Forecasting Chinese Stock Markets Volatility Based on Neural Network Combining

Wenbing Xiao
Institute of Systems Engineering, Huazhong University of Science & Technology, Wuhan 430074, China
xiaowenbing11@163.com

Qi Fei
Institute of Systems Engineering, Huazhong University of Science & Technology, Wuhan 430074, China
qfei@mail.hust.edu.cn

Weimin Chen
College of Business Administration, Hunan University, Changsha, 410082, China
wminc@126.com

Abstract

Volatility plays a key role in asset and portfolio management and derivatives pricing. As such, more accurate measures and better forecasts of volatility are crucial for the implementation and evaluation of asset and derivative pricing models in addition to trading and hedging strategies. However, whilst GARCH models are able to capture the observed clustering effect in asset price volatility， they appear to provide relatively poor out-of-sample forecasts. In order to improve volatility forecasts, we put forward three combining volatility forecasting methods through simple averaging, an ordinary least squares model and an artificial neural network using daily closing data of the Shanghai Stock Exchange Composite index. The empirical study reveals that combining volatility forecasting methods have better forecast performance, and the non-linear method simulated by neural network are especially superior.

1. Introduction

The accurate estimation and forecasting of volatility in financial markets is an essential issue. Variations in market returns and other economy-wide risk factors are main features of asset and portfolio management and play a key role in derivatives pricing models. Moreover, price movements are linked to the arrival of news at an intra-day level. Thus, accurate measures and forecasts of volatility are crucial for the implementation and evaluation of asset and derivative pricing models in addition to trading and hedging strategies. Whilst it has long been recognized that returns volatility exhibits ‘clustering’, such that large (small) returns follow large (small) returns of random sign, it is only since the introduction of the generalized autoregressive conditional heteroscedasticity (GARCH) model that financial economists have modeled these temporal dependencies using econometric techniques[1-2]. However, despite the empirical success of the GARCH model, and its subsequent developments, in modeling the volatility of asset prices, numerous studies report that standard volatility models provide poor forecasts [3-4]. In order to improve volatility forecasts, we put forward three volatility forecast combining methods through simple averaging, an ordinary least squares model and an artificial neural network using daily closing data of the Shanghai Stock Exchange Composite index. The empirical study reveals that volatility forecast combining methods have better forecast performance, and the non-linear method simulated by neural network are especially superior.

The rest of the paper is as follows: Section 2 presents the research methodology, starting with the conditional volatility models and continuing with the forecast combining models, one of which is an artificial neural network (ANN). The data set is described in Section 3. The results are presented in Section 4, starting with some diagnostic statistics and proceeding with a number of measures of the models’ out-of-sample forecasting performance. The conclusions are drawn in Section 5.

2. Methodology

2.1. General volatility forecasting model

2.1.1. Random walk model.

Under a random walk model, the best forecast of this period's volatility is last period's observed volatility:

$$\hat{\sigma}_T^2 = \sigma_{T-1}^2 \qquad (1)$$

978-0-7695-3304-9/08 $25.00 © 2008 IEEE
DOI 10.1109/ICNC.2008.416

2.1.2 Moving average models.

Under the assumption of a stationary mean, the best forecast of this period's volatility is an average of past observed volatilities. The moving average model can be expressed as:

$$\hat{\sigma}_T^2 = \frac{1}{M}\sum_{j=1}^{M} \sigma_{T-j}^2 \qquad (2)$$

2.1.3. Exponential smoothing model.

An exponential smoothing model is posited to be a function of the immediate past forecast and the immediate past observed volatility.

$$\hat{\sigma}_T^2 = \alpha\hat{\sigma}_{T-1}^2 + (1-\alpha)\sigma_{T-1}^2 \qquad (3)$$

The smoothing parameter α is constrained to lie between zero and one. The optimal value of α must be determined empirically.

