Scientific
Research

Recurrent Support and Relevance Vector Machines Based Model with Application to Forecasting Volatility of Financial Returns

Altaf Hossain[1*], Mohammed Nasser[2]

[1]Department of Statistics, Islamic University, Kushtia, Bangladesh; [2]Department of Statistics, Rajshahi University, Rajshahi, Bangladesh.
Email: [*]rasel_stat71@yahoo.com, mnasser.ru@gmail.com

ABSTRACT

In the recent years, the use of GARCH type (especially, ARMA-GARCH) models and computational-intelligence-based techniques—Support Vector Machine (SVM) and Relevance Vector Machine (RVM) have been successfully used for financial forecasting. This paper deals with the application of ARMA-GARCH, recurrent SVM (RSVM) and recurrent RVM (RRVM) in volatility forecasting. Based on RSVM and RRVM, two GARCH methods are used and are compared with parametric GARCHs (Pure and ARMA-GARCH) in terms of their ability to forecast multi-periodically. These models are evaluated on four performance metrics: MSE, MAE, DS, and linear regression R squared. The real data in this study uses two Asian stock market composite indices of BSE SENSEX and NIKKEI225. This paper also examines the effects of outliers on modeling and forecasting volatility. Our experiment shows that both the RSVM and RRVM perform almost equally, but better than the GARCH type models in forecasting. The ARMA-GARCH model is superior to the pure GARCH and only the RRVM with RSVM hold the robustness properties in forecasting.

Keywords: RSVM, RRVM, ARMA-GARCH, Outliers, Volatility Forecasting

1. Introduction

In financial markets, volatility is important as its forecasts on stock price are crucial for portfolio selection, pricing derivatives, calculating measure of risk and hedging strategy. A risk manager must know today the likelihood that his portfolio will decline in the future and he may want to sell it before it becomes too volatile [1]. According to Merton [2], expected market return is related to predictable stock market volatility. Due to the necessity of volatility prediction, a large number of time series based volatility models have been developed since the induction of ARCH model of Engle [3]. Later Bolleslev [4] generalized the model as GARCH to capture a higher order of ARCH; See Ref. [5] for review and references. To deal with the intricacy specially, Wong *et al.* [6] adopted the well-known GARCH model in the form of the so-called mixture of AR-GARCH model in exchange rate prediction. Again, Tang *et al.* [7] explored the mixture of ARMA-GARCH model for stock price prediction; See [8] for more details. Evidence on the forecasting ability of the GARCH model is somewhat mixed. An-

derson and Bollerslev [9] showed that the GARCH model provides good volatility forecast. Conversely, some empirical studies showed that the GARCH model tends to give poor forecasting performances [10-15].

To obtain more accurate predictions, recently, machine learning approaches have been successfully introduced to predict volatility based on various models of GARCH family. For example, Ref. [16] for Neural Network based GJR model, Ref. [17]: SVM based GARCH; Ref. [18]: RVM based GARCH, EGARCH and GJR; Ref. [19-20] for SVM based GARCH with wavelet and spline wavelet kernels, and Ref. [21] for Neural Network based on nine different models of GARCH family. The neural network suffers from overfitting problems and the algorithm can result in a local minima solution which is not unique [22]. In this regard, Support Vector Machine developed by Vapnik [23] is a novel neural network algorithm model with various applications to prediction problems [24-28]. The algorithm results in the globally optimum solution. The SVM algorithm, based on structural risk minimization, is equivalent to solving a convex programming problem

where the unique result is always obtained. Moreover, with the help of kernel function satisfying Mercer's conditions, the difficulty of working with a nonlinear mapping function in high dimensional space is removed [29].

The RVM, an alternative method of SVM, is a probabilistic model introduced by Tipping in 2000. The RVM has recently become a powerful tool for prediction problems. One of the main advantages is that the RVM has functional form identical to SVM and hence it enjoys various benefits of SVM based techniques: generalization and sparsity. On the other hand, RVM avoids some disadvantages faced by SVM such as the requirement to obtain optimal value of regularized parameter, C, and epsilon tube; SVM needs to use Mercer's kernel function and it can generate point prediction but not distributional prediction in RVM [30]. Tipping [30] illustrated the RVM's predictive ability on some popular benchmarks by comparing it with the SVM. The empirical analysis also proved that the RVM outperformed SVM; some other applications of RVM in prediction problems are referred to [31-33].

Chen *et al.* [34] applied SVM to model and forecast GARCH (1, 1) volatility based on the concept of recurrent SVM (RSVM) in Chen *et al.* [35], following from the recurrent algorithm of neural network and least square SVM of [36]. Accordingly, Ou and Wang [37] proposed RVM (as recurrent RVM) to model and forecast GARCH (1, 1) volatility based on the concept of recurrent SVM in [34,35]. The models were shown to be a dynamic process and capture long memory of past information than the feed-forward SVM and RVM which are just static.

Multi-period forecasts of stock market return volatilities are often used in many applied areas of finance where long horizon measures of risk are necessary. Yet, very little is known about how to forecast variances several periods ahead, as most of the focus has been placed on one-period-ahead forecasts. In this regard, only Chen *et al*. [34] considered multi-period-ahead forecasting with one-period-ahead forecasting. They showed that multi-period-ahead forecasting method performs better than the counterpart in forecasting volatility. Specifically, Ou and Wang [37] did not consider multi-period-ahead method in forecasting volatility by RRVM. Yet, none of them investigated the above models' (GARCH type, RSVM and RRVM) combination in the context of two Asian stock market (emerging) composite indices: BSE SENSEX and NIKKEI225. It is important for us to forecast the BSE and NIKKEI225 markets volatility more accurately for recent potential growth of the markets.

Our first contribution is to deal with the application of ARMA-GARCH with pure GARCH, RSVM and RRVM

in volatility forecasting of **multi-period-ahead**. Based on RSVM and RRVM, two GARCH methods are used and are compared with parametric GARCHs (Pure and ARMA-GARCH) in terms of their ability to forecast volatility of two Asian stock market (emerging) composite indices: BSE SENSEX and NIKKEI225.

Secondly, being inspired by Tang *et al.* [7], we put more emphasis on the comparison between the ARMA-GARCH and pure GARCH models in forecasting volatileity of emerging stock market returns.

Of increasing importance in the time series modeling and forecasting is the problem of outliers. Volatility of e-merging stock market returns poses especial challenges in this regard. In sharp contrast to the well developed stock markets, emerging markets are generally characterized by high volatility. In addition, high volatility in these markets is often marked by frequent and erratic changes, which are usually driven by various local events (such as political developments) rather than by the events of global importance [38,39]. Outliers in time series were first studied by Fox in 1972. The outliers, which are really independent, are the situations that cause the parameter estimation values in classical modeling (ARMA and GARCH type) to be subjective, they damage the processes even though they are set properly and it is an obligation to destroy or to eliminate the effects. They diminish the reliability of the results; see Ref. [40-43] for more details. Outliers may affect forecasts through the carryover effect on the ARCH and GARCH terms, and may have a permanent effect on the parameter estimates. There are different types of outliers (like innovational and additive outlier) with different criteria (like Likelihood Ratio and Lagrange Multiplier) for detecting them in conventional time series volatility (GARCH type) modeling; for example, [43-45], etc. But the outliers are not classified in this paper. Also the numerical tests (like Likelihood Ratio and Lagrange Multiplier) are not used to detect the outliers in this paper; rather we use a graphical (Quantile-Quantile) test to detect general outliers very simply. Despite the voluminous research that examines the effects of outliers on the properties of the GARCH type models, no attention has been given to the effects of outlying observations in the combination of GARCH type models and computational-intelligence-based techniques (SVM and RVM) in forecasting financial volatility of emerging stock market returns.

