

Short-Term Forecasting on Technology Industry Stocks Return Indices by Swarm Intelligence and Time-Series Models

Tien-Wen Sung^{1,2}, Cian-Lin Tu³, Pei-Wei Tsai⁴, and Jui-Fang Chang³(✉)

¹ College of Information Sciences and Engineering, Fujian University of Technology, Fuzhou 350118, Fujian Province, China

tienwen.sung@gmail.com

² Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian University of Technology, Fuzhou 350118, Fujian Province, China

³ Department of International Business, National Kaohsiung University of Applied Sciences, Kaohsiung 807, Taiwan
1104346116@gm.kuas.edu.tw, rose@kuas.edu.tw

⁴ Department of Computer Science and Software Engineering, Swinburne University of Technology, Victoria 3122, Australia
ptsai@swin.edu.au

Abstract. Forecasting is an important technique in many industries and business fields for reading the terrain. The category of technology industry stock, which includes 7 independent stocks, in Taiwan Stock Exchange (TWSE) is selected to be the study subject in this paper. The goal is to forecast the return index of the individual stocks base on the information observed from the trading historical data of the subjects. By including the trading volume, the number of trading records, the opening price, and the closing price in the inputs to the representative models in time-series and computational intelligence: EGARCH(1,1) and the In-teractive Artificial Bee Colony (IABC), respectively, the forecasting accuracy are compared by the Mean Absolute Percentage Error (MAPE) value. The experimental results indicate that the IABC forecasting model with the selected input variables presents superior results than the EGARCH(1,1).

Keywords: IABC · EGARCH · Return index · Technical industry stock

1 Introduction

An accurate forecasting result is an important and crucial support for business and industry managers to make decisions. The demand for the forecasting lies in many different places such as the customers' needs, the product production capacity, the trading costs, and the expected net profit, so on so forth. The foreign exchange rate forecasting is one of them directly related to the net profit and the trading costs for the companies involving in the international trading business. In addition, the business circle is a series of fluctuated economic experiences

© Springer International Publishing AG 2018

J.-S. Pan et al. (eds.), *Advances in Intelligent Information Hiding and Multimedia Signal Processing*, Smart Innovation, Systems and Technologies 81,

DOI 10.1007/978-3-319-63856-0_34

composed of four circulating phases: the expansion, the peak, the contraction, and the trough. Motions of the circulation provide the impact to the trend of the long-term growth in Gross Domestic Product (GDP), which is closely related to the economic cycle. During the past two decades, globe stock markets undergo two significant episodes that can use economic cycle to spell out a rough sketch about the current globe stock markets: the first event takes place in the US that the dot-com bubble took globe stock markets a nose dive in 2001, and the other is subprime mortgage crisis culminated in the financial crisis of 2007 to 2008 that lasted over seven years' bear market over the world. In accordance with the laws of the business circle, globe stock markets took a jump in 2016 and the prediction on the next tendency becomes a crucial problem to be monitored. Investors can obtain the flow of stock market by the industrial sectors' indices which commonly assemble stocks into groups and are the primary form to convey the level of prices. The indices provide more encyclopaedic comparability and the more wide-spread angle of industrial performances we can grasp.

Base on the ground truth mentioned above, a swarm intelligence algorithm called Interactive Artificial Bee Colony (IABC) is utilised to compose the forecasting model. It produces the forecasting result of the return index by receiving inputs from the trading volume, the number of trading records, the opening price, and the closing price. The same inputs are used in a time series model called EGARCH to produce the baseline for compare. The rest of the paper is composed as follows: the literature review is given in Sect. 2, the experiment design is discussed in Sect. 3, the experimental results are given in Sect. 4, and the conclusion is given in Sect. 5.

2 Literature Review

Time-series Model is a typical technique used in the financial analysis [1]. It can bring data to light and conjecture that the reactions are taking place by analysing the mass data. Autoregressive Conditional Heteroscedasticity (ARCH) is brought up by Robert F. Engle (1982) [5] and that included Mean and Variance equations. Mean equations denote that the Efficient-market, the long-term arbitrage, and the speculation are unworkable. The variance equations denote that the risks are foreseeable and handleable. In ARCH model, the conditional variance will be adjusted with the time, which provides the ability of enhancing the characteristics of accuracy of the model. In recent years, ARCH models are expanded in a bunch of linked time-series models called ARCH family under many improvements and developments made by the scholars. For example, Tim Bollerslev (1986) proposes the Generalised Auto-regressive Conditional Heteroskedastic model (GARCH) [2] based on the ARCH [3] fundamental to simplify the setting of parameters that presumes the conditional variance is affected by both the error sum of squares and the conditional variance in the previous period. The ARCH and the GARCH models are both constructed base on the symmetric in fluctuation. Different from these models, French et al. (1986) [10] consider the same subject in a different angle of view. They point out that the

impacts of the markets brought by the good news and the bad news are asymmetric. The bad news always trigger more variations than the good news when the event appears. In 1991, Nelson presents the Exponential GARCH (EGARCH) [11] model to include the asymmetric fluctuation phenomenon in the consideration. It has been proofed that the EGARCH model has better explanatory power in the analysis.

