Expert Systems with Applications 36 (2009) 6187–6191

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Volatility model based on multi-stock index for TAIEX forecasting

Ching-Hsue Cheng [a], Liang-Ying Wei [a,b,*]

[a] Department of Information Management, National Yunlin University of Science and Technology, 123, Section 3, University Road, Touliu, Yunlin 640, Taiwan, ROC

[b] Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan 128, Section 2, Academia Road, Nankang, Taipei 115, Taiwan, ROC

ARTICLE INFO

Keywords:
Time series analysis
Adaptive learning
Stock price index forecasting

ABSTRACT

Conventional time series models have been applied to handle many forecasting problems, such as financial, economic and weather forecasting. In stock markets, correct stock predictions will bring a huge profit for stock investors. However, conventional time series models produce forecasts based on some strict statistical assumptions about data distributions, and, therefore, they are not very proper to forecast financial datasets. This paper proposes a new forecasting model using adaptive learning techniques to predict TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock index) with multi-stock indexes (NASDAQ stock index and Dow Jones stock index). In verification, this paper employs seven year period of TAIEX stock index, from 1997 to 2003, as experimental datasets, and the root mean square error (RMSE) as evaluation criterion. The performance comparison results show that the proposed model outperforms the listing methods in forecasting Taiwan stock market. Besides, from statistical test results, it is showed that the volatility of Dow Jones and the NASDAQ affect TAIEX significantly.

1. Introduction

In stock market, practical experiments of investors tell that the relation between different stock markets is strong enough to be sensible and therefore many investors would utilize the volatility of other stock markets, which are highly affect native stock markets, as prevision for forecasting. The research advanced by Dickinson (2000) has revealed the stock price indexes of different stock markets, located in different countries, would influence mutually. Take the Taiwan stock market as example, it is common sense for investor that the volatility of NASDAQ affects the TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock index) conspicuously. The correlation may be explained by two facts: (1) The American is the leader nation for global economic and the volatility of American stock markets certainly influences the other stock markets worldwide. (2) There are many electronic companies in Taiwan such as TSMC (Taiwan Semiconductor Manufacturing Company) and CT (Chunghwa Telecom), which are main component of the TAIEX, that have issued ADR (American Depositary Receipt) in NASDAQ (the largest US electronic stock market). For the reason, there should be a relation between these two stock markets. Therefore, from the research of Huarng, Yu, and Hsu (2007), a fuzzy time series model was proposed to use the volatility of NASDAQ and Dow Jones (Dow Jones Industrial Average) as forecasting factors for Taiwan stock market. For the reasons above, it can be claimed that the volatility of two American stock markets (Dow Jones and NASDAQ) can play a prevision for the volatility of Taiwan stock market in the next day.

Time series have been applied to forecast stock markets, and various models have been proposed (Enders, 2004). In 1982, Engle (Engle, 1982) proposed the ARCH (p) (Autoregressive Conditional Heteroscedasticity) model that has been used by many financial analysts to forecast stock market. The following researcher, Bollerslev (Bollerslev, 1986), proposed the GARCH (Generalized ARCH) model to refine the ARCH model. In order to overcome the drawback of the GARCH model, two time series models were proposed sequentially: (1) Nelson provided the EGARCH (Nelson, 1991) (Exponential GARCH) model to overcome the drawback of the GARCH model, leverage effects; and (2) Morgan proposed the EWMA (Morgan, 1996) (exponentially weighted moving average) model to strengthen the initial GARCH model, which is a non-stationary GARCH (1,1) model.

Additionally, Box and Jenkins (1976) proposed the autoregressive moving average (ARMA) model which combines a moving average process with a linear difference equation to obtain an autoregressive moving average model, and the ARMA model performs forecasting at linear stationary condition. However, under non-stationary condition, the ARIMA (Box & Jenkins, 1976) model was proposed to describe such homogeneous non-stationary behavior.

* Corresponding author. Address: Department of Information Management, National Yunlin University of Science and Technology, 123, Section 3, University Road, Touliu, Yunlin 640, Taiwan, ROC. Tel.: +886 5 5342601x5312; fax: +886 5 531 2077.

E-mail addresses: chcheng@yuntech.edu.tw (C.-H. Cheng), g9523808@yuntech.edu.tw (L.-Y. Wei).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.07.020

From the reviewed literature above, it is found that there are some statistic assumptions about data distributions for the conventional financial time series models. However, in practical stock markets, the distributions of stock data do not obey these assumptions, and therefore these conventional models can not properly describe stock data.

