https://cibg.org.au/ P-ISSN: 2204-1990; E-ISSN: 1323-6903

FORECASTING USING NEURAL NETWORKS AND STOCHASTIC MODELS ON DAY OF THE WEEK EFFECT: A CASE STUDY OF KSE 100 INDEX

Azhar Ali Marri[1,2] and Mir G.H.Talpur[1]

[1]Department of Statistics, University of Sindh Jamshoro Pakistan

[2]Department of Statistics, University of Balochistan Quetta Pakistan

Abstract

The main objective of this research was to determine the day of the week effect in Karachi stock exchange (KSE) 100 Index. The problems of financial market structure is analyzed and forecasted by many statistical models. For example Auto regressive integrated moving average (ARIMA) model and artificial neural network (ANN) models. In this research day of the week effect was investigated Wednesday found significant and Monday was noted not significant. 15 step ahead prediction of Wednesday was observed through ARIMA model and ANN models. The coefficient of correlation for actual and forecasted values was perceived 0.8871 by ARIMA model and 0.924 by ANN model. Power of accuracy was displayed 88.765% by ARIMA model and 97.18% by ANN model.

Keywords; Anomaly, Wednesday, KSE 100 Index, ARIMA, ANN

1. Introduction

The time series is chronologically observed historical data including large in size high dimensionality and essential to update continuously. The cumulative usage of time series observations are initiated with a great deal of investigation and improvement to endeavor the time series study (Kumar and Murugan 2013). Time series modelling and prediction is a significant area of research where past recorded values of the same variable are collected and investigated (Zhang 2003). The prediction of the future event based on present and past observable events (Yao and Tan 2001). Numerious studies are reported in the literature for stock exchange prediction and it is still an active part of the study (Adebiyi, et al. 2012). Day of the week effect has attracted considerable attention since its discovery back in 1930 (Gharaibeh and Hammadi 2013). An adequate number of studies are conducted on day of the week effect. Ko, Li and Erickson (1997) reported that stock prices fluctuates with day of the week. Aly, Mehdian and Perry (2004) evaluated variation in daily stock exchange prices. Rossi and Gunardi (2018) examined calander anomaly (CA) and found recurring anomalies in stock market. The stock exchange prices prediction acquired attention of many researchers for private and institutional sectors. Selvan and Arun (2012) reported stock exchange prices are highly irregular with time and generally follows nonlinear pattern.

Adebiyi, Adewumi and Ayo (2014) forecasted ANN and ARIMA model using New York stock exchange (NYSE) daily closing price index. Emin (2007) forecasted daily and seasonal data of IMKB 100 Index with neural network models. Saiful and Yoshiki (2011) examined the applications of ANN model for forecasting of mudharabah time credit return. Olatunji, et al. (2011) predicted Saudi stock market with ANN model and observed lowest forecasting error with coefficient of correlation up to 99.9%.

https://cibg.org.au/ P-ISSN: 2204-1990; E-ISSN: 1323-6903

The ARIMA model prediction for time series data is crucial with uncertainty. Adebiyi, Adewumi and Ayo (2014); Falinouss (2007); Setty, Rangaswamy and Subramanya (2010) observed ARIMA model depends on the historical time series data as well as the pervious error terms. Wijaya, S. and Napitupulu (2010) proposed that the ARIMA model is a statistical technique of prediction which follows certain rules such as autocorrelation and linearity. Wijaya, S. and Napitupulu (2010) found ARIMA model is suitable for making short-term prediction. Falinouss (2007) reported ARIMA model is relatively more robust and effective than the other complex models. Wijaya, S. and Napitupulu (2010) noted ARIMA model based on the design of current and past price changes.

In this study day of the week effect was used to predict the KSE 100 Index by using ARIMA and ANN models. Stock exchange of Pakistan is a developing stock exchange its information structure is not robust (Shagufta and Siddiqui 2018 ; Farooq and Muhammad 2009). KSE is the biggest and most liquid stock exchange of Pakistan which returns the national economy of the country (Haroon 2012). KSE established on 18 September 1947 and incorporated at 10th March 1949. At present KSE has the four indexes, (i). KSE 100 Index, (ii). KSE all share index, (iii). KSE-30 share index and (iv). KMI-30 Index. Therefore current study was conducted on KSE 100 Index for forecasting. The performance of the ARIMA and ANN model is also inspected to reveal and conform the contradoctory reports of ARIMA model.

The rest of the paper is organized as: literature review is presnted in section 2. Methodology and data are displayed in section 3. Section 4 represents performance measurement and emprical analysis is given in section 5. While useful conclusion of the study is given in section 6.

2. Methodology

Stock prices are inherently noisy and non-stationary (Ju-Jie, et al. 2012). To convert the daily closing price index in to stationary series the following equation will be used.

$$R_t = \log(P_t/P_{t-1}) * 100 \tag{1}$$

R_t is the daily return in percentage of KSE 100 Index for day P_t , is the current closing price index and P_{t-1} is the previous closing price index.