Different from the time-series models, the swarm intelligence is a research field containing algorithms designed based on the observation and simulation of the collective behaviours in natural or artificial [12]. The Artificial Bee Colony Algorithm (ABC) algorithm [13], which is proposed by Karaboga (2005). A few years later, Tsai et al. (2009) embed the Newtonian law of Universal Gravitation in the ABC algorithm to reform the selection mechanism for broadening the development capability of onlooker bees, the trial probes the interactive affection of ABC algorithm with the Interactive Artificial Bee Colony (IABC). The IABC algorithms has been used in foreign exchange rate forecasting in 2014 and 2016 [8,9], renewable energy unit distribution [6] and pattern recognition [7]. And the outcomes obtained in the foreign exchange rate forecasting is satisfactory.

3 Experiment Design

In this work, the historical data of the trading volume, the number of trading records, the opening price, and the closing price in the past 30 days are collected for feeding into the EGARCH(1,1) and the IABC forecasting models, respectively. The experimental data includes those collected in January 10th in 2009 and December 31st in 2015. By the lubrication of the sliding window strategy, the forecasting models are capable to produce the outcomes for all targeting observational periods. According to the published classification in Taiwan Stock Exchange (TWSE), the category of technology industry stock contains seven independent stocks. The stocks in the technology industry category are the subject studied in this paper. The forecasting accuracy is measured via the Mean Absolute Percentage Error (MAPE) value. The MAPE tells the distance between the forecasted result and the actual return index in an objective way.

The experiment is composed of the steps listed as follows:

1. Fundamental statistical tests of the input data: The input data need to be pre-processed by the Ljung-Box Q test and the Augmented Dickey-Fuller (ADF) Unit Root Tests [4]. The details will be given in the following subsection.
2. Feeding the input data into both the EGARCH and the IABC forecasting models for training and testing.
3. Calculating the MAPE values corresponding to the forecasting results of both models.

3.1 Input Data for the Forecasting Models

The study subjects are the stocks in the technology industry category in TWSE market. This category includes seven stocks of computers and computer

peripheral equipment, the optoelectronics, the communication technology and internet, the electronic parts and components, the electronic product distribution, the information service, and the other electronics.

For every stock, the collected trading volume, the number of trading records, the opening price, and the closing price are examined to make sure they are with the expected statistical characteristics and distribution. In general, the stock price should fit in the normal distribution by logarithm as a log return. The rate of the return is calculated by Eq. (1):

$$p_r = \log \left(\frac{p_t}{p_{t-1}} \right) \tag{1}$$

where p_t refers to the price on day t , and p_r denotes the rate of the return.

If the outcome of Eq. (1) fits in the normal distribution, the original input data can be further processed by the Ljung-Box Q test for analyzing the residual correlation by Eq. (2):

$$Q = n(n + 2) \sum_{k=1}^h \frac{\hat{p}_k^2}{n - k} \tag{2}$$

where Q denotes the test result, n refers to the size of the sample, \hat{p}_k indicates the autocorrelation at lag k , and h is the test number of lags.

The ADF unit root test is applied after the Ljung-Box Q test for verifying the stationary of the series data by Eq. (3):

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \tag{3}$$

where α stands for a constant, β indicates the coefficient on a time trend, and p is the lag order of the Autoregressive (AR) process.

3.2 The Time-Series Model: EGARCH

In this study, the classical EGARCH(1,1) model is used in the experiments. The EGARCH model can be depicted as follows:

$$y_t = x_t b + \varepsilon_t \tag{4}$$

where y_t refers the function of exogenous variable x_t , $x_t b$ is the conditional mean of y_t , and ε_t denotes the residuals.

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0, h_t) \tag{5}$$

where Ω_{t-1} means all set of information to period $t - 1$ and h_t refers conditional variance of y_t .

$$\ln(h_t) = \alpha_0 + \sum_{i=1}^q \alpha_i \left\{ \gamma_i \frac{\varepsilon_{t-i}}{\sqrt{h_{t-1}}} + \vartheta_i \left[\frac{|\varepsilon_{t-i}|}{\sqrt{h_{t-1}}} - \sqrt{\frac{2}{\pi}} \right] \right\} + \sum_{j=1}^p \beta_j \ln(h_{t-j}),$$

subject to $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, \forall i, j$

(6)

where γ_i is the parameter of the asymmetric volatility effect, $\gamma_i \frac{\varepsilon_{t-i}}{\sqrt{h_{t-i}}}$ denotes the sign effect, ϑ_i indicates the parameter of scale of unexpected variation, and $\vartheta_i \left[\frac{|\varepsilon_{t-i}|}{\sqrt{h_{t-i}}} - \sqrt{\frac{2}{\pi}} \right]$ represents the magnitude effect.