2.1.4. Exponentially weighted moving average model.

An exponentially weighted moving average model (EWMA) is similar to the exponential smoothing model except that the past observed volatility in expression (3) is replaced by the moving average forecast which can be formally expressed as:

$$\hat{\sigma}_T^2 = \alpha\hat{\sigma}_{T-1}^2 + (1-\alpha)\frac{1}{M}\sum_{j=1}^{M}\sigma_{T-j}^2 \qquad (4)$$

2.1.5. GARCH models.

The GARCH model involves the joint estimation of a conditional mean and a conditional variance equation. As the GARCH (1, 1) model has generally been found to be the most appropriate of the standard ARCH family of models for stock return data, this model is employed, viz:

$$r_t = \mu + \varepsilon_t, \varepsilon_t \sim N(0, h_t) \qquad (5)$$

$$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1} \qquad (6)$$

Following Engle and Bollerslev [5], s-step ahead forecast can be

formed based on the GARCH(1,1) model as follows:

$$\hat{h}_{t+s} = \hat{\omega}\sum_{i=0}^{s-2}(\hat{\alpha}_1 + \hat{\beta}_1)^{i-1} + (\hat{\alpha}_1 + \hat{\beta}_1)^{s-1}\hat{h}_{t+s} \qquad (7)$$

A given periods volatility forecasts are then formed by aggregating the s-step ahead daily forecasts across trading days in each periods as follows:

$$\begin{aligned}\hat{\sigma}_T^2 &= \sum_{s=1}^{N_t}\hat{h}_{t+s} \\ &= \hat{\omega}\sum_{s=l}^{N_t}\sum_{i=0}^{s-2}(\hat{\alpha}_l + \hat{\beta}_l)^i + \sum_{s=1}^{N_t}(\hat{\alpha}_l + \hat{\beta}_l)^{s-1}\hat{h}_{t+1}\end{aligned} \qquad (8)$$

Where a given periods (T) has N_t daily observations.

2.1.6. GJR-GARCH and EGARCH model. The GJR-GARCH model augments the conditional variance equation in the GARCH (p, q) model with a variable equal to the product of I_{t-1} and ε_{t-1}^2 where is a dichotomous dummy variable that takes the value of unity if ε_{t-1} is negative and zero otherwise [6]. In the case of the GJR-GARCH (1, 1) version of the model, the conditional variance equation becomes:

$$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1} + \gamma\varepsilon_{t-1}^2 I_{t-1} \qquad (9)$$

EGARCH is exponential GARCH model ([7]). EGARCH(1, 1)model's conditional variance equation is as follows：

$$\begin{aligned}\ln(h_t) &= \omega + \beta_1\ln(h_{t-1}) \\ &+ \gamma\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \alpha_1\left(\frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} - \sqrt{\frac{2}{\pi}}\right)\end{aligned} \qquad (10)$$

A similar forecasting procedure to the GARCH (1, 1) model is used to obtain daily volatility forecasts for TGARCH(1, 1) and EGARCH(1, 1) specifications which are then summed over trading days to obtain a given periods volatility forecasts.

2.2. Traditional combining Methods

There are now many commonly employed methods for combining forecasts. The assumption that the conditional expectation of the variable being forecasted is a linear combination of the available forecasts is consistent across all combining methods.

The four individual volatility forecasts models (GARCH, EGARCH, TGARCH, and MAV) were combined through the following formula:

$$F_T = \beta_0 + \sum_{k=1}^{k} l_k f_k \qquad (11)$$

Where F_T is the combined volatility forecast for T period, f_k is the selected individual forecasts ($k = 4$ in our case), and β_0, l_k are parameters. The first simple forecast combining technique was simple averaging (Comb1(AVE)), that is, estimation of Eq. (11) by setting $\beta_0 = 0$ and $l_k = 1/4$.The second combining technique was based on ordinary least squares (Comb2(OLS)) to estimate the parameters β_0 and l_k 's value.

2.3. Artificial neural networks (ANNs)

Neural networks are non-linear non-parametric models that have their roots in biology and the research on neurons' behavior [8-10]. The artificial neural networks combining technique is as follows:(Comb3 (ANN)).

$$F_T = \beta_0 + \sum_{k=1}^{k} l_k f_k + \sum_{j=1}^{J} w_{1j} S(\beta_j + \sum_{k=1}^{K} v_{jk} f_k) \quad (12)$$

F_t is the combined volatility forecast for T period, $S(\cdot)$ is a functional form, f_k is the selected individual forecasts ($k=4$ in our case), and $\beta_0, \beta_j, l_k, w_{1j}$ and v_{jk} are parameters.

