

# VOLATILITY FORECAST BASED ON INTELLIGENT EGARCH ERROR CORRECTION MODEL

*Liqiang Hou, Shanlin Yang*

Original scientific paper

As the stock market volatility is highly nonlinear, coupling and time varying, it is difficult to predict by the traditional forecasting methods. For explaining the existing problems of the current volatility forecasting method, we use the model based on the weighted least squares support vector regression (WLS-SVR) method to predict the stock index volatility in this paper. After the prediction, there is the error sequence that is a random time series. Therefore, this paper proposes the use of EGARCH model to construct an error forecast model based on the returns of stock predicted error time series. Then, we use these results to correct the volatility of stock. Finally, we use the volatility of Shanghai Composite Index as the application object. The experimental results show that the prediction accuracy of this method has improved significantly with regard to other forecasting methods.

**Keywords:** error correction, forecasting, intelligent EGARCH model, SPA test, volatility

## Predviđanje promjenjivosti na temelju inteligentnog EGARCH modela korekcije pogreške

Izvorni znanstveni članak

Kako je promjenjivost dionica na tržištu vrlo nelinearno povezana i vremenski kolebljiva, teško ju je predvidjeti pomoću tradicionalnih metoda predviđanja. Objašnjavajući postojeće probleme trenutnih metoda predviđanja promjenjivosti, koristimo model zasnovan na metodi najmanjih kvadrata s težinama potpore vektorskoj regresiji (WLS-SVR) za predviđanje promjenjivosti indeksa dionica u ovom radu. Nakon predviđanja, postoji niz pogrešaka koji je slučajna vremenska serija. Stoga, ovaj članak predlaže uporabu EGARCH modela za stvaranje modela predviđanja grešaka zasnovanog na dobivenoj predviđenoj vremenskim nizom pogrešaka. Nakon toga, koristimo dobivene rezultate za ispravljanje nestabilnosti dionica. Konačno, koristimo nestabilnost Šangajskog složenog indeksa kao objekta primjene. Eksperimentalni rezultati pokazuju da je točnost predviđanja ove metode znatno poboljšana u odnosu na druge metode predviđanja.

**Ključne riječi:** inteligentni EGARCH model, korekcija pogrešaka, predviđanje, promjenjivost, SPA test

## 1 Introduction

The stock market of China has been undertaking a hugely important revolution since the beginning of right dividend of China Stock. If handled successfully, the behaviour of China Stock market will definitely become the economic barometer. On the other hand, as the nation economics is speeding up and taking incredible change, the dependence of nation economy on the financial market will be intensified which requires a healthy and high-efficient capital market as the support of China economy. As a result, we have all the right to do research on the prediction of stock returns both from theoretical and practical aspects.

Stock market is a complex dynamic system composed of lots of factors [1]. As the inner and outer factors are hard to work out, research on the model of stock returns attracts a lot of people involved. However the results are not so satisfactory, they even rendered lots of methods unsatisfactory in reality. For instance, the regression is suitable for linear time series prediction, but it is not so accurate in non-linear ones [2]. Neutral network has the advantage of drawing near to non-linear, but it turns out the disappointment to predict stock price [3]. Because rate of yield is linear in China Stock Market, we can transform the regression questions into convex quadratic programming based on the SVR algorithm that uses structure risk minimum principle. By using kernel function, we can efficiently let the nonlinear data map into high dimensional space and be disposed by linear learning machine, as the parameters of SVR model are hard to define while the effect of this model is determined to a great extent by those parameters. In practice, they are selected mostly by experience. Document [4] predicts the

important indicators of opening price by vector regression algorithm, but there is still big discrepancy between the predicted ones and the real ones.

The stock market is so complex, because the time series are hard to detect. Meanwhile, there are errors produced by the process of prediction. Then, the predicted error can form a time series which includes information missing from prediction model [5]. Actually, the prediction error belongs to a composite stochastic process due to two facts. One fact is that the model has the ability to predict the rate of yield in some certain moment (higher, lower or common). The other one is the reflection of prediction error in the model. The error is the emerging one. If we can dig out the rule of change in the model state series according to the historical data, we can predict the forecast error and adjust original prediction, so as to improve the accuracy [6]. As the forecast error usually has auto-correlation, we can figure out the inner information of error series and predict the follow-up in the series, and then use the result to modify the prediction of yield of the stock rate. Researches indicate that EGARCH is of great use in explaining daily change in stock price and can reflect the time-change conditional variance character of forecast error. It beats ARMA in forecast [7].

This paper focuses on the methods of combining intelligence algorithm and time series in order to predict the returns of stock based on its fluctuation and ordinary prediction methods. We use the WLS supported vector regression (WLS-SVR) model to implement initial forecast, and try to add the EGARCH model to the forecast error afterward process. By using EGARCH model we obtain the error and its regularity of distribution. Finally, we modify the rate of stock yield

which we got in the first prediction and thus can get the final one. By means of practical study, we draw the conclusion by the SPA test that the method rendered in this paper is better as compared to ANN, GM(1,1), EGRACH and LS-SVR.

## 2 LS-SVR model

LS-SVR expands standard SVR by optimizing the square of relaxation factors and converting the constraints of inequality to equality, so the quadratic programming problem in traditional SVR becomes linear simultaneous equations, thus the calculating difficulty reduces a lot in company with the solution high efficiency and convergence speeding up.

