Modern Economy, 2020, 11, 1329-1341
https://www.scirp.org/journal/me
ISSN Online: 2152-7261
ISSN Print: 2152-7245

The Forecasting Efficiency of Monthly Stock Indices between Macroeconomic Factors and Technical Indicators by Using Augmented Genetic Algorithm and Artificial Neural Network Model

Yeong-Jia Goo[1], Chia-Chen Chen[1,2*]

[1]Department of Business Administration, National Taipei University, Taiwan
[2]Department of Business Administration, China University of Technology, Taiwan
Email: goo@mail.ntpu.edu.tw, *chiachen@gm.cute.edu.tw

How to cite this paper: Goo, Y.-J., & Chen, C.-C. (2020). The Forecasting Efficiency of Monthly Stock Indices between Macroeconomic Factors and Technical Indicators by Using Augmented Genetic Algorithm and Artificial Neural Network Model. *Modern Economy, 11,* 1329-1341.
https://doi.org/10.4236/me.2020.117094

Received: June 8, 2020
Accepted: July 18, 2020
Published: July 21, 2020

Abstract

The purpose of this study is to compare the forecasting efficiency of stock indices between macroeconomics and technical analysis by using augmented Genetic Algorithm and Artificial Neural Network model. Monthly data of Taiwan stock index, electronic index, and financial index, from Jan. 2001 to Dec. 2019 are collected. Eight influential macroeconomic factors and seven commonly watched technical indicators are used as determinants. Three models are adopted for comparison. The models include the ARMA(p, q) model as the benchmark, GA_ANN with macroeconomic factors, and GA_ANN with technical indicators. The sliding window method with 24-, 30-, 36-, 42- and 48-month training base periods is simulated. Linear unit root tests of ADF, PP, and KPSS, and nonlinear unit root test of KSS are examined. Internal validity index of hit ratio and external validity indices of MAPE, HR, ARV and Theil U coefficients are compared. The empirical findings are summarized as follows. 1) The overall forecasting performance between MACRO and TECH models shows little difference. The electronic and financial stock indices have the out-of-sample hit ratios of 77.78% and 68.89%, respectively. Thus, these two stock indices may be suitable for making meaningful investment decisions. 2) The best training base observed from the market stock index is between 30 to 48 months. The best base observed from the electronic stock index is between 42 to 48 months. The best base observed from the financial stock index is between 42 to 48 months. Thus, the training base from 42 to 48 months exhibits better forecasting performance. 3) The optimal transformation parameter under ANN may range from 0.50

to 0.99 and may not be a constant parameter.

Keywords

The Forecasting Efficiency of Stock Indices, Genetic Algorithm Model, Artificial Neural Network Model, The ARMA Model

1. Introduction

Stock index forecasting has been empirically investigated over the past decades. The importance of stock index forecasting in making speculation, hedge, and arbitrage investment decisions is addressed by many practitioners, financial engineers, and academic researchers. Due to the stochastic and much like a random walk phenomenon nature of stock index movement, the task of making efficient forecast is challenging and requires innovative thinking in investment theory, model settings, and variable selection.

The stock market behavior is a typical financial time-series process which involves issues such as stationarity, serial correlation, heteroscedasticity, nonlinearity, and causality. While ARIMA models could be used to build a stock market index forecasting model, the results are usually unsatisfactory (Khandelwal et al., 2015; Ariyo et al., 2014; Zhang, 2003). Many researchers had tried to use traditional econometric model with macroeconomic variables in forecasting the stock returns, but the forecasting power is limited (Laichena & Obwogi, 2015; Ouma & Muriu, 2014, Flannery & Protopapadakis, 2002). Some of researchers utilized the technical indicators in forecasting the stock returns (Paluch & Jackowska-Strumiłło, 2018; Paluch & Jackowska-Strumiłło, 2012; Sutheebanjard & Premchaiswadi, 2010; Tilakaratne, Morris, Mammadov & Hurst, 2007).

On the other hand, dramatic development in statistical and heuristic computing algorithms such as genetic algorithm (GA) and artificial neural networks (ANN) have been seen in the past decades. The improvement of mathematical optimization capability for handling complicated, dynamic, and nonlinear functional forms with multivariate dataset could help researchers enhance the construction data classification, financial forecasting, and risk management models.