The ANN model of this study is a multilayer perceptron (MLP) and consists of one input layer with four input nodes (one for each of the GARCH, EGARCH, TGARCH, and MAV volatility forecasts), one hidden layer with four nodes, and one output layer with one node (the combined forecast). The architecture includes five bias nodes (one for each hidden node and one for the output node) and four skip connections that link the input nodes directly to the output node. This architecture results in a total of 29 parameters: four output-to-hidden-node connections (w_{1j}), 16 hidden-to-input-node connections (v_{jk}), five bias nodes (β_0, β_j), and four skip connections (l_k). The choice of the specific architecture was based on previous empirical evidence that found networks having the number of hidden nodes equal to the number of input nodes to exhibit superior forecasting performance. Adding more hidden nodes or layers can easily lead to over-fitting and poor forecasting performance. All initial values for the weights and biases were randomly generated from a uniform distribution in the range [-0.20, 0.20]. We experimented with several pairs of learning and momentum rates. Small values make the convergence process agonizingly slow while large values lead to oscillation around a local minimum. We ended up with the value of 0.002 which performed the best in our experiments. All inputs to the ANN were linearly normalized to [0, 1]. The weights were estimated using the standard back-propagation algorithm and epoch-based training. The actual volatility was approximated by the squared forecast error and was normalized to serve as the target value for the ANN.

3. Data

In order to investigate aggregate stock market volatility, a market index is required. The index used in our study is Shanghai Stock Exchange Composite closing index. Our sample consists of over 2410 observations encompassing the period from January 6, 1997 to December 29, 2006. Daily returns are identified as the difference in the natural logarithm of the closing index value for two consecutive trading days.

To ensure that the volatility forecasting models are robust, the full (working) weekly volatility series (on business time scale) [10], is defined as the sum of five squared daily returns, viz:

$$\sigma_T^2 = \sum_{t=1}^{5} r_t^2, T = 1,2,\cdots 482 \quad (13)$$

Where r_t is the daily rate of return.

The 10-year study period is split up into two subperiods: January 6, 1997-December 8, 2004 (I) (382 full working weeks), December 9, 2004 - December 29, 2006 (II) (100 full working weeks). Subperiod I is used to estimate the parameters of the parameters of individual volatility forecasts (GARCH, EGARCH, TGARCH, MAV). The forecast combing process worked in a similar way. The parameters of the OLS forecast combining model and the weights of the ANN model were trained using the volatility forecast over subperiods I. Subperiods II (100 full working weeks) are used to test sets for all models.

4. Out-of-sample model forecast results

4.1. Definition of the forecast statistical criteria

As is standard in the economic literature [11, 12], we compare the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE), Theil U-statistic (Theil-U), the Mean Error (ME) and Correct Directional Change (CDC). Let σ^2 , $\hat{\sigma}^2$ be the actual volatility, forecasting volatility respectively. The forecast error statistics are defined as follows:

$$ME = \frac{1}{100} \sum_{i=1}^{100} (\hat{\sigma}_i^2 - \sigma_i^2) \quad (14)$$

$$MAE = \frac{1}{100} \sum_{i=1}^{100} | \hat{\sigma}_i^2 - \sigma_i^2 | \quad (15)$$

$$RMSE = \sqrt{\frac{1}{100} \sum_{i=1}^{100} (\hat{\sigma}_i^2 - \sigma_i^2)^2} \quad (16)$$

$$MAPE = \frac{1}{100} \sum_{i=1}^{100} | (\hat{\sigma}_i^2 - \sigma_i^2) / \sigma_i^2 | \quad (17)$$

$$Theil-U=\frac{\sqrt{(1/100)\sum_{i=1}^{100}(\hat{\sigma}_i^2-\sigma_i^2)^2}}{\sqrt{(1/100)\sum_{i=1}^{100}\hat{\sigma}_i^4}+\sqrt{(1/100)\sum_{i=1}^{100}\sigma_i^4}} \quad (18)$$

$$CDC=\frac{100}{100}\sum_{i=1}^{100}D_i \quad , \quad \text{if} \quad (\sigma_i^2-\sigma_{i-1}^2)(\hat{\sigma}_i^2-\sigma_{i-1}^2)>0,$$

$D_i=1$, else $D_i=0$. (19)

4.2 Forecast results

Table 1 presents the actual forecast error statistics for each model across the six error measures. The ME does not allow for the offsetting of errors of different signs and as such, little credence should be placed upon it. However, the ME can be used as a general guide as to the direction of over or under–prediction. All models are found to over-predict volatility.