Thirdly, we are to re-examination the effects of outliers on the ACFs, descriptive statistics, and classical tests (Ljung-Box Q and ARCH-LM) in context of emerging stock markets.

Finally, we check the impact of outliers or unusual observations in the model estimation and forecasting, that is,

examining the robustness properties of RSVM and RRVM compared with GARCH type model, especially, in forecasting volatility in the presence of outliers.

The remainder of the paper is organized as follows. The next two sections review the SVM and RVM algorithms. Section 4 specifies the empirical model and forecasting scheme. Section 5 describes the BSE SENSEX and NIKKEI225 composite index data and discusses the volatility forecasting performance of all models. Finally the conclusion is made in section 6.

2. Support Vector Machines

The SVM deals with the classification and regression problems by mapping the input data into the higher-dimensional feature spaces. In this paper, the SVM deals only with the regression problems. Its central feature is that the regression surface can be determined by a subset of points or Support-Vectors (SV); all other points are not important in determining the surface of the regression. Vapnik introduced a ε-insensitive zone in the error loss function (**Figure 1**). Training vectors that lie within this zone are deemed correct, whereas those that lie outside the zone are deemed incorrect and contribute to the error loss function. As with classification, these incorrect vectors also become the support vector set. Vectors lying on the dotted line are SV, whereas those within the ε-insensitive zone are not important in terms of the regression function.

The SVM algorithm tries to construct a linear function such that training points lie within a distance ε (**Figure1**). Given a set of training data $\{(x_1,y_1),\cdots,(x_n,y_n)\}\subset X\times R$, where X denotes the space of the input patterns, the goal of SVM is to find a function $f(x)$ that has at most ε deviation from the targets y_i for all the training data and, at the same time, is as flat as possible.

Let the linear function f takes the form:

$$f(x)=\langle w,x\rangle+b; w\in X, b\in R \qquad (1)$$
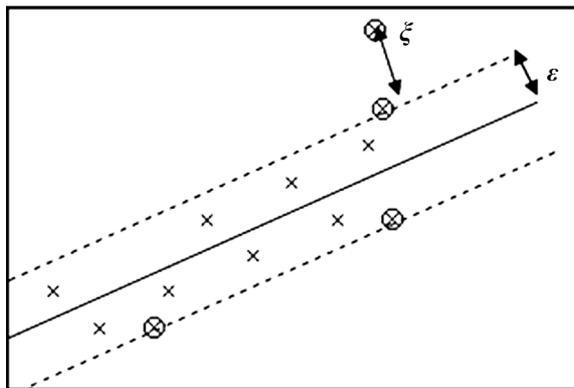
Figure 1. Approximation function (solid line) of SV regression using a ε-insensitive zone.

The optimal regression function is given by the minimum of the functional,

$$\Phi(w,\xi)=\frac{1}{2}\|w\|^2+C\sum_i\left(\xi_i^-+\xi_i^+\right) \qquad (2)$$

where C is pre-specified value, and ξ^-,ξ^+ are slack variables representing upper and lower constraints on the outputs of the system. Flatness in (1) means a smaller $\|w\|$. Using an ε-insensitive loss function,

$$L_\varepsilon(y)=\begin{cases}0 & \text{for } |f(x)-y|<\varepsilon \\ |f(x)-y|-\varepsilon & \text{otherwise}\end{cases} \qquad (3)$$

the solution is given by,

$$\max_{\alpha,\alpha^*}W(\alpha,\alpha^*)=\max_{\alpha,\alpha^*}-\frac{1}{2}\sum_{i,j=1}^n(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\langle x_i,x_j\rangle$$
$$+\sum_{i=1}^n\alpha_i(y_i-\varepsilon)-\alpha_i^*(y_i+\varepsilon)$$

$$(4)$$

with constraints,

$$0\le\alpha_i,\alpha_i^*\le C,\quad i=1,2,\cdots,l$$
$$\sum_{i=1}^n(\alpha_i-\alpha_i^*)=0 \qquad (5)$$

Solving equation of (4) with constraints Equation (5) determine the Lagrange multipliers, α,α^* and the regression function is given by (1), where

$$\bar{w}=\sum_{i=1}^n(\alpha_i-\alpha_i^*)x_i$$
$$\bar{b}=-\frac{1}{2}\langle\bar{w},(x_r+x_s)\rangle \qquad (6)$$

w is determine by training patterns x_i, which are SVs. In a sense, the complexity of the SVM is independent of the dimensions of the input space because it only depends on the number of SV. To enable the SVM to predict a non-linear situation, we map the input data into a *feature space*. The mapping to the feature space F is denoted by

$$\phi:\Re^n\to F$$
$$x\mapsto\phi(x)$$

The optimization Equation (4) can be written as

$$\max_{\alpha,\alpha^*}W(\alpha,\alpha^*)=$$

$$\max_{\alpha,\alpha^*}-\frac{1}{2}\sum_{i,j=1}^n(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\langle\phi(x_i),\phi(x_j)\rangle$$

$$+\sum_{i=1}^n\alpha_i(y_i-\varepsilon)-\alpha_i^*(y_i+\varepsilon)$$

The decision can be computed by the inner products,

$\langle \phi(x_i), \phi(x_j) \rangle$ without explicitly mapping to a higher dimension which is a time-consuming task. Hence the kernel function is as follows:

$$K(x,z) \equiv \langle \phi(x), \phi(z) \rangle$$

By using a kernel function, it is possible to compute the SVM without explicitly mapping in the feature space.

3. Relevance Vector Machines

Let $\{x_i, t_i\}_{i=1}^n$ be a training data. The goal is to model the data by a function indexed by parameters defined as

$$y(x;w) = \sum_{j=1}^m w_j \varphi_j(x) = w^T \phi(x), \tag{7}$$

where basis function $\phi(x) = (\phi_1(x), \cdots, \phi_m(x))^T$ is non-linear, $w = (w_1, \cdots, w_m)^T$ is weight vector and x in input vector. Hence the target is sum of the function and error term:

$$t_i = y_i + \varepsilon_i \tag{8}$$

and vector form of the target is written as

$$t = y + \varepsilon \tag{9}$$

For simplicity, ε_i are assumed to follow independent Gaussian process with mean zero and variance σ^2. The likelihood of the complete dataset corresponding to (8) is obtained as the following

$$p(t_i/w, \sigma^2) = N(y_i, \sigma^2)$$
$$= (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma^2}(t_i - y_i)^2 \right\}.$$

For the n simultaneously training points,

$$p(t/w, \sigma^2) = \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma^2}(t_i - y_i)^2 \right\}$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \|t - \Phi w\|^2 \right\}, \tag{10}$$

where $t = (t_1, \cdots, t_n)^T$, $w = (w_1, \cdots, w_m)^T$ and Φ is (n × m) design matrix with $\Phi = [\phi(x_1), \cdots, \phi(x_n)]^T$ and $\phi(x_i) = [1, K(x_i, x_1), \cdots, K(x_i, x_n)]^T$. As [46], to avoid overfitting problems which may be caused by the Maximum likelihood estimation of w and σ^2, zero mean Gaussian prior over the weights w is introduced,

$$P(w/\alpha) = \prod_{i=0}^n N(w_i/0, \alpha_i^{-1}) \tag{11}$$

where α_i is the ith element of vector hyperparameter α assigned to each model parameter w_i.

