2\tau}\right) q_y(\theta|\mathbf{x}). \quad (2.8)$$

3. SVQR, RSVQR and LS-SVM

In this section we review SVQR, RSVQR and LS-SVM, which will be used for estimating ES in the next section. For other different types of kernel machines, see Bae *et al.* (2012), Hwang and Shim (2011, 2012) and Shim and Seok (2012).

3.1. SVQR

Let the training data set be denoted by $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, with each input vector $\mathbf{x}_i \in R^d$ and the output $y_i \in R$ which is linearly or nonlinearly related to the input vector \mathbf{x}_i . Here the feature mapping function $\phi : R^d \rightarrow R^h$ maps the input space to the higher dimensional feature space where the dimension h is defined in an implicit way. An inner product in feature space has an equivalent kernel in input space, $\phi(\mathbf{x}_i)^t \phi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$. Several choices of the kernel $K(\cdot, \cdot)$ are possible. We consider the nonlinear case, in which the θ th quantile function of y given \mathbf{x} , $q_y(\theta|\mathbf{x})$ for $\theta \in (0, 1)$, can be regarded as a nonlinear function of input vector \mathbf{x} .

With a check function ρ_θ , the θ th quantile regression function can be defined as a function of any solution to the optimization problem,

$$\min \frac{1}{2} \mathbf{w}^t \mathbf{w} + C \sum_{i=1}^n \rho_\theta (y_i - q_y(\theta|\mathbf{x}_i)), \quad (3.1)$$

where C is a penalty parameter penalizing the training errors and $\rho_\theta(\cdot)$ is the check function defined as

$$\rho_\theta(r) = \theta r I(r \geq 0) + (\theta - 1)r I(r < 0) \text{ for } \theta \in (0, 1).$$

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

$$\min \frac{1}{2} \mathbf{w}^t \mathbf{w} + C \sum_{i=1}^n (\theta \xi_i + (1 - \theta) \xi_i^*) \tag{3.2}$$

subject to

$$\begin{cases} y_i - \mathbf{w}^t \phi(\mathbf{x}_i) - b \leq \xi_i \\ \mathbf{w}^t \phi(\mathbf{x}_i) + b - y_i \leq \xi_i^* \\ \xi_i, \xi_i^* \geq 0. \end{cases}$$

We construct a Lagrange function as follows:

$$\begin{aligned} L = & \frac{1}{2} \mathbf{w}^t \mathbf{w} + C \sum_{i=1}^n (\theta \xi_i + (1 - \theta) \xi_i^*) - \sum_{i=1}^n \alpha_i (\xi_i - y_i + \mathbf{w}^t \phi(\mathbf{x}_i) + b) \\ & - \sum_{i=1}^n \alpha_i^* (\xi_i^* + y_i - \mathbf{w}^t \phi(\mathbf{x}_i) - b) - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*). \end{aligned} \tag{3.3}$$

We notice that the non-negative constraints $\alpha_i^{(*)}, \xi_i^{(*)} \geq 0$ should be satisfied. After taking partial derivatives of the equation (3.3) with regard to the primal variables $(\mathbf{w}, \xi_i^{(*)}, b)$ and plugging them into the equation (3.3), we have the optimization problem below,

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*} -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^n \alpha_i y_i - \sum_{i=1}^n \alpha_i^* y_i \tag{3.4}$$

with constraints $\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0$, $\alpha_i \in [0, \theta C]$ and $\alpha_i^* \in [0, (1 - \theta)C]$.

Solving the above optimization problem with the constraints determines the optimal Lagrange multipliers, $\hat{\alpha}_i$ and $\hat{\alpha}_i^*$. Thus, the estimated θ th quantile function given the input vector \mathbf{x} is obtained as

$$\hat{q}_y(\theta | \mathbf{x}) = \sum_{i=1}^n (\hat{\alpha}_i - \hat{\alpha}_i^*) K(\mathbf{x}_i, \mathbf{x}) + \hat{b}, \tag{3.5}$$

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

$$\hat{b} = \frac{1}{n_s} \sum_{i \in I_s} \left(y_i - \sum_{k=1}^n (\hat{\alpha}_k - \hat{\alpha}_k^*) K(\mathbf{x}_k, \mathbf{x}_i) \right),$$

where n_s is a size of the set $I_s \equiv \{i = 1, \dots, n \mid C(\theta - 1) < \hat{\alpha}_i - \hat{\alpha}_i^* < C\theta\}$.

