Journal of the Korean Data & Information Science Society
2011, **22**(3), 589–596

한국데이터정보과학회지

Forecasting volatility via conditional autoregressive value at risk model based on support vector quantile regression†

Jooyong Shim[1] · Changha Hwang[2]

[1]Department of Data Science, Inje University
[2]Department of Statistics, Dankook University

Received 1 April 2011, revised 26 April 2011, accepted 1 May 2011

Abstract

The conditional autoregressive value at risk (CAViaR) model is useful for risk management, which does not require the assumption that the conditional distribution does not vary over time but the volatility does. But it does not provide volatility forecasts, which are needed for several important applications such as option pricing and portfolio management. For a variety of probability distributions, it is known that there is a constant relationship between the standard deviation and the distance between symmetric quantiles in the tails of the distribution. This inspires us to use a support vector quantile regression (SVQR) for volatility forecasts with the distance between CAViaR forecasts of symmetric quantiles. Simulated example and real example are provided to indicate the usefulness of proposed forecasting method for volatility.

Keywords: Conditional autoregressive value at risk model, cross validation, support vector quantile regression, volatility.

1. Introduction

Volatility forecasting is important for many financial market applications, including option pricing and investment decisions. The empirical finding that series of returns often exhibit volatility clustering has led to the development of a variety of univariate time series methods for volatility forecasting. The generalized autoregressive conditional heteroscedasticity (GARCH) model and the stochastic volatility model rely on the assumption that the conditional distribution does not vary over time but the volatility does. Parameters in GARCH model are usually estimated by maximum likelihood procedures which are optimal when the data set is drawn from a Gaussian distribution. If there is a variation of the distribution

† This work was supported by the Korea Research Foundation Grant funded by the Korean Government(MOEHRD, Basic Research Promotion Fund) (KRF-2008-313-C00159).
[1] Department of Data Science, Institute of Statistical Information, Inje University, Kimhae 621-749, Korea.
[2] Corresponding author, Department of Statistics, Dankook University, Yongin 448-701, Korea.
E-mail: chwang@dankook.ac.kr

over time, this may produce errors for the volatility forecasts produced by GARCH model. Perez-Cruz *et al.* (2003) proposed a volatility forecasting procedure by applying the support vector machine (SVM; Vapnik, 1998) to GARCH(1,1) model. The applications of SVM can be found in Hwang (2007), Hwang (2008), Kim *et al.* (2008), Shim and Lee (2010). CAViaR model proposed by Engle and Manganelli (2004) requires no distributional assumptions, which allow quantiles to be modelled directly in the autoregressive framework. The θ th quantile of a financial return r_t, is known as the value at risk (VaR), and is defined as $q_t(\theta)$ with $\theta = P(r_t \leq q_t(\theta))$. As VaR is a risk management tool, the quantiles of interest are in the tails of the distribution.

Pearson and Tukey (1965) showed that, for a variety of probability distributions, there is a constancy of the ratio of the standard deviation to the distance between symmetric tail quantiles, $q(\theta)$ and $q(1-\theta)$. For example, they showed that a simple approximation to the standard deviation is provided by the distance between $q(0.025)$ and $q(0.975)$ divided by 3.92 if the distribution is Gaussian. From this, we can see that even though conditional volatility and distribution of financial returns may vary over time, the conditional volatility can be approximated by a linear or nonlinear function of the distance between symmetric conditional quantiles. This provides us with a basis for constructing volatility forecasting from quantile forecasts produced by CAViaR model. GARCH model uses only an autoregressive model for the variance. For the case that the left and right tails of the conditional distribution are driven differently over time, our method can capture the evolution of the variance better than GARCH model.

In Sections 2 and 3, we briefly review CAViaR model and SVQR, respectively. Section 4 describes the volatility forecasting method by SVQR. In Section 5 we perform the numerical studies through the simulated example and the real example, and provides concluding remarks.