The genetic algorithm (GA) uses the biological evolutionary rule for finding optimal number of variables and weighting schemes. Specifically, the optimal final outcomes can be found by using reproduction, crossover, and mutation procedure with a fitness function and a certain amount of iterative generations. Past literatures have disclosed the application of the GA techniques for forecasting stock price (Armano, Marchesi, & Murru, 2005; Kim & Han, 2000; Kai & Wenhua, 1997). The artificial neural networks (ANN) imitate the bio-neural processing system with hidden layers and hidden units for finding better solutions. Specifically, the ANN model can be used in making a forecasting model by searching optimal hidden layers, hidden units, transformation, and learning coefficient. Past literatures have disclosed the application of the ANN techniques for forecasting stock price (Nayak, Misra & Behera, 2017; Kwon & Moon, 2007; Chen, Leung, & Daouk, 2003).

According to past literatures, past researches had focused on many issues regarding stock index forecasting. However, this study intends to re-examine some issues which may not have been addressed in the past studies. First, the GA and ANN models are integrated in such a way that allows GA method to randomly select proper sets of variables through crossover and mutation, the ANN methodology is applied in each simulation to find optimal simulated parameters, and a forecast for one-period ahead stock index is made. Second, randomly selected transforming and learning rates in both hidden layers and final outcome stages are simulated. Third, the stock index forecasting efficiency between macroeconomic factors and technical indicators are compared. Fourthly, the focus is placed on the monthly stock index rather than the daily stock index.

The rest of the paper is organized as follows: Section 2 discusses data and methodology; Section 3 provides the empirical results; and Section 4 summarizes the discussion and concludes the paper.

2. Data and Methodology

2.1. Data Description

Monthly data of Taiwan stock index, electronic index and financial index from Jan. 2001 to Dec. 2019 are collected as dependent variables. Eight influential macroeconomic factors and seven commonly watched technical indicators are used as independent variables. The total number of months is 228. All of the dependent and independent variables are lagged $t-1$ thru lagged $t-6$. Thus, there are 54 and 48 predetermined variables for maroeconomic and technical analysis data set, respectively.

The stock index return (RET) is computed as the natural log of (Price/lagged_Price). The eight macroeconomic variables are as follows: (Kvainickas & Stankevičienė, 2019; Laichena & Obwogi, 2015; Ouma & Muriu, 2014)

1) GDP: the growth rate of gross national product.

2) M1B: the government defined M1B money supply.

3) BOND: the monthly 10-year Long-term government bonds.

4) UMR: the monthly Unemployment rate.

5) Wage: the average monthly salary of manufacturing industry.

6) IPI: the industrial production index.

7) CPI = the monthly consumer price index.

8) WPI = the monthly wholesale price index.

The seven technical indicators are as follows: (Paluch & Jackowska-Strumiłło, 2018; Paluch & Jackowska-Strumiłło, 2012; Sutheebanjard & Premchaiswadi, 2010; Tilakaratne, Morris, Mammadov, & Hurst, 2007)

1) MA5: the 5-month moving average.

2) MA10: the 10-month moving average.

3) MA20: the 20-month moving average.

4) OSC: the Oscillator indicator, i.e., DIF − MACD.

Where,

$$\text{DIF} = \text{EMA12} - \text{EMA26};$$

$$\text{MACD} = \text{EMA9};$$

$$\text{EMA12}_t = (2 \times P_t + 11 \times \text{EMA12}_{t-1})/13$$

5) BIAS5: the 5-month BIAS, i.e. PRICE/MA5.

6) BIAS10: the 10-month BIAS, i.e. PRICE/MA10.

7) BIAS20: the 20-month BIAS, i.e. PRICE/MA20.

