Journal of Forecasting
J. Forecast. **29**, 406–433 (2010)
Published online 9 September 2009 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/for.1134

Forecasting Volatility with Support Vector Machine-Based GARCH Model

SHIYI CHEN,[1]* WOLFGANG K. HÄRDLE[2] AND KIHO JEONG[3]

[1] *China Center for Economic Studies, School of Economics, Fudan University, Shanghai, China*
[2] *Center for Applied Statistics and Economics, Humboldt University, Berlin, Germany*
[3] *School of Economics and Trade, Kyungpook National University, Daegu, Republic of Korea*

ABSTRACT

Recently, support vector machine (SVM), a novel artificial neural network (ANN), has been successfully used for financial forecasting. This paper deals with the application of SVM in volatility forecasting under the GARCH framework, the performance of which is compared with simple moving average, standard GARCH, nonlinear EGARCH and traditional ANN-GARCH models by using two evaluation measures and robust Diebold–Mariano tests. The real data used in this study are daily GBP exchange rates and NYSE composite index. Empirical results from both simulation and real data reveal that, under a recursive forecasting scheme, SVM-GARCH models significantly outperform the competing models in most situations of one-period-ahead volatility forecasting, which confirms the theoretical advantage of SVM. The standard GARCH model also performs well in the case of normality and large sample size, while EGARCH model is good at forecasting volatility under the high skewed distribution. The sensitivity analysis to choose SVM parameters and cross-validation to determine the stopping point of the recurrent SVM procedure are also examined in this study. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS (recurrent) support vector machine; GARCH model; volatility forecasting; Diebold–Mariano test

INTRODUCTION

Volatility is important in financial markets since it is a key variable in portfolio optimization, securities valuation and risk management. Much attention of academics and practitioners has been focused on modeling and forecasting volatility in the last few decades (see Franses and McAleer, 2002, and Poon and Granger, 2003, for a comprehensive review). So far in the literature, the predominant model of the past is the GARCH model by Bollerslev (1986), who generalizes the seminal idea on

*Correspondence to: Shiyi Chen, China Center for Economic Studies, School of Economics, Fudan University, Guoquan Road 600, Shanghai, China 200433. E-mail: shiyichen@fudan.edu.cn

Copyright © 2009 John Wiley & Sons, Ltd.

ARCH by Engle (1982), and its various extensions; see Li *et al.* (2002) for recent surveys of the models. The GARCH family models, together with the simplest historical price model prevalent in the pre-GARCH era[1] and stochastic volatility model studied a decade later than GARCH development,[2] comprise one of the two broad categories of methods widely used in volatility forecasting, the so-called time series volatility model; another is the market determined option implied volatility model.[3] This paper limits itself mainly to the analysis within the GARCH framework.

The popularity of the GARCH model is due to its ability to capture volatility persistence or clustering, supported by many studies (Akgiray, 1989; Bollerslev *et al.*, 1992; West and Cho, 1995; Andersen and Bollerslev, 1998; Marcucci, 2005). However, some empirical studies report that the GARCH model provides poor forecasting performance (Jorion, 1995, 1996; Brailsford and Faff, 1996; Figlewski, 1997; McMillan *et al.*, 2000; Choudhry and Wu, 2008). To improve the forecasting ability of the GARCH model, some alternative approaches have been advocated by innovating the model specification and estimation,[4] by using different evaluation metrics and definitions of realized volatility,[5] or by enriching the informational content of the model.[6]

As for GARCH model specification and estimation, for example, many financial returns are skewed distributed and nonlinearly dependent such that the linear GARCH model cannot cope with them and therefore forecast of symmetric GARCH model would be biased (Pagan and Schwert, 1990; Bollerslev *et al.*, 1992). To deal with this problem the regime-switching (RS) volatility model is proposed to detect nonlinear behavior in the variance by various tests for asymmetry or threshold

[1]This includes simple moving average method, exponential smoothing method, random walk model, ARMA model, exponentially weighted moving average (EWMA) method and its current extension of Riskmetrics™ model, etc.

[2]The stochastic volatility (SV) model has an additional innovative term in the volatility dynamics (Taylor, 1986). For a detailed discussion on the SV model and its relation to the GARCH class models, see the survey articles by Ghysels *et al.* (1996) and Chib *et al.* (2002), among others.

[3]The time series volatility model is based on historical price information only, while the option implied volatility (IV) model uses market traded option information alone or in addition to historical price sets to forecast volatility. Many studies examine the relative performance of the IV model to forecasting volatility (Day and Lewis, 1992; Lamoureux and Lastrapes, 1993; Pong *et al.*, 2004; Dotsis *et al.*, 2007; Becker *et al.*, 2009; Neely, 2009). This paper limits itself mainly to analysis within the GARCH framework.

[4]Except for the introduction below, other relatively sophisticated GARCH models and estimations include the multivariate GARCH model (Bauwens *et al.*, 2006; Rosenow, 2008), outlier-corrected GARCH model (Park, 2002; Zhang and King, 2005; Ané *et al.*, 2008), Markov chain Monte Carlo (MCMC) sampling techniques to estimate the GARCH model (Gerlach and Tuyl, 2006), other semiparametric or nonparametric specification and estimation such as genetic algorithm, wavelet smoother, kernel density etc. (Franke *et al.*, 2004; Lux and Schornstein, 2005; Renò, 2006; Chen *et al.*, 2008; Feng and McNeil, 2008; Corradi *et al.*, 2009) and combination forecasts from competing approaches (Hu and Tsoukalas, 1999; Dunis and Huang, 2002).

[5]Many studies find that the relative accuracy of various models is also highly sensitive to the measures used to evaluate them (Taylor, 1999; Brooks and Persand, 2003). Most comparisons are based on the average figure of mean absolute error (MAE) and mean square error (MSE) etc. Diebold and Mariano (1995) and West (1996) show how standard errors for MAE and MSE are derived taking into account serial correlation in the forecast errors for statistical inference. Lehar *et al.* (2002) applies value-at-risk (VaR)-oriented evaluation measures to compare the out-of-sample performance. In addition to the symmetric measures of MAE and MSE, Balaban (2004) also uses asymmetric evaluation criteria such as mean mixed error statistics to compare the forecasting performance, penalizing under/over-predictions of volatility more heavily. Recent research has also suggested that this relative failure of GARCH models arises not from a failure of the model but a failure to specify correctly the true volatility measure against which forecasting performance is measured. It is argued that the standard approach of using *ex post* daily squared returns as the measure of true volatility includes a large noisy component. An alternative measure for true volatility has therefore been suggested based on the cumulative squared returns from intra-day data, also referred to as realized, or integrated volatility (Andersen and Bollerslev, 1998; Andersen *et al.*, 2003; Meddahi, 2003; McMillan and Speight, 2004; Galbraith and Kisinbay, 2005; Ghysels *et al.*, 2006).

[6]In many instances, the researchers find the inclusion of implied volatility or trade volume as an exogenous variable in the framework of the GARCH model to be beneficial (Brooks, 1998; Fleming, 1998; Blair *et al.*, 2001; Koopman *et al.*, 2005; Gospodinov *et al.*, 2006; Becker *et al.*, 2007).

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

nonlinearity (Franses and Dijk, 2000). The first class of RS volatility model assumes that the regime can be determined by an observable variable, including the nonlinear exponential GARCH (EGARCH) model of Nelson (1991), threshold GJR-GARCH model of Glosten *et al.* (1992) and quadratic GARCH model of Engle *et al.* (1993) and Sentana (1995). The second class of RS model for volatility implements GARCH with a Hamilton (1989) type framework that assumes the regime is the realization of a hidden Markov chain, such as (double) Markov switching GARCH model of Gray (1996), Klaassen (2002) and Chen *et al.* (2008).

Both the linear and nonlinear GARCH model described above are parametric and normally estimated jointly by maximum likelihood estimation (MLE). That is, they make specific assumptions about the functional form of the data generation process and the distribution of error terms that is necessary for MLE. Such parametric models are easy to estimate and readily interpretable, but these advantages may come at a cost. Perhaps nonparametric models are better representations of the underlying data generation process. Instead of specifying a particular functional form and making *a priori* distributional assumption, the nonparametric model will search for the best fit over a large set of alternative functional forms. Thus, in the literature, many nonlinear nonparametric GARCH models are developed and still developing fast, among which the artificial neural network (ANN) is extensively used. This paper focuses on one of the neural network algorithms, the support vector machine (SVM), and investigates its forecasting ability of volatility as compared with the simplest moving average method, standard linear GARCH model, nonlinear EGARCH model and traditional recurrent ANN-based nonlinear GARCH model. The moving average method is chosen as the benchmark because some studies find that it provides more accurate forecasts than GARCH models (Dimson and Marsh, 1990; Tse and Tung, 1992; Figlewski, 1997). Among the number of nonlinear parametric GARCH models the EGARCH model is also the most commonly used (Cao and Tsay, 1992; Cumby *et al.*, 1993; Heynen and Kat, 1994; Chong *et al.*, 1999; Hu and Tsoukalas, 1999; Gokcan, 2000; Balaban, 2004).

In recent years, ANN has been successfully used for forecasting financial time series; for recent work, see Fernandez-Rodriguez *et al.* (2000), Qi and Wu (2003), and Pantelidaki and Bunn (2005). The studies in favor of ANN-based GARCH model as opposed to parametric GARCH model in forecasting conditional volatility include Donaldson and Kamstra (1997), Schittenkopf *et al.* (2000), Taylor (2000), Dunis and Huang (2002), Hamid and Iqbal (2004), Ferland and Lalancette (2006), Tseng *et al.* (2008). However, the traditional ANN algorithm also suffers from its own weaknesses such as the need for many controlling parameters, difficulty in obtaining a global solution and the danger of over-fitting (Tay and Cao, 2001). Thus, SVM that can obtain a unique global solution by solving a quadratic programming is developed by Vapnik and his coworkers (1995, 1997). Naturally, SVM also keeps the advantages of conventional ANN such as the flexibility in approximating any nonlinear function arbitrarily well, without *a priori* assumptions about the properties of the data and without the requirement of large sample size that MLE-based parametric GARCH models have. Unlike traditional ANN implementing the empirical risk minimization (ERM) principle, the most particular principle of SVM is to implement the structural risk minimization (SRM), which seeks to achieve a balance between the training error and generalization error, leading, theoretically, to better forecasting performance than traditional ANN (Gunn, 1998; Haykin, 1999). Recently, SVM has gained popularity in predicting financial variables owing to such attractive features (Cao and Tay, 2001; Härdle *et al.*, 2005, 2007; Chen *et al.*, 2009). Pérez-Cruz *et al.* (2003) also propose an SVM-based GARCH (1, 1) model and shows that it provides better volatility forecasts than the standard GARCH model. However, they use the feedforward SVM procedure, which has the same structure as the autoregressive (AR) process and has poor ability

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

to model a long-time memory. Inspired by the merit of recurrent ANN (Kuan and Liu, 1995; Dunis and Huang, 2002; Bekiros and Georgoutsos, 2008), in this paper we propose a recurrent SVM procedure which can model the ARMA process and apply it to forecast the conditional variance equation of the GARCH model in real data analysis.