By Bayes rule,

$$P(w, \alpha, \sigma^2/t) = \frac{P(t/w, \alpha, \sigma^2) P(w, \alpha, \sigma^2)}{P(t)} \tag{12}$$

The posterior in (12) cannot calculated directly as denominator of (12) contain normalizing integral *i.e.*,

$$P(t) = \int P(t/w, \alpha, \sigma^2) P(w, \alpha, \sigma^2) dw d\alpha d\sigma^2 .$$

However, the posterior can be decomposed as

$$P(w, \alpha, \sigma^2/t) = P(w/t, \alpha, \sigma^2) P(\alpha, \sigma^2/t) \tag{13}$$

Now, the first term of (13) can be written as below

$$P(w/t, \alpha, \sigma^2) = \frac{P(t/w, \sigma^2) P(w/\alpha)}{P(t/\alpha, \sigma^2)} .$$

Noticeably, $P(t/\alpha, \sigma^2) = \int P(t/w, \sigma^2) P(w/\alpha) dw$ is convolution of Gaussians, $P(w/\alpha)$ is Gaussian prior and $P(t/w, \sigma^2)$ is also Gaussian likelihood by (10), implying the posterior $P(w/t, \alpha, \sigma^2)$ is Gaussian which obtained as [47],

$$P(w/t, \alpha, \sigma^2) =$$
$$(2\pi)^{-\frac{(n+1)}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(w - \mu)^T \Sigma^{-1}(w - \mu) \right\} \tag{14}$$

with covariance

$$\Sigma = (\sigma^{-2} \Phi^T \Phi + A)^{-1} \tag{15}$$

and mean

$$\mu = \sigma^{-1} \Sigma \Phi^T t \tag{16}$$

where $A = \text{diag}(\alpha_o, \alpha_1, \cdots, \alpha_n)$.

In order to evaluate μ and Σ we need to find the hyperparameters α and σ^2 which maximize the second term of (13):

$$P(\alpha, \sigma^2/t) \propto P(t/\alpha, \sigma^2) P(\alpha) P(\sigma^2).$$

For uniform hyperprior, it is just required to maximize the term $P(\alpha, \sigma^2/t)$ with respect to α and σ^2 by ignoring $P(\alpha)$ and $P(\sigma^2)$. Then, the problem becomes

$$P(t/\alpha, \sigma^2) = \int P(t/w, \sigma^2) P(w/\alpha) dw$$
$$= (2\pi)^{-\frac{n}{2}} |\sigma^2 I + \Phi A^{-1} \Phi^T|^{-\frac{1}{2}} \tag{17}$$
$$\exp\left\{ -\frac{1}{2} t^T (\sigma^2 I + \Phi A^{-1} \Phi^T)^{-1} t \right\}.$$

This is called marginal likelihood which needs to maximize with respect to α and σ^2. The maximization process is known as type II maximum likelihood method or evidence procedure. The hyperparameters are estimated by iterative method as it cannot be obtained in closed form.

As from [30], the solutions are obtained,

$$\alpha_i^{\text{new}} = \frac{\gamma_i}{\mu_i^2}. \tag{18}$$

by differentiating (17) and equating to zero. μ_i is the ith posterior mean weight from (16) and $\gamma_i \equiv 1 - \alpha_i$. $\Sigma_{ii} \in [0,1]$ can be interpreted as a measure of well determinedness of each parameter w_i. Σ_{ii}. is the ith diagonal element of the posterior weight covariance in (15) computed with the current α and σ^2 values.

Another differentiation with respect to σ^2 leads to

$$\left(\sigma^2\right)^{\text{new}} = \frac{\|t - \Phi\mu\|^2}{n - \Sigma_i \gamma_i}. \tag{19}$$

The learning algorithm is applied repeatedly to (18) and (19) with updating of Σ and μ in (15) and (16) until suitable convergence criteria are obtained.

During the re-estimation, many α_i tend to infinity such that w will have a few nonzero weights that will be considered as relevance vectors and analogous to the support vectors of SVM. Thus the resulting model enjoys the properties of SVM such as sparsity and generalizetion.

Given a new input x^*, the probability distribution of the output is given by the predictive distribution

$$P\left(t^*/t, \alpha_{MP}, \sigma_{MP}^2\right) = \int P\left(t^*/w, \sigma_{MP}^2\right) P\left(w/t, \alpha_{MP}, \sigma_{MP}^2\right) dw$$
$$= N\left(t^*/y^*, \sigma_*^2\right)$$

which has the Gaussian form with mean $y^* = \mu^T \phi\left(x^*\right)$ and variance $\sigma_*^2 = \sigma_{MP}^2 + \phi\left(x^*\right)^T \Sigma \phi\left(x^*\right)$. So the predicttive mean is $y\left(x^*; \mu\right)$ and the predictive variance composes of two variance components.

4. Empirical Modeling and Forecasting Scheme

In this paper, the data we analyze is just the daily financial returns, y_t, converted from the corresponding price or index, p_t, using continuous compounding transformation as

$$y_t = 100 \cdot \left(\ln p_t - \ln p_{t-1}\right) \tag{20}$$

A GARCH (1, 1) specification is the most popular form for modeling and forecasting the conditional variance of return of volatility, [48]. Therefore, we consider GARCH (1, 1) model throughout our paper.

4.1. The Linear Pure GARCH/ARMA-GARCH Model

The basic Linear "pure" GARCH (1, 1) model

$$y_t = \sigma_t \varepsilon_t \tag{21}$$

$$\sigma_t^2 = w + \beta_1 \sigma_{t-1}^2 + \alpha_1 y_{t-1}^2 \tag{22}$$

$$\varepsilon_t \sim IID(0,1)$$

where σ_t^2 is their conditional variance.

The basic Linear ARMA(p, q)-GARCH(1, 1) model

$$y_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \theta_1 u_{t-1} + \ldots + \theta_q u_{t-q} + u_t \tag{23}$$

$$u_t = \sigma_t \varepsilon_t = y_t - \left(\mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \theta_1 u_{t-1} + \ldots + \theta_q u_{t-q} + u_t\right)$$

$$\sigma_t^2 = w + \beta_1 \sigma_{t-1}^2 + \alpha_1 u_{t-1}^2 \tag{24}$$

when $p = 1$ and $q = 0$, then it is reduced to AR(1)-GARCH(1, 1) process.

The important point is that the conditional variance of u_t is given by $\sigma_t = E_{t-1} u_t^2 = \hat{u}_{t/t-1}^2$. Thus, the conditional variance of u_t is the ARMA process given by the expression σ_t^2 in the equation (22 or 24) [4,49,50].

$$u_t^2 = w + \left(\alpha_1 + \beta_1\right) u_{t-1}^2 + w_t - \beta_1 w_{t-1} \tag{25}$$

$$w_t \equiv u_t^2 - \hat{u}_{t/t-1}^2 = u_t^2 - \sigma_t^2$$

where w_t is white noisy errors. The parameters w, β_1 and α_1 must satisfy $w > 0$, $\beta_1 \geq 0$, and $\alpha_1 \geq 0$ to ensure that the conditional variance is positive. Together with the nonnegative assumption, if $\alpha_1 + \beta_1 < 1$, then u_t^2 is covariance stationary.