3.3 The Interactive Artificial Bee Colony (IABC) Model

IABC is a branch of ABC algorithm, which involves the universal gravitation into the onlooker movement process in the conventional ABC algorithm. The IABC model can be depicted as follows:

1. Initialization: Randomly spread n_e percent of the population into the solution space, where n_e refers the ratio of employed bees to the total population.
2. Move the onlookers: Calculating the probability of selecting a food source by Eq. (7) and choose G of them as the reference by the roulette wheel scheme for every on-looker bees.

$$P_i = \frac{F(\Theta_i)}{\sum_{k=1}^S F\Theta_k} \tag{7}$$

where P_i denotes the probability of selecting the i^{th} employed bee, S stands for the total number of the employed bees, $F(\Theta_k)$ and $F(\Theta_i)$ represent the fitness value of employed bees k and i , respectively. Move the onlookers by Eq. (8) based on the selected G reference employed bees.

$$x_i(t+1) = \Theta_i + \sum_{k=1}^G \{ \tilde{F}_{ik} \cdot [\Theta_i(t) - \Theta_k(t)] \} \tag{8}$$

where x_i is the coordinate of the i^{th} onlooker bee, t indicates the iteration number, Θ_i and Θ_j denote the coordinates of the i^{th} employed bee and the randomly chosen employed bee, respectively, G is the number of the selected employed bees, and \tilde{F}_{ik} refers to the normalized universal gravitation.

3. Move the scouts: The employed bee will become the scout and be moved again if none of the onlooker bees selected it as a reference after a predefined *Limit* iterations. Nevertheless, there are two randomly selected employee bees can be exempted from the rule. The rest of the employee bees, which matches the condition to become the scouts will be moved by Eq. (9):

$$\Theta_{ij} = \Theta_{jmin} + r(\Theta_{jmax} - \Theta_{jmin}) \tag{9}$$

where Θ_{jmax} denotes the maximum value over all dimensions of Θ_i , Θ_{jmin} indicates the minimum value appears in all dimensions of Θ_i , and r is a random variable in the range of $[0, 1]$.

4. Update the near best solution: Keep the best fitness value and the corresponding co-ordinates of the bee.
5. Termination checking: As the termination conditions are satisfied, terminate the algorithm and output the kept near best solution; otherwise, go back to Step 2 and repeat the processes.

In the IABC forecasting model, the input data is merged from four variables into two variables by Eqs. (10) and (11):

$$TP = \frac{T_V}{V_R} \quad (10)$$

where TP denotes the trading power, T_V is the trading volume, and T_R is the number of the trading records.

$$PD = \frac{P_O}{P_C} \quad (11)$$

where PD stands for the price difference of a trading day, P_O and P_C represents the opening price and the closing price of a trading day, respectively.

In the training phase, 30-day historical data (denoted by $Data_{(t-31:t-1)}$) is feed into the IABC model for training the corresponding weighting mask (denoted by W). The desired output should be as much close to the actual return index on day $t - 1$. The trained W is later used in the testing phase with the input data $Data_{(t-30:t)}$ to produce the forecasting return index for day t .

3.4 Calculation of the MAPE Value

The MAPE is calculated by Eq. (12):

$$MAPE = \frac{1}{m} \sum_{t=1}^m \frac{|\hat{S}_t - S_t|}{S_t} \times 100\% \quad (12)$$

where \hat{S}_t and S_t stand for the prediction and the actual value, respectively, and m is the total number of the data.

4 Experiments and Experimental Results

The experimental results include seven independent stocks in the technology industry category in 2009 to 2015. Totally, 1731 observations are included in the data. The experimental results are presented base on the stocks in Fig. 1.

According to the experimental results obtained with 7 independent stocks, we can find that the IABC forecasting model produces much smaller MAPE values than the EGARCH(1,1) model. Nevertheless, it is still suffered from trapping in the local optimum, sometimes, and produces a significant jump of the MAPE value.