The forecasting accuracy of the recurrent SVM-based GARCH model in one-period-ahead volatility forecasting is compared with the competing models in terms of two evaluation metrics of mean absolute error (MAE) and directional accuracy (DA). The statistical hypothesis of equal forecasting accuracy between pairwise models is also investigated by using the Diebold and Mariano (1995) test, calculated according to the Newey–West procedure (Newey and West, 1987). The Diebold and Mariano (DM) test is one of the most important contributions to the study of out-of-sample forecasting accuracy evaluation over the past two decades, and has been further generalized and extensively used in many studies since then (Corradi and Swanson, 2004; Awartani and Corradi, 2005; Preminger and Franck, 2007; Taylor, 2008; Groen *et al.*, 2009; Wong and Tu, 2009).

This paper is organized as follows. The next section briefly introduces the theory of SVM. The third section specifies the empirical model and forecasting scheme. The fourth section uses the Monte Carlo simulation to evaluate how the models perform under controlled conditions. The fifth section describes the GBP exchange rates and NYSE composite index data and discusses the volatility forecasting performance of all models for the real data. The paper concludes with the sixth section.

SUPPORT VECTOR MACHINE

The support vector machine (SVM) originates from Vapnik's statistical learning theory (Vapnik, 1995, 1997), which has the design of a feedforward network with an input layer, a single hidden layer of nonlinear units and an output layer, and formulates the regression problem as a quadratic programming (QP) problem. SVM estimates a function by nonlinearly mapping the input space into a high-dimensional hidden space and then running the linear regression in the output space. Thus, the linear regression in the output space corresponds to a nonlinear regression in the low-dimensional input space. The theory denotes that if the dimensions of feature space (or hidden space) are high enough, SVM may approximate any nonlinear mapping relations. As the name implies, the design of the SVM hinges upon the extraction of a subset of the training data that serves as support vectors, which represent a stable characteristic of the data.

Given a training dataset $(\mathbf{x_t}, y_t)$, where input vector $\mathbf{x_t} \in \mathbb{R}^p$ and output scalar $y_t \in \mathbb{R}^1$. Indeed, the desired response y, known as a 'teacher', represents the optimum action to be performed by the SVM. We aim at finding a sample regression function $f(\mathbf{x})$, or denoted by $\hat{y}$, as below to approximate the latent, unknown decision function $g(\mathbf{x})$:

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \tag{1}$$

where the superscript T is a transposing operator that should be differentiated from the sample size T of the time series used later in this paper. In equation (1), $\phi(\mathbf{x}) = [\phi_1(\mathbf{x}), \ldots, \phi_l(\mathbf{x})]^T$, $\mathbf{w} = [w_1, \ldots, w_l]^T$. The $\phi(\mathbf{x})$ is known as the nonlinear transfer function which represents the features of the input space and projects the inputs into the feature space. The dimension of the feature space is l, which is directly related to the capacity of the SVM to approximate a smooth input–output mapping; the higher the dimension of the feature space, the more accurate the approximation will be. Parameter

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

w denotes a set of linear weights connecting the feature space to the output space, and b is the threshold.

To get the function $f(\mathbf{x})$, the optimal $\mathbf{w}^*$ and b^* have to be estimated from the data. First, we define a linear ε-insensitive loss function, L_ε, originally proposed by Vapnik (1995):

$$L_\varepsilon(\mathbf{x}, y, f(\mathbf{x})) = \begin{cases} |y - f(\mathbf{x})| - \varepsilon & \text{for } |y - f(\mathbf{x})| \geqslant \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

This function indicates the fact that it does not penalize errors below ε. The training points within the ε-tube have no loss and do not provide any information for decision. Therefore, these points do not appear in the decision function $f(\mathbf{x})$. Only those data points located on or outside the ε-tube will serve as the support vectors and are finally used to construct the $f(\mathbf{x})$. This property of sparseness algorithm results only from the ε-insensitive loss function and greatly simplifies the computation of SVM. The non-negative slack variables, ξ and ξ' (below or above the ε-tube, or denoted together by $\xi^{(\prime)}$; see Figure 1) are employed to describe this kind of ε-insensitive loss.

The derivation of SVM follows the principle of structural risk minimization (SRM) that is rooted in the Vapnik–Chervonenkis (VC) dimension theory (Haykin, 1999). Structural risk is the upper boundary of empirical loss, denoted by ε-insensitive loss function, plus the confidence interval (or called margin), which is constructed in equation (3). The primal constrained optimization problem of SVM is obtained below:

$$\min_{\mathbf{w}\in\mathbb{R}^l, \xi^{(\prime)}\in\mathbb{R}^{2T}, b\in\mathbb{R}} \mathbf{C}(\mathbf{w}, b, \xi_t, \xi_t') = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{t=1}^{T}(\xi_t + \xi_t') \tag{3}$$

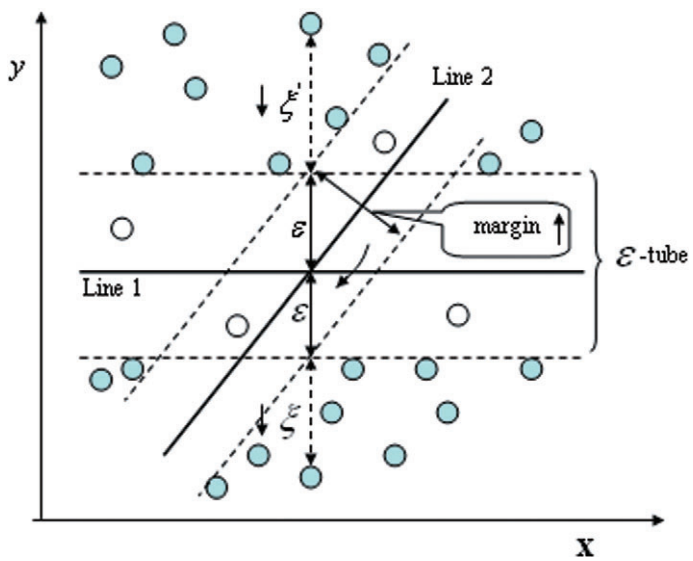

Figure 1. Principle of structural risk minimization (SRM) of SVM

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

such that

$$\mathbf{w}^T\phi(\mathbf{x_t}) + b - y_t \leqslant \varepsilon + \xi_t \tag{4}$$

$$y_t - \mathbf{w}^T\phi(\mathbf{x_t}) - b \leqslant \varepsilon + \xi_t' \tag{5}$$

$$\xi_t \geqslant 0, \xi_t' \geqslant 0, t = 1, 2, \ldots, T \tag{6}$$

The formulation of the cost function $\mathbf{C}(\cdot)$ in equation (3) is in perfect accord with the SRM principle, which is illustrated in Figure 1 (in which the dark circles are data points extracted as support vectors). In equation (3), the first term indicates the Euclidean norm of the weight vector $\mathbf{w}(\|\mathbf{w}\|^2 = \mathbf{w}^T\mathbf{w})$ and measures the function flatness; to minimize it is equivalent to maximizing the separation margin $(2/\|\mathbf{w}\|)$, that is, maximizing the generalization ability. The second term represents the empirical risk loss determined by the ε-insensitive loss function and is similar to the sum of residual squares in the objective function of ANN. Finally, SVM obtains the tradeoff between the two terms; as a result, it not only fits the historical data well but also forecasts the future data excellently. As shown in Figure 1, both regression lines 1 and 2 can classify the data points correctly and then minimize the empirical loss; however, the separation margin of the two lines are different, in which the regression line 1 has the larger margin. It is the special design of minimizing the structural risk that endows SVM with the excellent forecasting ability among all candidates. In addition, the convex quadratic programming and linear restrictions in the above primal problem ensure that SVM can always obtain the global unique optimal solution, which is different from the usual networks that easily get trapped in local minima. The penalty parameter $C > 0$ controls the penalizing extent on the sample points which lie outside ε-tube. Both ε and C, the free parameter of SVM, must be selected by the user.

The corresponding dual problem of the SVM can be derived from the primal problem by using the Karush–Kuhn–Tucker conditions as follows:

$$\min_{\alpha_t^{(\prime)} \in \mathbb{R}^{2T}} \frac{1}{2}\sum_{s=1}^{T}\sum_{t=1}^{T}(\alpha_s' - \alpha_s)(\alpha_t' - \alpha_t)K(x_s \cdot x_t) + \varepsilon\sum_{t=1}^{T}(\alpha_t' + \alpha_t) - \sum_{t=1}^{T} y_t(\alpha_t' - \alpha_t) \tag{7}$$

such that

$$\sum_{t=1}^{T}(\alpha_t - \alpha_t') = 0 \tag{8}$$

$$0 \leqslant \alpha_t, \alpha_t' \leqslant Cs, t = 1, 2, \ldots, T \tag{9}$$

where α_t and α_t' (or $\alpha_t^{(\prime)}$) are the Lagrange multipliers. The dual problem can be solved more easily than the primal problem (Scholkopf and Smola, 2001; Deng and Tian, 2004). Making use of any solution of α_t and α_t', the optimal solutions of the primal problem can be calculated in which $\mathbf{w}^*$ is unique and expressed as follows:

$$\mathbf{w}^* = \sum_{t=1}^{T}(\alpha_t' - \alpha_t)\phi(\mathbf{x_t}) \tag{10}$$

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

However, b^* is not unique and formulated in terms of different cases. If $i \in \{t | \alpha_t \in (0, C)\}$, then

$$b^* = y_t - \sum_{t=1}^{T} (\alpha_t' - \alpha_t) K(\mathbf{x_t} \cdot \mathbf{x_i}) + \varepsilon \tag{11}$$

If $j \in \{t | \alpha_t' \in (0, C)\}$, then

$$b^* = y_j - \sum_{t=1}^{T} (\alpha_t' - \alpha_t) K(\mathbf{x_t} \cdot \mathbf{x_j}) - \varepsilon \tag{12}$$

The cases of both $i, j \in \{t | \alpha_t^{(\prime)} = 0\}$ and $i, j \in \{t | \alpha_t^{(\prime)} = C\}$ rarely occur in reality.

Thus the regression decision function $f(\mathbf{x})$ will be computed by using $\mathbf{w}^*$ and b^* in the following forms:

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{w}^{*\mathbf{T}} \phi(\mathbf{x}) + b^* \\ &= \sum_{t=1}^{T} (\alpha_t' - \alpha_t) \phi^T(\mathbf{x_t}) \phi(\mathbf{x}) + b^* \\ &= \sum_{t=1}^{T} (\alpha_t' - \alpha_t) K(\mathbf{x_t}, \mathbf{x}) + b^* \end{aligned} \tag{13}$$

where $K(\mathbf{x_t}, \mathbf{x}) = \phi^T(\mathbf{x_t})\phi(\mathbf{x})$ is the inner-product kernel function. In fact, the SVM theory considers only the form of $K(\mathbf{x_t}, \mathbf{x})$ in the feature space without specifying explicitly $\phi(\mathbf{x})$ and without computing all corresponding inner products. Therefore, the kernel function greatly reduces the computational complexity of high-dimensional hidden space and becomes the crucial part of SVM. The function which satisfies the Mercer theorem can be chosen as the SVM kernel. No analytical method is currently available to determine the most suitable kernel for a particular dataset. This paper experiments with three different kernels to investigate the effect of a kernel type in Monte Carlo simulation:

$$\text{Linear:} \quad K(\mathbf{x_t}, \mathbf{x}) = \mathbf{x}_t^T \mathbf{x} \tag{14}$$

$$\text{Polynomial:} \quad K(\mathbf{x_t}, \mathbf{x}) = \left(\mathbf{x}_t^T \mathbf{x} + 1\right)^d \tag{15}$$

$$\text{Gaussian:} \quad K(\mathbf{x_t}, \mathbf{x}) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{x_t}\|^2}{2\sigma^2}\right) \tag{16}$$

where d and σ^2 are the parameters for the polynomial and Gaussian kernel. Before implementation of the SVM, the appropriate values of the coefficients ε, C, d and σ^2 must be determined in advance through cross-validation. The sensitivity analysis of the parameters and the kernel type will be illustrated by using the simulated data below ('Monte Carlo Simulation').

