Comparison of GARCH, Neural Network and Support Vector Machine in Financial Time Series Prediction

Altaf Hossain[1], Faisal Zaman[2], M. Nasser[1] and M. Mufakhkharul Islam[3]

[1] Department of Statistics, Rajshahi University, Rajshahi-6205, Bangladesh
Email: rasel_stat71@yahoo.com, mnasser.ru@gmail.com
[2] Department of System Design and Informatics, Kyushu Institute of Technology,
680-4 Kawazu, Iizuka, 820-8502 Fukuka, Japan. Email: faisal.zaman80@gmail.com
[3] Department of Computer Science & Engineering, BUET, Dhaka-1000, Bangladesh
Email: nasif192@yahoo.com

Abstract. This article applied GARCH model instead AR or ARMA model to compare with the standard BP and SVM in forecasting of the four international including two Asian stock markets indices.These models were evaluated on five performance metrics or criteria. Our experimental results showed the superiority of SVM and GARCH models, compared to the standard BP in forecasting of the four international stock markets indices.

Index Terms-Generalized Autoregressive Conditional Heteroskedastic (GARCH), Neural Network (NN),Back Propagation (BP), Artificial BPNN (BPANN), Support Vector Machine(SVM)

1 INTRODUCTION

There is a powerful method, available in the literature for univariate time series forecasting, known as Box-Jenkins [2] Autoregressive Moving Average (ARMA) approach on stationary time-series. In such conventional econometric models, the variance of the disturbance is assumed to be constant. But many economic and financial time series such as exchange rates, indices, market returns, inflation rate etc exhibit periods of unusual large volatility, followed by periods of relative tranquility. In such circumstance, Engle [5] showed that it is possible to model the mean and the variance of a series simultaneously by a model named Autoregressive Conditional Heteroskedastic (ARCH). In latter, Bollerslev [1] extended Engle's original work by developing a technique that allows the conditional variance to be an ARMA process and that extended process is known as the GARCH process.The classical methods are based on some specific assumptions, such as linearity; or on error distributions, such as normality. In this circumstance, ANNs were developed to meet the increasing demand that can predict, detect, classify and summarize the structure of variables and define the relationships between them - without relying too much on such assumptions. Not all relationships in economics and finance are direct. But the hidden layer of

an ANN can capture all non-direct relationships between input and output variables. On the other hand, the SVM, originally developed as an implementation of Vapnik's Structural Risk Minimization (SRM) principle (Vapnik [12]), is now being used to solve a variety of learning, classification and prediction problems. In many ways, an SVM performs the same function as an ANN. It has the following advantages over ANN: (1)it can obtain the global optimum and (2) the overfitting problem can be easily controlled.

Steiner and Wittkemper [8] investigated the performance of several ANN models that forecast the return of a single stock. Kamruzzaman and Sarker [7] modeled and forecasted currency of exchange rates using three ANNs and a comparison was made with traditional ARIMA model. Tay and Cao [9] examined the feasibility of SVM in financial time series forecasting by comparing it with a multilayer BPNN and they showed that SVM outperforms the BPNN. Kim and Han [14] showed that SVM provides a promising alternative to stock market prediction comparing it with BPNNs. In most of studies like the above, they showed that SVM outperformed both BPNN and traditional statistical models. Again the simple neural learning procedures, such as BP algorithms easily outperformed the best practice of traditional statistical models. Chen and Wu [3] compared SVMs and BPs taking Autoregressive (AR) as a benchmark in forecasting the six major Asian stock markets. They showed that both the SVMs and BPs outperform the traditional models, ARs. They did prediction of transformed data, but not level data.

In this paper, our main contribution is to examine the capacity of the GARCH model comparing with the standard BPNN and simple SVM in forecasting financial time series at level data. We are going to use the monthly data in our empirical study for its Aggregational Gaussianity property.

The methodology including two prediction methods in this paper is discussed in Section 2. The results and discussion are presented in Section 3. Finally, our conclusion is in Section 4.

2 METHODOLOGY

2.1 The GARCH Modeling

Engle [5] showed that the serial correlation in squared returns, or conditional heteroskedasticity, can be modeled using an ARCH model. An important extension of the ARCH model proposed by Bollerslev [1] replaces the AR (p) representation in the variance series of ARCH process with an ARMA (p,q) formulation of the form

$$y_t = E_{t-1}[y_t] + \epsilon_t \tag{1}$$

$$\epsilon_t = z_t \sigma_t \tag{2}$$

$$\sigma_t^2 = a_0 + \sum_{j=1}^{p} b_j \sigma_{t-j}^2 + \sum_{i=1}^{q} a_i \epsilon_{t-i}^2 \tag{3}$$

where $E_{t-1}[.]$ represents expectation conditional on information available at time $t-1$ and z_t is a sequence of iid random variables with mean zero and unit variance. In the GARCH model z_t is assumed to be iid standard normal. And the coefficients $a_i(i=0,...,q)$ and $b_j(j=1,...,p)$ are all assumed to be positive to ensure that the conditional variance σ_t^2 is always positive. The model in (3) together with (1)-(2) is known as the generalized ARCH or GARCH (p,q) model.