2. CAViaR model

In this paper, we denote the conditional variance of the log return r_t, at time t given D_{t-1} which is a set of informations gathered up to time t by $\sigma_t^2 = Var(r_t|D_{t-1})$. Error term ϵ_t is denoted by $\epsilon_t = r_t - E(r_t|D_{t-1})$, where $E(r_t|D_{t-1})$ is the conditional mean of r_t given D_{t-1}. CAViaR model involves direct autoregressive modelling of the conditional quantiles and thus does not involve distributional assumptions. Engle and Manganelli (2004) presented the following four CAViaR models:

$$\text{Adaptive CAViaR}: q_t(\theta) = q_{t-1}(\theta) + \alpha(\theta - I(\epsilon_{t-1} \leq q_t(\theta))) \tag{2.1}$$

$$\text{Asymmetric slope CAViaR}: q_t(\theta) = \omega + \alpha q_{t-1}(\theta) + \beta_1(\epsilon_{t-1})^+ + \beta_2(\epsilon_{t-1})^- \tag{2.2}$$

$$\text{Indirect GARCH(1,1) CAViaR}: q_t(\theta) = (1 - 2I(\theta < 0.5))\sqrt{\omega + \alpha q_{t-1}^2(\theta) + \beta\epsilon_{t-1}^2} \tag{2.3}$$

$$\text{Symmetric absolute value CAViaR}: q_t(\theta) = \omega + \alpha q_{t-1}(\theta) + \beta|\epsilon_{t-1}| \tag{2.4}$$

where $q_t(\theta)$ is the θ th conditional quantile, $I(\cdot)$ is the indicator function, $(x)^+ = \max(x, 0)$ and $(x)^- = \min(x, 0)$. Parameters involved in CAViaR models are estimated by minimizing

the sum of check functions introduced by Koenker and Bassett (1978),

$$\min \sum_{\mathrm{t}=1}^{\mathrm{n}} \rho_\theta(\mathrm{r_t} - \mathrm{q_t}(\theta)), \tag{2.5}$$

where $0 < \theta < 1$ and $\rho_\theta(e) = \theta e I(e \geq 0) + (1-\theta) e I(e < 0)$.

3. Support vector quantile regression

Let the training data set denoted by $\{\boldsymbol{x_t}, y_t\}_{t=1}^{n}$, with each input $\boldsymbol{x_t} \in R^d$ and the response $y_t \in R$, where the output variable y_t is related to the input vector $\boldsymbol{x_t}$. Here the feature mapping function $\boldsymbol{\phi}(\cdot) : R^d \to R^{d_f}$ maps the input space to the higher dimensional feature space where the dimension d_f is defined in an implicit way. An inner product in feature space has an equivalent kernel in input space, $\boldsymbol{\phi}(\boldsymbol{x_s})'\boldsymbol{\phi}(\boldsymbol{x_t}) = K(\boldsymbol{x_s}, \boldsymbol{x_t})$ (Mercer, 1909). Several choices of the kernel $\boldsymbol{K}(\cdot,\cdot)$ are possible. We consider the nonlinear regression case, in which the quantile regression function $q(\boldsymbol{x})$ of the response given $\boldsymbol{x}$ can be regarded as a nonlinear function of input vector $\boldsymbol{x}$.

With a check function $\rho_\theta(\cdot)$, the estimator of the θ th quantile regression function can be defined as any solution to the optimization problem,

$$\min \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + \mathrm{C}\sum_{\mathrm{t}=1}^{\mathrm{n}} \rho_\theta(\mathrm{y_t} - \mathrm{q}(\boldsymbol{x_t})) \tag{3.1}$$

where $\rho_\theta(e) = \theta e I(e \geq 0) + (1-\theta) e I(e < 0)$. We can express the regression problem by formulation for SVQR as follows,

$$\min \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + \mathrm{C}\sum_{\mathrm{t}=1}^{\mathrm{n}} (\theta\xi_\mathrm{t} + (1-\theta)\xi_\mathrm{t}^*) \tag{3.2}$$

subject to

$$y_t - \boldsymbol{w}'\boldsymbol{\phi}(\boldsymbol{x_t}) - b \leq \xi_t, \quad \boldsymbol{w}'\boldsymbol{\phi}(\boldsymbol{x_t}) + b - y_t \leq \xi_t^*, \xi_t, \xi_t^* \geq 0,$$

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

$$\begin{aligned} L =& \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{t=1}^{n}(\theta\xi_t + (1-\theta)\xi_t^*) - \sum_{t=1}^{n}\alpha_i(\xi_t - y_t + \boldsymbol{w}'\phi(\boldsymbol{x_t}) + b) \\ &- \sum_{t=1}^{n}\alpha_t^*(\xi_t^* + y_t - \boldsymbol{w}'\phi(\boldsymbol{x_t}) - b) - \sum_{t=1}^{n}(\eta_t\xi_t + \eta_t^*\xi_t^*). \end{aligned} \tag{3.3}$$