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

EMPIRICAL MODELING

In this study, the forecasts are obtained first by applying the Monte Carlo Simulation, following the suggestions in Andersen and Bollerslev (1998) and Clements and Smith (1999, 2001). The main motivation for conducting a simulation experiment is that, since the true volatility is known, the candidate volatility measures can be compared with certainty. We then fit each of the models to the daily returns on the GBP exchange rate and NYSE stock indexes and forecast their respective volatility. The empirical modeling and forecasting scheme described below are employed for both simulation and real data.

Model specification

In this paper the real data we analyze are the daily financial returns, y_t, converted from the corresponding price or index, I_t, using continuous compounding transformation as

$$y_t = 100 \times (\log I_{t+1} - \log I_t) \tag{17}$$

Empirical findings suggest that GARCH is a more parsimonious model than ARCH, and GARCH (1, 1) specification is sufficient to model the variance changing over long sample periods and has become the most popular structure when capturing financial volatility (Akgiray, 1989; Franses and Dijk, 1996; Brooks, 1998; Gokcan, 2000; Andersson, 2001; Brooks and Persand, 2003; Poon and Granger, 2003; Gerlach and Tuyl, 2006). As such, throughout the paper, the analysis is restricted to the case of the GARCH (1, 1) process for the second conditional variance function and the AR(1)[7] process for the conditional mean equation, for the sake of candidate comparison under the same conditions.

Thus the linear standard GARCH (1, 1) model is specified as follows:

$$y_t = c + \phi_1 y_{t-1} + u_t \quad u_t \sim N(0, h_t) \tag{18a}$$

$$h_t = \kappa + \delta_1 h_{t-1} + \alpha_1 u_{t-1}^2 \tag{18b}$$

where c, ϕ_1, κ, δ_1 and α_1 are constant parameters. Such restrictions on the parameters that κ, δ_1 and α_1 are non-negative and $\delta_1 + \alpha_1 < 1$ prevent negative variances (Bollerslev, 1986).

All odd moments of u_t in the standard GARCH model equal zero, and hence u_t and y_t are symmetric time series. The nonlinear EGARCH (1, 1) model that is able to capture the asymmetry is similar to the linear GARCH model but the h_t process is given by

$$\log(h_t) = \kappa + \delta_1 \log(h_{t-1}) + \alpha_1 \left(\frac{|u_{t-1}|}{\sqrt{h_{t-1}}} - \sqrt{2/\pi} \right) + \beta_1 \frac{u_{t-1}}{\sqrt{h_{t-1}}} \tag{19}$$

where κ, δ_1, α_1 and β_1 are the constant parameters. The EGARCH model is fundamentally different from the standard GARCH model in that the standardized innovation serves as the forcing variable for the conditional variance. Also, there are no restrictions on the parameters to ensure non-negativity

[7] Franses and Dijk (1996) also denote that the order of autoregression in the first conditional mean equation of the GARCH framework is usually 0 or small. Thus, the order 1 is specified for this study.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

of the variances. The coefficient β_1 is introduced to capture the asymmetry. If $\beta_1 = 0$, a positive return shock has the same effect on h_t as the negative return shock of the same amount; if $\beta_1 < 0$, a positive return shock actually reduces h_t; if $\beta_1 > 0$, then a positive return shock increases h_t. Previous studies have viewed this coefficient as typically negative, indicating that negative return shocks normally generate more volatility than positive return shocks, so generating the so-called leverage effect.

The conditional variance of u_t is given by $h_t = E_{t-1}u_t^2 = \hat{u}_{t|t-1}^2$. Roughly speaking, in a GARCH process the conditional variances can be modeled by an ARMA type process (Franses and Dijk, 1996). For instance, the ARMA process of the conditional variance of u_t in a linear GARCH model can be expressed as below (Hamilton, 1997; Enders, 2004):

$$u_t^2 = \kappa + (\delta_1 + \alpha_1)u_{t-1}^2 + w_t - \delta_1 w_{t-1} \tag{20}$$

where $w_t \equiv u_t^2 - \hat{u}_{t|t-1}^2 = u_t^2 - h_t$, which is white noisy error. Inspired by this, the nonparametric recurrent ANN and SVM based nonlinear GARCH (1, 1) model is specified as the following form:

$$y_t = f(y_{t-1}) + u_t \tag{21a}$$

$$u_t^2 = g(u_{t-1}^2, w_{t-1}) + w_t \tag{21b}$$

where $f(\cdot)$ and $g(\cdot)$ are nonlinear nonparametric function forms for conditional mean and variance equations, respectively. Note that equation (21b) is adopted for the analysis of real data because the actual volatility h_t is unobservable, while in the case of simulation the conditional variance equation is just specified as $h_t = f(h_{t-1}, u_{t-1}^2)$ due to h_t being known. Because of the way GARCH (1, 1) class models are constructed, the volatility is known at time $t - 1$. Thus the one-step-ahead forecast of volatility is readily available.

The moving average method uses weighted moving averages of past squared innovations to forecast volatility (Niemira and Klein, 1994). For simulated data, the moving average forecast for the next-day volatility, using the five most recent observations, is expressed as

$$\hat{u}_{t+1}^2 = \frac{1}{5}\sum_{j=t-4}^{t} u_j^2 \tag{22}$$

For real data, the moving average forecast for the next-day volatility is expressed as (Engle *et al.*, 1993)

$$\hat{u}_{t+1}^2 = \frac{1}{5}\sum_{j=t-4}^{t} (y_j - \bar{y}_{5,t})^2 \tag{23}$$

where

$$\bar{y}_{5,t} = \frac{1}{5}\sum_{j=t-4}^{t} y_j$$

The recurrent ANN used in this study is the feedback multilayer perceptrons (MLP) network with the addition of a global feedback connection from the output layer to its input space. We specify

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

this kind of recurrent back-propagation network with the following architecture: one nonlinear hidden layer with four neurons, each using a tan-sigmoid differentiable transfer function to generate the output, and one linear output layer with one neuron. As a training algorithm, the fast training Levenberg–Marquardt algorithm is chosen. The value of the learning rate parameter used in the training process is set to be 0.05. These specifications and choices are standard in the neural network literature.

Recurrent SVM procedure

As Haykin (1999) said, the standard SVM described above usually appears in the design of a simple network in which an input layer of source nodes projects onto an output layer of computation node, but not vice versa (see Figure 2(a)). This process is known as feedforward SVM and could be easily employed to estimate such AR process as the first conditional mean function (21a), $y_t = f(y_{t-1}) + u_t$, and the second conditional variance function in the situation of simulation, $h_t = f(h_{t-1}, u_{t-1}^2)$. However, because the unobservable error term w_t is introduced into the GARCH model which indeed exhibits the nonlinear ARMA process, how to estimate the conditional volatility model (21b) for real data?

To estimate the nonlinear ARMA model, a feedback process of SVM with unobservable moving average part as inputs, not addressed before our application[8], has to be described, which distinguishes itself from feedforward SVM in that it has at least one feedback loop (see Figure 2(b)). In this paper, we abuse terminology and refer to this process as 'recurrent SVM'. The feedback loops involve the use of particular branches composed of *one-delay operator*, z^{-1}, which result in nonlinear dynamical behavior and have a profound impact on the learning capability of SVM. Thus the recurrent SVM will capture more dynamic characteristics of y_t than does feedforward SVM.

To overcome the problem that the series of error term w_t is unavailable, we employ the model residuals as estimates of the errors in an iterative way, which is similar to the way that the linear ARMA model is iteratively estimated by MLE (Box *et al.*, 1994; Hamilton, 1997). Likewise, the

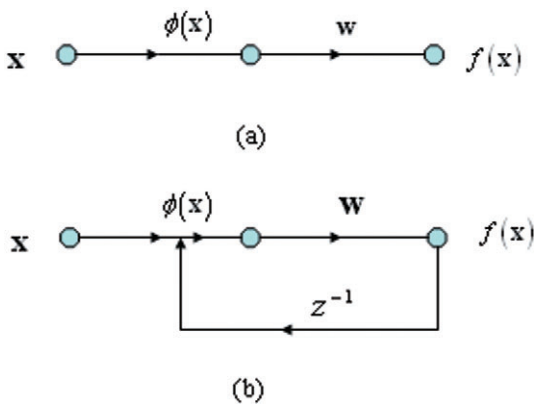

Figure 2. Signal-flow graphs of feedforward and recurrent SVM. (a) Signal-flow graph of a feedforward SVW. (b) Signal-flow graph of a single-loop recurrent SVW

[8] Suykens and Vandewalle (2000) proposed the algorithm of recurrent least squares SVM. The difference between the two recurrent SVM algorithms is their sparseness solutions.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

error term is initially set to be its expectation: zero. The empirical procedure of the recurrent SVM executed during the training phase is described as follows. The letter i indicates the iterative epoch and t denotes the period:

- Step 1: Set $i = 1$ and star with all residuals at zero: $w_t^{(1)} = 0$.
- Step 2: Run an SVM procedure to get the decision function $f^{(i)}$ to the points $\{x_t, y_t\} = \{u_{t-1}^2, u_t^2\}$ with all inputs $x_t = \{u_{t-1}^2, w_{t-1}^{(i)}\}$.
- Step 3: Compute the new residuals $w_t^{(i+1)} = u_t^2 - f^{(i)}$.
- Step 4: Terminate the computational process when the stopping criterion is satisfied; otherwise, set $i = i + 1$ and go back to Step 2.

Note that the first iterative epoch is in fact a feedforward SVM process and results in an AR (1) model and that the following epochs provide results of the ARMA (1, 1) model, being estimated by the recurrent SVM.

In general, the procedure cannot be shown to converge, and there are no well-defined criteria for stopping its operation. Rather, some reasonable criteria can be found, although with its own practical drawback, which may be used to terminate the computational process.

To formulate such a criterion, it is logical to think in terms of the properties of the estimated residual series. After sufficiently long iterative steps, the autocorrelation displayed behind the residuals during the first AR epoch should disappear, and the information in the residual behavior has been completely adopted and the final residual series should be white noisy. Accordingly, we may suggest a sensible convergence criterion for the recurrent SVM procedure as follows:

> *The recurrent SVM procedure is considered to have converged when the corresponding residuals become white noisy, or has no autocorrelation.*

To quantify the measurement of white noise, we use the formal hypothesis test, the Ljung–Box–Pierce Q-test, to investigate a departure from randomness based on the ACF of the residuals. Under the null hypothesis of no autocorrelation in residuals, the Q-test statistic is asymptotically distributed as chi-square. In fact, we just check the actual p-values (exact level of significance) of the Q-test of lag 1. It is reasonable to think there is no higher-order autocorrelation if there is no one-order autocorrelation in residuals. Only if the p-values of the Q-test for five consecutive epochs are simultaneously higher than 0.1 is the iterative computational process stopped. To overcome the drawback of this convergence criterion, we use cross-validation to avoid the possible over-fitting problem; see 'Real data analysis' below for the iterative process in detail.