For the model selection Yuan (2006) proposed the generalized approximate cross validation function as follows,

$$GACV(\boldsymbol{\lambda}) = \frac{1}{n - d_{eff}} \sum_{i=1}^n \rho_{\theta}(y_i - \hat{q}_y(\theta|\mathbf{x}_i)), \tag{3.6}$$

where $\boldsymbol{\lambda}$ is the set of hyperparameters, d_{eff} is a measure of the effective dimensionality of the fitted model and Yuan (2006) used $d_{eff} = \sum_{i=1}^n \partial \hat{q}_y(\theta|\mathbf{x}_i) / \partial y_i$ with a differentiable modified check function. Li *et al.* (2007) showed that d_{eff} is equal to n_s .

3.2. RSVQR

We now introduce RSVQR proposed by Shim and Lee (2010). RSVQR is based on the following nonlinear heteroscedastic model,

$$y_i = \text{med}(\mathbf{x}_i) + s(\mathbf{x}_i)\epsilon_i, \tag{3.7}$$

where $s(\mathbf{x}_i)$ is assumed to be positive, ϵ_i is assumed to have median 0 and $|\epsilon_i|$ is assumed to have median 1. For noncrossing quantile regression, we employ the basic ideas of restricted regression quantile of He (1997) with support vector median regression as follows:

1. Apply support vector median regression (SVQR with $\theta = 0.5$) on $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ to obtain the median function $\hat{\text{med}}(\mathbf{x}_i)$ of y given \mathbf{x}_i and residuals $\hat{r}_i = y_i - \hat{\text{med}}(\mathbf{x}_i)$.
2. Apply support vector median regression on $\{(\mathbf{x}_i, |\hat{r}_i|)\}_{i=1}^n$ to obtain the estimated median function of $|\hat{r}_i|$ given \mathbf{x}_i , \hat{s}_i , which is the estimate of $s(\mathbf{x}_i)$ since the median of $|\hat{\epsilon}_i|$ is assumed to be 1.
3. Find the θ th quantile of \hat{r}_i , $\beta_{\theta}\hat{s}_i$, by minimizing $\sum_{i=1}^n \rho_{\theta}(\hat{r}_i - \beta_{\theta}\hat{s}_i)$. Since $r_i = y_i - \text{med}(\mathbf{x}_i) = s(\mathbf{x}_i)\epsilon_i$, the quantile of r_i depends on \mathbf{x}_i through $s(\mathbf{x}_i)$ whose estimate is \hat{s}_i .

Then the estimated θ th quantile regression function of y given \mathbf{x}_t is obtained as

$$\hat{q}_y(\theta|\mathbf{x}_t) = \hat{\text{med}}(\mathbf{x}_t) + \beta_{\theta}\hat{s}_t, \tag{3.8}$$

where \hat{s}_t is the estimated median function of $|\hat{r}|$ given \mathbf{x}_t . Here quantiles of \hat{r}_i 's are noncrossed since the linear quantile regression is performed, which leads quantile functions of y given \mathbf{x}_i noncrossed. For model selection see Shim and Lee (2010).

3.3. LS-SVM

The LS-SVM, a modified version of SVM in a least squares sense, has been proposed for the classification and the regression by Suykens and Vanderwalle (1999). The LS-SVM model for function estimation has the following representation in feature space

$$f(\mathbf{x}) = \mathbf{w}^t \boldsymbol{\phi}(\mathbf{x}) + b \text{ with } \mathbf{w} \in R^h, b \in R,$$

where superscript t represents the transpose of a vector. Given a training set $\{y_i, \mathbf{x}_i\}_{i=1}^n$ we define now the following optimization problem to get optimal \mathbf{w} and b

$$\min_{\mathbf{w}, b, e} \frac{1}{2} \mathbf{w}^t \mathbf{w} + \frac{C}{2} \sum_{i=1}^n e_i^2 \quad (3.9)$$

subject to the equality constraints

$$y_i = \mathbf{w}^t \phi(\mathbf{x}_i) + b + e_i, \quad i = 1, \dots, n. \quad (3.10)$$