4.2. Recurrent SVM/RVM Based GARCH Model

For recurrent SVM or RVM methods, the nonlinear AR (1)-GARCH (1, 1) model has the following form:

$$y_t = f\left(y_{t-1}\right) + u_t \tag{26-1}$$

$$u_t^2 = g\left(u_{t-1}^2, w_{t-1}\right) + w_t \tag{26-2}$$

The algorithm of the recurrent SVM or RVM-based GARCH model is described as follows:

Step 1: fit SVM (or RVM) to the return y_t as AR (1) format in the full sample period N,

$$y_t = f\left(y_{t-1}\right) + u_t; \quad t = 1, 2, \ldots, N,$$

to obtain residuals, $u_1, u_2, \ldots, u_N$.

Step 2: run the recurrent SVM (or RVM) for squared residuals, $u_1^2, u_2^2, \ldots, u_n^2$ ($N_1 < N$) without updating,

$$u_t^2 = g\left(u_{t-1}^2, w_{t-1}\right) + w_t$$

to obtain n multi-period-ahead forecasted volatilities:

$$\hat{u}_{N1+1}^2, \hat{u}_{N1+2}^2, \ldots, \hat{u}_{N1+n}^2.$$

For estimations, set the residuals of w_t to be zero at the first time in the Step 2, and then run the feed-forward SVM (or RVM) to obtain estimated residuals. Using the estimated residuals as new w_{t-1} inputs, this process can be carried out repeatedly until the stopping criterion is satisfied. Unlike the parametric case, by using the proposed

approach we don't need any assumption on the model parameters for stationary condition. We use R packages: "e1071" and "kernlab" to model and forecast SVM and RVM, respectively in the experiment.

4.3. Evaluation Measures and Proxy of Actual Volatility

Although the Mean Square Error

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(u_i^2 - \hat{u}_i^2\right)^2$$

is a perfectly acceptable measure of performance, in practice the ultimate goal of any testing strategy is to confirm that the results of models are robust and capable of measuring the profitability of a system. It is important, therefore, to design a test from the outset. According to [51-53], the prediction performance is also evaluated using the following statistics: Mean Absolute Error (MAE) and Directional Symmetry (DS), expressed as follows [54,55]:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|u_i^2 - \hat{u}_i^2\right|$$

$$\mathrm{DS}(\%) = \frac{100}{n}\sum_{i=1}^{n}a_i,$$

$$\text{where, } a_i = \begin{cases} 1, \left(u_i^2 - u_{i-1}^2\right)\left(\hat{u}_i^2 - \hat{u}_{i-1}^2\right) \\ 0, \qquad\qquad \text{Otherwise} \end{cases}$$

MAE measures the average magnitude of forecasting error which disproportionately weights large forecast errors more gently relative to MSE; and DS measures the correctness of the turning points forecasts, which gives a rough indication of the average direction of the forecasted volatility.

Also linear regression technique is employed to evaluate the forecasting performance of the volatility models. We simply regress squared return on the forecasted volatility for out-of-sample time point; the squared correlation R^2 is a measure of forecasting performance. We report the proportion of the sample variation explained by the forecasts with the R^2 statistic [56] defined by

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(u_i^2 - \hat{u}_i^2\right)^2}{\sum_{i=1}^{n}\left(u_i^2 - \frac{1}{n}\sum_{i=1}^{n}u_i^2\right)^2}$$

The fundamental problem with the evaluation of volatility forecasts of real data is that volatility is unobservable and so actual values, with which to compare the forecasts, do not exist. Therefore, researchers are necessarily required to make an auxiliary assumption about how the actual ex post volatility is calculated. In this pa-

per, we use square of the return assuming its mean value equal to zero as the proxy of actual volatility against which MSE, MAE, DE and R^2 can be calculated; because this approach is the standard one, following from the previous research of [17,18,37]. The proxy of actual volatility in real data is expressed as $u_t^2 = y_t^2$, where y_t: returns.

5. Empirical Results

5.1. Data Description

We examine Bombay Stock Exchange (BSE) SENSEX Index of India Stock Market and NIKKEI225 of Japan Stock Market in the experiment. It is important for us to forecast the BSE SENSEX and NIKKEI225 markets volatility more accurately. Recently the potential growth of these two markets has attracted foreign and local investors. The BSE index has increased by over ten times from June 1990 to the present. Using information from April 1979 onwards, the long-run rate of return on the BSE SENSEX works out to be 18.6% per annum, which translates to roughly 9% per annum after compensating for inflation. The NIKKEI225 average has deviated sharply from the textbook model of stock averages which grow at a steady exponential rate. The average hit its all-time high on December 29, 1989, during the peak of the Japanese asset price bubble, when it reached an intra-day high of 38957.44 before closing at 38915.87, having grown six-fold during the decade. Subsequently it lost nearly all these gains, closing at 7054.98 on March 10, 2009—81.9% below its peak twenty years earlier.

The stock index prices are collected from Yahoo Finance and are transformed into log returns before making analysis. For BSE, the whole sample of size 1000, spanned from 05 Oct. 2006 to 01 Nov. 2010, is used in the experiment to check the predictive capability and reliability of the proposed models. First 900 data are taken for the in-sample estimation and last 100 data are reserved for out of sample forecasting. For NIKKEI225, the whole sample of size 2411, spanned from 04 Jan. 2001 to 01 Nov. 2010, is used in the experiment to serve the same purposes. First 2171 data are for the in-sample estimation and the last 240 data are reserved for out of sample forecasting.

The daily series for the log-levels and the returns of the BSE and NIKKEI225 are depicted in **Figures 2** and **3**, respectively. Both figures show that the returns series are mean-stationary, and exhibit the typical volatility clustering phenomenon with periods of unusually large volatility followed by periods of relative tranquility.

The autocorrelation functions (ACFs) of the return and squared return series for both markets are depicted in **Figures 4** and **5**, respectively. In **Figure 4** (non-squared

ACFs), almost all the spikes are within the boundary (formed by standard errors), that is, ACFs decay very quickly toward zero; whereas almost all the spikes go out of the boundary in **Figure 5** (squared ACFs), that is, it produces slowly decreasing positive autocorrelation functions of the squared returns, **especially for NIKKEI225**. **Figure 5** indicates that the volatility clustering is reflected in the significant correlations of squared returns.