The basic method of SVR:

Define  $x \in R^n$  and  $y \in R$ , Let  $R^2$  be the input space, by nonlinear transformation  $\phi(\cdot)$ , we let in the input space  $x$  map into a high dimensional characteristic space where we use the linear function to fit sample data while making sure the generalization.

In the characteristic space, the linear estimation function is defined as:

$$y = f(x, \omega) = \omega^T \phi(x) + b, \tag{1}$$

where  $w$  is the weight and  $b$  is the skewness.

The aim function is:

$$\min_{\omega, b, \xi} J(\omega, b, \xi) = \frac{1}{2} \omega^T \omega + \frac{1}{2} C \sum_{i=1}^N \xi_i^2, \tag{2}$$

s. t.

$$y_i = \phi(x_i) \omega + b + \xi_i, \tag{3}$$

$i = 1, \dots, N$

where  $\omega \in R^h$  is the weight vector and  $\phi(\cdot)$  is non-linear mapping function,  $\xi_i \in R^{N \times 1}$ , is relaxation factor,  $b \in R$  is the skewness while  $C > 0$  is penalty factor.

Including factors  $\alpha_i \in R^{N \times 1}$ , we can easily get the function as:

$$L(\omega, b, \xi_i, \alpha_i) = \frac{1}{2} \|\omega\|^2 + \frac{1}{2} C \sum_{i=1}^N \xi_i^2 - \sum_{i=1}^N \alpha_i [\phi(x_i) \omega + b + \xi_i - y_i] \tag{4}$$

According to the KTT we get

$$\begin{cases} \frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^N \alpha_i \phi(x_i) = 0 \\ \frac{\partial L}{\partial b} = \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial L}{\partial \xi_i} = \alpha_i - C \xi_i = 0 \\ \frac{\partial L}{\partial \alpha_i} = \phi(x_i) \omega + b + \xi_i - y_i = 0 \end{cases} \tag{5}$$

$$\begin{bmatrix} 0 & E^T \\ E & \phi \phi^T + C^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}, \tag{6}$$

where  $E$  is the matrix whose elements are all 1,  $I$  is an  $N \times N$  identity matrix.

Inner product of regression in non-linear function can be replaced by kernel function of Mercer. Let  $\Omega_{ij} = \phi \phi^T$ , then

$$\Omega_{ij} = \phi(x_i)^T \phi(x_j) = K(x_i, x_j). \tag{7}$$

We then have the LS-SVR regression function model

$$f(x) = \sum_{i=1}^N \alpha_i K(x_i, x_j) + b. \tag{8}$$

Kernel functions commonly used in practice are linear kernel, polynomial kernel, and RBF kernel. We use the RBF kernel as our kernel function for its better generalization. The form is as follows:

$$K(x_i, x_j) = \phi(x_i) \phi(x_j) = \exp(-\|x_i - x_j\| / 2\delta^2) \tag{9}$$

in which the regularization parameter  $C$  and the kernel breadth  $\delta$  are the crucial parameters of LS-SVR.

## 3 Modified Model Building

This section presents the fundamental knowledge of LS-SVR. Suppose a set of data

$$T = \{(x_i, y_i) | x_i \in X \subset R^n, y_i \in Y \subset R, i = 1, \dots, N\} \tag{10}$$

where  $x_i \in R^n$  are given as inputs,  $y_i \in R^n$  are the corresponding outputs. The SVR theory is applied to find a nonlinear map from input space to output space, and map the data to a higher dimensional feature space. Then the following estimate function is used:

$$f(x) = \omega^T \cdot \phi(x) + b, \tag{11}$$

where the  $\phi(x)$  maps the input data to a higher dimensional feature space,  $w$  is a weight vector, and  $b$  is the threshold value.  $f(x)$  is the regression estimate function which is constructed through learning of the sample set. In the function estimation of LS-SVR, the objective function of optimization problem is defined as

$$\min_{\omega, b, \xi} J(\omega, \xi_i) = \frac{1}{2} \|\omega\|^2 + \frac{1}{2} C \sum_{i=1}^N \xi_i^2. \tag{12}$$

Subject to the constraints

$$f(x) = \omega^T \cdot \phi(x_i) + b + \xi_i, \quad i = 1, 2, \dots, N \tag{13}$$

Where  $\|\omega\|^2$  is the weights vector norm, which is used to constrain the model structure capacity in order to obtain better generalization performance. The  $C$  is the user-

defined regularization constant which balances the model's complexity and approximation accuracy, and  $\xi_i$  is the approximation error.

Estimation of support values in the LS-SVR is optimal only when there is a Gaussian distribution of error variables. However, a Gaussian assumption for error variables is not realistic. This is because the SSE cost function of the LS-SVR assigns an equal weight to error at all times, treating all data equally. To obtain a robust estimate when the distribution is not a Gaussian one, the correction must be made by defining weights based on the error distribution by the so-called weighted LS-SVR method.

We can modify these weights and obtain a robust estimate based on the previous LS-SVR solution. In the main space  $R^2$ , a new objective function of the optimization problem is shown as follows:

$$\min_{\omega, b, \xi, v} J(\omega, b, \xi, v) = \frac{1}{2} \|\omega\|^2 + \frac{1}{2} C \sum_{i=1}^N v_i \xi_i^2 \quad (14)$$

$$\text{s. t. } y_i = \phi(x_i)\omega + b + \xi_i, \quad i = 1, \dots, N \quad (15)$$

where  $J$  is loss function,  $\xi_i \in R^{N \times 1}$  is relaxation factor,  $\phi(\cdot)$  is the nonlinear mapping function,  $\phi(\cdot)$  map the input vectors  $x \in X$  of system to the input vectors  $\phi(x) \in H$  which has the high dimensional feature space.