The cost function with squared error and regularization corresponds to a form of ridge regression. We construct the Lagrangian

$$L = \frac{1}{2} \mathbf{w}^t \mathbf{w} + \frac{C}{2} \sum_{i=1}^n e_i^2 - \sum_{i=1}^n \alpha_i (\mathbf{w}^t \phi(\mathbf{x}_i) + b + e_i - y_i), \quad (3.11)$$

where α_i 's are Lagrange multipliers. The conditions for optimality $\frac{\partial L}{\partial \mathbf{w}} = \mathbf{0}$, $\frac{\partial L}{\partial b} = 0$, $\frac{\partial L}{\partial e_i} = 0$ and $\frac{\partial L}{\partial \alpha_i} = 0$ yield the linear equation

$$\begin{bmatrix} 0 & \mathbf{1}^t \\ \mathbf{1} & K + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \quad (3.12)$$

with $\mathbf{y} = (y_1, \dots, y_n)^t$, $\mathbf{1} = (1, \dots, 1)^t$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^t$ and where

$$\begin{aligned} K_{kl} &= K(\mathbf{x}_k, \mathbf{x}_l), \quad k, l = 1, \dots, n \\ &= \phi(\mathbf{x}_k)^t \phi(\mathbf{x}_l) \end{aligned}$$

is a kernel function obtained from the Mercer's condition (Mercer, 1909). Several choices of the kernel function are possible. Here \mathbf{w} and $\phi(\mathbf{x})$ are not calculated.

The resulting LS-SVM model for function estimation becomes

$$\hat{f}(\mathbf{x}_t) = \sum_{i=1}^n \hat{\alpha}_i K(\mathbf{x}_i, \mathbf{x}_t) + \hat{b}, \quad (3.13)$$

where $\hat{\alpha}_i$'s and \hat{b} are the solution to the linear system.

The GCV function for model selection can be obtained as follows,

$$GCV(\boldsymbol{\lambda}) = \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \hat{m}(\mathbf{x}_i))^2}{(1 - \text{tr}(S)/n)^2}, \quad (3.14)$$

where S is the hat matrix such that $(\hat{f}(\mathbf{x}_1), \dots, \hat{f}(\mathbf{x}_n))^t = S\mathbf{y}$, which is obtained from (3.12).

4. ES estimation via kernel machines

In this section we describe four methods for estimating ES using SVQR, RSVQR and LS-SVM explained in the previous section. First we consider two methods that estimate ES employing IQRF (2.5) after computing VaR using SVQR or RSVQR. Hereafter, we call these methods as IQRF/SVQR and IQRF/RSVQR, respectively. These methods can be explained in more detail as follows:

$$\hat{ES}(\theta|\mathbf{x}) = \frac{1}{\theta} \sum_{j=1}^J w_j \tilde{q}_y(p_j|\mathbf{x}) \tag{4.1}$$

with

$$w_j = p_j - p_{j-1}, \quad j \geq 1 \quad \text{and} \quad \tilde{q}_y(p_j|\mathbf{x}) = \frac{\hat{q}_y(p_j|\mathbf{x}) + \hat{q}_y(p_{j-1}|\mathbf{x})}{2}, \quad j \geq 1$$

where $\hat{q}_y(p_j|\mathbf{x})$ is an estimator of the quantile regression function $q_y(p_j|\mathbf{x})$, the p_j 's are grid points such that $0 = p_0 < p_1 < \dots < p_J = \theta$ and J is the number of w_j 's. Here $\hat{q}_y(p_j|\mathbf{x})$ is obtained by SVQR or RSVQR, and $J = \lceil n\theta \rceil$, which is the nearest integer to $n\theta$.

Next we consider the method using SVQR and LS-SVM, which is explained below.

1. Find $\hat{q}_y(\theta|\mathbf{x}_i)$ using SVQR from $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$.
2. Find $\hat{ES}(\theta)$ using LS-SVM from $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ such that $y_i \leq \hat{q}_y(\theta|\mathbf{x}_i)$.

Hereafter, we call this method as SVQR/LS-SVM. Note that LS-SVM uses the training data consisting of (\mathbf{x}_i, y_i) such that $y_i \leq \hat{q}_y(\theta|\mathbf{x}_i)$. Therefore, the estimated ES for \mathbf{x}_t is given by $\hat{ES}(\theta|\mathbf{x}_t) = \sum_{i: y_i \leq \hat{q}_\theta(y|\mathbf{x}_i)} K(\mathbf{x}_i, \mathbf{x}_t) \hat{\alpha}_i + \hat{b}$.