Forecasting scheme

To illustrate the forecasting scheme, the SVM-GARCH model is also exemplified. First, estimate the conditional mean equation (21a) by using the feedforward SVM in the full sample period $T(1, 2, \ldots, T)$ to obtain residuals, $u_1, u_2, \ldots, u_T$. Then, recursively run the SVM-GARCH (1, 1) model for squared residuals thus obtained to forecast the one-period-ahead volatility. The recursive forecasting scheme is employed with an updating sample window; the estimating and forecasting process is carried out recursively by updating the sample with one observation each time, rerunning the SVM approach and recalculating the model parameters and corresponding forecasts. Here, the SVM approach to estimate the conditional volatility is feedforward for simulation and recurrent, as described in the above subsection, for real data. The first training sample is $u_1^2, u_2^2, \ldots, u_{T_1}^2$ ($T_1 < T$). The observations of $T - T_1$ are retained as a forecasting or test sample.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Therefore, we can estimate and forecast the SVM-based conditional volatility equation for $n = T - T_1$ times. We set $n = 60$ for both simulation and real data in this study. Thus, 60 one-period-ahead forecast volatilities, $\hat{u}^2_{T-59}$, $\hat{u}^2_{T-58}, \ldots, \hat{u}^2_{T-1}$, $\hat{u}^2_T$, will be acquired for out-of-sample forecasting evaluation.

Evaluation measures and pairwise comparison of competing models

We evaluate the forecasting performance using two standard statistical criteria: mean absolute forecast error (MAE) and directional accuracy (DA), expressed as follows (Brooks, 1998; Moosa, 2000):

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=T_1}^{T-1}\left|u^2_{t+1} - \hat{u}^2_{t+1}\right| \tag{24}$$

$$\mathrm{DA}(\%) = \frac{100}{n}\sum_{t=T_1}^{T-1} a_t \tag{25}$$

where

$$a_t = \begin{cases} 1 & \left(u^2_{t+1} - u^2_t\right)\left(\hat{u}^2_{t+1} - \hat{u}^2_t\right) \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

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

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 paper, we use the square of the return minus its mean value as the surrogate of actual volatility against which MAE and DA can be calculated. This approach is similar to the standard one, squared returns, because the mean of returns is usually close to zero. The proxy of actual volatility in real data is expressed as

$$u^2_t = (y_t - \bar{y})^2 \tag{26}$$

where y_t is returns and $\bar{y}$ is mean of returns. This proxy has been used in many recent papers, such as Pagan and Schwert (1990), Day and Lewis (1992), Chan *et al.* (1995), West and Cho (1995), Chong *et al.* (1999), Brooks (2001) and Brooks and Persand (2003).

To test for equal forecasting accuracy of two competing models, we use the two-sided DM test statistic proposed by Diebold and Mariano (1995) for the difference of MAE loss function. The null and alternative hypotheses in this case are

$$H_0: \mathrm{MAE}_1 - \mathrm{MAE}_0 = 0 \textit{ versus } H_1: \mathrm{MAE}_1 - \mathrm{MAE}_0 \neq 0$$

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

where the subscript 0 denotes the benchmark model and 1 the competing model. The DM statistic in a robust form is then based on the following large sample statistic:

$$\mathrm{DM} = \frac{1}{\sqrt{n}} \frac{1}{\sqrt{\hat{S}^2}} \sum_{t=T_1}^{T-1} \left(\left| u_{t+1}^2 - \hat{u}_{1,t+1}^2 \right| - \left| u_{t+1}^2 - \hat{u}_{0,t+1}^2 \right| \right) \sim N(0,1) \tag{27}$$

where $\hat{S}^2$ denotes a heteroscedasticity and autocorrelation consistent (HAC) robust (co)variance matrix which is estimated according to the Newey–West procedure (Newey and West, 1987). We use Andrews' (1991) approximation rule to automatically select the number of lags for the HAC matrix. If n grows at a rate such that as $T \to \infty$, $n \to \infty$ and $n/T_1 \to 0$, then the DM statistic converges in distribution to a standard normal.

MONTE CARLO SIMULATION

Data-generating process

In this section we investigate the forecasting performance of all candidates using artificial simulated data under controlled conditions. To generate the data, we first need to parameterize the GARCH (1, 1) model in equation (18) with the following settings (c, ϕ_1, κ, δ_1, α_1) = (0, 0.5, 0.0005, 0.8, 0.1) for medium persistence and a disturbance term u_t distributed first as Gaussian and then as a Student's t with five degrees of freedom (kurtosis = 5). The second distribution tries to model the skewness and excess of kurtosis that usually appears in real financial series. Using the same specified models, two artificial samples of size 500 and 1000 are created under a two-distributions assumption, giving a total of four situations. To limit the computational burden, each situation is replicated only 50 times. Then the multiple simulated y_t and h_t are 500 × 50 and 1000 × 50 element matrices for different distribution.

Parameter selection

The use of cross-validation is appealing particularly when we have to design a somewhat complex approach with good generalization as the goal. For example, here we may use cross-validation to determine the values of free parameters of SVM with the best performance. One series of 50 simulated returns and volatility of 1000 size and Student's t distribution, one of the four situations, is exemplified as below. The first training data, that is, the former 940 observations, are used to determine the appropriate values taken by the free parameters. The training data are further randomly partitioned into two disjoint subsets: estimating sample and validating sample (700 and 240 observations, respectively).

As shown above, two free parameters (ε and C) and two kernel coefficients (d and σ^2) have to be selected by users before running the SVM procedure. The motivation for using cross-validation here is to validate the model on a dataset different from the one used for parameter estimation. In this way we may use the training set to assess the performance of various values of parameters, and thereby choose the best one. The sensitivity investigation of SVM (represented by the generalization error, MAE) with respect to four parameters is illustrated in Figures 3 and 4 for conditional mean and variance estimation, respectively.

Figure 3 describes the sensitivity analysis for the conditional mean equation. Parameter C varies from a very small value of 0.0001 to infinity, with ε being fixed at 0.0001 and σ^2 0.4. Clearly, when

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

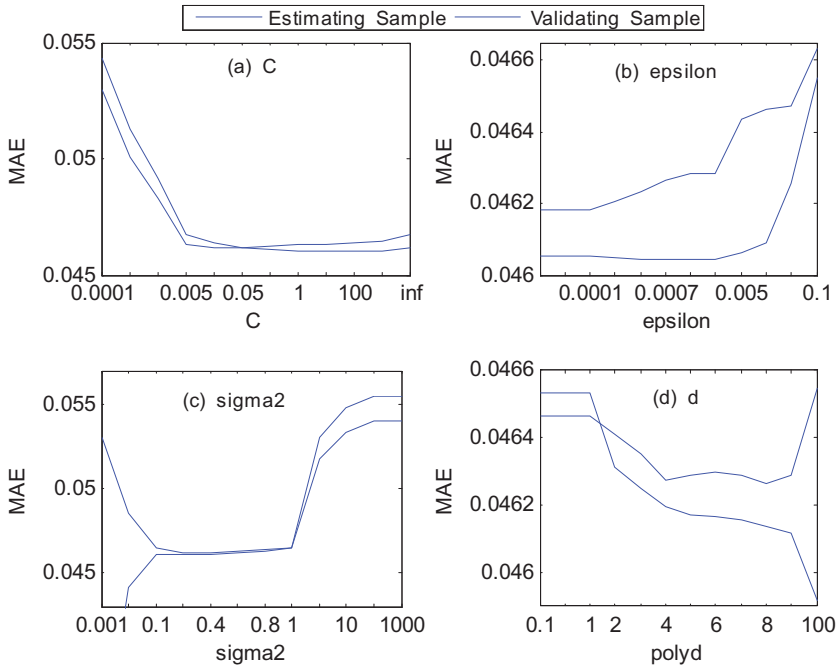

Figure 3. Sensitivity analysis of SVM in conditional mean estimation

$C = 0.05$, MAE of the validation sample obtains the lowest value, 0.046. Parameter ε takes values in the range [0.00001, 0.00005, 0.0001, 0.0003, 0.0005, 0.0007, 0.0009, 0.001, 0.005, 0.01, 0.05, 0.1], with $C = 0.05$ and $\sigma^2 = 0.4$. The values of ε to the left of the point = 0.0001 have no influence on the performance of SVM. Coefficient σ^2 varies from values of 0.001 to 1000, with C being 0.05 and 0.0001. Obviously, the value of $\sigma^2 = 0.4$ leads to the best validation performance. If we set $C = 0.05$ and 0.0001 and the polynomial kernel parameter d = [0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 10, 100], the validating MAE attains the minima when $d = 8$; after that, over-fitting the training set occurs. Note that the polynomial kernel with $d = 1$ is similar to the linear kernel. Thus, the appropriate parameters of SVM for the conditional mean returns are: $C = 0.05$, $\varepsilon = 0.0001$, $\sigma^2 = 0.4$ and $d = 8$.

Figure 4 describes the parameter selection process for conditional variance series. Similar to the return series, the MAE of both estimating and validating sample decreases as the values of C increase and become stable when C takes a value greater than 10; in contrast to C, as the values of ε increase, both MAE of SVM are considerably more stable before the point of $\varepsilon = 0.0001$ and increase slowly, and sharply after $\varepsilon = 0.001$. The value of $\sigma^2 = 0.01$ results in the best validation performance; namely, its MAE reaches the minimum value, about 0.000065. The values of d taken between 100 and 1000 have not much effect on the performance of SVM but after that range the over-fitting phenomenon becomes serious. Likewise, when one parameter is analyzed, the others are set to be fixed. Therefore, the correct parameters chosen for the conditional variance series are $C = 10$, $\varepsilon = 0.00005$, $\sigma^2 = 0.01$ and $d = 250$, respectively.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

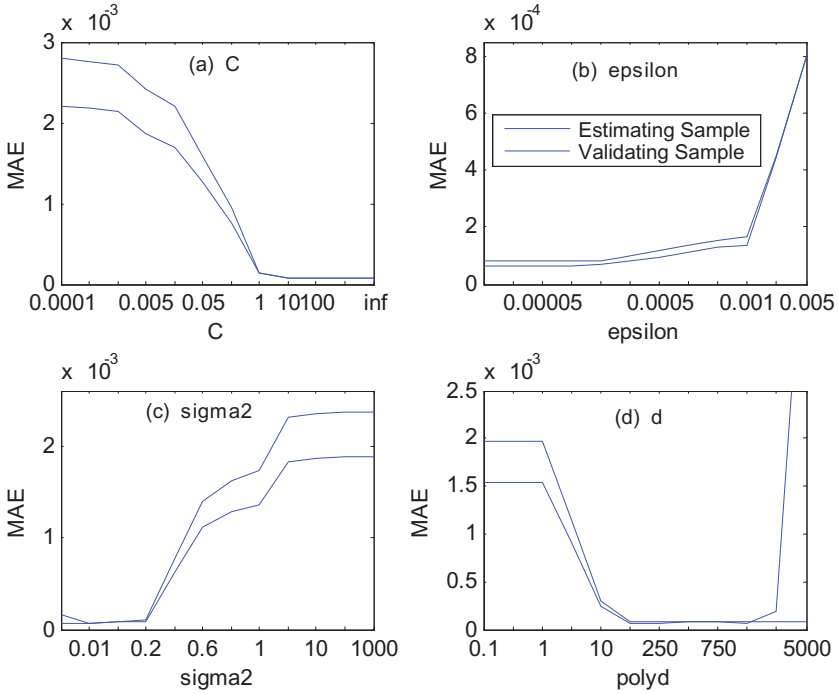

Figure 4. Sensitivity analysis of SVM in conditional variance estimation

Thus far we discuss the sensitivity investigation of parameters by using the simulated data with 1000 observations and *t* distribution. The parameter selection for the other three random samples is similar to this and not reported here to save space.