With the Lagrange multipliers  $\alpha_i \in R^{N \times 1}$  are introduced, the Lagrange function is given by

$$L(\omega, b, \xi_i, \alpha_i, v_i) = \frac{1}{2} \|\omega\|^2 + \frac{1}{2} C \sum_{i=1}^N v_i \xi_i^2 - \sum_{i=1}^N \alpha_i [\phi(x_i)\omega + b + \xi_i - y_i] \quad (16)$$

The Karush-Kuhn-Tucker (KKT) conditions for optimality are given by

$$\begin{cases} \frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^N \alpha_i \phi(x_i) = 0 \\ \frac{\partial L}{\partial b} = \sum_{i=1}^N \alpha_i = 0 \\ \frac{\partial L}{\partial \xi_i} = \alpha_i - \frac{C v_i \xi_i}{N} = 0 \\ \frac{\partial L}{\partial \alpha_i} = \omega \cdot \phi(x_i) + b + \xi_i - y_i = 0 \end{cases}$$

After elimination of  $\xi_i$  and  $\omega$ , it is reduced to a matrix form

$$\begin{bmatrix} 0 & \mathbf{E}^T \\ \mathbf{E} & \mathbf{\Omega} + V_C \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}, \quad (17)$$

There  $V_C = \text{diag}\{1/Cv_1, \dots, 1/Cv_N\}$ ,  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]^T$ ,  $y = [y_1, y_2, \dots, y_N]^T$ ,  $\mathbf{E}$  is the weight value variable, and  $\mathbf{E}$  is

an  $N \times 1$  identity matrix,  $\mathbf{\Omega}$  is  $N \times N$  Hessian matrix, and  $\mathbf{\Omega}$  follows Mercer's condition.

$$\mathbf{\Omega}_{ij} = \phi(x_i)^T \phi(x_j) = K(x_i, x_j), \quad i, j = 1, 2, \dots, N. \quad (18)$$

Eqs. (10) and (15) provide the final result of the LS-SVR model for function estimation, and then the following estimate function is the regression function model of WLS-SVR

$$f(x) = \sum_{i=1}^N \alpha_i K(x_i, x_j) + b. \quad (19)$$

As the choices of kernel function, there are several possibilities. The Kernel function is commonly used by the linear kernel, polynomial kernel, and RBF kernel. In this paper, we select the RBF kernel which has better generalization. Its expression form is the following:

$$K(x_i, x_j) = \phi(x_i)\phi(x_j) = \exp(-\|x_i - x_j\| / 2\delta^2), \quad (20)$$

where the regularization parameter  $C$  and kernel breadth  $\delta$  are the crucial parameters of the WLS-SVR.

The WLS-SVR reflects the behaviour of the random errors in the model, through introducing weight factor  $v_i (i = 1, \dots, N)$  to correct the approximation error vector  $\xi$  of the LS-SVR, and then its algorithm is more robust. One common choice for  $v_i$  has been given by Suykens et al.

$$v_i = \begin{cases} 1 & \text{if } |\xi_i/\hat{s}| \leq c_1 \\ \frac{c_2 - |\xi_i/\hat{s}|}{c_2 - c_1} & \text{if } c_1 \leq |\xi_i/\hat{s}| \leq c_2, \\ 10^{-4} & \text{otherwise} \end{cases}, \quad (21)$$

$$\hat{s} = 1,483MAD(\xi), \quad (22)$$

where  $\hat{s}$  is a robust of the standard deviation of the LS-SVR error variable  $\xi_i = \alpha_i/C$ , which denotes how much the estimated error distribution deviates from a Gaussian distribution. MAD is the middle of error absolute value. The constants  $C_1$  and  $C_2$  typically are chosen to be  $C_1=2,5$  and  $C_2=3$

#### 4 EGARCH error correction model

As the return of stock is a time-series, we set the time space as  $t \in D^- + D^+$ . The  $D^-$  represents the past time space, so the returns are the foundation in constructing the model as they are known to all.

$D^+$  represents the prediction time space, where we have no idea of the rates. Assumed that the historical data can reflect the time change feature of returns, we are able to construct a prediction model that reports the history data variation tendency.

Thus, we get predicted value series  $\{\hat{y}_t : t = 1, \dots, N\}$  and corresponding error value series which is

$\{\varepsilon_t : t=1, \dots, N\}$ . What we care about is the key factors while ignoring the secondary factors, which leads to prediction error. In order to fix it, we bring in the modification process. In fact, we can let the product error be a series  $\{\varepsilon_t : t=1, \dots, N\}$  and consider  $\varepsilon_t$  being a new stochastic process, so we can build the prediction model for  $\varepsilon_t$  and get its residual error series  $\{\hat{\varepsilon}_t : t=2, \dots, N+1\}$ . The adjusted initial predicted value is

$$y'_t = \hat{y}_t + \hat{\varepsilon}_t. \tag{23}$$

The model of EGARCH is used to describe the stable stochastic time series. By logarithmic transformation, conditional variance may turn out to be negative. Besides, during the calculation there are no parameter constraints, so the complexity of the calculation is trimmed while the efficiency goes up. Meanwhile, it can easily be used in auto-correlation elimination of fluctuation ration in asymmetric information circumstances.