Finally, we consider the method using SVQR and SVER, which is explained below.

1. Find $\hat{q}_y(\theta|\mathbf{x}_i)$ using SVQR and $\hat{E}(y|\mathbf{x})$ using LS-SVM from $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$.
2. Find $\tau \in (0, 1)$ such that $\hat{q}_y(\theta|\mathbf{x}) = \hat{e}_y(\tau|\mathbf{x})$, where $\hat{e}_y(\tau|\mathbf{x})$ is obtained using SVER.
3. Compute $\hat{ES}(\theta|\mathbf{x})$ using (2.7) .

Hereafter, we call this method as SVQR/SVER.

5. Numerical studies

We now illustrate the performance of four ES estimation methods through two simulated and two real data sets. We compare four methods on the prediction performance. The numerical studies are conducted in MATLAB environment. In Example 5.3 we consider SVQR as a function of the length of the holding period and the one-step-ahead volatility forecast, and compare the performance of ES estimation methods for return data of KOSPI 200 and S&P 500 index data. The Gaussian kernel is utilized in every example, which is

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{1}{\sigma^2} \|\mathbf{x}_i - \mathbf{x}_j\|^2\right).$$

Optimal values of hyperparameters of SVQR, RSVQR and LS-SVM are selected by the corresponding model selection criterion.

Example 5.1 Fifty data sets are generated to present the prediction performance of the proposed methods. Each data set is divided into a training data set and test data set, and each data set consists of 100 (x, y) 's and 40 (x, y) 's. Here x 's are generated from the uniform distribution $U(0, 1)$, y 's are generated from a normal distribution $N(\sin(2\pi x), 1)$. The true θ th quantile regression function is given by

$$q_y(\theta|x) = \sin(2\pi x) + \Phi^{-1}(\theta) \text{ for } \theta \in (0, 1),$$

and the true θ th ES is given by

$$ES(\theta|x) = \sin(2\pi x) - \frac{1}{\theta} \phi(\Phi^{-1}(\theta)).$$

The left panel of Figure 5.1 shows the true θ th quantile regression functions (solid lines) and ESs (dotted lines) imposed on the scatter plot of a test data set for $\theta = 0.1, 0.15, 0.2, 0.25$. From 100 test data sets we obtain the average of mean squared errors (MSEs) of $\hat{ES}(\theta|x)$ for $\theta = 0.1, 0.15, 0.2, 0.25$ to compare the prediction performance, which are shown in Table 5.1. The boldfaced values indicate best performance/result. From the results we can see that IQRF/RSVQR and SVQR/LS-SVM provide better prediction performance than other methods for this normal case.

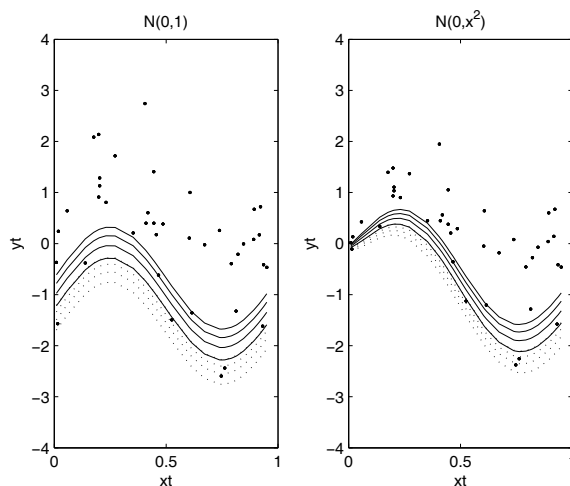


Figure 5.1 The true quantile regression functions (solid lines) and ESs(dotted lines) superimposed on the scatter plot of a test data set for four values of θ

Table 5.1 Average and standard error of 50 MSEs of $\hat{ES}(\theta|x)$ for Example 5.1