2.2. Methodology

2.2.1. Linear and Nonlinear Unit Root Tests

Financial time series often exhibit trending behavior or non-stationarity in the mean. The study conducts the linear unit root tests of the three stock index series by applying the augmented Dickey-Fuller (ADF) test (Dickey & Fuller, 1979; Dickey and Fuller, 1981), the Phillips-Perron (PP) test (Phillips & Perron, 1988), the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski, Phillips, Schmidt, & Shin, 1992), as well as the nonlinear Kapetanios-Shin-Snell (KSS) test (Kapetanios, Shin, & Snell, 2003). The ADF test's regression includes lags of the first differences of Y_t, and the corresponding three models are expressed in the following equations:

$$\Delta Y_t = \phi Y_{t-1} + \sum_{i=1}^{k} \beta_i \Delta Y_{t-i} + \varepsilon_t \tag{1}$$

$$\Delta Y_t = \alpha + \phi Y_{t-1} + \sum_{i=1}^{k} \beta_i \Delta Y_{t-i} + \varepsilon_t \tag{2}$$

$$\Delta Y_t = \alpha + \phi Y_{t-1} + \lambda t + \sum_{i=1}^{k} \beta_i \Delta Y_{t-i} + \varepsilon_t \tag{3}$$

where t is the time index, α is an intercept constant called a drift, λ is the coefficient on a time trend, ϕ is the coefficient presenting the process root, i.e., the focus of testing, k is the lag order of the first-differences autoregressive process, and ε_t is an independent identically distributed residual term.

Model (1) is a pure random walk with the lag terms. Model (2) possesses a drift. Model (3) includes a drift and a time trend. The null hypothesis for the ADF test is: $H_0: \phi = 0$, with the alternative $H_1: -2 < \phi < 0$. The ADF t-test statistic is $\hat{\phi} \big/ \text{se}\left(\hat{\phi}\right)$.

The PP test differs from the ADF test mainly in how PP test deals with serial correlation and heteroscedasticity in the error term. The PP test does not require the specification of the form of the serial correlation of ΔY_t under the null, nor the errors ε_t be conditionally homoscedastic. The ADF and PP unit root tests are for the null hypothesis that a time series Y_t is integrated of order one, $I(1)$. On the other hand, the KPSS unit root test is for the null that Y_t is integrated of order zero, $I(0)$. In addition, the KSS test is applied since the above linear unit root tests may suffer from important power distortions in the presence of nonlinearities in the data generating process.

2.2.2. The ARMA(p, q) Model as the Benchmark

In this study, the ARMA(p, q) model is used as the benchmark model. The stationarity of the returns series is checked using the unit root tests. The estimation of the ARMA models for three stock index returns includes the checking of appropriate ARMA(p, q) orders, the sliding window of the training sample, and one-month ahead forecasting.

2.2.3. Development of Augmented GA_ANN (AGA_ANN) Model

The traditional genetic algorithm estimation procedure includes Initialization, reproduction, genetic operations (including crossover and mutation), heuristics, and termination. As shown in Figure 1, the ANN model consists of three stages, i.e. input, hidden layer, and output. The components of ANN includes neurons, connections and weights, propagation function, ANN parameters (including learning rate, the number of hidden layers and batch size), weights adjustment, backpropagation, and self-learning.

The rationale of the newly proposed augmented GA_ANN (namely, AGA_ANN) model is to adopt the advantages of GA and ANN so as to improve the forecasting accuracy. The transformation functions from the input node, the hidden layer node, to the output node are as follows: (the λ_h and λ_o are transformation parameters.)

$$H_j = \frac{1}{1+\mathrm{e}^{-\left(\lambda_h \sum W_{ji} X_i\right)}}$$

$$\hat{Y} = \frac{1}{1+\mathrm{e}^{-\left(\lambda_o \sum W_j H_j\right)}}$$

where

H_j is the j^{th} hidden unit; $\hat{Y}$ is the forecasted output; X_i is the input variable. W_{ji} is the weight of input variable; W_j is the weight of hidden unit.