This paper applies the EGARCH model to construct an error forecast model based on the returns of stock predicted error time series.

The chosen EGARCH model is:

$$\begin{cases} y_t = \theta_0 + \theta_1 x_{1t} + \dots + \theta_k x_{kt} + r_t \\ r_t = \sigma_t \varepsilon_t \\ \ln(\sigma_t^2) = \alpha_0 + \alpha \left| \frac{r_{t-1}}{\sigma_{t-1}} \right| + \beta \ln(\sigma_{t-1}^2) + \gamma \frac{r_{t-1}}{\sigma_{t-1}} \\ \varepsilon_t \sim I.I.D \end{cases} \tag{24}$$

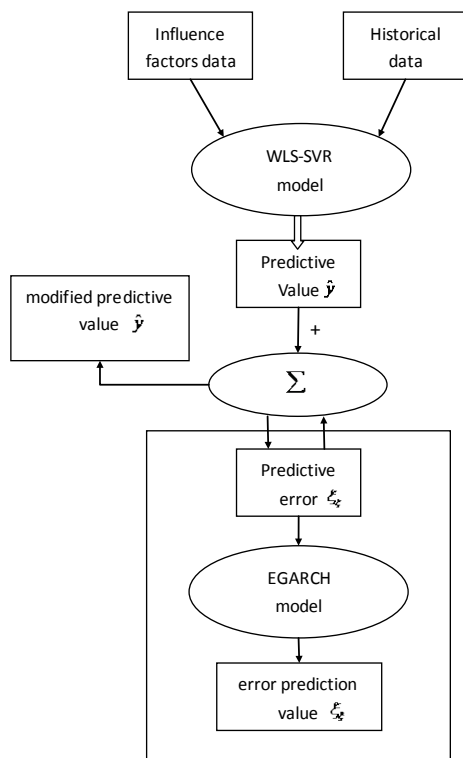


Figure 1 The flow path of constructed model including error correction

The EGARCH model is used to annul the stability and not pure stochastic series. Thus, after we get the error series of returns of stock, we will undertake stability test

of root of utility. Then, we use difference method to make it stable. Next, we can get the EGARCH order  $(p, d, q)$  after knowing Lagrange multiplier (LM) and Q statistic for correlation test which takes step after ACF and PACF. We estimated the parameters by some methods to choose the best EGARCH model to fit error series.

The specific model constructing and prediction flow path are shown in Fig. 1.

### 5 Testing the adequacy of the EGARCH model

Consider a time series model with the following structure:

$$y_t = f(w_t; \varphi) + \varepsilon_t, \tag{25}$$

where  $f$  is at least twice continuously differentiable with respect to  $\varphi \in \phi$ , for all  $w_t = (y'_{t-1}, u'_t)'$  with  $y_{t-1} = (1-y_{t-1}, \dots, y_{t-n})' \in \mathbb{R}^{n+1}$  and exogenous  $u_t = (u_{1t}, \dots, u_{kt})' \in \mathbb{R}^k$ , everywhere in  $\phi$ . The error is parameterized as

$$\varepsilon_t = \xi_t h_t^{1/2}, \tag{26}$$

where  $\{\xi_t\}$  is a sequence of independent identically distributed random variables with mean zero, unit variance and  $E\xi^3 = 0$ . The conditional variance  $h_t = \eta' s_t$  such that and  $\eta = (\alpha_0, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)'$  with  $\alpha_0 > 0$ , whereas  $\alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p$  satisfy the conditions in Nelson and Cao [14] that ensure the positivity of  $h_t$ . These conditions allow some of the parameters to be negative, unless  $p=q=1$ .

Eq. (26) is thus the standard GARCH  $(p, q)$  model. We assume regularity conditions hold so that the central limit theorem and the law of large numbers apply whenever required. For such conditions in the multivariate GARCH  $(p, q)$  case see Comte and Lieberman [8].

In the univariate case, their conditions require the density of  $\xi_t$  to be absolutely continuous with respect to the Lévesque measure and positive in a neighborhood of the origin. Furthermore, it is required that  $E\xi^3 < \infty$ , which of course implies further restrictions on the density of  $\xi_t$ .

The assumption  $E\xi^3 = 0$  that Comte and Lieberman do not need guarantees block diagonality of the information matrix of the log-likelihood function. However, it is not just a technical simplification. We shall consider, among other things, a test that has power against asymmetric response to shocks. In deriving such a test it is appropriate to assume that the conditional distribution of  $\varepsilon_t$  given  $h_t$  is not skewed.

In order to consider the adequacy of the GARCH model, we formulate a parametric alternative to the model. Assume that in Eq. (26)

$$\xi_t = z_t g_t^{1/2}, \tag{27}$$

where  $\{z_t\}$  is a sequence of independent, identically distributed random variables with zero mean, unit variance and  $Ez_t^3 = 0$ . Furthermore,  $g_t = 1 + \pi' v_t$ , where

$v_t = (\xi_{t-1}^2, \dots, \xi_{t-m}^2)'$  and  $\pi = (\pi_1, \dots, \pi_m)'$ ,  $\pi_j \geq 0, j = 1, \dots, m$   
 Eq. (25) may thus be written as

$$\varepsilon_t = z_t(h_t g_t)^{1/2}. \tag{28}$$

It could be called an "ARCH nested in GARCH" model as  $\xi_{t-j}^2 = \varepsilon_{t-j}^2 / h_{t-j}, j = 1, \dots, m$ . We want to test  $H_0 : \pi = 0$  against  $\pi \neq 0$  and thus follow the standard practice of choosing a two-sided alternative although the elements of  $\pi$  are constrained to be non-negative. Under this hypothesis,  $g_t \equiv 1$ , and the model collapses into a GARCH(p,q) model. (For ways of testing  $H_0 : \pi = 0$  against  $\pi > 0$  when  $h_t \equiv \alpha_0$ , see Lee and King [9] and Demos and Sentana [10]).