EFFECT OF KERNEL TYPE AND FORECASTING EVALUATION

There is still the possibility of over-fitting after training. Therefore, the generalization performance of the competing models is further measured and evaluated on the test set, which is different from the validation subset. For the simulated data, the forecasting sample is the last 60 observations. For each replication, the SVM-based GARCH (1, 1) model and the others are estimated, and the forecasting errors are calculated using the forecasting schemes described above. The results of out-of-sample one-period-ahead volatility forecasting measures for four situations are shown in Table I. The reported results are the mean values of 50 independent replications. Table II presents the *p*-values of Diebold-Mariano (DM) test for the MAE difference, which are defined as the significance levels at which the null hypothesis under investigation can be rejected. In calculating the DM statistic, the null hypothesis of equal forecasting ability is related to the four benchmark models: moving average, standard GARCH, EGARCH and traditional ANN models. We report the results of the DM test, say DM1, in the third and seventh columns for two simulated series, respectively, under the null hypothesis that the absolute forecast error produced by the moving average method is equal to those obtained using the other models. DM2, DM3 and DM4 are organized in the same manner and show the test results when the benchmark models are respectively the standard GARCH, EGARCH and recurrent ANN models. The DM tests in this study are investigated in a robust form, by simply

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Table I. Diebold–Mariano test for the MAE difference on real data

Models	Sample Size = 500				Sample Size = 1000			
	Normality		Student's *t*		Normality		Student's *t*	
	MAE	DA	MAE	DA	MAE	DA	MAE	DA
Moving Average	0.0001276	44.07	0.0001747	59.32	0.0001198	54.24	0.0002130	40.68
Standard GARCH	0.0000972	76.27	0.0001765	55.93	0.0000488	79.66	0.0001083	59.32
EGARCH	0.0001312	67.80	0.0002075	64.41	0.0000730	57.63	0.0001864	74.58
ANN-GARCH	0.0001517	72.88	0.0002481	57.63	0.0000904	62.71	0.0001442	67.80
SVMl-GARCH	0.0000960	76.27	0.0001369	71.19	0.0000501	74.58	0.0000715	72.88
SVMp-GARCH	0.0000924	76.27	0.0001371	71.19	0.0000479	71.19	0.0000714	77.97
SVMg-GARCH	0.0000796	86.44	0.0001397	81.36	0.0000456	83.05	0.0000769	98.31

Note: SVMl, SVMp and SVMg represent the SVM with linear, polynomial and Gaussian kernel, respectively, for short.

scaling the numerator by a heteroscedasticity and autocorrelation consistent (HAC) (co)variance matrix calculated according to Newey-West procedures (Newey and West, 1987).

Table I firstly shows the effect of kernel functions on out-of-sample forecasting performance of SVM. The linear kernel behaves better in the sample with 500 sizes and *t* distribution based on DA measure. The polynomial kernel is the most suitable for forecasting the *t*-distributed 1000 sample size also based on DA. For all the other six cases, the Gaussian kernel looks promising, however, which is not a general conclusion but only true for the case we are studying. As a whole, three types of kernel-based SVM have a similar volatility forecasting performance and almost behave better than the benchmarks. Since no single kernel function dominates all volatility predictions, practitioners could try any kernel function. In the real data analysis later, for example, we only investigate the performance of the Gaussian kernel-based SVM-GARCH model.

Now, based on Table I, we revert to comparing the volatility forecasting ability among all competing models. In terms of the average ranking of MAE measures, the order of the forecasting ability of the different methods from highest to lowest is displayed in turn as follows: SVMp-GARCH, SVMg-GARCH, SVMl-GARCH[9], standard GARCH, EGARCH, moving average and ANN-GARCH model. Concretely, in the situation of normal distribution, the standard GARCH model behaves not badly, which is ranked fourth (only inferior to three SVM models) in the 500 sizes and even ranked third (only defeated by Gaussian and polynomial SVM models) in the series of 1000 sizes. Even though the data satisfy the normality assumption that is required for MLE in the standard GARCH model, the SVM-GARCH models still outperform it in forecasting the magnitude of the volatility error. Nonlinear EGARCH and ANN-GARCH models perform worse than the linear GARCH model. In the situation of *t* distribution, the forecasting performance of the linear GARCH model grows poorer and the difference of MAE values between SVM-GARCH and standard MLE-GARCH models becomes larger than that under normality. Possibly this results from the fact that the normality assumption required for MLE is violated but it is not necessary for the SVM method. Not as expected, the asymmetric EGARCH model is weak in reducing the forecasting error even in the case of skewed distribution.

Based on the DA measures in Table I, on average, the Gaussian SVM-GARCH model ranks highest (for all four situations) in forecasting volatility directions, followed by polynomial and linear

[9]That is, corresponding to SVM-based GARCH models with polynomial, Gaussian and linear kernel function, respectively.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

SVM-GARCH models, linear GARCH model, EGARCH model, ANN-GARCH model and moving average, in turn. In the situation of the normal distribution, the standard GARCH model behaves even better than forecasting error magnitude—ranked second for both the series of 500 sizes (only inferior to Gaussian but equal to linear and polynomial SVM models) and 1000 sizes (worse than Gaussian but better than the other two SVM type models). In the case of normality and large sample sizes, particularly favorable for MLE, the standard GARCH model still cannot defeat the Gaussian-based SVM-GARCH model. It is not surprising for EGARCH to behave badly in this case. As for the situation of *t* distribution, the linear GARCH model is ranked last for the 500 sizes (55.93%) and second last for the 1000 sizes (59.32%); while the asymmetric EGARCH model is good at forecasts of volatility turning points—ranked fourth for short series (only behind the three SVM models) and even third for long series (inferior to Gaussian and polynomial but better than the linear SVM-GARHC model). This time the ANN-GARCH model defeats the linear GARCH model. As for the linear GARCH model and moving average method, in the situation of 500 sizes and *t* distribution the standard GARCH model performs worse than the moving average, the simplest time series method, in terms of both MAE and DA measures. The conclusions described above are obtained on average based on 50 replications.

Table II displays the *p*-values of the DM test when the moving average method, standard GARCH, EGARCH and ANN models are compared with each of the other models considered in the study. We denote these tests DM1, DM2, DM3 and DM4, respectively. For instance, DM1 presents the test results for the simple moving average, where a *p*-value no greater than 0.05 indicates that the moving average method yields a higher forecast error (in terms of absolute error) relative to the competing model at 5% significance level, a *p*-value no smaller than 0.95 means that the moving average produces a lower forecast error at the 5% level, while a *p*-value between 0.05 and 0.95 implies that the benchmark and competing model have equivalent forecasting accuracy from the viewpoint of statistics. The same interpretation applies to the *p*-values reported for DM2-DM4.

Table II. Diebold–Mariano test for the MAE difference on Monte Carlo simulation

Distribution	Models	Sample size = 500				Sample size = 1000			
		DM1	DM2	DM3	DM4	DM1	DM2	DM3	DM4
Normality	Moving average		0.976	0.401	0.070		1.000	0.999	0.875
	Standard GARCH	0.024		0.001	0.000	0.000		0.001	0.000
	EGARCH	0.600	0.999		0.005	0.001	0.999		0.033
	ANN-GARCH	0.930	1.000	0.995		0.125	1.000	0.967	
	SVMl-GARCH	0.018	0.460	0.002	0.000	0.000	0.574	0.002	0.000
	SVMp-GARCH	0.023	0.413	0.004	0.000	0.000	0.420	0.003	0.000
	SVMg-GARCH	0.002	0.097	0.000	0.000	0.000	0.354	0.000	0.000
Student's *t*	Moving average		0.480	0.036	0.000		1.000	0.822	0.984
	Standard GARCH	0.520		0.054	0.003	0.000		0.000	0.001
	EGARCH	0.964	0.946		0.021	0.178	1.000		0.966
	ANN-GARCH	1.000	0.997	0.979		0.016	0.999	0.034	
	SVMl-GARCH	0.043	0.037	0.002	0.000	0.000	0.019	0.000	0.000
	SVMp-GARCH	0.056	0.043	0.001	0.000	0.000	0.025	0.000	0.000
	SVMg-GARCH	0.070	0.050	0.000	0.000	0.000	0.033	0.000	0.000

Note: DM1, DM2, DM3 and DM4 are the robust Diebold and Mariano (1995) test by using the Newey–West procedures (Newey and West, 1987) when the benchmark models are the moving average, linear GARCH model, EGARCH model and traditional ANN-GARCH model, respectively. For each test we consider the MAE loss functions.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Under the normal distribution, DM1 tests indicate that there is equivalent forecasting ability between moving average and EGARCH for short series, and between moving average and ANN-GARCH for long series. Such models as standard GARCH and the three SVM-GARCH all have higher volatility forecasting accuracy than moving average for both series at least at the 5% significance level. Moving average outperforms the ANN-GARCH model at the 10% level for a series of 500 size and EGARCH outperforms moving average at the 0.1% significance level for long series. According to DM2, three SVM type models have statistically equivalent forecasting ability to standard GARCH model for both series, with only one exception that the Gaussian SVM-GARCH model behaves better than the standard GARCH model at 10% significance level for short series. For both series, the standard GARCH model outperforms EGARCH and ANN-GARCH models at extremely low significance level. The DM3 statistic reveals that, for two series, three SVM-GARCH models perform better than the EGARCH model and EGARCH better than the ANN-GARCH model all at extremely significant levels. Finally, the ANN-GARCH model is found statistically and consistently inferior to the three SVM models for any series based on DM4 tests.

In the case of Student's t distribution, the out-of-sample performance of the standard GARCH model deteriorates. Now, according to DM2, the three SVM-GARCH models forecast volatility significantly better than the standard GARCH model at the 5% level for both series. The standard GARCH model cannot statistically defeat the moving average, either, for short series. However, both EGARCH and ANN-GARCH models are still statistically inferior to the standard GARCH model. In fact, according to DM1, DM3 and DM4, the three SVM-GARCH models all consistently outperform such benchmarks as moving average, EGARCH and ANN-GARCH models in forecasting volatility for any series. In terms of DM1, furthermore, the null hypothesis of equal forecasting accuracy between moving average and EGARCH cannot be rejected for a series of 1000 size rather 500 size. Moving average is significantly better than the ANN-GARCH model for short series, but the case is reversed for long series. In a series of 500 sizes, the ANN-GARCH model is significantly outperformed by the EGARCH model, while for the series of 1000 size the ANN type model statistically defeats the EGARCH model.