The detailed AGA_ANN estimation procedure is as follows:

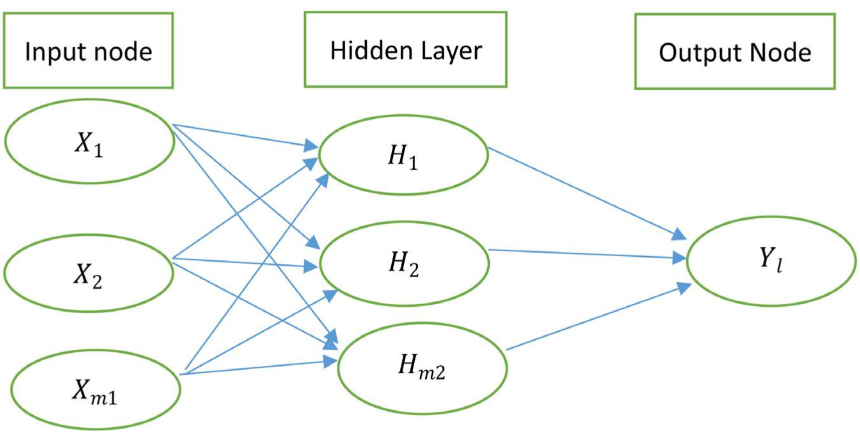

Figure 1. The AGA_ANN modle.

1) Variables transformation

a) Dependent variables

To improve simulated performance, the three stock index returns series are transformed by using the following logistic function. The transformed series (Y_1) is then converted into 0 or 1 series (Y).

$$Y_1 = 1/\left(1+\exp\left(-\mathrm{RET}\right)\right)$$

Y is one when Y_1 is greater than or equal to 0.5; otherwise Y is zero.

b) Independent variables

The independent variables are standardized with mean equal to zero and standard deviation equal to one. The transformed series is then logisticalized to within zero and one.

2) The sliding window span parameters

In this study, the sliding window spans are simulated by 24-, 30-, 36-, 42-, and 48-months as the training base. The base data is then used for simulating the AGA_ANN model. The best simulated parameters are then adopted for making the one-month ahead forecast. Then the sliding window moves one period ahead and performs next AGA_ANN model until the end of observations.

3) The initialization of W_{ji} and W_{lj} parameters

The coefficient weights of W_{ji} and W_{lj} are randomly and uniformly simulated having values within zero and one.

4) The selection of simulated IV and hidden units

In this study, the number of simulated independent variables (M) ranges from 6 to NVAR/2. The NVAR is the total number of predetermined variables. For each simulation, 100 sets of random selection are made. The number of hidden units (J) ranges from $M/2$ to M.

5) The GA procedure

By using the core ANN estimation, the hit ratios of the 100 sets are ranked. The top 10 sets are kept. The variables in the middle 80 sets are switched according to crossover method. The worst 10 sets are wiped off and additional new 10 sets are created. Thus, the newly created 100 sets are used for the next run.

6) The randomization of transformation and learning parameters

In this study, the transformation and learning Parameters are uniformly simulated from 0.5 to 1.0. For each simulation, 10 sets of random selection are made.

7) The one-month ahead forecast

For each simulation, the best simulated parameters are used to make a one-month ahead forecast until the end of observation.

8) The computation of performance indices

In this study, the proposed four performance indices are as follows:

a) MAPE

The forecasted value Y is converted into a forecasted stock index $\hat{P}_t$. The equation of the mean absolute percentage error (MAPE) is listed below: (P_t is the actual stock index at time t)

$$\text{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\frac{\left|P_t - \hat{P}_t\right|}{P_t} * 100\%$$

b) HR

The equation of the hit ratio (HR) is listed below:

$$\text{HR} = \frac{\sum_{t=1}^{N}\text{HIT}_t}{N} * 100\%$$

where $\text{HIT}_t = 1$ if RET × PRET > 0; $\text{HIT}_t = 0$ otherwise.

c) ARV

The equation of the average relative variance (ARV) is listed below:

$$\text{ARV} = \frac{\sum_{t=1}^{N}\left(\hat{P}_t - P_t\right)^2}{\sum_{t=1}^{N}\left(\hat{P}_t - \bar{P}\right)^2}$$

where $\bar{P}$ is the monthly average stock index.

d) Theil's U

The equation of the Theil's U is listed below: (The U2 measure)

$$\text{Theil's U} = \frac{\sum_{t=2}^{N}\left(\left(\hat{P}_t - P_t\right)/P_{t-1}\right)^2}{\sum_{t=2}^{N}\left(\left(P_t - P_{t-1}\right)/P_{t-1}\right)^2}$$

3. Empirical Results

3.1. Descriptive Statistics

The descriptive statistics is shown in Table 1. There are three subjects under study, namely, market, electronic, and financial stock indices. Monthly data is listed from Jan. 2000 to Dec. 2019. A total of 240 months of data are used for each subject. Seven technical indicators and eight macroeconomic variables are listed. In order to create the lagged values of predetermined variables including the lagged dependent and independent variables, year 2000 is used as the extra year for creating lagged values. The actual simulation starts from Jan. 2001.