We introduce the following notation. Let  $\hat{\varepsilon}_t$  and  $\hat{h}_t$  be the error which is  $\varepsilon_t$  and the conditional variance which is  $h_t$ , respectively, estimated under  $H_0$ , and  $\hat{\xi}_t^2 = \hat{\varepsilon}_t^2 / \hat{h}_t$ .

Then we set  $\hat{x}_t = \hat{h}_t^{-1} \partial \hat{h}_t / \partial \eta'$  and  $\hat{v}_t = (\hat{\xi}_{t-1}^2, \dots, \hat{\xi}_{t-m}^2)'$ .

The quasi maximum likelihood approach leads to the following result:

Theorem: Consider the Esq. (28) where  $g_t = 1 + \pi' v_t$  and  $\{z_t\}$  is a sequence of independent identically distributed random variables with zero mean; unit variance and  $Ez_t^3 = 0$ . Under  $H_0 : \pi = 0$ , the statistic

$$LM_\pi = (1/4T) \left\{ \sum_{t=1}^T (\hat{\varepsilon}_t^2 / \hat{h}_t - 1) \hat{v}_t' \right\} V(\hat{\eta})^{-1} \left\{ \sum_{t=1}^T (\hat{\varepsilon}_t^2 / \hat{h}_t - 1) \hat{v}_t \right\},$$

where

$$cV(\hat{\eta})^{-1} = (4T/\hat{k}) \left\{ \sum_{t=1}^T \hat{v}_t \hat{v}_t' - \sum_{t=1}^T \hat{v}_t \hat{v}_t' \left( \sum_{t=1}^T \hat{x}_t \hat{x}_t' \right)^{-1} \sum_{t=1}^T \hat{x}_t \hat{v}_t \right\}^{-1}.$$

With  $\hat{k} = (1/T) \sum_{t=1}^T (\hat{\varepsilon}_t^2 / \hat{h}_t - 1)^2$  is a consistent estimator of the inverse of the covariance matrix of the partial score under the null hypothesis; has an asymptotic  $\chi^2$  distribution with m degrees of freedom.

The test may also be carried out using an artificial regression as follows:

1. Estimate the parameters of the conditional variance model under the null, compute  $\hat{\varepsilon}_t^2 / \hat{h}_t - 1, t = 1, 2, \dots, T$ , and  $SSR_0 = \sum_{t=1}^T (\hat{\varepsilon}_t^2 / \hat{h}_t - 1)^2$ .
2. Regress  $\hat{\varepsilon}_t^2 / \hat{h}_t - 1$  on  $\hat{x}_t', \hat{v}_t'$  and compute the sum of squared residuals,  $SSR_1$ .
3. Compute the test statistic  $LM = T(SSR_0 - SSR_1)/SSR_0$  or the F-version:

$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T - p - q - 1 - m)}.$$

For the sample sizes relevant in GARCH modelling, there is no essential difference between the properties of LM and F.

## 6 The SPA test

It is generally accepted that squared daily returns provide a poor approximation of actual daily volatility. Andersen and Bollerslev [11] pointed out that more accurate estimates can be obtained by summing all squared intraday returns. If we were to apply their method directly in this paper, we would define the RV measurement as

$$RV'_t = \sum R_{t,d}^2. \tag{29}$$

However, this definition ignores the information contained in overnight returns. To address this problem, Hansen and Lunde [12] suggested scaling the RV measurement in the following way.

$$RV_t = \gamma RV'_t, \tag{30}$$

where the so-called scale parameter  $\gamma$  is defined as

$$\gamma = \frac{\frac{1}{N} \sum_{t=1}^N R_t^2}{\frac{1}{N} \sum_{t=1}^N RV'_t}. \tag{31}$$

Andersen et al. [13] found the distribution of RV to be highly non-normal and skewed, but its logarithms to be approximately normal. Accordingly, they suggested that the natural logarithms of a RV measurement series, were denoted as  $\ln RV$  that could be modelled by a Gaussian dynamic process.

Various forecasting criteria or loss functions can be considered in assessing the predictive accuracy of the volatility model, although, as Lopez [14] noted, it is not obvious which loss function is most appropriate for the evaluation of such models. Rather than making a single choice, we thus employ the following four accuracy statistics or loss functions as our forecasting criteria.

$$MSE = \frac{1}{N} \sum_{t=1}^n (RV_t - \hat{\sigma}_t^2)^2, \tag{32}$$

$$MAE = \frac{1}{N} \sum_{t=1}^n |RV_t - \hat{\sigma}_t^2|, \tag{33}$$

$$HMSE = \frac{1}{N} \sum_{t=1}^n (1 - \hat{\sigma}_t^2 / RV_t)^2, \tag{34}$$

$$HMAE = \frac{1}{N} \sum_{t=1}^n |1 - \hat{\sigma}_t^2 / RV_t|, \tag{35}$$

where  $n$  is the number of forecasting data points; MSE and MAE are the mean square error and mean absolute error; HMSE and HMAE are the MSE and MAE, respectively, adjusted for heteroskedasticity. Different

criteria serve different practical purposes. For example, in the case of Value-at-Risk applications, greater interest may lie in the accurate forecasting of a high rather than a low level of volatility, which implies that the *MSE* criterion is the most relevant loss function in risk management applications. Additional discussion of these criteria can be found in Ref. [15].