Besides, most of the conventional time series models utilize only one variable (stock price) in forecasting processes and ignore multiple variables for stock market such as native macro economic factors, global economic environments, and highly related foreign stock markets. Therefore, in stock markets forecasting, more than one variable should be considered.

In the evolution of time series models, many researchers have applied data mining techniques in financial analysis. In 1990, Kimoto, Asakawa, Yoda, and Takeoka (1990) developed a prediction system for stock market by using neural network. The following researchers, Nikolopoulos and Fellrath (1994), combined genetic algorithms (GA) and neural network to develop a hybrid expert system for investment advising. Kim and Han (2000) proposed genetic algorithms approach to feature discretization and the determination of connection weights for artificial neural networks (ANNs) to predict the stock price index. Recently, hybrid several artificial intelligence (AI) techniques have become major methods mining time series data. In 2004, Pai and Lin's (2004) proposed a hybrid model which combines ARIMA and SVM (support vector machines) to forecast stock price. Tseng, Tzeng, Yu, and Yuan (2001) proposed a fuzzy ARIMA model for forecasting exchange market. Huarng and Yu (2006) applied backpropagation neural network to establish fuzzy relationships in fuzzy time series for forecasting stock price. Chen and Chung (2006) presented a new method to deal with enrollments forecasting problems based on high-order fuzzy time series and genetic algorithms, where the length of each interval in the universe of discourse is tuned by using genetic algorithms. And Roh (2007) integrated neural network and time series model for forecasting the volatility of stock price index. However, the rules mined from time series data with the AI data mining techniques such as GA and ANN are not easy to be understood.

Based on the drawbacks of time series models reviewed from the literatures above, this paper proposes a simple time series model, which considers the volatility of the NASDAQ and Dow Jones index together as forecasting variables, and utilizes the adaptive learning technique to adapt the linear parameters of the proposed model to reach optimal predicting performance. To overcome the drawback the assumption about data distributions, adaptive learning is applied to the proposed model. From the literatures of econometrics, Kmenta (1986) has proposed the adaptive expectations model for financial analysis. Then adaptive learning approach has been applied in several studies (Giannitsarou, 2006; Marcet & Nicolini, 2003; Williams, 2003). The adaptive learning is a nonparametric method which can learn how to do tasks based on data given for training or initial experience without any strict theoretical assumption.

In the following sections, the proposed model is discussed further (Section 2); the performance of the proposed model is evaluated (Section 3), findings and discussions are provided (Section 4), and the conclusions are given in last section.

2. Proposed model

From the reviewed literatures, there are two major drawbacks in the conventional time series models: (1) most statistical methods rely upon some assumptions about the variables used in the analysis, so it is limited to be applied to all datasets. And (2) most conventional time series models utilize only one variable in forecasting. However there are many noises that are caused by changes in market conditions and environments, therefore, financial analysts should consider many market variables in forecasting. For this reason above, forecasting models should utilize more variables to improve forecasting accuracy.

In order to reconcile the drawbacks above, this paper considers that the volatility of American stock indexes play an important role to effect the volatility of TAIEX. Because the forecasting models utilize the relation between the volatility of American stock index and the volatility of TAIEX, the analytical results would approximate to the real world. Based on the concept above, this paper proposes a new volatility model to forecast Taiwan stock index. Firstly, this paper calculates the volatility of NASDAQ stock index and Dow Jones stock index by Eqs. (1), (2). Then use the proposed forecasting equation (as Eq. (3)) to forecast Taiwan stock index, it considers multi-stock index (NASDAQ stock index and Dow Jones stock index and TAIEX (t)) to forecast TAIEX $(t+1)$. Secondly, to optimize the degree of influential between independent variables (the volatility of NASDAQ stock index and Dow Jones stock index, and the TAIEX (t)) and dependent variable (TAIEX $(t+1)$) by adaptive learning, which can overcome the limitations of statistical methods (data need obey some mathematical distribution). Then, the overall procedure of the proposed model is shown in Fig. 1.

For easy understanding, this section uses some numerical data as the example and step by step shows the core concept in proposed algorithm.

Step 1: Collect data set

Collect the TAIEX from TSEC (Taiwan Stock Exchange Corporation) as experimental datasets. In this section, we choose 2000-year TAIEX which contains 271 transaction days as an example to illustrate the proposed model. The data from January to October are used for training whereas those from November and December are used for testing.