Methods	$\theta = 0.1$	$\theta = 0.15$	$\theta = 0.2$	$\theta = 0.25$
IQRF/SVQR	0.2275 (0.0190)	0.1809 (0.0190)	0.1530 (0.0145)	0.1356 (0.0130)
IQRF/RSVQR	0.1681 (0.0163)	0.1544 (0.0153)	0.1473 (0.0151)	0.1430 (0.0151)
SVQR/LS-SVM	0.2571 (0.0305)	0.1749 (0.0154)	0.1455 (0.0139)	0.1323 (0.0134)
SVQR/SVER	0.2605 (0.0265)	0.2822 (0.0345)	0.2452 (0.0244)	0.2994 (0.0383)

where $\hat{\sigma}_{t+1}$ is a 248×1 or 254×1 vector with elements of one-step-ahead volatility forecast $\hat{\sigma}_{t+1}$ for each corresponding index. The t -GARCH(1,1) model is utilized to obtain $\hat{\sigma}_{t+1}$ for r_1 of each corresponding index.

Then $VaR(\theta|\mathbf{x})$ using SVQR can be expressed as $q_y(\theta|\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b$, where $\mathbf{x} = (\mathbf{k}, \hat{\sigma})$. Gaussian kernel function is utilized in this example. The confidence level of VaR and ES given portfolio over a prescribed holding period is typically chosen to be 1% or 5%. After we calculate the estimates of ESs at 1% and 5% level using four estimation methods, we evaluate their performance through the average difference between the observed log returns and estimated ESs. The average difference is defined as follows:

$$BT_k(\theta) = \sum_{t=1}^n \frac{(r_{t,k} - \hat{E}S_k(\theta|\mathbf{x}_t)) I(r_{t,k} < \hat{V}aR_k(\theta|\mathbf{x}_t))}{I(r_{t,k} < \hat{V}aR_k(\theta|\mathbf{x}_t))},$$

where the subscript k indicates the holding period, $I(\cdot)$ is the indicator function and n is the sample size. Given this formula, the negative value of the statistic indicates the underestimation of risk losses and the positive value, the risk losses' overestimation. ES measures are accurate when the value of the $BT_k(\theta)$ tends to zero.

Table 5.3 and 5.4 show the average differences of four estimates of 1% and 5% ESs for KOSPI 200 index. Table 5.5 and 5.6 are for S&P 500 index. The boldfaced values indicate best performance/result for a given holding period. For every different holding period, overall SVQR/LS-SVM shows better performance in accuracy than the other methods.

Table 5.3 Average difference $\times 10^2$ for four estimates of 1% ES for KOSPI 200

Methods	Holding period						
	1	3	5	7	10	12	15
IQRF/SVQR	-0.03	0.02	0.07	0.07	0.06	0.11	0.06
IQRF/RSVQR	-0.16	0.64	0.56	-0.28	1.10	2.64	1.27
SVQR/LS-SVM	0.00	0.00	0.00	-0.01	-0.01	-0.01	0.00
SVQR/SVER	0.55	0.37	1.22	2.59	2.53	2.34	2.48

Table 5.4 Average difference $\times 10^2$ for four estimates of 5% ES for KOSPI 200

Methods	Holding period						
	1	3	5	7	10	12	15
IQRF/SVQR	-0.01	0.01	0.00	-0.02	-0.01	0.03	-0.01
IQRF/RSVQR	-0.27	0.18	0.15	-0.36	0.26	0.94	0.68
SVQR/LS-SVM	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SVQR/SVER	0.01	0.20	0.20	-0.04	0.03	0.42	0.21

Table 5.5 Average difference $\times 10^2$ for various estimates of 1% ES for S&P 500

Methods	Holding period						
	1	3	5	7	10	12	15
IQRF/SVQR	0.07	0.05	0.02	0.06	0.08	0.11	0.09
IQRF/RSVQR	1.76	1.86	-0.18	-0.02	0.43	2.22	1.48
SVQR/LS-SVM	0.00	0.01	-0.02	-0.01	-0.01	-0.01	-0.02
SVQR/SVER	0.97	0.48	2.19	2.13	2.29	2.24	2.79

Table 5.6 Average difference $\times 10^2$ for various estimates of 5% ES for S&P 500

Methods	Holding period						
	1	3	5	7	10	12	15
IQRF/SVQR	-0.04	-0.02	0.05	-0.28	0.08	-0.04	0.10
IQRF/RSVQR	0.55	1.64	-0.97	-0.40	0.25	0.21	0.94
SVQR/LS-SVM	0.01	0.01	0.01	0.00	0.00	-0.07	0.00
SVQR/SVER	0.49	0.85	1.20	-0.98	-0.04	1.45	0.08