In summary, it appears that the three SVM-GARCH models do a better job of forecasting volatility than the moving average, standard GARCH, EGARCH and ANN-GARCH models in terms of MAE measures, which is statistically supported by the DM1, DM3, DM4 tests and DM2 in the case of t distribution. The DM2 test reveals that under the normal distribution the three SVM-GARCH models and standard GARCH model have similar volatility forecasting ability. Based on DA measures, the standard GARCH model too has a better ability in forecasting volatility turning points under normality and large sample sizes, while the asymmetric EGARCH model behaves better under the skewed t distribution. But both linear GARCH and nonlinear EGARCH cannot defeat all SVM-type models, at least the Gaussian-based SVM-GARCH model, in forecasting volatility directions.

REAL DATA ANALYSIS

In this section, we investigate the volatility forecasting performance of all candidates by using real data for two kinds of financial variables: GBP/USD exchange rates and NYSE average index.

Data description

The first dataset consists of the daily nominal bilateral exchange rates of British pounds (GBP) against the US dollar for the period January 5, 2004 to December 31, 2007. The data are obtained

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

from a database provided by Policy Analysis Computing and Information Facility in Commerce (PACIFIC) at the University of British Columbia, which contains the closing rates for a total of 81 currencies and commodities. The second dataset consists of the daily closing price of the New York Stock Exchange (NYSE) composite stock index for the period January 8, 2004 to December 31, 2007. The data are downloaded directly from the Market Information section of the NYSE web page.

It has been widely accepted that a variety of financial variables including foreign exchange rates and stock prices are integrated of order one. To avoid the issue of possible nonstationarity, both sets of raw real data are transformed into daily returns via equation (17), giving a returns series of 1001 observations and then a residual series is obtained from a fitted conditional mean equation of the GARCH class models. For the squared residuals of 1000 observations, the recursive estimating samples for the conditional volatility function are updated from the former 940 observations through the former 999 and then 60 numbers of one-period-ahead volatility forecasts are obtained, corresponding to an evaluation sample spanned from the 941st through the 1000th data points, that is, out-of-sample period of October 3, 2007 to December 31, 2007 for GBP and October 5, 2007 to December 31, 2007 for NYSE data.

The daily series for the log-levels and the returns of the GBP and NYSE are depicted in Figure 5. This figure shows 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. Table III reports the summary of the descriptive statistics for the GBP and NYSE returns. Both series are typically characterized by excessive kurtosis and asymmetry. The Bera and Jarque (1981) tests all strongly reject the normality hypothesis. For GBP series, the Ljung–Box Q(6) statistic

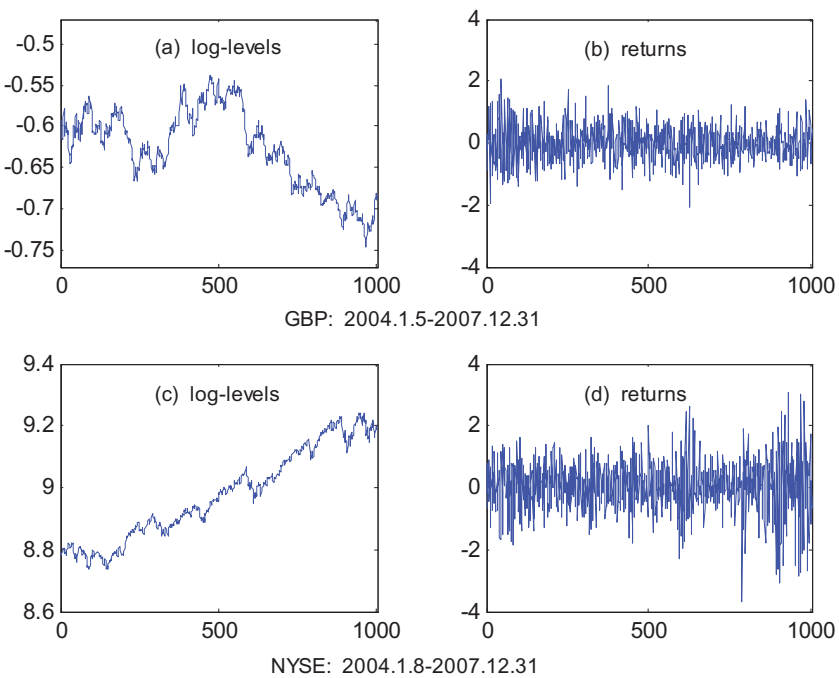

Figure 5. Log levels and returns of GBP exchange rates and NYSE stock index

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Table III. Descriptive statistics for daily financial returns

Returns	GBP		NYSE	
	Statistics	*p*-value	Statistics	*p*-value
Mean	–0.0092		0.0393	
Variance	0.2827		0.6197	
Skewness	0.1206		–0.3489	
Kurtosis	3.7130		4.9343	
Normality	23.1860	0.00001	174.7200	0.00000
Q(6)	3.0313	0.80490	12.7100	0.04788
Q(6)*	31.6390	0.00002	150.2400	0.00000
ARCH(6)	28.9280	0.00006	101.8400	0.00000

Notes: Normality is the Bera-Jarque (1981) normality test; Q(6) is the Ljung-Box Q test at 6 order for raw returns; Q(6)* is LB Q test for squared returns; ARCH(6) is Engle's (1982) LM test for ARCH effect.

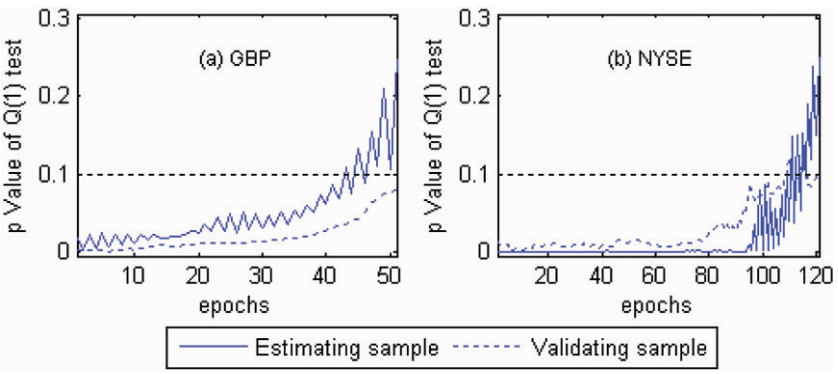

Figure 6. Iterative epochs of recurrent SVR procedure for real data

of raw returns indicates no significant correlation, but the Q(6) value of the squared returns reveals that there is significant autocorrelation in the squared returns. The Q(6) tests of both raw and squared returns of NYSE are all significant. Engle's (1982) LM tests for ARCH effect show significant evidence in support of GARCH effects (i.e., heteroscedasticity) for both series. Note that the number in parentheses indicates testing at 6 lag order. This examination of daily returns on the GBP and NYSE data reveals that returns can be characterized by heteroscedasticity and time-varying autocorrelation; therefore, we expect the GARCH class models to capture it adequately. Furthermore, as seen from Figure 5 and Table III, it seems that NYSE returns exhibit more variability, skewness, kurtosis and volatility clustering than GBP series such that nonlinear asymmetric EGARCH model should fit it more accurately.

Iterative epochs of recurrent SVM

Because the actual volatility h_t is unobservable for real data analysis, the second conditional variance equation (21b) of the GARCH (1, 1) model should be estimated by using the recurrent SVM procedure, as proposed above. Again, we use cross-validation to determine when the procedure is stopped.

With good forecasting performance as the goal, it is very difficult to figure out when it is best to stop training only in terms of fitting performance. It is possible for the procedure to end up

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

over-fitting the training data if the training session is not stopped at the right point. We can identify the onset of over-fitting and the stopping point through the use of cross-validation. Figure 6(a) and (b) describes the iterative epochs for volatility prediction of the first training sample of GBP and NYSE, respectively. For the GBP series, the iterative process of recurrent SVM procedure is stopped at the 51st epoch; while, for NYSE, the iterative process is longer and stopped after 121 iterative steps, possibly due to higher kurtosis and more variability and noise behind the NYSE series. Now, we could say, at about the 10% level of significance, the final residuals of equation (21b) obtained from the recurrent SVM procedure have no autocorrelation. In addition, the *p*-value curves of both estimating and validating samples exhibit a similar pattern (namely, increase with an increasing number of epochs) and point to almost the same stopping point. That is to say, there is no over-fitting phenomenon for the examples illustrated here; the recurrent SVM model does as well on the validating subset as it does on the estimating subset, on which its design is based.

The values taken by the free parameter of SVM and kernel coefficients are also selected according to the sensitivity investigation, similar to that done in Monte Carlo simulation. We do not report the parameter selection process here but present the formal results throughout the real data analysis. For both conditional mean and variance estimation of GBP and NYSE series, fortunately, similar parameter values of feedforward and recurrent SVM procedure could be found as follows: $C = 0.005$, $\varepsilon = 0.05$ and $\sigma^2 = 0.2$. Note that in the analysis of financial returns only the Gaussian kernel is employed for the sake of simplicity due to its best performance among linear, polynomial and Gaussian kernels, as described in Monte Carlo simulation.

Comparing the forecasting ability

The results of out-of-sample volatility forecasting accuracy for each model by using real data are presented in Table IV. Table V reports the *p*-values of the Diebold– Mariano (DM) test for the difference of MAE loss function in a robust HAC form from Newey–West procedures. In calculating the DM statistic, the null hypothesis of equal forecasting accuracy is related to the four benchmark

Table IV. Measure of volatility forecasting performance for real data

Models	Measures	Moving average	Standard GARCH	EGARCH	ANN-GARCH	SVM-GARCH
GBP	MAE	0.28895	0.24713	0.25719	0.24691	0.23257
	DA	37.29	38.98	49.15	38.98	45.76
NYSE	MAE	1.69610	1.51000	1.44880	1.62980	1.50410
	DA	32.20	42.37	55.93	32.20	57.63

Table V. Diebold–Mariano test for the MAE difference on real data

Models	GBP				NYSE			
	DM1	DM2	DM3	DM4	DM1	DM2	DM3	DM4
Moving average		0.990	0.970	0.981		0.935	0.970	0.813
Standard GARCH	0.010		0.017	0.583	0.065		0.902	0.061
EGARCH	0.030	0.983		0.980	0.030	0.098		0.044
ANN-GARCH	0.019	0.417	0.020		0.187	0.939	0.956	
SVM-GARCH	0.001	0.076	0.000	0.067	0.047	0.054	0.885	0.042

Note: DM1, DM2, DM3 and DM4 are the robust Diebold and Mariano (1995) test by using the Newey–West procedures (Newey and West, 1987) when the benchmark models are the moving average, linear GARCH model, EGARCH model and traditional ANN-GARCH model, respectively. For each test we consider the MAE loss functions.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

models: moving average, standard GARCH, EGARCH and ANN models. We specify them as DM1, DM2, DM3 and DM4, respectively. A *p*-value no greater than 0.05 indicates that the benchmark model yields a higher forecast error (in terms of absolute error) relative to the competing model at the 5% significance level, a *p*-value no smaller than 0.95 means that benchmark model produces a lower forecast error at 5% level, while a *p*-value between 0.10 and 0.90 implies that the benchmark and competing models have the equal forecasting accuracy at 10% significance level.