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

$$\max -\frac{1}{2}\sum_{\mathrm{s,t}=1}^{\mathrm{n}} \beta_\mathrm{s}\beta_\mathrm{t}\mathrm{K}(\boldsymbol{x_s}, \boldsymbol{x_t}) + \sum_{\mathrm{t}=1}^{\mathrm{n}} \beta_\mathrm{t}\mathrm{y_t} \tag{3.4}$$

with constraints

$$\beta_t = \alpha_t - \alpha_t^* \in [(\theta-1)C, \theta C],\ t = 1, \cdots, n.$$

Solving the above equation with the constraints determines the optimal Lagrange multipliers, β_t. The bias b_0 is obtained from

$$b_0 = \frac{1}{n_s}\sum_{t\in I_s}(y_t - \boldsymbol{k_t\beta}), \tag{3.5}$$

where $I_s = \{t = 1, \cdots, n|(\theta-1)C < \beta_t < \theta C \text{ and } \beta_t \neq 0\}$, n_s is the size of I_s and $\boldsymbol{k_t}$ is the tth row of the $n \times n$ kernel matrix $\boldsymbol{K}$ with the (i,j)th elements $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \boldsymbol{\phi}(\boldsymbol{x_i})'\boldsymbol{\phi}(\boldsymbol{x_j})$.

The estimator of the θth quantile given the input vector $\boldsymbol{x}_0$ are obtained as follows:

$$\widehat{q_0}(\theta) = \boldsymbol{k_0\beta} + b_0, \tag{3.6}$$

where $\boldsymbol{k_0} = (K(\boldsymbol{x_0}, \boldsymbol{x_1}), K(\boldsymbol{x_0}, \boldsymbol{x_2}), \cdots, K(\boldsymbol{x_0}, \boldsymbol{x_n}))$. In the nonlinear case, $\boldsymbol{w}$ is no longer explicitly given. However, it is uniquely defined in the weak sense by the dot products. Here the linear regression model can be regarded as the special case of the nonlinear regression model by using identity feature mapping function, that is, $\phi(\boldsymbol{x}) = \boldsymbol{x}$ which implies the linear kernel such that $K(\boldsymbol{x_1}, \boldsymbol{x_2}) = \boldsymbol{x_1}'\boldsymbol{x_2}$. Note that SVQR is called support vector median regression (SVMR) when $\theta = 0.5$.

The functional structures of SVQR is characterized by the hyperparameters consisting of the penalty parameter C and the kernel parameters. To select the hyperparameters of SVQR we consider the cross validation (CV) function as follows:

$$CV(\lambda) = \sum_{t=1}^{n}\rho_\theta(y_t - \widehat{q}_t^{(-t)}(\theta)), \tag{3.7}$$

where λ is the set of parameters and $\widehat{q}_t^{(-t)}(\theta)$ is the quantile function estimated without the t th observation. Since for each candidates of parameters, $\widehat{q}_t^{(-t)}(\theta)$ for $t = 1, \cdots, n$, should be evaluated, selecting parameters using CV function is computationally formidable. Yuan (2006) proposed GACV function to select the set of parameters λ for SVQR as follows:

$$GACV(\lambda) = \frac{\sum_{t=1}^{n}\rho_\theta\left(y_t - \widehat{q}_t^{(-t)}(\theta)\right)}{n - trace(H)}, \tag{3.8}$$

where H is the hat matrix such that $\widehat{\boldsymbol{q}}(\theta) = H\boldsymbol{y}$ with the (i, j)th element $h_{ij} = \partial\widehat{q}_i(\theta)/\partial y_j$. From Li *et al.* (2007) we have that the trace of the hat matrix H equals to the size of $I_s = \{t = 1, \cdots, n|(\theta-1)C < \alpha_t < \theta C \text{ and } \alpha_t \neq 0\}$.