3.2. The Results of Linear and Nonlinear Unit Root Tests

A nonstationary time series might lead to spurious regression. Linear unit root tests of the ADF, PP, and KPSS, and nonlinear KSS unit root tests are conducted for the MKT, ELEC, and FINA returns. Tables 2-4 show the results and conclude that all three series are stationary statistically. Notice that an insignificant t value of KPSS test verifies the series is stationary.

3.3. The Simulated Parameters of the Three Models

Using the SAS-IML and FARMAFIT functional call, the estimation and sliding window simulation of ARMA(p, q) model reveals that AR(p) = 3 and MA(q) = 2 throughout entire simulation process.

In Table 5, the simulated parameters of the technical indicators (TECH) and macroeconomic factors (MACRO) shows that the total number of forecasted

Table 1. Descriptive statistics of the variable.

Variable	Label	N	Mean	Std Dev	Min	Max
IND		720	-	-	1	3
YM		720	-	-	200,001	201,912
PRICE	PRICE	720	2993.5	3526.36	165.72	11,997.14
RET (%)	RET	720	0.05	3.02	−11.86	11.73
$X1$	MA5	720	2982.31	3500.06	178.842	11,258.63
$X2$	MA10	720	2970.49	3471.48	191.462	10,995.51
$X3$	MA20	720	2944.79	3416.27	217.9205	388.2042
$X4$	OSC	720	4.667216	93.22655	−535.8124	388.204
$X5$	BIAS5	720	0.371465	8.077604	−33.47615	39.70125
$X6$	BIAS10	720	0.826026	11.96482	−44.46172	36.44888
$X7$	BIAS20	720	1.67106	15.53288	−48.9913	65.5809
$M1$	GDP%	240	0.002807	0.006384	−0.025674	0.019054
$M2$	M1B%	240	8.028167	6.595321	−6.51	30.51
$M3$	BOND	240	2.0045	1.205906	0.65	6.06
$M4$	UMR	240	4.247917	0.675399	2.73	6.13
$M5$	WAGE	240	44370.11	11281.05	34294	95165
$M6$	IPI	240	80.86992	19.93818	42.17	117.44
$M7$	CPI	240	93.43458	6.031846	84.19	103.02
$M8$	WPI	240	101.7391	10.82556	75.81	124.84

Note: IND = 1 for Market; IND = 2 for ELEC; IND = 3 for FINA.

Table 2. Unit root test results for the MKT returns.

	Linear test			Nonlinear test
Lags	ADF t-Stat	PP Adj. t-Stat	KPSS Adj. t-Stat	KSS t-Stat
5	−7.8239***	−14.1032***	0.0186	−2.6931***
10	−5.3143***	−14.0209***	0.0254	0.3998
20	−4.6522***	−14.4616***	0.0414	2.3383**

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 3. Unit root test results for the ELEC returns.

	Linear test			Nonlinear test
Lags	ADF t-Stat	PP Adj. t-Stat	KPSS Adj. t-Stat	KSS t-Stat
5	−7.7278***	−13.5935***	0.0308	−3.3413***
10	−4.9719***	−13.5009***	0.0412	0.4726
20	−4.6071***	−13.9192***	0.0670	1.5211

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 4. Unit root test results for the FINA returns.

Lags	Linear test ADF t-Stat	Linear test PP Adj. t-Stat	Linear test KPSS Adj. t-Stat	Nonlinear test KSS t-Stat
5	−7.2758***	−16.3478***	0.0260	−5.01902***
10	−6.2166***	−16.5218***	0.0344	−2.08615**
20	−3.9557***	−17.5037***	0.0555	0.08891

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 5. The simulated parameters for TECH and MACRO.