According to MAE measures in Table IV, the SVM-GARCH model is the best one for the GBP series and second for the NYSE series in forecasting the magnitude of volatility error. DM tests in Table V almost statistically favor the SVM-GARCH model as the best model, too, at least at 10% significance level. Even though the MAE metric reveals that the EGARCH model outperforms the SVM-GARCH model for the NYSE series, it is not supported by the DM3 test, which means both models have equal forecasting ability. The better performance of the EGARCH model for NYSE is perhaps due to its ability to capture higher skewness and asymmetry occurring in the SYSE series than in GBP. The standard GARCH model performs modestly in terms of MAE measures, statistically inferior to EGARCH and superior to the ANN-GARCH model for NYSE and significantly better than EGARCH and similar to the ANN-GARCH model for GBP according to DM2 tests. The moving average method is always ranked last in forecasting the magnitude of volatility error, the evidence being significantly supported at least at the 10% level by the DM1 tests in Table V with just one exception, that for NYSE series moving average and ANN-GARCH model have equal forecasting ability. MAE measures and DM3 and DM4 tests denote that the EGARCH model also significantly outperforms the ANN-GARCH model for highly skewed NYSE series but the case is totally reverse for the GBP sample.

Based on DA measures in Table IV, on average, the moving average method is still ranked last, the ANN-GARCH model is ranked second last and the standard GARCH model is ranked at the middle position in forecasting volatility directions. For the GBP series, EGARCH performs best with DA value to be highest 49.15%, followed closely by the SVM-GARCH model; while, for the NYSE model, the best model to forecast volatility turning points is the SVM-GARCH model, with the asymmetric EGARCH model is ranked second, their DA values being 57.63% and 55.93%, respectively.

The empirical evidence of real data also confirms the conclusion obtained in Monte Carlo simulation and favors the theoretical advantage of the SVM-GARCH model. Due to high skewness in financial returns, the asymmetric EGARCH model normally behaves better than the standard GARCH model, particularly in the case of higher skewness or in forecasting volatility turning points. The moving average method always behaves worst and the ANN-GARCH model sometimes good in forecasting one-period-ahead financial volatilities among all candidates.

CONCLUSIONS

In many applications, SVM has shown excellent forecasting performance due to its particular structural design of SRM principle rather than ERM employed by conventional ANN and MLE methods. This inspires us to use it to improve the volatility forecasting ability of the parametric GARCH models. Empirical applications are made for forecasting the simulated data and the real data of daily GBP exchange rates and NYSE stock index.

To avoid the problem that the actual volatility for real data is unobservable, we propose a recurrent SVM procedure with a global feedback loop from the output layer to the input, as opposed to

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

the feedforward one for simulation, to estimate the conditional volatility equation, that is the ARMA process in nature, of the nonlinear GARCH model. The forecasting performance of the SVM-GARCH model is compared with the moving average, standard GARCH, asymmetric EGARCH and traditional ANN-GARCH models based on two quantitative evaluation measures and robust Diebold–Mariano tests following the Newey–West procedure.

The real data results, together with the simulation evidence, consistently and significantly support the use of the feedforward and recurrent SVM-based GARCH (1, 1) models in forecasting the one-period-ahead volatility error magnitude and direction. The standard GARCH model also performs well in the case of normality and large sample size, while the asymmetric EGARCH model is good at forecasting volatility under the high skewed distribution; but they rarely exceed SVM-GARCH models, at least the Gaussian-type SVM. The recurrent ANN-GARCH model and moving average method behave well only in a few cases. Overall, empirical analysis is in favor of the theoretical advantage of the SVM.

How to choose the appropriate values of free parameters and kernel coefficients and what effect of kernel type in the SVM procedure are investigated by using the sensitivity analysis in Monte Carlo simulation. The iterative process of the proposed recurrent SVM procedure in real data analysis is also examined in detail by the cross-validation method, which is shown to be implemented very easily and could be adopted as another standard SVM construction procedure in other applications.

ACKNOWLEDGEMENTS

The authors acknowledge the editor, Derek Bunn, and the referees for their constructive comments. Thanks also goes to the production editor, Ivry Tan, and my student, Qian Feng, who print and proofread the manuscript. This work is sponsored by Deutsche Forschungsgemeinschaft through the SFB 649 'Economic Risk'. Shiyi Chen is also supported by Kyungpook National University Graduate Scholarship for Excellent International Students, Shanghai Leading Academic Discipline Project (No. B101) and State Innovative Institute of Project 985 at Fudan University.

REFERENCES

Akgiray V. 1989. Conditional heteroskedasticity in time series models of stock returns: evidence and forecasts. *Journal of Business* **62**(1): 55–80.

Andersen T, Bollerslev T. 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* **39**: 885–905.

Andersen T, Bollerslev T, Diebold F, Labys P. 2003. Modeling and forecasting realized volatility. *Econometrica* **71**: 579–625.

Andersson J. 2001. On the normal inverse gaussian stochastic volatility model. *Journal of Business and Economic Statistics* **19**: 44–54.

Andrews D. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**: 817–858.

Ané T, Ureche-Rangau L, Gambet J, Bouverot J. 2008. Robust outlier detection for Asia-Pacific stock index returns. *Journal of International Financial Markets, Institutions and Money* **18**(4): 326–343.

Awartani B, Corradi V. 2005. Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries. *International Journal of Forecasting* **21**(1): 167–183.