When a particular loss function is smaller for model A than it is for model B, it is impossible to conclude that the forecasting performance of the former is superior to that of the latter. Such a conclusion cannot be made on the basis of a single loss function and a single sample. Recent work has focused on a testing framework that can determine whether one particular model is outperformed by another. As discussed in the Introduction, the SPA test, an extension of the White framework proposed by Hansen and Lunde [16], has been shown to possess good power properties and to be more robust than previous approaches.

In contrast to other evaluation techniques, the SPA test can be used to compare the performance of two or more forecasting models at the same time. Forecasts are evaluated employing a pre-specified loss function, and the "best" forecasting model is the one that produces the smallest expected loss. In the SPA test, the loss function relative to the benchmark model is defined as  $x_{i,t}^{(0,i)} = L_{i,t}^{(0)} - L_{i,t}^{(i)}$ , where  $L_{i,t}^{(0)}$  is the value of loss function  $l$  at time  $t$  for benchmark model  $M_0$ , and  $L_{i,t}^{(i)}$  is the value of loss function  $l$  at time  $t$  for competing model  $M_i$ , for  $i = 1, \dots, k$ .

The SPA test is used to compare the forecasting performance of a benchmark model against its  $K$  competitors. The null hypothesis that the benchmark or base model is not outperformed by any of the competing models can be expressed as  $H_0 : \max_{i=1, \dots, k} E(X_{i,t}^{(0,i)}) \leq 0$ . It is tested with the statistic  $T_l^{SPA} = \max_{i=1, \dots, k} (\sqrt{n} \bar{X}_{i,l} / \sqrt{\lim_{n \rightarrow \infty} Var(\sqrt{n} \bar{X}_{i,l})})$ , where  $n$  is the number of forecast data points and  $\bar{X}_{i,l} = \frac{1}{N} \sum_{t=1}^n X_{i,t}^{(0,i)}$ . The estimation of  $\lim_{n \rightarrow \infty} Var(\sqrt{n} \bar{X}_{i,l})$  and the  $p$ -value of  $T_l^{SPA}$  are obtained using the stationary bootstrap procedure discussed by Politis and Romano [17].

**7 Case study**

**7.1 Selection of trained sample data**

We choose the closing prices from 2010.7.1 to 2011.3.31 of The Shanghai Composite Index as the sample data to analyze the Shanghai Composite Index recently change situation. There are 182 valid sample data (data from Security Star Website). We choose the closing prices as the sample data series to analyze. And then, we predict the yield rate of the 40 trading days. After getting  $\hat{y}_t, t = 1, 2, \dots, 40$ , we can use those predicted errors  $\varepsilon_t, t = 1, 2, \dots, 40$  in accordance with  $\hat{y}_t, t = 1, 2, \dots, 40$  to construct time series model, meanwhile we predict the afterward 10 errors and get  $\hat{\varepsilon}_t, t = 1, 2, \dots, 10$ . Finally, we get the eventual predicted value by means of Eq. (23).

As the closed point can be considered the reflection of the day's ending information, we suppose the daily closing price as  $k_t$ , at the same time, we define the rate of The Shanghai Composite Index as  $y_t$ , whose method of calculating is as below:

$$y_t = R_t = \ln(k_t/k_{t-1}). \tag{36}$$

The great length of the above definition is reducing the influence of skewness and kurtosis of Single index. Unlike stock price, the returns can accurately reflect the fluctuate tendency in stock market. Besides, we choose SSE 50 Index growth rate of equity and volume of trade as the influential factors on yield. We preprocess those three kinds of data just such as  $y_t$ .

That means the average growth rate per month in  $t$  day is:

$$X_{1t} = \ln(x_{1,t}/x_{1,t-1});$$

$$X_{2t} = \ln(x_{2,t}/x_{2,t-1});$$

$$X_{3t} = \ln(x_{3,t}/x_{3,t-1}).$$

$X_{2t}$  is the growth rate of market equity rate,  $X_{3t}$  is the growth rate of volume of trade.

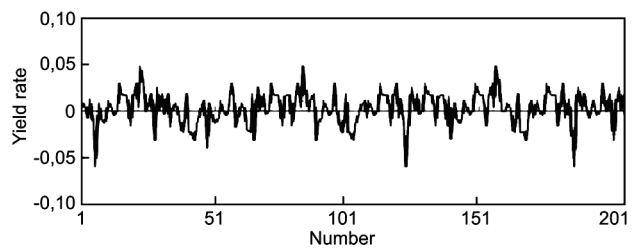


Figure 2 The scatter diagram of  $y$  produced by standardization of Shanghai Composite Index

**7.2 WLS-SVR prediction model using trained sample data**

Here we implement WLS-SVR model to the historical data and meanwhile to get the predicted error. The effect is shown as the picture below.