ITEM	N	Technical Indicators				Macroeconomic Factors			
		Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
IND	2880	-	-	1	3	-	-	1	3
YM	2880	-	-	200,301	201,912	-	-	200,301	201,912
SDATE	2880	-	-	25	228	-	-	25	228
M	2880	-	-	6	24	-	-	6	24
J	2880	-	-	3	24	-	-	3	24
BASE	2880	-	-	24	48	-	-	24	48
HR	2880	71.49%	5.85%	52.78%	89.58%	71.62%	5.43%	52.78%	90.00%
LAMH	2880	0.8113	0.1331	0.5001	0.9999	0.8052	0.1339	0.5011	0.9999
LAMO	2880	0.7989	0.142	0.5000	0.9999	0.7894	0.1418	0.5004	0.9999
ETAH	2880	0.7603	0.1463	0.5003	0.9999	0.7552	0.1447	0.5002	0.9997
ETAO	2880	0.7322	0.1445	0.5001	0.9999	0.7286	0.1458	0.5001	0.9996
PY	2880	0.5537	0.2437	0.0007	0.9997	0.5497	0.2356	0.0024	0.9996

Note: N = 3 sectors in 5-base forecasted series of 204, 198, 192, 186, 180; IND = 1 for Market; IND = 2 for ELEC; IND = 3 for FINA. YM is the year-month; SDATE is the date of simulated series; M = # of Indep. Var; J = # of hidden units; HR is the training sample's hit ratio; LAMH and LAMO are the transformation coefficients for hidden and output transformation; ETAH and ETAO are the learning rates for the hidden and output weights; PY is the predicted Y.

observations is 2880. The predetermined variable (M) ranges from 6 to 24. The number of hidden units range from 3 to 24. The mean value of training sample's hit ratios for TECH and MACRO are 71.49% and 71.62%, respectively. The mean values of transformation parameters LAMH and LAMO for TECH and MACRO are (0.8113, 0.7989) and (0.8052, 0.7894), respectively.

3.4. The Performance Comparison of the Three Models

In Table 6 and Table 7, the results of the forecasting performances of the three proposed models are as follows:

1) The TECH model has the best overall MAPE. The MACRO model has the best overall HR and ARV. The ARMA model has the best THEIL's U.

2) In terms of the market stock index, the ARMA model has the best MAPE.

Table 6. The MAPE and HR performance measures.

IND	BASE	MAPE (%)			HR (%)		
		ARMA	TECH	MACRO	ARMA	TECH	MACRO
MKT	24	4.00	4.19	4.19	50.98	51.47	50.98
MKT	30	3.83	4.05	4.08	48.48	51.52	56.06
MKT	36	3.73	4.02	4.25	53.13	52.08	49.48
MKT	42	3.71	4.04	4.13	51.61	50.00	49.46
MKT	48	3.70	3.98	4.06	52.22	51.11	55.56
ELEC	24	4.54	4.24	4.22	51.96	64.22	63.24
ELEC	30	4.40	4.10	4.04	47.98	66.67	69.70
ELEC	36	4.37	3.95	3.92	51.56	69.79	74.48
ELEC	42	4.39	4.00	3.89	51.61	71.51	73.12
ELEC	48	4.36	3.85	3.90	51.67	77.78	73.89
FINA	24	4.97	4.66	4.59	51.47	62.25	63.24
FINA	30	4.55	4.39	4.49	52.02	65.66	65.66
FINA	36	4.48	4.29	4.29	52.60	63.02	65.10
FINA	42	4.42	4.21	4.44	53.76	67.20	61.29
FINA	48	4.50	4.19	4.13	46.11	67.78	68.89
AVG		4.2628	4.1436	4.1755	51.1452	62.1368	62.6755

Table 7. The ARV and THEIL_U Performance measures.