Balaban E. 2004. Comparative forecasting performance of symmetric and asymmetric conditional volatility models of an exchange rate. *Economics Letters* **83**(1): 99–105.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Bauwens L, Sebastien L, Jeroen R. 2006. Multivariate GARCH models: a survey. *Journal of Applied Econometrics* **21**: 79–109.
Becker R, Clements A, White S. 2007. Does implied volatility provide any information beyond that captured in model-based volatility forecasts? *Journal of Banking and Finance* **31**(8): 2535–2549.
Becker R, Clements A, McClelland A. 2009. The jump component of S&P 500 volatility and the VIX index. *Journal of Banking and Finance* **33**(6): 1033–1038.
Bekiros S, Georgoutsos D. 2008. Direction-of-change forecasting using a volatility-based recurrent neural network. *Journal of Forecasting* **27**(5): 407–417.
Bera A, Jarque C. 1981. An efficient large-sample test for normality of observations and regression residuals. *Australian National University Working Papers in Econometrics*, 40. Canberra.
Blair B, Poon S,-H, Taylor S. 2001. Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics* **105**: 5–26.
Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307–327.
Bollerslev T, Chou R, Kroner K. 1992. Arch modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics* **52**: 5–59.
Box G, Jenkins G, Reinsel G. 1994. *Time Series Analysis: Forecasting and Control*. Prentice Hall: Englewood Cliffs, NJ.
Brailsford T, Faff R. 1996. An evaluation of volatility forecasting techniques. *Journal of Banking and Finance* **20**: 419–438.
Brooks C. 1998. Predicting stock index volatility: can market volume help? *Journal of Forecasting* **17**: 59–80.
Brooks C. 2001. A double-threshold GARCH model for the french franc/deutschmark exchange rate. *Journal of Forecasting* **20**: 135–143.
Brooks C, Persand G. 2003. Volatility forecasting for risk management. *Journal of Forecasting* **22**: 1–22.
Cao C, Tsay R. 1992. Nonlinear time-series analysis of stock volatilities. *Journal of Applied Econometrics* **1**: 165–185.
Cao L, Tay F. 2001. Financial forecasting using support vector machines. *Neural Computation and Application* **10**: 184–192.
Chan K, Christie W, Schultz P. 1995. Market structure and the intraday pattern of bid-ask spreads for NASDAQ securities. *Journal of Business* **68**(1): 35–40.
Chen G, Choi Y, Zhou Y. 2008. Detections of changes in return by a wavelet smoother with conditional heteroscedastic volatility. *Journal of Econometrics* **143**(2): 227–262.
Chen S, Härdle W, Moro R. 2009. Modeling default risk with support vector machines. *Quantitative Finance* (accepted for publication).
Chib S, Nardari F, Shephard N. 2002. Markov chain Monte Carlo methods for generalized stochastic volatility models. *Journal of Econometrics* **108**: 281–316.
Chong C, Ahmad M, Abdullah M. 1999. Performance of GARCH models in forecasting stock market volatility. *Journal of Forecasting* **18**: 333–343.
Choudhry T, Wu H. 2008. Forecasting ability of GARCH vs kalman filter method: evidence from daily UK time-varying beta. *Journal of Forecasting* **27**(8): 670–689.
Clements M, Smith J. 1999. A Monte Carlo study of the forecasting performance of empirical SETAR models. *Journal of Applied Econometrics* **14**: 123–141.
Clements M, Smith J. 2001. Evaluating forecasts from SETAR models of exchange rates. *Journal of International Money and Finance* **20**: 133–148.
Corradi V, Swanson N. 2004. Some recent developments in predictive accuracy testing with nested models and (generic) nonlinear alternatives. *International Journal of Forecasting* **20**(2): 185–199.
Corradi V, Distaso W, Swanson N. 2009. Predictive density estimators for daily volatility based on the use of realized measures. *Journal of Econometrics* **150**(2): 119–138.
Cumby R, Figlewski S, Hasbrouck J. 1993. Forecasting volatility and correlations with EGARCH models. *Journal of Derivatives* Winter: 51–63.
Day T, Lewis C. 1992. Stock market volatility and the information content of stock index options. *Journal of Econometrics* **52**: 267–287.
Deng N, Tian Y. 2004. *New Methods in Data Mining: Support Vector Machine*. Science Press: Beijing.
Diebold F, Mariano R. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–265.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Dimson E, Marsh P. 1990. Volatility forecasting without data-snooping. *Journal of Banking and Finance* **44**: 399–421.
Donaldson R, Kamstra M. 1997. An artificial neural network-GARCH model for international stock return volatility. *Journal of Empirical Finance* **4**: 17–46.
Dotsis G, Psychoyios D, Skiadopoulos G. 2007. An empirical comparison of continuous-time models of implied volatility indices. *Journal of Banking and Finance* **31**: 3584–3603.
Dunis C, Huang X. 2002. Forecasting and trading currency volatility: an application of recurrent neural regression and model combination. *Journal of Forecasting* **21**: 317–354.
Enders W. 2004. *Applied Econometric Time Series* (2nd edn). Wiley: New York.
Engle R. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica* **50**: 957–1008.
Engle R, Hong C-H, Kane A, Noh J. 1993. *Advances in Futures and Options Research*, Vol. 6. JAI Press: Greenwich, CT; 393–415.
Feng Y, McNeil A. 2008. Modelling of scale change, periodicity and conditional heteroskedasticity in return volatility. *Economic Modelling* **25**(5): 850–867.
Ferland R, Lalancette S. 2006. Dynamics of realized volatilities and correlations: an empirical study. *Journal of Banking and Finance* **30**(7): 2109–2130.
Fernandez-Rodriguez F, Gonzalez-Martel C, Sosvilla-Rivero S. 2000. On the profitability of technical trading rules based on artificial neural networks: evidence from the Madrid stock market. *Economics Letters* **69**(1): 89–94.
Figlewski S. 1997. Forecasting volatility. *Financial Markets, Institutions and Instruments* **6**: 1–88.
Fleming J. 1998. The quality of market volatility forecasts implied by S&P 100 index option prices. *Journal of Empirical Finance* **5**: 317–345.
Franke J, Neumann M, Stockis J. 2004. Bootstrapping nonparametric estimators of the volatility function. *Journal of Econometrics* **118**: 189–218.
Franses P, Dijk DV. 1996. Forecasting stock market volatility using (non-linear) GARCH models. *Journal of Forecasting* **15**(3): 229–235.
Franses P, Dijk DV. 2000. *Nonlinear Time Series Models in Empirical Finance*. Cambridge University Press: Cambridge, UK.
Franses P, McAleer M. 2002. Financial volatility: an introduction. *Journal of Applied Econometrics* **17**: 419–424.
Galbraith J, Kisinbay T. 2005. Content horizons for conditional variance forecasts. *International Journal of Forecasting* **21**: 249–260.
Gerlach R, Tuyl F. 2006. MCMC methods for comparing stochastic volatility and GARCH models. *International Journal of Forecasting* **22**(1): 91–107.
Ghysels E, Harvey A, Rebault E. 1996. *Handbook of Statistics: Statistical Methods in Finance*, Vol. 14. Elsevier Science: Amsterdam; 119–191.
Ghysels E, Santa-Clara P, Valkanov R. 2006. Predicting volatility: how to get most out of returns data sampled at different frequencies. *Journal of Econometrics* **131**: 59–95.
Glosten L, Jagannathan R, Runkle D. 1992. On the relation between the expected value and the volatility of nominal excess return on stocks. *Journal of Finance* **46**: 1779–1801.
Gokcan S. 2000. Forecasting volatility of emerging stock markets: linear versus non-linear GARCH models. *Journal of Forecasting* **19**(6): 499–504.
Gospodinov N, Gavala A, Jiang D. 2006. Forecasting volatility. *Journal of Forecasting* **25**(6): 381–340.
Gray S. 1996. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* **42**: 27–62.
Groen J, Kapetanios G, Price S. 2009. A real time evaluation of bank of England forecasts of inflation and growth. *International Journal of Forecasting* **25**(1): 74–80.
Gunn S. 1998. Support vector machines for classification and regression. *Isis-1-98*. Technical report, Image Speech and Intelligent Systems Group, University of Southampton, UK.
Hamid S, Iqbal Z. 2004. Using neural networks for forecasting volatility of S&P 500 index futures prices. *Journal of Business Research* **57**: 1116–1125.
Hamilton J. 1989. *A* new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**: 357–384.
Hamilton J. 1997. *Time Series Analysis*. Princeton University Press: Princeton, NJ.
Härdle, W, Moro R, Schäfer D. 2005. *Statistical Tools for Finance and Insurance*. Springer: Berlin.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Härdle W, Moro R, Schäfer D. 2007. *Handbook for Data Visualization*. Springer: Berlin.
Haykin S. 1999. *Neural Networks: A Comprehensive Foundations* (2nd edn). Prentice Hall: Englewood Chiffs, NJ.
Heynen R, Kat H. 1994. Volatility prediction: a comparison of stochastic volatility, GARCH(1, 1) and EGARCH(1, 1) models. *Journal of Derivatives* 50–65.
Hu M, Tsoukalas C, 1999. Combining conditional volatility forecasts using neural networks: an application to the EMS exchange rates. *Journal of International Financial Markets, Institution and Money* **9**: 407–422.
Jorion P. 1995. Predicting volatility in the foreign exchange market. *Journal of Finance* **50**: 507–528.
Jorion P. 1996. *The Microstructure of Foreign Exchange Markets*. Chicago University Press: Chicago, IL.
Klaassen F. 2002. Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics* **27**: 363–394.
Koopman S, Jungbacker B, Hol E. 2005. Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* **12**: 445–475.
Kuan C, Liu T. 1995. Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics* **10**: 347–364.
Lamoureux C, Lastrapes W. 1993. Forecasting stock-return variance: understanding of stochastic implied volatilities. *Review of Financial Studies* **6**: 293–326.
Lehar A, Scheicher M, Schittenkopf C. 2002. GARCH vs. stochastic volatility: option pricing and risk management. *Journal of Banking and Finance* **26**: 323–345.
Li W, Ling S, McAleer M. 2002. Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys* **16**: 245–269.
Lux T, Schornstein S. 2005. Genetic learning as an explanation of stylized facts of foreign exchange markets. *Journal of Mathematical Economics* **41**: 169–196.
Marcucci J. 2005. *Studies in Nonlinear Dynamics and Econometrics*, Vol. 9. Berkeley Electronic Press: Berkeley, CA; 1145.
McMillan D, Speight A. 2004. Daily volatility forecasts: reassessing the performance of GARCH models. *Journal of Forecasting* **23**(6): 449–460.
McMillan D, Speight A, Gwilym O. 2000. Forecasting UK stock market volatility: a comparative analysis of alternate methods. *Applied Financial Economics* **10**: 435–448.
Meddahi N. 2003. ARMA representations of integrated and realized variances. *Econometrics Journal* **6**: 334–355.
Moosa I. 2000. *Exchange Rate Forecasting: Techniques and Applications*. Macmillan Press: London.
Neely C. 2009. Forecasting foreign exchange volatility: why is implied volatility biased and inefficient? And does it matter? *Journal of International Financial Markets, Institutions and Money* **19**(1): 188–205.
Nelson D. 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**: 347–370.
Newey W, West K. 1987. A simple positive, semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**(3): 703–708.
Niemira M, Klein P. 1994. *Forecasting Financial and Economic Cycles*. Wiley: New York.
Pagan A, Schwert G. 1990. Alternative models for conditional stock volatility. *Journal of Econometrics* **45**: 267–290.
Pantelidaki S, Bunn D. 2005. Development of a multifunctional sales response model with the diagnostic aid of artificial neural networks. *Journal of Forecasting* **24**: 505–521.
Park B. 2002. An outlier robust GARCH model and forecasting volatility of exchange rate returns. *Journal of Forecasting* **21**(5): 381–393.
Pérez-Cruz F, Afonso-Rodriguez J, Giner J. 2003. Estimating GARCH models using SVM. *Quantitative Finance* **3**: 163–172.
Pong S, Shackleton M, Taylor S, Xu X. 2004. Forecasting currency volatility: a comparison of implied volatilities and AR(FI) MA models. *Journal of Banking and Finance* **28**: 2541–2563.
Poon S-H, Granger C. 2003. Forecasting volatility in financial markets: a review. *Journal of Economic Literature* **41**: 478–539.
Preminger A, Franck R. 2007. Forecasting exchange rates: a robust regression approach. *International Journal of Forecasting* **23**(1): 71–84.
Qi M, Wu Y. 2003. Nonlinear prediction of exchange rates with monetary fundamentals. *Journal of Empirical Finance* **10**: 623–640.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Renò R. 2006. Nonparametric estimation of stochastic volatility models. *Economics Letters* **90**(3): 390–395.
Rosenow B. 2008. Determining the optimal dimensionality of multivariate volatility models with tools from random matrix theory. *Journal of Economic Dynamics and Control* **32**(1): 279–302.
Schittenkopf C, Dorffner G, Dockner E. 2000. Forecasting time-dependent conditional densities: a semi-non-parametric neural network approach. *Journal of Forecasting* **19**: 355–374.
Scholkopf B, Smola A. 2001. *Learning with Kernels*. MIT Press: Cambridge, MA.
Sentana E. 1995. Quadratic ARCH models. *Review of Economic Studies* **62**: 639–661.
Suykens J, Vandewalle J. 2000. Recurrent least squares support vector machines. *IEEE Transactions on Circuits and Systems I* **47**(7): 1109–1114.
Tay F, Cao L. 2001. Application of support vector machines in financial time series forecasting. *Omega* **29**: 309–317.
Taylor J. 1999. Evaluating volatility and interval forecasts. *Journal of Forecasting* **18**: 111–128.
Taylor J. 2000. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *Journal of Forecasting* **19**: 299–311.
Taylor N. 2008. Can idiosyncratic volatility help forecast stock market volatility? *International Journal of Forecasting* **24**(3): 462–479.
Taylor S. 1986. *Modelling Financial Time Series*. Wiley: Chichester.
Tse Y, Tung S. 1992. Forecasting volatility in the singapore stock market. *Asia Pacific Journal of Management* **9**: 1–13.
Tseng C, Cheng S, Wang Y, Peng J. 2008. Artificial neural network model of the hybrid EGARCH volatility of the taiwan stock index option prices. *Physica A: Statistical Mechanics and its Applications* **387**(13): 3192–3200.
Vapnik V. 1995. *The Nature of Statistical Learning Theory*. Springer: New York.
Vapnik V. 1997. *Statistical Learning Theory*. Wiley: New York.
West K. 1996. Asymptotic inference about predictive ability. *Econometrica* **64**: 1067–1084.
West K, Cho D. 1995. The predictive ability of several models of exchange rate volatility. *Journal of Econometrics* **69**: 367–391.
Wong W, Tu A. 2009. Market imperfections and the information content of implied and realized volatility. *Pacific Basin Finance Journal* **17**(1): 58–79.
Zhang X, King M. 2005. Influence diagnostics in generalized autoregressive conditional heteroscedasticity processes. *Journal of Business and Economic Statistics* **23**: 118–129.

Authors' biographies:

Shiyi Chen started teaching at the School of Economics of Fudan University in China as an Assistant Professor after receiving his PhD degree in econometrics from the School of Economics and Trade at Kyungpook National University in the Republic of Korea in February 2006. From November 2008, Shiyi Chen became an Associate Professor of Econometrics. His research interests are time series forecasting, nonparametric econometrics, and energy and emission economics. One of his articles, Modeling Default Risk with Support Vector Machines, co-authored with Wolfgang K. Härdle and Rouslan A. Moro, was accepted by the *Journal of Quantitative Finance* in January 2009.

Wolfgang K. Härdle gained his Dr rer. nat. in mathematics at Universität Heidelberg in 1982 and his Habilitation at Universität Bonn in 1988. He is currently Chair Professor of Statistics at the Department of Economics and Business Administration, Humboldt-Universität zu Berlin. He is also director of CASE (Center for Applied Statistics and Economics) and of the Collaborative Research Center 'Economic Risk'. His research focuses on dimension reduction techniques, computational statistics and quantitative finance. He has published 34 books and more than 200 papers in leading statistical, econometrics and finance journals. He is one of the 'Highly Cited Scientists' according to the Institute of Scientific Information.

Kiho Jeong received a PhD degree in econometrics from the University of Wisconsin at Madison in 1991. After working for two years at the Korea Energy Economic Institute as an economist, he joined the School of Economics and Trade at Kyungpook National University in 1994 as an assistant professor, where he is now a full professor. His research interests are forecasting energy/financial markets, modelling climate change effects and nonparametric kernel methods.

Copyright © 2009 John Wiley & Sons, Ltd.

DOI: 10.1002/for

Authors' addresses:
Shiyi Chen, School of Economics, Fudan University, 600 Guoquan Road, Shanghai 200433, China.

Wolfgang K. Härdle, Center for Applied Statistics and Economics, Humboldt University, Berlin, Germany.

Kiho Jeong, School of Economics and Trade, Kyungpook National University, Daegu, Republic of Korea.

Copyright © 2009 John Wiley & Sons, Ltd.

J. Forecast. **29**, 406–433 (2010)
DOI: 10.1002/for