4. Volatility forecasting via CAViaR based on SVQR

Pearson and Tukey (1965) found that the ratio of the standard deviation to the distance between symmetric quantiles in the tails of the distribution is constant for a variety of

distributions. They considered 98%, 95% and 90% intervals, and proposed the following simple approximations for the standard deviation in terms of estimated symmetric quantiles,

$$\widehat{\sigma} = \frac{q(0.975) - q(0.025)}{3.92}, \ \widehat{\sigma} = \frac{q(0.95) - q(0.05)}{3.29}$$

For a Gaussian distribution, the denominators above become $2 \times 1.960 = 3.92$ and $2 \times 1.645 = 3.29$, respectively. Pearson and Tukey (1965) showed that the accuracy of these approximations depends on the values of skewness and kurtosis of the given distribution. They found that the approximation based on the 90% interval was the most robust to different values of skewness and kurtosis.

Pearson and Tukey (1965) inspires us to consider that the squared volatility σ_t^2 can be obtained as a function of $(q_t(\theta) - q_t(1-\theta))^2$ if we have found $q_t(\theta)$ and $q_t(1-\theta)$. Thus the forecasting procedure of volatility by SVQR consists of two stages - forecasting conditional quantiles and forecasting volatility from obtained conditional quantiles.

First, we assume that conditional quantile $q_t(\theta)$ is an unknown nonlinear function of r_{t-1} and $q_{t-1}(\theta)$ such as $q_t(\theta) = f(r_{t-1}, q_{t-1}(\theta))$, which is similar to CAViaR model. Here r_t denotes the log return of a portfolio at time t. But the estimate of conditional quantile $\widehat{q_t}(\theta)$ is obtained as $\widehat{q_t}(\theta) = \boldsymbol{k_t}\boldsymbol{\beta} + b$ by the iteration method using SVQR is given as follows,

(i) Find $\hat{q}_t^{(0)}(\theta) = K(\boldsymbol{x}_t, \boldsymbol{x})\boldsymbol{\beta} + b$ from $\boldsymbol{y} = \{r_t\}_{t=1}^n$ and $\boldsymbol{x} = \{r_{t-1}\}_{t=1}^n$.
(ii) Find $\hat{q}_t^{(l+1)}(\theta) = K(\boldsymbol{x}_t, \boldsymbol{x})\boldsymbol{\beta} + b$ from $\boldsymbol{y} = \{r_t\}_{t=1}^n$ and $\boldsymbol{x} = \{r_{t-1}, \widehat{q}_{t-1}^{(l)}(\theta)\}_{t=1}^n$.
(iii) Iterate until $\sum_t |\hat{q}_t^{(l+1)}(\theta) - \hat{q}_t^{(l)}(\theta)| <$ tolerance.

The estimate of conditional quantile $\widehat{q_t}(1-\theta)$ also can be obtained by the procedure above.

Next, we assume that σ_t^2 is a nonlinear function of e_{t-1}^2 and $(q_{t-1}(\theta) - q_{t-1}(1-\theta))^2$ such as $\sigma_t^2 = f(e_{t-1}^2, (q_{t-1}(\theta) - q_{t-1}(1-\theta))^2)$ where $e_{t-1} = r_{t-1} - \hat{q}_{t-1}(0.5)$. Here the estimate of volatility is obtained by SVQR as $\widehat{\sigma_t} = (K(\boldsymbol{x_t}, \boldsymbol{x})\boldsymbol{\beta} + b))^{1/2}$ where the response $\widetilde{\sigma_t^2} = \frac{1}{5}\sum_{k=0}^4 (e_{t-k})^2$ and $\boldsymbol{x_t} = (e_{t-1}^2, (q_{t-1}(\theta) - q_{t-1}(1-\theta))^2)$.

5. Numerical studies and conclusions

We illustrate the performance of the volatility forecasting method based on SVQR through one simulated data set and one real data set by comparing with Perez-Cruz *et al.* (2003) and GARCH(1,1) model.