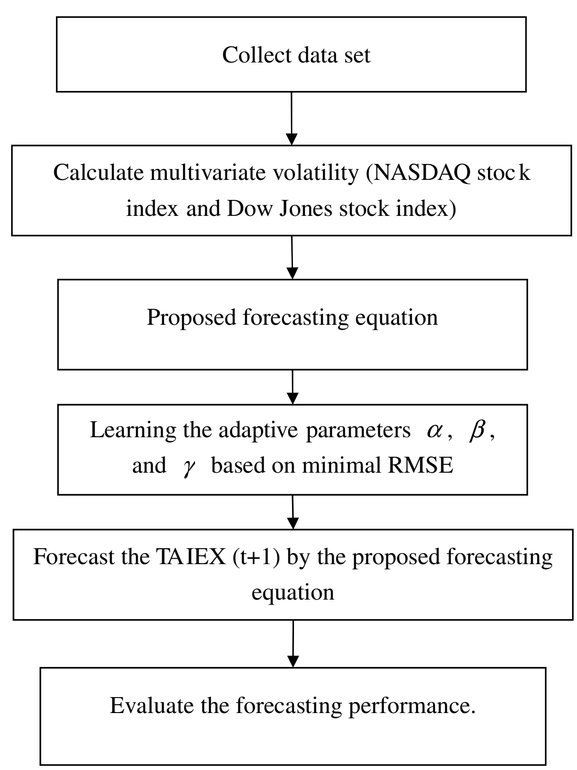

Fig. 1. Flowchart of the proposed model.

Step 2: Calculate multivariate volatility (NASDAQ stock index and Dow Jones stock index)

In this section, we apply two variables namely (1) the NASDAQ (N) and (2) the Dow Jones (D), and calculate the volatility of the two variables by Eqs. (1), (2). The differences in the variables NASDAQ and Dow Jones are listed in Table 1. In (Table 1), some data under the NASDAQ and Dow Jones are empty; because of there were no transactions on those days. For this reason, this paper fills in the last volatility as the differences.

$$\mathrm{diff}(N(t)) = N(t) - N(t-1) \tag{1}$$

$$\mathrm{diff}(D(t)) = D(t) - D(t-1) \tag{2}$$

Step 3: Proposed forecasting equation

Firstly, this paper proposes a new forecasting equation based on the best adaptive parameter (see Eq. (3)) for forecasting TAIEX ($T+1$) is defined as follows:

$$T(t+1) = \alpha \times T(t) + \beta \times \mathrm{diff}(N(t)) + \gamma \times \mathrm{diff}(D(t)) \tag{3}$$

where α, β, and γ denote adaptive parameters.

Secondly, let α denote the degree of influential for TAIEX between time $t+1$ and t, let β represent the degree of influential between TAIEX at time $t+1$ ($T(t+1)$) and the volatility of NASDAQ (diff($N(t)$)), and let γ represent the degree of influential between $T(t+1)$ and the volatility of Dow Jones (diff($D(t)$)). The range of coefficient values α is from 0.5 to 1.5 (because of the trading limits of TAIEX (±7%), then we expend the range of α for forecasting other stock markets), and the ranges of coefficient values (β,γ) are from -1 to 1 (-1 denotes perfect negative relation and +1 denotes perfect positive relation).

For general equation, we can rewrite Eq. (3) as Eq. (4):

Table 1
Differences in variables

Date	NASDAQ	diff($N(t)$)	Dow Jones	diff($D(t)$)
2000/1/3	4131.15		11357.51	
2000/1/4	3901.69	−229.46	10997.93	−359.58
2000/1/5	3877.54	−24.15	11122.65	124.72
2000/1/6	3727.13	−150.41	11253.26	130.61
2000/1/7	3882.62	155.49	11522.56	269.3
2000/1/8		155.49		269.3
2000/1/9		155.49		269.3
2000/1/10	4049.67	167.05	11572.2	49.64
2000/1/11	3921.19	−128.48	11511.08	−61.12
2000/1/12	3850.02	−71.17	11551.1	40.02
2000/1/13	3957.21	107.19	11582.43	31.33
2000/1/14	4064.27	107.06	11722.98	140.55
2000/1/15		107.06		140.55
2000/1/16		107.06		140.55
2000/1/17		107.06		140.55
2000/1/18	4130.81	66.54	11560.72	−162.26
2000/1/19	4151.29	20.48	11489.36	−71.36
2000/1/20	4189.51	38.22	11351.3	−138.06
2000/1/21	4235.4	45.89	11251.71	−99.59
2000/1/22		45.89		−99.59
2000/1/23		45.89		−99.59
2000/1/24	4096.08	−139.32	11008.17	−243.54
2000/1/25	4167.41	71.33	11029.89	21.72
2000/1/26	4069.91	−97.5	11032.99	3.1
2000/1/27	4039.56	−30.35	11028.02	−4.97
2000/1/28	3887.07	−152.49	10738.87	−289.15
2000/1/29		−152.49		−289.15
2000/1/30		−152.49		−289.15
2000/1/31	3940.35	53.28	10940.53	201.66

$$T(t+1) = \alpha \times T(t) + \beta \times \mathrm{diff}(N(t)) + \gamma \times \mathrm{diff}(D(t)) + \cdots + \zeta \times \mathrm{diff}(X(t)) \tag{4}$$

where α, β, γ and ζ denote adaptive parameters, diff($X(t)$) is the volatility of another stock index, ζ represents the degree of influential between $T(t+1)$ and the volatility of another stock index (diff($X(t)$)).