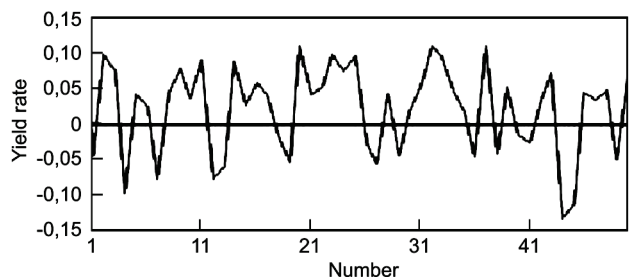


Figure 3 The error of prediction using WLS-SVR

**7.3 Error predictive model and SPA test**

As shown in Fig. 3, we need to have a further analysis on the predictive error which is  $\varepsilon_t$ .

We can see from Tab. 1 that the kurtosis is 5,37425, much bigger than the kurtosis of normal distribution. The skewness is -0,33733, J-B statistics is 479,3341 and  $p$  value is so much close to 0. It is clear that the time series

of predictive error of returns do not obey the normal distribution. It is equipped with 'High Kurtosis and Fat Tail', anti-symmetric, Mean 0 and left skewness.

**Table 1** Output of descriptive statistics of Predictive error  $\varepsilon_t$

<i>N</i>	50
Mean	0,02673
Std Deviation	2,24623
Skewness	-0,33733
Kurtosis	5,37425
Maximum	0,08575
Minimum	-0,09346
Jarque-Bera	479,3341

**Table 2** The Stability Test Result of predictive error  $\varepsilon_t$

	ADF	Magnificent level
$\ln \varepsilon_t$	-39,3018	(1 %) 3,426985
		(5 %) 2,876217
		(10 %) 2,567311

We can see from Tab. 2 that the ADF statistic of  $\varepsilon_t$  is -39,30182, smaller than the marginal value which is -3,426985 in significance level 1 %. Absolutely it is smaller than the other marginal value. So the assumption that the time series of  $\varepsilon_t$  is stable is of sound ground.

**Table 3** Testing the ARCH impact on  $\varepsilon_t$

F- statistic	18,996	<i>p</i> -value	0,000
$R^2$	17,468	<i>p</i> -value	0,000

From Tab. 3, we can see the feature aspect of  $\varepsilon_t$ . The fluctuation of the series is gregarious and consistent. So we can infer that the fluctuation has ARCH effect. However, up to now we still have to test from the quantity aspect. In the context, we apply LM to test its ARCH effect. When the  $q=4$ , we get the result as shown in Tab. 3.

The concomitant probability of both  $F$  and  $R^2$  is very low which indicates that only if  $q \geq 4$ , so there is a high order ARCH effect. As a result, it is not a smart idea to apply ARCH ( $q$ ) model. So we should consider using GARCH to construct our model and predict values.

Given by estimation of parameters of EGARCH model, we use the estimation into our EGARCH regression function.

(1) Mean value function:

$$\ln \varepsilon_t = 1,00375 \cdot \ln \varepsilon_{t-1}.$$

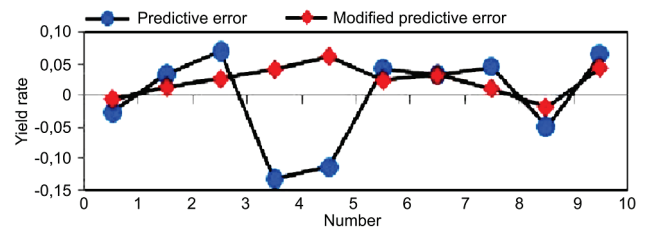
(2) Variance function:

$$\ln \sigma_t^2 = -0,312587 + 0,147986 \cdot \left| \frac{r_{t-1}}{\sigma_{t-1}} \right| + 0,001725 \cdot \frac{r_{t-1}}{\sigma_{t-1}} + 0,861578 \cdot \ln \sigma_{t-1}^2,$$

$R^2 = 0,9976$ ,  $DW = 1,97$ , Log likelihood = 2569,3,  $AIC = -4,96$ ,  $AIC = -4,96$ .

The above results indicate that every parameter is significant. The  $R^2$  of the model is as big as 0,9976 with  $\gamma = 0,001725 \neq 0$ . That proves there is asymmetric error information. From all the above results, we can safely believe in the success of fitting data by the method rendered in this paper.

After a further step on the prediction of error  $\varepsilon_t$  by using EGARCH model, we make a comparison between the modified  $\hat{\varepsilon}_t$  and the original  $\varepsilon_t$  as shown in Fig. 4.



**Figure 4** Modified  $\hat{\varepsilon}_t$  vs original  $\varepsilon_t$

The modified errors have the tendency of whole shrinking as it is obvious from Fig. 4. We use Eqs. (32) ÷ (35) to calculate the loss function and the SPA value which is shown in Tab. 4. This chart presents the *p*-values of the SPA tests on the six models. The first column in the table lists the name of the base model ( $M_0$ ) of the SPA test, and the seven remaining models are treated as competing models ( $M_i$ ).

**Table 4** SPA *p*-values for out-of-sample daily volatility forecasts

$L_i$	$M_0$	$M_i$					
		GARCH	GARCH-t	GM	ANN	EGARCH	LS-SVR
<i>MSE</i>	SVR-EGARCH	1,000	0,991	0,899	0,873	0,872	0,876
<i>MAE</i>		0,992	0,932	0,903	0,899	0,874	0,898
<i>HMSE</i>		0,904	0,934	0,922	0,867	0,859	0,861
<i>HMAE</i>		0,933	0,911	0,871	0,845	0,834	0,843

Notes: The values in bold are the largest *p*-values under a specific loss function. The larger the *p*-value, the less likely it is that the SPA null hypothesis "the base model is not outperformed by all competing models" can be rejected.