Example 5.1. For the simulated example, we consider the autoregressive heteroscedastic model,

$$y_1 = -\sigma_1 \Phi^{-1}(\theta) + e_1, \ y_t = q_t(\theta) - \sigma_t \Phi^{-1}(\theta) + e_t, \ t = 2, \cdots, 100$$

where σ_1 is generated from $U(0,1)$, e_t is generated from $N(0, \sigma_t^2)$, $\sigma_t = \exp(0.5\sin(y_{t-1}\sigma_{t-1}))$, $q_t(\theta) = \sin(\pi y_{t-1}) + \cos(\pi q_{t-1}(\theta))$. We can see that $q_t(0.5) = q_t(\theta) - \sigma_t \Phi^{-1}(\theta)$ and $q_t(\theta) - q_t(1-\theta) = \sigma_t \Phi^{-1}(\theta)$. We set $\theta = 0.1$ and utilize the RBF kernel functions in this example. Optimal values of penalty constant and kernel parameter are chosen by GACV function (3.8). The plot on the left panel of Figure 5.1 illustrates the true volatilities (solid line) and the estimated volatilities (dotted line) by the proposed method. The plot on the

right panel of Figure 5.1 illustrates the estimated volatilities (solid line) by Perez-Cruz *et al.* (2003) and the estimated volatilities (dotted line) by GARCH(1,1), both of which are imposed on the scatter plots of 100 data points of y_t's in a data set. Note that since $E(y_t)$ is not constant, $y_t - \widehat{q}_t(0.5)$ is used as the log return r_t and the response for Perez-Cruz *et al.* (2003) and GARCH(1,1) model. In Figure 5.1 we can see that the estimated volatilities by the proposed method seem to represent better the behavior of volatilities of given data than other methods.

We repeated the above procedure 100 times to obtain the root mean squared errors (RMSE) for the performance metric as follows,

$$RMSE = \sqrt{\frac{1}{100}\sum_{t=1}^{100}(\widehat{\sigma}_t - \sigma_t)^2}.$$

For the proposed method we obtained the average of 100 *RMSE*'s and their standard error as 0.5849 and 0.0095, respectively. For Perez-Cruz *et al.* (2003) we obtained the average of 100 *RMSE*'s and their standard error as 0.7329 and 0.0098, respectively. For GARCH(1,1) we obtained the average of 100 *RMSE*'s and their standard error as 0.6220 and 0.0109, respectively. The smaller value of average of *RMSE*'s indicates that the proposed method works better than other methods in this example.

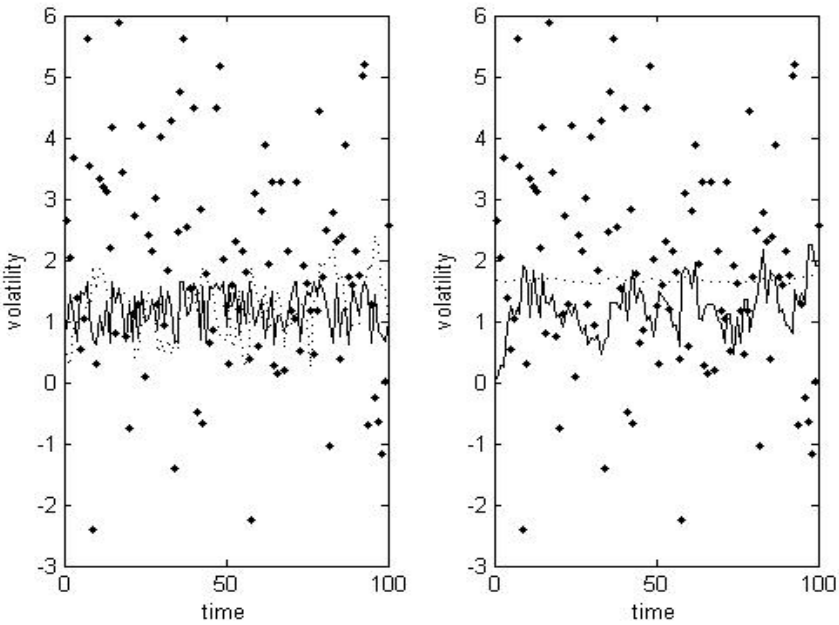

Figure 5.1 (Left): True volatilities (solid) and their estimates by the proposed method (dotted). (Right): volatility estimates by Perez-Cruz *et al.* (2003) (solid) and GARCH(1,1) model (dotted) for a data set in Example 5.1.