Step 4: Learning the adaptive parameters α, β, and γ based on minimal RMSE

Based on minimal RMSE (as Eq. (5)), adjust the values (α, β and γ) by the range for each parameter in the training dataset, each iterative increment with 0.001, then this step can get the best parameter set (α, β and γ), so as to make our forecasts more reliable.

$$\mathrm{RMSE} = \sqrt{\sum_{t=1}^{n} |\mathrm{actual}(t) - \mathrm{forecast}(t)|^2 / n} \tag{5}$$

where actual (t) denotes the real TAIEX value, forecast (t) denotes the predicting TAIEX value and n is the number of data.

Step 5: Forecast the TAIEX ($t+1$) by the proposed forecasting equation.

From step 4, the best parameters are determined when the forecasting performance reach best condition (minimal RMSE) in the training dataset. Then the proposed model uses the best parameters to forecast $T(t+1)$ for the target testing datasets by Eq. (3).

Step 6: Evaluate the forecasting performance.

Finally, calculate RMSE in testing dataset by Eq. (5), then the value of RMSE is taken as evaluation criterion to compare with different models.

3. Experiment and comparison

In order to verify and compare with the listing models, this paper collects TAIEX from 1997 to 2003 (7 sub-datasets) as experimental datasets. The sub-datasets for previous 10-month are used for training and those from November to December are selected for testing.

After experiments using the 7 sub-datasets of TAIEX, we generate 7 forecasting performances for all testing datasets and the best adaptive parameters in different sub-datasets are shown in Table 2. To evaluate whether the proposed model surpasses the conventional time series model, this paper compares the performance of the proposed model with AR (1) model. Furthermore, because of fuzzy time series have been applied to several domains with better performance. We also employ fuzzy time series models (Chen's, 1996; Huarng & Yu's, 2006) as comparison models. From the performance comparison table (see Table 3), we can see that the proposed model outperforms the performances of listing models in six testing period (exclude 2000). From the 2000-year experimental result, we can see that only the performance of AR (1) model (RMSE = 130) slightly surpasses the proposed model (RMSE = 134).

4. Findings and discussions

From verification in Section 4, we can see that the proposed model outperforms Chen's (1996) and Huarng and Yu's (2006) models. Furthermore, the performance of the proposed model is better or comparable to the conventional time series model (AR (1)). In this empirical analysis, there are three findings provided as follows:

Table 2
The best adaptive parameter in different sub-datasets (TAIEX)

Adaptive parameter	Year						
	1997	1998	1999	2000	2001	2002	2003
α	1.000	0.999	1.000	0.999	1.000	1.000	1.001
β	0.499	0.014	−0.036	0.260	0.421	0.499	0.312
γ	0.243	0.355	0.169	0.080	0.010	0.107	0.180

Table 3
The performance comparisons of different models (TAIEX)

Models	Year						
	1997	1998	1999	2000	2001	2002	2003
Chen's model (Chen, 1996)	154	134	120	176	148	101	74
Huarng's model (Huarng & Yu, 2006)	141	121	109	152	130	84	56
AR (1)	141	114	102	130[a]	115	66	54
Proposed model	135[a]	111[a]	101[a]	134	110[a]	64[a]	50[a]

[a] Performs best among four models.