6. Conclusions

In this paper we considered four ES estimation methods constructed through combinations of SVQR, RSVQR, LS-SVM and SVER, which solve the optimization problem in a mapped high-dimensional feature space. Two artificial data sets and two real data sets were used to test their effectiveness. The results clearly show their novelty in solving nonlinear estimation problem. Our finding suggests that the SVQR/LS-SVM would have great potentials in estimating ESs. However, unfortunately we could not provide the theoretical validation for the performances of these methods in this paper. We think this is a limitation of this research.

References

- Artzner, P., Delbaen, F., Eber, J. M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**, 203–28.
- Bae, J., Hwang, C. and Shim, J. (2012). Two-step LS-SVR for censored regression. *Journal of the Korean Data & Information Science Society*, **23**, 393–401.
- Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of returns. *Review of Economics and Statistics*, **69**, 542–547.
- Cai, Z. and Wang, X. (2008). Nonparametric estimation of conditional VaR and expected shortfall. *Journal of Econometrics*, **147**, 120–130.
- Chen, S.X. (2008). Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, **6**, 87–107.
- Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, **1**, 93–125.
- He, X. (1997). Quantile curves without crossing. *The American Statistician*, **51**, 86–192.
- Hwang, C. and Shim, J. (2011). Cox proportional hazard model with L1 penalty. *Journal of the Korean Data & Information Science Society*, **22**, 613–618.
- Hwang, C. and Shim, J. (2012). Mixed effects least squares support vector machine for survival data analysis. *Journal of the Korean Data & Information Science Society*, **23**, 739–748.
- Jorion, P. (2007). *Value at risk: The new benchmark for managing financial risk*, McGraw-Hill, New York.
- Kato, K. (2012). Weighted Nadaraya-Watson estimation of conditional expected shortfall. *Journal of Financial Econometrics*, **10**, 265–291.
- Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of 2nd Berkeley Symposium*, University of California Press, Berkeley, 481–492.
- Leorato, S., Peracchi, F. and Tanase, A. (2012). Asymptotically efficient estimation of the conditional expected shortfall. *Computational Statistics and Data Analysis*, **56**, 768–784.
- Li, Y., Liu, Y. and Ji, Z. (2007). Quantile regression in reproducing kernel Hilbert spaces. *Journal of the American Statistical Association*, **102**, 255–268.
- Mercer, J. (1909). Functions of positive and negative type and their connection with theory of integral equations. *Philosophical Transactions of Royal Society A*, 415–446.
- Shim, J. and Lee, J. (2010). Restricted support vector quantile regression without crossing. *Journal of the Korean Data & Information Science Society*, **21**, 1319–1325.
- Shim, J., Kim, Y., Lee, J. and Hwang, C. (2012). Estimating value at risk with semiparametric support vector quantile regression. *Computational Statistics*, **27**, 685–700.
- Shim, J. and Seok, K. (2012). Semiparametric kernel logistic regression with longitudinal data. *Journal of the Korean Data & Information Science Society*, **23**, 385–392.

- Suykens, J. A. K. and Vanderwalle, J. (1999). Least square support vector machine classifier. *Neural Processing Letters*, **9**, 293–300.
- Takeuchi, I., Le, Q. V., Sears, T. D. and Smola, A. J. (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research*, **7**, 1231–1264.
- Taylor, J. W. (2000). A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *Journal of Forecasting*, **19**, 299–311.
- Taylor, J. W. (2008a). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, **6**, 231–252.
- Taylor, J. W. (2008b). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics*, **6**, 382–406.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*, Springer, New York.
- Wang, Y., Wang, S. and Lai, K. (2011). Measuring financial risk with generalized asymmetric least squares regression. *Applied Soft Computing*, **11**, 5793–5800.
- Yuan, M. (2006). GACV for quantile smoothing splines. *Computational Statistics & Data Analysis*, **50**, 813–829.
- Zhu, D. and Galbraith, J. W. (2011). Modeling and forecasting expected shortfall with the generalized asymmetric Student-t and asymmetric exponential power distributions. *Journal of Empirical Finance*, **18**, 765–778.