The autocorrelation coefficients of squared returns are larger and last longer (persistent) than those of the return series (non-squared). We must point out that the return


(a)

(b)
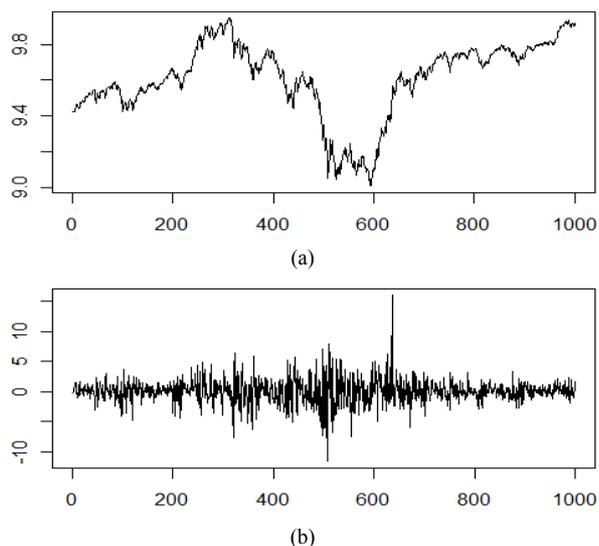
**Figure 2. Bombay stock exchange (BSE) index: 2006.10.5-2010.11.1. (a) Log-levels; (b) Returns.**

(a)

(b)
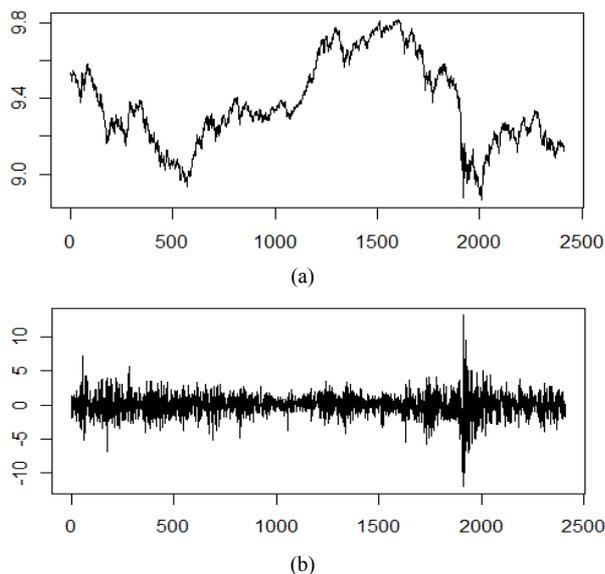
Figure 3. Japan stock exchange (NIKKEI 225) index: 2001.1.4-2010.11.1. (a) Log-levels; (b) Returns.
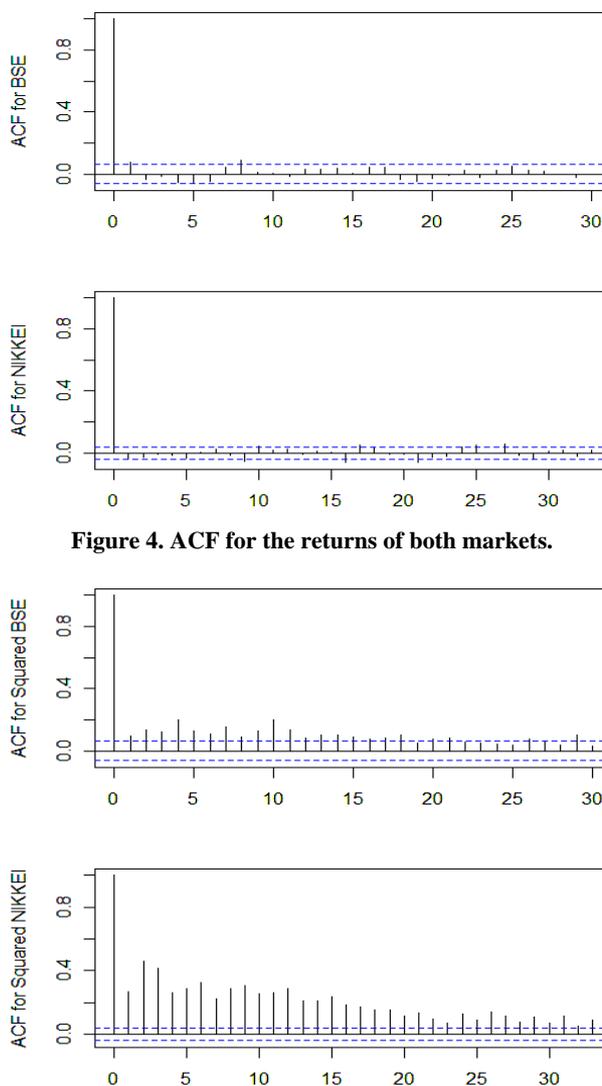
**Figure 4. ACF for the returns of both markets.**

Figure 5. ACF for squared returns of both markets.

series' show little or no correlation, but its squares show high correlation, which indicate the ARCH or GARCH effect, **especially for NIKKEI225**. It is not clear—why the volatility clustering is not clearly/remarkably reflected in the significant correlations of squared return of BSE. It may happen due to influencing outlying observations.

Figures 6 and **7** show the Q-Q plots of returns and squared returns, respectively, for each market. **Figure 6** (of return series) shows that some observations question the assumption of normality, that is, they may be outliers/unusual observations. It is clear from **Figure 7** (of squared return) that the assumption of normality is violated showing two outliers or unusual observations for BSE and a group of outliers for NIKKEI255.

Table 1 reports the summary statistics and diagnostics for the total sample of BSE and NIKKEI225 returns. From

the table, we can see that means of the returns are not far from zero as expected. Both the series are typically characterized by excessive kurtosis and asymmetry. The Bera-Jargue [57] test strongly rejects the normality hypothesis on the returns for each market. For both series, the Ljung-Box Q (10) statistics of returns indicate no significant correlation at 1% and 5% level of significances; but at 10% and more level of significance, there is relevant autocorrelation in the return series of BSE. The Q (10)* values of the squared returns reveal that there is no significant correlation in the squared returns at

Table 1. Descriptive statistics for the returns of BSE and NIKKEI.

Returns	BSE SENSEX	NIKKEI 225
Minimum	–11.60444	–12.11103
Maximum	15.98998	13.23458
Mean	0.04965155	–0.01669452
Variance	4.058489	2.682107
Skewness	0.171791	–0.2858808
Kurtosis	6.093471	6.220848
Normality	272.5166 [0.0000]	52.794 [3.434e-12]
Q(10)	18.26964 [0.0505]	13.48905 [0.19759]
Q(10)*	7.091121 [0.7168]	16.04430 [0.09837]
ARCH-LM	7.187405 [0.8449]	16.22137 [0.18130]

Note: Kurtosis quoted is excess kurtosis; Normality is the Bera-Jargue (1981) normality test; Q (10) is the Ljung-Box Q test at 10 order for raw returns; Q (10)* is LB Q test for squared returns; ARCH-LM is Engle's (1982) LM test for ARCH effect. Significance levels (p-values) are in brackets.

1% and 5% level of significances; even at 10% level of significance. Engle's (Engle, 1982) ARCH tests show that there is no significant evidence in support of GARCH effects (*i.e.*, heteroscedasticity) for both series (strongly for BSE).

This numerical examination of daily returns on the BSE and NIKKEI225 data reveals that returns are not characterized by heteroscedasticity and time-varying autocorrelation in spite of having the following pre-evidences and stylized facts: 1) The graphical test (**Figure 5**) indicates the presence of time varying volatility in BSE and NIKKEI225, 2) the statistics (Maximum, Variance and Kurtosis) of BSE and NIKKEI225 are comparatively higher than those of the other markets used in the previous research with different periods, and 3) generally, the return series exhibit volatility clustering and leptokurtic pattern for most of the market in the world. This situation (or problem) is created due to the outliers or unusual observations detected by the graphical test (**Figure 7**: Q-Q). To check the robustness properties of the used models, the unusual observations are kept in the data set. However, we assume that the return series of BSE and NIKKEI225 exhibit volatility clustering and leptokurtic pattern. Therefore, it is very suitable to model and forecast the return series by GARCH (1, 1).