Hansen and Lunde [6], provided compelling evidence that it is difficult to find a volatility model that outperforms the simple GARCH (1,1). So we use GARCH(1,1) in this paper. For testing ARCH/GARCH effects, assessing the fit, diagnostic checks for model adequacy and estimation see Enders [4].

2.2 Support-vector Regression (SVR)

The SVM deals with the classification and regression problems by mapping the input data into the higher-dimensional feature spaces. Our problem is regression.Firstly, we discuss the simplest model of SVR, a linear regression in the feature space. The SVR algorithm tries to construct a linear function such that training points lie within a distance ε(i.e., ε-insensitive zone in the error loss function).

Given a set of training data $\{(X_1,Y_1),...,(X_l,Y_l)\} \subset X \times \mathbb{R}$, where X denotes the space of the input patterns, the goal of SVR is to find a function $f(x)$ that has at most ε deviation from the targets y_i for all the training data and, at the same time, is as flat as possible. Let the linear function f takes the form:

$$f(x) =< w, x > +b; w \in X, b \in \mathbb{R} \qquad (4)$$

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

$$\Phi(w,\xi) = \frac{1}{2}\|w\|^2 + C \sum_i (\xi_i^- + \xi_i^+) \qquad (5)$$

where C is pre-specified value, and ξ_i^-, ξ_i^+ are slack variables representing upper and lower constraints on the outputs of the system. Flatness in (5) means a smaller $\|w\|$. Using ϵ -insensitive loss function, we get our required solutions.

To enable the SVR to predict a non-linear situation, we map the input data into a feature space. By using a kernel function, it is possible to compute the SVR without explicitly mapping in the feature space. See Scholkopf and Smola [13], and Chen and Wu [3] for more details. We use the radial basis kernel function $k(x,y) = \exp{(-\gamma|x-y|^2)}$ from several typical kernel functions as it performs well under general smoothness assumptions.

3 RESULTS AND DISCUSSION

3.1 Data Source, Nature and Preprocessing of the Models

For analysis we have used the monthly data of the four international stock market indices collected from the Financial Forecast Center Home Page. We have used

here the monthly data for its Aggregational Gaussianity property. The names of the stock markets are Japan (Nikkei 225 from January, 1984 to January, 2006), Hong Kong (Hang Seng from December, 1986 to January, 2006), U.K. (FTSE 100 from April, 1984 to January, 2006), and Germany (DAX from November, 1990 to January, 2006). The time periods used and the indices' statistical data are listed in Table 1. The time periods cover many important economic events, which we believe are sufficient for the training models.

Table 1. DESCRIPTION OF DATA SETS

$INDEX$	$TIMEPERIOD$	N	$HIGH$	LOW	$MEAN$	SD
Nikkei 225 (NK)	1984.01- 2006.01	265	38920	7837	18020	6643.7
Hang Seng (HS)	1986.12- 2006.01	230	17410	2138	8977	4411.6
FTSE 100 (FT)	1984.04- 2006.01	262	6930	1009	3559	1612.3
DAX (DX)	1990.11- 2006.01	183	7645	1398	3637	1690.4

Here N indicates the number of realizations for each market

We transform the original closing price into white noise series differencing two times at lag one for each series except the market NK, where we take one time difference. According to Thomason [10], [11] and Tay and Cao [9], this transformation has many advantages. The modification is known, so we can reverse it to compare the forecasts with the original series. Each of the four data sets is partitioned into three subsets according to the time sequence in the ratio 95: 2.5: 2.5. The first part is used for training; the second is a validating set that selects optimal parameters for the GARCH and SVR, and prevents the overfitting found in the BP neural networks selecting its number of inputs and units for the hidden layer, and the last is used for testing.

3.2 Comparison and Discussion

For each market, we also design different ordered ARMA models as to show the necessity of GARCH model. The Autocorrelation Function (ACF) of squared residuals for each fitted ARMA indicates the presence of GARCH effects. The forecasting results of the ARMA, GARCH, BP and SVM for the test set are collated in Table 2. According to Tay and Cao [9] and Thomason [10] [11], the prediction performance is evaluated using the following statistics: Mean Square Error (MSE), Normalized Mean Squared Error (NMSE), Mean Absolute Error (MAE), Directional Symmetry (DS) and Weighted Directional Symmetry (WDS). See Chen and Wu [3] for more details. In general, GARCH, BP and SVM models outperform the ARMA model in the deviation performance criteria except only one market, HS. However, in the either direction and weighted direction performance criteria, the ARMA model sometimes shows better results comparing with BP and SVM models, possibly because SVMs and BPs are trained in terms deviation performance; the former to find the error bound and

the latter to minimize MSE. The results of ARMA model are not added in the Table due to space constrain.