Example 5.2. We consider 372 daily log-returns (in percentages) from the US S&P500 index between January 1990 and June 1991. We set $\theta = 0.05$ and utilize the RBF kernel

function in this example. In forecasting quantiles, the optimal values of hyperparameters are chosen by GACV function (3.8) such as $(C, s) = (100, 1)$ for SVQR($\theta = 0.05$), $(C, s) = (100, 0.5)$ for SVQR($\theta = 0.95$), and $(C, s) = (100, 1)$ for SVMR. In forecasting the squared volatility the optimal values of hyperparameters are chosen by GACV function (3.8) such as $(C, s) = (100, 1)$ for SVMR. The plot on the left panel of Figure 5.2 illustrates the true volatilities (solid line) and the estimated volatilities (dotted line) by the proposed method. The plot on the right panel of Figure 5.2 illustrates the estimated volatilities (solid line) by Perez-Cruz *et al.* (2003) and the estimated volatilities (dotted line) by GARCH(1,1), both of which are imposed on the scatter plots of 372 data points of r_t's in a data set. We use the root mean squared errors (RMSE) for the performance metric follows,

$$RMSE = \sqrt{\frac{1}{3}72\sum_{t=1}^{372}(\widehat{\sigma}_t - \sigma_t)^2}.$$

For the proposed method we obtained *RMSE* as 4.3577. For Perez-Cruz *et al.* (2003) and GARCH(1,1) we obtained *RMSE*'s as 5.1134 and 4.4316, respectively. The smaller value *RMSE* indicates that the proposed method works better than other methods in this example.

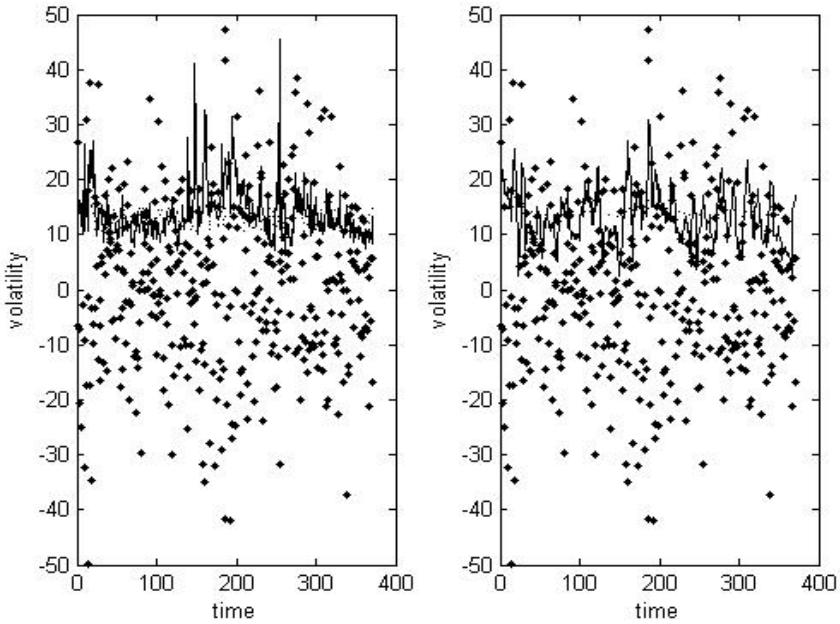

Figure 5.2 (Left): True volatilities (solid) and their estimates by the proposed method (dotted). (Right): volatility estimates by Perez-Cruz *et al.* (2003) (solid) and GARCH(1,1) model (dotted) for a data set in Example 5.2.

In this paper we dealt with forecasting the volatility based on SVQR. Through the examples we showed that the proposed method derives more satisfying results on forecasting the volatility than other methods. We also found that the optimal values of hyperparmeters in the proposed method can be obtained by model selection method such as GACV function.

References

Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, **22**, 367-381.

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

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

Kim, M., Park, H., Hwang, C. and Shim, J. (2008). Claims reserving via kernel machine. *Journal of the Korean Data & Information Science Society*, **19**, 1419 -1427.

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

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 the theory of integral equations. *Philosophical Transactions of the Royal Society*, **A**, 415-446.

Pearson, E. S. and Tukey, J. W. (1965). Approximate means and standard deviations based on distances between percentage points of frequency curves. *Biometrika*, **52**, 533-546.

Perez-Cruz, F., Afonso-Rodrıguez, J. A. and Giner, J. (2003). Estimating GARCH models using support vector machines. *Quantitative Finance*, **3**, 1-10.

Shim, J. and Lee, J. T. (2010). Restricted support vector quantile regression without crossing. *Journal of the Korean Data & Information Science Society,* **21**, 1319 -1325.

Vapnik, V. (1998). *Statistical learning theory*, Wiley, New York.

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