Figure 6. Quantile-Quantile plots for the returns of BSE and NIKKEI225.
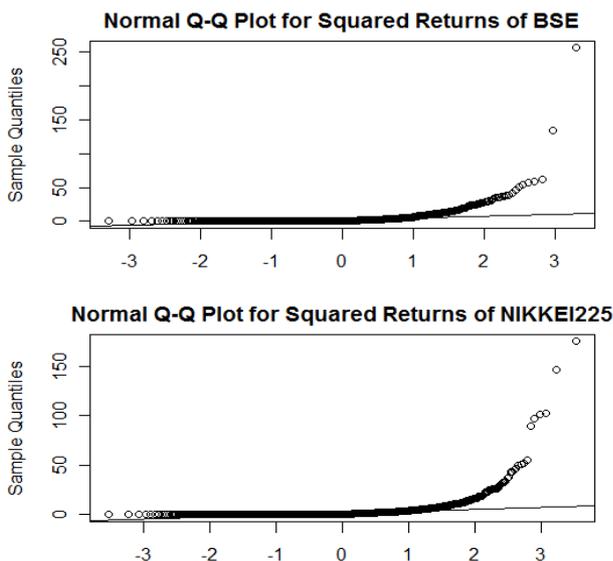
Figure 7. Quantile-Quantile plots for the squared returns of BSE and NIKKEI225.

5.2. In Sample Estimation or Training Results

We first fit the in-sample returns series to GARCH (1, 1) and ARMA-GARCH (1, 1) models in (22) and (24) to obtain their Maximum Likelihood Estimates. The estimation results and the diagnostic test results of GARCH (1, 1) and ARMA-GARCH (1, 1) volatility models for the BSE and NIKKEI225 returns are not reported here as the main focus is given in out of sample forecasting. It is

seen that based on log likelihood (LL), AIC, and BIC criteria, ARMA-GARCH (1, 1) model is more adequate to the data than pure GARCH (1, 1) model.

Now we turn to consider our models, recurrent support vector machine and recurrent relevance vector machine. The considered models must be trained using the above algorithm stated in Step 1 and Step 2. While training RSVM, two parameters γ and C are considered since they are sensitive for modeling the SVM. ε is assumed to be 0.005 and used for all cases. We apply ten-fold-cross-validation technique to tune the values of γ and C with the range $[2^{-5}, 2^5]$ and $[2^{-5}, 2^5]$, respectively. The optimal parameters $(C, \gamma) = (2^2, 2^{-2})$ and $(2^5, 2^2)$ for training BSE and NIKKEI225, respectively. Table 2 illustrates the training results (only the number of support vectors and relevance vectors) by RSVM and RRVM for both markets. From the **Table 2**, we can see that RRVM is more adequate to the in-sample series compared to RSVM for each market since RRVM produces smallest number of relevance vectors compared to the number of support vectors of RSVM.

5.3. Out of Sample Volatility Forecasting Results

Table 3 summarizes the forecasting performance based on four measures defined in section 4.3, MSE, MAE, DS and R square. From the table 3, we can see that ARMA-GARCH generates smaller values of SqrtMSE (3.1742) and MAE (3.0199) but larger value of R^2 (0.00553) than those of pure GARCH for BSE. For NIKKEI225, ARMA-GARCH generates smaller values of SqrtMSE (2.8094) and MAE (2.2401) than those of pure GARCH. Both the ARMA-GARCH and pure GARCH produce the same value of DS for each market and R^2 for NIKKEI225. Hence the ARMA-GARCH model outperforms the pure GARCH model.

Whereas RSVM and RRVM, they provide better performance than GARCH type models (pure GARCH and

ARMA-GARCH) for all cases except the RSVM for NI-KKEI225 based on MSE and DS, where GARCH type models perform better than RSVM. If we make comparison between RSVM and RRVM, the RSVM is better than RRVM based on MAE and R^2 only; but in term of MSE and DS, the RRVM is better than RSVM for both markets. The forecasting performances of GARCH type models are very poor compared to that of RSVM and RRVM due to outliers affect on traditional GARCH type model in forecasting; that is, both the RSVM and RRVM (not GARCH type) hold the robustness properties in forecasting through estimation.

Figures 8 and **9** plot multi-period-ahead forecasts by the machine learning models (RSVM and RRVM) and GARCH type models (pure GARCH and ARMA-GAR-

Table 2. Training results for RSVM and RRVM.

	BSE	NIKKEI225
No. of S.V.s	661	2049
No. of R.V.s	77	49
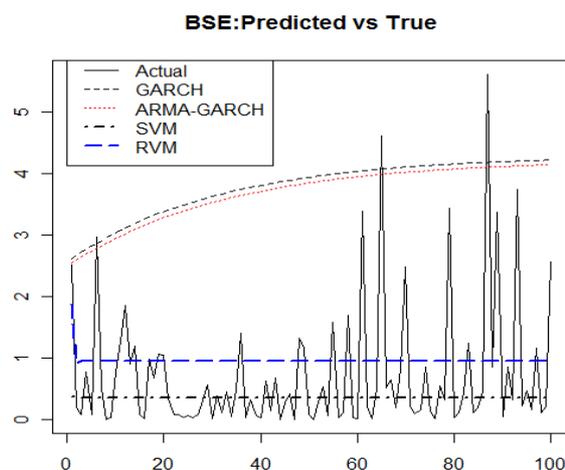
Figure 8. Volatility forecasts of BSE index returns.

Table 3. Multi-period-ahead forecasting accuracy by different models for real data.

Models	BSE				NIKKEI225			
	Sqrt MSE	MAE	DS	R^2	Sqrt MSE	MAE	DS	R^2
GARCH	3.2568	3.1024	51	0.00485	2.8124	2.2464	52.5	0.00071
ARMA-GARCH	3.1742	3.0199	51	0.00553	2.8094	2.2401	52.5	0.00071
RSVM	1.1160	**0.6258**	65	**0.03011**	3.0073	**1.6864**	50.4	**0.00417**
RRVM	**1.0670**	0.8104	**87**	0.02973	**2.7466**	1.7677	**95.8**	4.58E-5
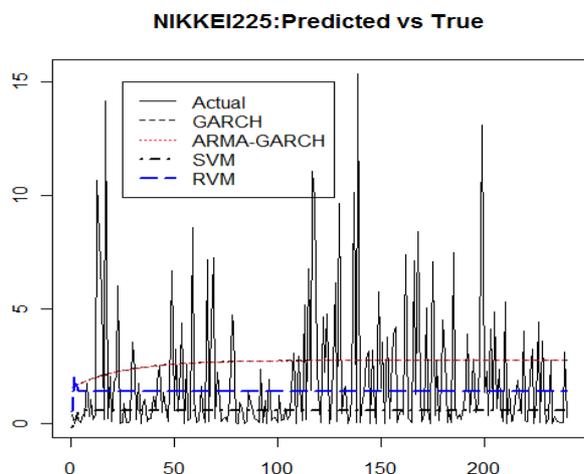
Figure 9. Volatility forecasts of NIKKEI225 index returns.

CH) against actual values for BSE and NIKKEI225, respectively. From the plots we can see that the machine learning techniques generate better forecasting performances than the GARCH type models. The forecasted series' by GARCH type models are pushed up due to outliers affect; where the ARMA-GARCH model is less affected than pure GARCH. The series by RSVM is not affected due to its robustness properties; where RRVM is slightly affected by outliers for BSE. For NIKKEI, no model is remarkably affected by outliers; because of having a group of outliers which may not be very much influential.