Every number in Tab. 4 is a SPA *p*-value which is obtained through 10 000 times of bootstraps under a specified loss function. The larger the *p*-value, the less likely it is that the SPA null hypothesis "the base model is not outperformed by all competing models" can be

rejected, that is, the better the forecasting performance of the base model ( $M_0$ ) is relative to its alternatives ( $M_i$ ). The values in bold are the largest *p*-values under a specific loss function. The precision of the intelligent EGARCH prediction model which is established by this article is

significantly better than the other models. The  $p$  values were close to 1.

## 8 Conclusion

As the volatility of stocks is stochastic, stable and asymmetric in information, single predictive method turns out to be insufficient. In this paper, we combine WLS-SVR and EGARCH to predict the volatility of stocks due to their superiority. We put the predictive errors in our model to improve the accuracy. At first, we use a known trained sample data to map into a higher-dimensional characteristic space non-linearly while at the same time using WLS-SVR predictive model to predict the volatility of stocks, we analyze the formed error time series. Next, we implement EGARCH to predict the error time series. The corrected error accuracy improved a lot, especially to the dots with large error. The effects of EGARCH correction are even more obvious.

China's stock market has been undergoing 20 years during which a great change has taken place. We cannot ignore that there exist certain uneconomic factors influencing the behavior of the market, such as banker operators, massive speculative behaviors. However, after removing those uneconomic factors, the method presented in the paper is a convincing way to predict the volatility of stocks especially those short-time ones.

## Acknowledgements

This research was partially supported by National Natural Science Foundation of China, grant No. 71101041, National 863 Project; grant No. 2011AA05A116 and National Innovative Experiment Program No. 111035954.

## 9 References

- [1] Tan Hua, Xie Chi. The prediction model of uncertainty mining in the stock market. *Social Scientist*. 4, 4(2008), pp. 65-67.
- [2] James S. Ang, Yulong Ma. Transparency in Chinese stocks: A study of earnings forecasts by professional analysts. // *Pacific-Basin Finance Journal*. 7, (1999), pp. 129-155.
- [3] Angelos Kanas, Andreasyannopoulos. Comparing linear and nonlinear forecasts for stock return. // *International Review of Economics and Finance*. 10, (2001), pp. 383-398.
- [4] Cheng-Lung Huang, Cheng-Yi Tsai. A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting. // *Expert Systems with Applications*. 3, 2(2009), pp. 1529-1539.
- [5] Zhou Ming, Yan Zheng, Ni Yixin, et al. A novel ARIMA approach on electricity price for forecasting with the improvement of predicted error. // *Proceedings of the CSEE*. 24, 12(2004), pp. 63-68.
- [6] Wenmao, Liu; Kun, Yang; Da, Liu et al. Day-ahead Electricity Price Forecasting with Error Calibration by Hidden Markov Model. // *Automation of Electric Power Systems*. 10, 33(2009), pp. 34-37.
- [7] Yuan-sheng, Huang; Jiajia, Deng; Zhenzhen, Yuan. SVM short-term load forecasting based on ARMA error calibration and the adaptive particle swarm optimization. // *Power System Protection and Control*. 39, 14(2011), pp. 26-32.
- [8] Comte, F.; Lieberman, O. Asymptotic theory for multivariate GARCH processes. // *Journal of Multivariate Analysis*, forthcoming, 2000.
- [9] Lee, J.; King, M. A locally most mean powerful based score test for ARCH and GARCH regression disturbances. // *Journal of Business and Economic Statistics*. 11, 3(1993), pp. 17-27.
- [10] Demos, A.; Sentana, E. Testing for GARCH effects: a one-sided approach. // *Journal of Econometrics*. 86, 3(1998), pp. 97-127.
- [11] Andersen, T. G.; Bollerslev, T. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. // *International Economic Review*. 39(1998), pp. 885-905.
- [12] Hansen, P. R.; Lunde, A. A forecast comparison of volatility models: does anything beat a GARCH(1,1). // *Journal of Applied Econometrics*. 20(2005), pp. 873-889.
- [13] Andersen, T. G.; Bollerslev, T.; Diebold, F. X.; Ebens, H. The distribution of realized stock return volatility. // *Journal of Financial Economics*. 61(2001), pp. 43-76.
- [14] Lopez, J. A. Evaluation of predictive accuracy of volatility models. // *Journal of Forecasting*. 10(2001), pp. 87-109.
- [15] Bollerslev, T.; Engle, R. F.; Nelson, D. ARCH models. // R. F. Engle, D. L. McFadden (Eds.), *Handbook of Econometrics*, vol. IV, Elsevier Science B.V., Amsterdam, 1994, pp. 2961-3038.
- [16] Hansen, P. R.; Lunde, A. A forecast comparison of volatility models: does anything beat a GARCH(1,1). // *Journal of Applied Econometrics*. 20(2005), pp. 873-889.
- [17] Politis, D. N.; Romano, J. P. The stationary bootstrap, *Journal of the American Statistical Association*, 89(1994), pp. 1303-1313.

## Authors' addresses

*Li qiang Hou*

*Shanlin Yang, Professor*

School of Management of Hefei University of Technology,  
Anhui Hefei 230009, China

E-mail: liqianghou1020@gmail.com

E-mail: shanlinyang1948@gmail.com