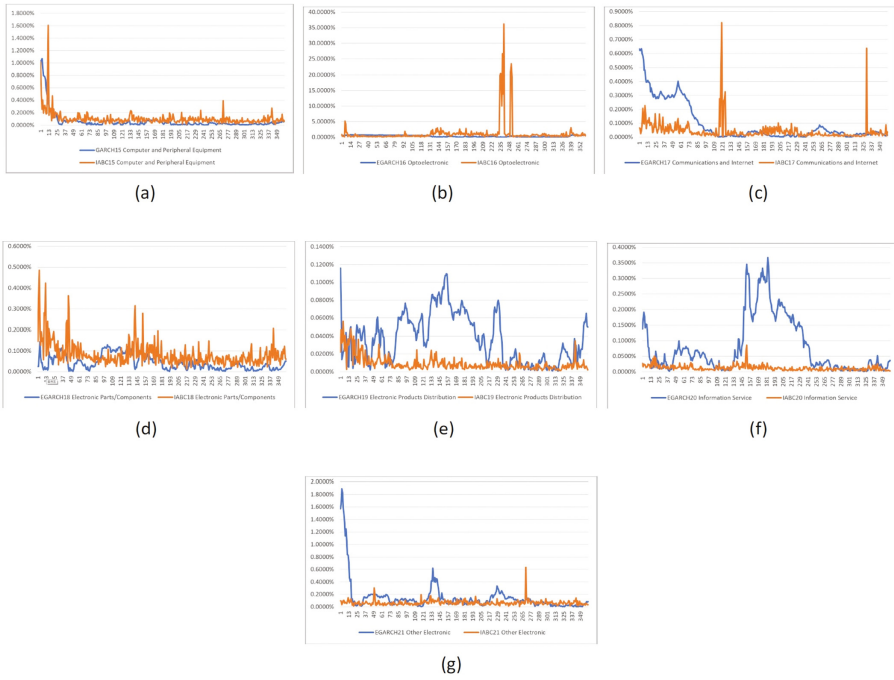


Fig. 1. MAPE values of all stocks in 2009–2015.

5 Conclusions and Future Works

The goal of this work is to forecast the return index of the individual stocks base on the information observed from the trading historical data of the subjects. The category of technology industry stock, which includes 7 independent stocks, in Taiwan Stock Exchange (TWSE) is selected to be the study subject. The EARCH(1,1) and the IABC forecasting models are utilized to generate the forecasting results. The forecasting accuracy is measured by the MAPE value. The experimental results indicate that the IABC forecasting model presents superior results than the EGARCH(1,1). In the future work, we will focus on increasing the stability of the IABC forecasting model to reduce the unsteadiness of the outcome. The forecasting models and the forecasting return index provides the shareholders additional references for operating the investments.

Acknowledgement. This work is funded by the Key Project of Fujian Provincial Education Bureau (JA15323).

References

1. Chen, S.-S.: Applied Time-series Econometrics for Macroeconomics and Finance. Tunghua, Taipei (2013)
2. Bollerslev, T.R.: Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**(3), 307–327 (1986)
3. Bollerslev, T.R.: ARCH modeling in finance: A review of the theory and empirical evidence. *J. Econometrics* **52**(1–2), 5–59 (1992)
4. Dickey, D.A.: Likelihood ratio statistic for autoregression time series with a unit root. *Econometrica* **49**, 1057–1072 (1981)
5. Engle, R.F.: Autoregressive conditional heteroskedasticity with estimates of the variance of united kindom inflation. *Econometrica* **50**(4), 987–1007 (1982)
6. Abedinia, O., Barazandeh, E.S.: Interactive artificial bee colony based on distribution planning with renewable energy units. In: 2013 IEEE PES Innovative Smart Grid Technologies (ISGT), pp. 1–6. IEEE Press, Washington, DC (2013)
7. Tsai, P.-W., Pan, J.-S., Liao, B.-Y., Chu, S.-C.: Enhanced artificial bee colony optimization. *J. Innov. Comput. Inf. Control* **5**(12B), 5081–5092 (2009)
8. Tsai, C.-F., Hsiao, C.-T., Tsai, P.-W., Chang, J.-F.: Applying interactive artificial bee colony algorithm and the time-series methods in the foreign exchange rate forecasting. *Econ. Manage.* **36**(Z1), 147–150 (2014)
9. Tsai, P.-W., Liu, C.-H., Zhang, J., Chang, J.-F.: Structuring interactive artificial bee colony forecasting model in foreign exchange rate forecasting with consumer confidence index and conventional microeconomics factors. *ICIC Expr. Lett. Part B: Appl.* **7**(4), 895–902 (2016)
10. French, K.R., Roll, R.: Stock return variance: the arrival of information and the reaction of traders. *J. Financ. Econ.* **17**(1), 5–26 (1986)
11. Nelson, D.B.: Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**(2), 347–370 (1991)
12. Beni, G., Wang, J.: Swarm intelligence in cellular robotic systems. In: Dario, P., Sandini, G., Aebischer, P. (eds.) *Robots and Biological Systems: Towards a New Bionics?*, 1993. NATO ASI Series (Series F: Computer and Systems Sciences), vol. 102, pp. 703–712. Springer, Berlin, Heidelberg (1993)
13. Karaboga, D.: An Idea Based on Honey Bee Swarm for Numerical Optimization. In: Technical Report-TR06, 2005 (2005)