6. Conclusions

To measure the model performances, we apply recurrent SVM and RVM to model GARCH as hybrid approaches comparing with traditional pure GARCH and ARMA-GARCH models to forecast (multi-periodically) volatility of Asian (emerging) stock markets, BSE SENSEX and NIKKEI225. The above models are evaluated using the criteria: MSE, MAE, DS and linear regression R^2 in out-of-sample forecasting. In the parallel way, we examine the robustness properties of the used models in forecasting volatility in the presence of outliers, where outliers are detected very simply using Q-Q plot. Using Q-Q plots of the observed volatility, two outliers are clearly detected for BSE and a group of outliers for NIKKEI225. Due to the affect of these outliers, the ACFs (especially, of BSE), descriptive statistics and classical tests (Ljung-Box Q and ARCH-LM) give the misleading results, which agree with the previous research on outliers in time series analysis. From the experimental results, we can come to the conclusion that 1) the outliers significantly affect the parameter estimates of the pure GARCH and ARMA-GARCH models, 2) the RRVM produces small-

lest number of relevance vectors compared to the number of support vectors of RSVM, 3) the computational-intelligence-based techniques (RSVM and RRVM) perform better than the GARCH type models in out-of-sample forecasting, 4) the ARMA-GARCH model is superior to the pure GARCH model in out-of-sample forecasting, 5) both the RSVM and RRVM perform almost equally in out-of-sample forecasting—the RSVM is better than RRVM based on MAE and R^2, but in terms of MSE and DS, the RRVM is better than RSVM, and 6) RRVM with RSVM holds the robustness properties in forecasting through estimation, however, RRVM is slightly affected by outliers for being Bayesian approach. Theoretically, RVM is a probabilistic model having its functional form identical to SVM, where there is no requirement of free parameters and Mercer's kernel function for RVM like SVM. Considering the above empirical results and theoretical properties of RVM and SVM, we are in favor of recurrent RVM (like the previous research) in forecasting volatility of emerging stock markets, even in the presence of outliers.

REFERENCES

[1] R. F. Engle and A. J. Patton, "What Good Is a Volatility Model?" *Journal of Quantitative Finance*, Vol. 1, No. 2, 2001, pp. 237-245. doi:10.1088/1469-7688/1/2/305

[2] R. C. Merton, "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, Vol. 8, 1980, pp. 323-361. doi:10.1016/0304-405X(80)90007-0

[3] R. F. Engle, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, Vol. 50, No. 2, 1982, pp. 987-1007. doi:10.2307/1912773

[4] T. Bollerslev, "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometric*, Vol. 31, No. 3, 1986, pp. 307-327. doi:10.1016/0304-4076(86)90063-1

[5] S. H. Poon and C. Granger, "Forecasting Volatility in Financial Markets: A Review," *Journal of Economic Literature*, Vol. 41, No. 2, 2003, pp. 478-539. doi:10.1257/002205103765762743

[6] W. C. Wong, F. Yip and L. Xu, "Financial Prediction by Finite Mixture GARCH Model," *Proceedings of Fifth International Conference on Neural Information Processing*, Kitakyushu, 21-23 October 1998, pp. 1351-1354.

[7] H. Tang, K. C. Chun and L. Xu, "Finite Mixture of ARMA-GARCH Model for Stock Price Prediction," *Proceedings of 3rd International Workshop on Computational Intelligence in Economics and Finance* (*CIEF* 2003), North Carolina, 26-30 September 2003, pp. 1112-1119.

[8] A. Hossain and M. Nasser, "Comparison of Finite Mixture of ARMA-GARCH, Back Propagation Neural Networks and Support-Vector Machines in Forecasting Financial Returns," *Journal of Applied Statistics*, Vol. 38,

No. 3, 2011, pp. 533-551.

[9] T. G. Andersen and T. Bollerslev, "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, Vol. 39, No. 4, 1998, pp. 885-905. doi:10.2307/2527343

[10] T. J. Brailsford and R. W. Faff, "An Evaluation of Volatility Forecasting Techniques," *Journal of Banking and Finance*, Vol. 20, No. 3, 1996, pp. 419-438. doi:10.1016/0378-4266(95)00015-1

[11] R. Cumby, S. Figlewski and J. Hasbrouck, "Forecasting Volatility and Correlations with EGARCH Models," *Journal of Derivatives Winter*, Vol. 1, No. 2, 1993, pp. 51-63.

[12] S. Figlewski, "Forecasting Volatility," *Financial Markets*, *Institutions and Instruments*, Vol. 6, No. 1, 1997, pp. 1-88. doi:10.1111/1468-0416.00009

[13] P. Jorion, "Predicting Volatility in the Foreign Exchange Market," *Journal of Finance*, Vol. 50, No. 2, 1995, pp. 507-528. doi:10.2307/2329417

[14] P. Jorion, "Risk and Turnover in the Foreign Exchange Market," In: J. A. Franke, G. Galli and A. Giovannini, Eds., *The Microstructure of Foreign Exchange Markets*, Chicago University Press, Chicago, 1996.

[15] D. G. McMillan, A. E. H. Speight and O. Gwilym, "Forecasting UK Stock Market Volatility: A Comparative Analysis of Alternate Methods," *Applied Financial Economics*, Vol. 10, No. 4, 2000, pp. 435-448. doi:10.1080/09603100050031561

[16] R. G. Donaldson and M. Kamstra, "An Artificial Neural Network-GARCH Model for International Stock Returns Volatility," *Journal of Empirical Finance*, Vol. 4, No. 1, 1997, pp. 17-46. doi:10.1016/S0927-5398(96)00011-4

[17] F. Perez-Cruz, J. A. Afonso-Rodriguez and J. Giner, "Estimating GARCH Models Using Support Vector Machines," *Journal of Quantitative Finance*, Vol. 3, 2003, pp. 163-172.

[18] P. H. Ou and H. Wang, "Predicting GARCH, EGARCH, GJR Based Volatility by the Relevance Vector Machine: Evidence from the Hang Seng Index," *International Research Journal of Finance and Economics*, No. 39, 2010, pp. 46-63.

[19] L. B. Tang, H. Y. Sheng and L. X. Tang, "Forecasting Volatility Based on Wavelet Support Vector Machine," *Expert Systems with Applications*, Vol. 36, No. 2, 2009, pp. 2901-2909.

[20] L. B. Tang, H. Y. Sheng and L. X. Tang, "GARCH Prediction Using Spline Wavelet Support Vector Machine," *Journal of Neural Computing and Application*, Vol. 18, No. 8, 2009, pp. 913-917.

[21] M. Bildirici and Ö. Ö. Ersin, "Improving Forecasts of GARCH Family Models with the Artificial Neural Networks: An Application to the Daily Returns in Istanbul Stock Exchange," *Expert Systems with Applications*, Vol. 36, No. 4, 2009, pp. 7355-7362. doi:10.1016/j.eswa.2008.09.051

[22] L. J. Cao and F. Tay, "Application of Support Vector Machines in Financial Time Series Forecasting," *International Journal of Management Science*, Vol. 29, No. 4, 2001, pp. 309-317.

[23] V. N. Vapnik, "The Nature of Statistical Learning Theory," 2nd Edition, Sringer-Verlag, New York, 1995.