IND	BASE	ARV			THEIL U		
		ARMA	TECH	MACRO	ARMA	TECH	MACRO
MKT	24	0.0505	0.0490	0.0503	1.0459	1.0136	1.0362
MKT	30	0.0537	0.0572	0.0569	0.8016	1.0485	1.0455
MKT	36	0.0558	0.0593	0.0621	0.8340	1.0337	1.0492
MKT	42	0.0572	0.0611	0.0641	0.7862	1.0488	1.0534
MKT	48	0.0618	0.0716	0.0670	0.8130	1.0795	1.0653
ELEC	24	0.0569	0.0533	0.0504	0.6466	0.9762	0.9668
ELEC	30	0.0619	0.0498	0.0541	0.8304	0.9249	0.9543
ELEC	36	0.0624	0.0516	0.0528	0.8477	0.9232	0.9275
ELEC	42	0.0632	0.0519	0.0485	0.7908	0.9088	0.8924
ELEC	48	0.0679	0.0524	0.0539	0.8010	0.8914	0.9058
FINA	24	0.2438	0.1050	0.1060	1.2783	0.9881	0.9816
FINA	30	0.1110	0.1041	0.1010	0.8897	0.9779	0.9739
FINA	36	0.1134	0.1010	0.0990	0.9542	0.9587	0.9530
FINA	42	0.1077	0.0915	0.0913	0.9302	0.9375	0.9326
FINA	48	0.1081	0.0906	0.0912	0.9693	0.9359	0.9326
AVG		0.0850	0.0700	0.0699	0.8812	0.9765	0.9781

The MACRO model has the best HR. The TECH model has the best ARV and THEIL_U.

3) In terms of the electronic stock index, the TECH model has the best MAPE and HR. The MACRO model has the best ARV. The ARMA model has the best THEIL_U.

4) In terms of the financial stock index, the MACRO model has the best MAPE and HR. The TECH model has the best ARV. The ARMA model has the best THEIL_U.

5) In terms of the training base in MAPE and HR, the best base observed from the market stock index shows is between 30 to 48 months. The best base observed from the electronic stock index is between 42 to 48 months. The best base observed from the financial stock index is between 42 to 48 months. Thus, the training base from 42 to 48 months exhibits better forecasting performance.

In sum, previous study shows that daily stock index forecast is quite satisfactory. However, the monthly stock index forecasts tell the story otherwise, which indicates monthly data forecast might be even more difficult than that of daily data. The overall forecasting performance between TECH and MACRO models show little difference. The electronic and financial stock indices have the out-of-sample hit ratios of 77.78% and 68.89%, respectively. Thus, these two stock indices might be suitable for making meaningful investment decisions.

4. Conclusion and Discussion

The study attempted to compare the forecasting efficiency of Stock Indices between macroeconomic factors and technical indicators by using augmented GA and ANN Models. Three models are proposed including the ARMA model as the benchmark, GA_ANN with macroeconomic factors (MACRO), and GA_ANN with technical indicators (TECH). The empirical findings are summarized as follows:

1) The overall forecasting performance between MACRO and TECH models shows little difference. The electronic and financial stock indices have the out-of-sample hit ratios of 77.78% and 68.89%, respectively. Thus, these two stock indices may be suitable for making meaningful investment decisions.

2) The best training base observed from the market stock index is between 30 to 48 months. The best base observed from the electronic stock index is between 42 to 48 months. The best base observed from the financial stock index is between 42 to 48 months. Thus, the training base from 42 to 48 months exhibits better forecasting performance.

3) The optimal transformation parameters under ANN may range from 0.50 to 0.99 and may not be a constant parameter.

Due to the complexity of the augmented GA_ANN model, tremendous computing time and efforts are involved. The study found that monthly stock index forecasts may be more challenging than daily data. Further theoretical and empirical works are needed. Specifically, previous researches have adopted many different types of models, variables, and data frequency. All aspects require extensive and prudent investigations.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock Price Prediction Using the ARIMA Model. In *UKSim-AMSS* 16*th International Conference on Computer Modelling and Simulation* (pp. 105-111).