Table 2. COMPARISON OF THE RESULTS OF GARCH, NN AND SVM MODELS ON THE TEST SET

		$SQRTMSE$	MAE	$NMSE$	DS	WDS
NK	GARCH	824.82	686.75	0.21	85.7	0.00
	BP	900.91	765.64	0.25	71.4	0.42
	SVM	792.57	659.18	0.91	85.7	0.00
HS	GARCH	656.85	525.59	1.88	0.00	0.00
	BP	778.94	864.11	2.65	33.3	1.18
	SVM	632.15	505.45	1.74	16.6	5.06
FT	GARCH	196.91	170.64	1.04	42.8	1.23
	BP	236.81	220.12	1.51	28.5	5.87
	SVM	158.47	147.05	0.67	57.1	0.45
DX	GARCH	234.99	229.32	0.55	40.0	0.66
	BP	191.05	156.14	0.36	60.0	0.79
	SVM	218.31	204.54	0.47	40.0	0.50

The deviation criteria, such as NMSEs are less than 1, except for the HS index in the SVMs. The NMSEs, both of GARCH and BP are greater than 1 for the two markets, HS and FT. It is remarkable that the NMSE of the BP is not only greater than 1 but also 2 for the market, HS. In the NK and HS data sets, the SVM models perform better than the GARCH and BP models. On the other hand, the GARCH method performs better than SVM and BP in only one market, FT; the BP models also perform better than the GARCH and SVM in only one market, DX. If we make comparison only between GARCH and NN, then GARCH shows better results than NN in three markets out of four.

4   CONCLUSION

In this article, we examine the feasibility of applying two Artificial Intelligence (AI) models, SVM and standard BP, and one classical statistical model, GARCH to financial time-series forecasting. Our experiments demonstrate that both models, SVM and BP perform better than the ARMA model in the deviation measurement criteria. Previous research also claims that the NN method is superior to the classical statistical method, ARMA or AR. If we consider the classical statistical method GARCH (just extended version of ARMA process), it is superior to the NN method except in the only one market DX of indices. It is not only superior to the NN method but also it shows close performance to the SVM method. The classical methods, ARMA and GARCH, require large number of sample size for better forecasting and these drastically reduce the original

sample size when the high order model is fitted. The SVMs and GARCH models are more parsimonious than the NN models since the SVMs and GARCH models take smaller number of parameters for better forecasting. The SVMs and GARCH models are also more interpretable than that of NN because the ANN structures contain hidden layers which are not interpretable. In general, the GARCH, BP and SVM models perform well in the prediction of indices behavior in terms of the deviation criteria. Our experiments also show that both the statistical (GARCH) and the AI techniques can assist the stock market trading and the development of the financial decision support systems. But if we consider predictability along with interpretability, parsimonious and intrinsic properties, then we recommend the GARCH models with SVMs in financial time series prediction.

References

1. Tim. Bollerslev: Generalized autoregressive conditional heteroscedasticity. Journal of Econometric 31, 307-27, (1986)
2. G. E. P. Box and G. M. Jenkins: Time Series Analysis Forecasting and Control. San Francisco: Holden-Day, 1976.
3. W-H. Chen, J-Y. Shih, and S. Wu: Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets. Int. J. Electronic Finance, Vol. 1, No. 1, pp.49-67, 2006.
4. Walter Enders: Applied Econometric Time Series. John Wiley & Sons, Inc., 2004.
5. Robert F Engle: Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50, 987-1007, July 1982.
6. P. Hansen and A. Lunde: A Forecast comparison of volatility models: does anything beat a GARCH (1,1) model? Journal of Applied Econometrics, 20, 873-889, 2004.
7. J. Kamruzzaman, and R. Sarker: Forecasting of currency exchange rates using ANN: a case study. 2003 Proc. IEEE Intl. Conf. on Neur. Net. & Sign. Process. (ICNNSP03), pp. 793-797, China, 2003.
8. M. Steiner and H.G. Wittkemper: Neural Networks as an alternative stock market model. Neural Networks in the Capital Markets, pp.135-147, 1995.
9. F.E.H. Tay and L. Cao: Application of support vector machines in financial time-series forecasting. Omega, Vol. 29, pp.309-317, 2001.
10. M. Thomason: The practitioner method and tools: a basic neural network-based trading system project revisited (parts 1 and 2). Journal of Computational Intelligence in Finance, Vol. 7, No. 3, pp. 36-45, 1999a.
11. M. Thomason: The practitioner method and tools: a basic neural network-based trading system project revisited (parts 3 and 4). Journal of Computational Intelligence in Finance, Vol. 7, No. 3, pp. 35-48, 1999b.
12. V.N. Vapnik: The Nature of Statistical Learning Theory. 2nd ed. New York: Sringer-Verlag, 1995.
13. Bernhard Scholkopf and Alexander J. Smola: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. Cambridge: Massachusetts, 2002.
14. K-S. Kim, and I. Han: The cluster-indexing method for case-base reasoning using self-organizing maps and learning vector quantization for bond rating cases. Expert Systems with Applications, Vol. 21, pp. 147-156, 2001.