[24] L. J. Cao and F. Tay, "Modified Support Vector Machines in Financial Time Series Forecasting," *Journal of Neurocomputing*, Vol. 48, No. 1-4, 2002, pp. 847-861.

[25] K. J. Kim, "Financial Time Series Forecasting Using Support Vector Machines," *Journal of Neurocomputing*, Vol. 55, No. 1-2, 2003, pp. 307-319.

[26] W. Huang, Y. Nakamori and S. Y. Wang, "Forecasting Stock Market Movement Direction with Support Vector Machine," *Journal of Computers & Operational Research*, Vol. 32, No. 10, 2005, pp. 513-522.

[27] C. J. Lu, T. S. Lee and C. C. Chiu, "Financial Time Series Forecasting Using Independent Component Analysis and Support Vector Regression," *Journal of Decision Support Systems*, Vol. 47, No. 2, 2009, pp. 115-125.

[28] H. S. Kim and S. Y. Sohn, "Support Vector Machines for Default Prediction of SMEs Based on Technology Credit," *European Journal of Operational Research*, Vol. 201, No. 3, 2010, pp. 938-846.

[29] A. J. Smola and B. Scholkopf, "A Tutorial on Support Vector Regression," *Journal of Statistics and Computing*, Vol. 14, No. 3, 2004, pp. 199-222.

[30] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, Vol. 1, 2001, pp. 211-244.

[31] C. M. Bishop and M. E. Tipping, "Variational Relevance Vector Machine," In: C. Boutilier and M. Goldszmidt, Eds., *Uncertainty in Artificial Intelligence*, Morgan Kaufmann, Waltham, 2000, pp. 46-53.

[32] S. Ghosh and P. P. Mujumdar, "Statistical Downscaling of GCM Simulations to Stream Flow Using Relevance Vector Machine," *Advances in Water Resources*, Vol. 31, No. 1, 2008, pp. 132-146.

[33] D. Porro, N. Hdez, I. Talavera, O. Nunez, A. Dago and R. J. Biscay, "Performance Evaluation of Relevance Vector Machines as a Nonlinear Regression method in Real World Chemical Spectroscopic Data," 19*th International Conference on Pattern Recognition* (*ICPR* 2008), Tampa, 8-11 December 2008, pp. 1-4.

[34] S. Chen, K. Jeong and W. Härdle, "Support Vector Regression Based GARCH Model with Application to Forecasting Volatility of Financial Returns," SFB 649 Discussion Paper 2008-014. http://edoc.hu-berlin.de/series/sfb-649-papers/2008-14/PDF/14.pdf

[35] S. Chen, K. Jeong and W. Härdle, "Recurrent Support Vector Regression for a Nonlinear ARMA Model with Applications to Forecasting Financial Returns," SFB 649 Discussion Paper 2008-051.

[36] J. A. K. Suykens and J. Vandewalle, "Recurrent Least Squares Support Vector Machines," *IEEE Transactions on Circuits and Systems* I, Vol. 47, No. 7, 2000, pp. 1109-1114. doi:10.1109/81.855471

Recurrent Support and Relevance Vector Machines Based Model with Application to Forecasting
Volatility of Financial Returns

241

[37] P. H. Ou and H. Wang, "Predict GARCH Based Volatility of Shanghai Composite Index by Recurrent Relevant Vector Machines and Recurrent Least Square Support Vector Machines," *Journal of Mathematics Research*, Vol. 2, No. 2, 2010.

[38] G. Bekaert and C. R. Harvey, "Emerging Equity Market Volatility," *Journal of Financial Economics*, Vol. 43, No. 1, 1997, pp. 29-78. doi:10.1016/S0304-405X(96)00889-6

[39] R. Aggarwal, C. Inclan and R. Leal, "Volatility in Emerging Stock Markets," *Journal of Financial and Quantitative Analysis*, Vol. 34, No. 1, 1999, pp. 33-55. doi:10.2307/2676245

[40] J. Ledolter, "The Effects of Additive Outliers on the Forecasts from ARIMA Models," *International Journal of Forecasting*, Vol. 5, No. 2, 1989, pp. 231-240. doi:10.1016/0169-2070(89)90090-3

[41] R. F. Engle and G. G. J. Lee, "A Permanent and Transitory Component Model of Stock Returns Volatility," Discussion Paper 92-44R, University of California, San Diego, 1993.

[42] P. H. Franses, D. van Dijk and A. Locas, "Short Patches of Outliers, ARCH and Volatility Modeling," Discussion Paper 98-057/4, Tinbergen Institute, Erasmus University, Rotterdam, 1998.

[43] P. H. Franses and H. Ghijsels, "Additive Outliers, GARCH and Forecasting Volatility," *International Journal of Forecasting*, Vol. 15, No. 1, 1999, pp. 1-9. doi:10.1016/S0169-2070(98)00053-3

[44] A. J. Fox, "Outliers in Time Series," *Journal of the Royal Statistical Society*, *Series B*, Vol. 34, No. 3, 1972, pp. 350-363.

[45] C. Chen and L. Liu, "Joint Estimation of Model Parameters and Outlier Effects in Time Series," *Journal of American Statistical Association*, Vol. 88, No. 421, 1993, pp. 284-297. doi:10.2307/2290724

[46] M. E. Tipping, "Relevance Vector Machine," Microsoft Research, Cambridge, 2000.

[47] M. E. Tipping, "Bayesian Inference: An Introduction to Principles and Practice in Machine Learning," *Advanced Lectures on Machine Learning*, Vol. 3176/2004, 2004, pp. 41-62. doi:10.1007/978-3-540-28650-9_3

[48] P. R. Hansen and A. Lunde, "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH (1, 1)?" *Journal of Applied Econometrics*, Vol. 20, No. 7, 2005, pp. 873-889.

[49] J. D. Hamilton, "Time Series Analysis," Princeton University Press, Saddle River, 1997.

[50] W. Enders, "Applied Econometric Time Series," 2nd Edition, John Wiley & Sons, New York, 2004. doi:10.1016/S0305-0483(01)00026-3

[51] F. E. H. Tay and L. Cao, "Application of Support Vector Machines in Financial Time-Series Forecasting," *Omega*, Vol. 29, No. 4, 2001, pp. 309-317.

[52] M. Thomason, "The Practitioner Method and Tools: A Basic Neural Network-Based Trading System Project Revisited (Parts 1 and 2)," *Journal of Computational Intelligence in Finance*, Vol. 7, No. 3, 1999, pp. 36-45.

[53] M. Thomason, "The Practitioner Method and Tools: A Basic Neural Network-Based Trading System Project Revisited (Parts 3 and 4)," *Journal of Computational Intelligence in Finance*, Vol. 7, No. 3, 1999, pp. 35-48.

[54] C. Brooks, "Predicting Stock Index Volatility: Can Market Volume Help?" *Journal of Forecasting*, Vol. 17, No. 1, 1998, pp. 59-80. doi:10.1002/(SICI)1099-131X(199801)17:1<59::AID-FOR676>3.0.CO;2-H

[55] I. A. Moosa, "Exchange Rate Forecasting: Techniques and Applications," Macmillan Press LTD, Lonton, 2000.

[56] H. Theil, "Principles of Econometrics," Wiley, New York, 1971.

[57] A. K. Bera and C. M. Jarque, "An Efficient Large-Sample Test for Normality of Observations and Regression Residuals," Australian National University Working Papers in Econometrics, 40, Canberra, 1981.