Armano, G., Marchesi, M., & Murru, A. (2005). A Hybrid Genetic-Neural Architecture for Stock Indices Forecasting. *Information Science, 170,* 3-33. https://doi.org/10.1016/j.ins.2003.03.023

Chen, A., Leung, M., & Daouk, H. (2003). Application of Neural Networks to an Emerging Financial Market: Forecasting and Trading the Taiwan Stock Index. *Computers & Operations Research, 30,* 901-923. https://doi.org/10.1016/S0305-0548(02)00037-0

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association, 74,* 427-431. https://doi.org/10.1080/01621459.1979.10482531

Dickey, D. A., & Fuller, W. A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica, 49,* 1057-1072. https://doi.org/10.2307/1912517

Flannery, M. J., & Protopapadakis, A. A. (2002). Macroeconomic Factors Do Influence Aggregate Stock Returns. *The Review of Financial Studies, 15,* 1-32. https://doi.org/10.1093/rfs/15.3.751

Kai, F., & Wenhua, X. (1997). Training Neural Network with Genetic Algorithms for Forecasting the Stock Price Index. In *Proceedings of the 1997 IEEE International Conference on Intelligent Processing Systems* (pp. 401-403).

Kapetanios, G., Shin, Y., & Snell, A. (2003). Testing for a Unit Root in the Nonlinear STAR Framework. *Journal of Economics, 112,* 359-379. https://doi.org/10.1016/S0304-4076(02)00202-6

Khandelwal, I., Adhikari, R., & Verma, G. (2015). Time Series Forecasting Using Hybrid ARIMA and ANN Models Based on DWT Decomposition. *Procedia Computer Science, 48,* 173-179. https://doi.org/10.1016/j.procs.2015.04.167

Kim, K. J., & Han, I. (2000). Genetic Algorithms Approach to Feature Discretization in Artificial Neural Networks for the Prediction of Stock Price Index. *Expert Systems with Applications, 19,* 125-132. https://doi.org/10.1016/S0957-4174(00)00027-0

Kvainickas, T. S., & Stankevičienė, J. (2019). Regional Limitations of Stock Indices Prediction Models Based on Macroeconomic Variables. *Economics and Culture, 16,* 1-16. https://doi.org/10.2478/jec-2019-0018

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root? *Journal of Economics, 54,* 159-178. https://doi.org/10.1016/0304-4076(92)90104-Y

Kwon, Y. K., & Moon, B. R. (2007). A Hybrid Neuro Genetic Approach for Stock Forecasting. *IEEE Transactions on Neural Networks, 18,* 851-864. https://doi.org/10.1109/TNN.2007.891629

Laichena, K. E., & Obwogi, T. N. (2015). Effects of Macroeconomic Variables on Stock Returns in the East African Community Stock Exchange Market. *International Journal of Education and Research, 3,* 305-320.

Ouma, W. N., & Muriu, P. (2014). The Impact of Macroeconomic Variables on Stock Market Returns in Kenya. *International Journal of Business and Commerce, 3,* 1-31.

Paluch, M., & Jackowska-Strumiłło, L. (2012). Prediction of Closing Prices on the Stock Exchange with the Use of Artificial Neural Networks. *Image Processing & Communications, 17,* 275-282. https://doi.org/10.2478/v10248-012-0056-5

Paluch, M., & Jackowska-Strumiłło, L. (2018). Hybrid Models Combining Technical and Fractal Analysis with ANN for Short-Term Prediction of Close Values on the Warsaw Stock Exchange. *Applied Sciences, 8,* 2473. https://doi.org/10.3390/app8122473

Phillips, P. C., & Perron, P. (1988). Testing for a Unit Root in Time Series Regression. *Biometrika, 75,* 335-346. https://doi.org/10.1093/biomet/75.2.335

Sutheebanjard, P., & Premchaiswadi, W. (2010). Stock Exchange of Thailand Index Prediction Using Back Propagation Neural Networks. In *Proceedings of the Second International Conference on Computer and Network Technology* (pp. 377-380). https://doi.org/10.1109/ICCNT.2010.21

Tilakaratne, C. D., Morris, S. A., Mammadov, M. A., & Hurst, C. P. (2007). Predicting Stock Market Index Trading Signals Using Neural Networks. In *Proceedings of the 14th Annual Global Finance Conference* (pp. 171-179).

Zhang, G. P. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing, 50,* 159-175. https://doi.org/10.1016/S0925-2312(01)00702-0