(1) From Table 3, the proposed model outperforms the conventional statistical model (AR (1)). We can confirm that adaptive learning technique can automatically generate the optimal parameters (α, β, γ) to forecast TAIEX. The proposed model can overcome the limiting of AR model such as the assumption about the error $\varepsilon \sim N(0, 1)$.
(2) In this paper, the proposed model considers the volatility of multi-stock index (NASDAQ stock index and Dow Jones stock index) in the forecasting processes. In this empirical analysis (see Table 2), both of the influential values between TAIEX and the volatility of NASDAQ and the influential values between TAIEX and the volatility of Dow Jones (β, γ) are not equal to zero. Furthermore, we conduct statistic method to verify whether the volatility of Dow Jones and the NASDAQ affect Taiwan stock index. Using the data of Table 2, the statistic approaches, t-test is utilized to test the hypothesis: (a) H_0: $\beta = 0$ and (b) H_0: $\gamma = 0$. The results of t-test for hypothesis (a) and (b), which reject $\beta = 0$ and $\gamma = 0$ under 95% confidence interval (see Table 4).
(3) From Table 2, most of the influential values between TAIEX and the volatility of American stock (β, γ) are positive (exclude 1999), so we argue that the relation between TAIEX and the volatility of American stock is positive relation. For this reason above, we claim that NASDAQ and Dow Jones stock index would play an important role in TAIEX volatility. And Dickinson's paper (Dickinson, 2000) also shows that the stock price indexes of different countries would influence each other.

Additionally, the forecasting results may provide investors with important information for making investment decisions in stock markets. The forecasting index going upside or downside will guide the investors to buy or sell stocks in the future.

Table 4
One-sample t-test for (a) H_0:$\beta = 0$ and (b) H_0:$\gamma = 0$

Parameter	t	df	Sig. (2-tailed)	Mean difference	95% Confidence interval of the difference	
					Lower	Upper
β	3.398	6	0.015	0.28129	0.0787	0.4838
γ	3.816	6	0.009	0.16343	0.0586	0.2682

Note: t is the t statistics, df denotes degree of freedom.

5. Conclusions

The paper has proposed a model based on adaptive learning technique for forecasting the Taiwan stock index by considering the volatility of the NASDAQ and Dow Jones stock index. Two conclusions are given: (1) the proposed model can forecast stock market without any statistical assumptions; and (2) the results (see Table 3) show that the proposed model outperforms the listing models. In the future works, more stock markets (such as China and Japan) should be considered in forecasting processes and other artificial intelligence techniques can be utilized to adjust the linear parameters to exam forecasting performance.

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics, 31*, 307–327.
Box, G., & Jenkins, G. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
Chen, S. M. (1996). Forecasting enrollments based on fuzzy time-series. *Fuzzy Sets Systems, 81*, 311–319.
Chen, S. M., & Chung, N. Y. (2006). Forecasting enrollments using high-order fuzzy time series and genetic algorithms. *International Journal of Intelligent Systems, 21*, 485–501.
Dickinson, D. G. (2000). Stock market integration and macroeconomic fundamentals: An empirical analysis, 1980–95. *Applied Financial Economics, 10*(3), 261–276.
Enders, W. (2004). *Applied econometric time series*. Wiley.
Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimator of the variance of United Kingdom inflation. *Econometrica, 50*(4), 987–1008.
Giannitsarou, C. (2006). Supply-side reforms and learning dynamics. *Journal of Monetary Economics, 53*, 291–309.
Huarng, K., & Yu, H. K. (2006). The application of neural networks to forecast fuzzy time series. *Physica A, 363*, 481–491.
Huarng, K. H., Yu, H. K., & Hsu, Y. W. (2007). A multivariate heuristic model for fuzzy. *IEEE Transactions on Systems, Man, and Cybernetics-part B: Cybernetics, 37*(4), 836–846.
Kim, K., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for prediction of stock index. *Expert System with Application, 19*, 125–132.
Kimoto, T., Asakawa, K., Yoda, M., & Takeoka, M. (1990). Stock market prediction system with modular neural network. In *Proceedings of the international joint conference on neural networks, San Diego, California* (pp. 1–6).
Kmenta, J. (1986). *Elements of econometrics* (second ed.). New York: Macmillan Publishing Co.
Marcet, A., & Nicolini, J. P. (2003). Recurrent hyperinflations and learning. *American Economic Review, 93*, 1476–1498.
Morgan, J. P. (1996). Reuters. RiskMetrics – Technical Document, 4th ed., New York.

Nelson, D. B. (1991). Conditional heterosdasticity in asset returns: A new approach. *Econometrica, 59*(2), 347–370.

Nikolopoulos, C., & Fellrath, P. (1994). A hybrid expert system for investment advising. *Expert Systems, 11*(4), 245–250.

Pai, P. F., & Lin, C. S. (2004). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega, 33*, 497–505.

Roh, T. H. (2007). Forecasting the volatility of stock price index. *Expert Systems with Applications, 33*, 916–922.

Tseng, F. M., Tzeng, G. H., Yu, H. C., & Yuan, J. C. (2001). Fuzzy ARIMA model for forecasting the foreign exchange market. *Fuzzy Sets and Systems, 118*, 9–19.

Williams, N. (2003). *Adaptive learning and business cycles*. Mimeograph.