The MAE statistic indicates that Comb3 (ANN) model provides the most accurate forecasts, the Comb2 (OLS) model ranks a close second. The RMSE statistic also indicates that Comb3 (ANN) model provides the most accurate forecasts. The Comb2 (OLS) model ranks a close second. The MAPE statistic gives a relative indication of overall forecasting performance. The Comb3 (ANN) model has the best forecasts. The Comb1 (AVE) model ranks a close second and is only marginally less accurate than the best model. The Comb2 (OLS) model provides the best forecast of directional change, achieving a remarkable directional forecasting accuracy of around 75%. The Comb3 (ANN) model ranks a close second in terms of directional forecasting accuracy around 73%.

In summary, the ranking of the forecasting models varies depending upon the choice of error statistic for individual forecast. However, the empirical study reveals that combining volatility forecasting methods have better forecast performance, and the non-linear method simulated by neural network are especially superior.

Table 1 Error statistics from forecasting volatility

Forecast model	ME*10^5	MAE*10^4	RMSE*10^3	MAPE	Theil-U	CDC
RW	5.55	7.49	1.250	1.353	0.481	50
MAV(M=5)	4.57	6.11	1.010	1.261	0.424	73
MAV(M=20)	6.31	6.07	0.999	1.430	0.437	72
MAV(M=40)	2.85	5.75	0.980	1.470	0.442	71
Exp smooth	4.14	5.69	0.968	1.380	0.432	70
EWMA(M=5)	4.32	5.80	0.982	1.452	0.439	71
EWMA(M=20)	2.70	5.92	0.995	1.521	0.448	72
GARCH(1,1)	25.6	6.10	0.979	1.350	0.371	70
GJR(1,1)	20.9	5.91	0.967	1.291	0.369	71
EGARCH(1,1)	23.2	6.13	0.953	1.320	0.368	68
Comb1(AVE)	8.09	5.69	0.957	1.382	0.409	71
Comb2(OLS)	0.85	5.46	0.948	1.260	0.365	75
Comb3(ANN)	4.32	5.20	0.923	1.254	0.362	72

5. Conclusions

The presented study examined the out-of-sample volatility forecasting performance of Shanghai Stock Exchange Composite index. The ranking of any one forecasting model varies depending upon the choice of error statistic for individual forecast. However, the empirical study reveals that combining volatility forecasting methods have better forecast performance, and the nonlinear method simulated by neural network are especially superior.

6. Reference

[1] R.F., Engle, Autoregressive conditional heteroscedasticity with estimatesof the variance of United Kingdom inflation, Econometrics, 50(4), 1982, pp.987-1007.

[2] T. Bollerlsev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 31, 1986, pp. 307-327.

[3] S. Figlewski, Forecasting volatility. Financial Markets, Institutions and Instruments, 6, 1997, pp. 1–88.

[4] P. Jorion, Predicting volatility in the foreign exchange market. Journal of Finance, 50, 1995, pp. 507–528.

[5] L. R. Glosten, R. Jagannathan, and D.E. Runkle, On the relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, The Journal of Finance, 48(5), 1993, pp. 1779-1801.

[6] D.B. Nelson, Conditional Heteroskedasticity in Asset Returns: A New Approach, Econometrica, 59(2), 1991, pp. 347-70.

[7] I. Ginzburg, and D. Horn, Combined neural networks for time series analysis, Neural information processing systems, 6, 1994, pp. 224-31.

[8] G. P. Zhang, and V. L. Berardi, Times series forecasting with neural network ensembles: an application for exchange rate prediction, Journal of the Operational Research Society, 52, 2001, pp. 652-64.

[9] Y. Michael and T. Christos, Combining conditional volatility forecasts using neural networks: an application to the EMS exchange rates, Journal of International Financial Markets, Institutions and Money, 9, 1999, pp. 407-422.

[10] V. Valeriy, and B. Supriya, Volatility forecasting from multiscale and high-dimensional market data, Neurocomputing, 55, 2003, pp. 285-305.

[11] J. Timothy, and W. Robert, An evaluation of volatility forecasting techniques, Journal of Banking & Finance, 20, 1996, pp. 419-438.

[12] L. Christian, and H. Xuehuan, Forecasting and trading currency volatility: An application of recurrent neural regression and model combination, Journal of Forecasting, 21, 2002, pp. 317-354.