2.1 Day of the week effect

The study investigated the day of the week effect with the equation (1)

$$R_t = \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \beta_5 D_5 + \varepsilon_t \tag{2}$$

Where R_t is defined as dependent variable the stock returns, D_1 through D_6 are the dummy variables such that if t is Monday, then D_1=1 and all other trading days D_1=0, If it is Tuesday then D_2 = 0 and for all other trading days D_2 = 0 and onward, ε_t is the random error term and $\beta_1$through β_5 are the variable coefficients and to be estimated. If the KSE 100 Index reveals Monday effect then (i) β_1 likely to become negative and significant, (ii) Significantly rest of the week returns should be greater than the Monday return.

2.2 ARIMA model

ARIMA model (Box and Jenkins 1970) is the most prevalent time series approach as a traditional model can be described as, the ARIMA model assumes the future values of a variable to be a linear function of numerous historical observations and plus a random error term (Ju-Jie, et al. 2012). The linear function constructed on three parametric components. Auto- regression AR, integration I, and moving average (MA) (Box and Jenkins 1970) and can be written as ARIMA (p,d,q), where p is the number of auto-regressive terms, d is the no. of non-seasonal differences and q is the number of lagged forecast errors in the prediction equation. An example of ARIMA (p,q) stochastic model depends on past values for auto regressive part (p), moving average part (q) and a random error term ε_t then the complete auto regressive moving average model can be written as

$$Y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots\cdots\cdots + \phi_p y_{t-p} + \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

Determining p, d and q in an ARIMA model is an important term and typically several times repeated until a final suitable model is selected on the basis of smallest values of Akaike information criteria (AIC), Bayesian information criteria (BIC) and largest value of Log likelihood values and the minimum value of mean absolute error (MAE) and mean absolute percentage error (MAPE).

2.3 ANN model

Data selection and pre-processing are critical phases of any time series modelling exertion in order to generalize the new projecting model (Adebiyi, et al. 2012). The data set was scaled to a range of (0, 1) for normalization procedure of minimum – maximum and used in the ANN model with the equation given below

$$x_{ti} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (4)$$

Where x_i is the actual stock exchange data, x_{ti} scaled data input value of actual stock exchange value x_i, x_{max} and x_{min} are unscaled actual stock exchange data maximum and manimum values. The forecasted values of the neural network model were changed in range (0,1) to actual values with the following equation.

$$x_i = x_{ti}(x_{max} - x_{min}) + x_{min} \quad (5)$$

Artificial neural network (ANN) is a part of machine learning methodology and attempt to simulate the learning method of human brain. Its function mimics biological neurons in which the construction of ANN contains a group of artificial neurons which are linked with the network. Collected literature have revealed that ANN is better than traditional statistical models in constructing forecast for the non-linear time series observations. Most of the neural network models topology are involved in neuron collection and constructions for two or more layers. Our study has combined some neurons into multi-layer structures to have the power of pattern recognition and prediction. For this purpose most commonly used type of neural network currently in use multi-layer feed-forward neural network model that is a three layer (one hidden layer) multilayer perceptron feed-forward neural network model and trained with back propagation algorithm was employed in this study. This model composed of input layer, hidden layer and output layer. ANN model mathematically can be written as

$$y_t = f(x,\theta) + \varepsilon \tag{6}$$

Where x is the explanatory variable, θ is the weight vector or parameter and ε is the random error term.

$$y = f\left[w_0 + \sum_{i=1}^{p} w_i . g(\lambda_j + \sum_{j=1}^{q} x_i w_{ij})\right] \tag{7}$$

Where y is the output vector, p number of hidden units, q is the number input units, w_0 is the output bias, λ_i hidden unit biases, w_{ij} weight from input unit i to hidden unit j, w_j weight from hidden unit, g hidden layer activation function. The activation function used in this network is the logistic function.

$$g(x) = \frac{\exp(x)}{1+\exp(x)} \tag{8}$$

The ANN model (4) plots non-linear functional form of the past observations $(y_{t-1}, y_{t-2}, y_{t-3}, \cdots\cdots\cdots, y_{t-n})$ to the predicted output y_t i.e.

$$y_t = f(y_{t-1}, y_{t-2}, y_{t-3}, \cdots\cdots\cdots, y_{t-n}, \theta) + \varepsilon \tag{9}$$

Practically simple neural network model (2) are unexpectedly powerful and is able to estimate the arbitrary function as the number of hidden nodes p is sufficiently large (Kurt, Maxwell and Halbert 1990). The structure of simple neural network model with the small number of hidden nodes repeatedly works better in out of the sample prediction (Zhang 2003). The over fitting effect usually observed in development of neural network model. An over fitted model develops better fit to the used sample but unfortunately has poor generalization ability for out of the sample data (Zhang 2003). The selection of input units q is dependent and no any systematic rule for deciding these units. Five lagged values are used as input data from estimated day of the week effect Wednesday in this study.

3. Performance measurement

We used following three methods to evaluate the performance of the study

i. Graphical comparison of the actual and predicted values

ii. Comparison of the statistical parameters such as correlation coefficient r_{xy}, MAE, MAPE and normalized mean squared error (NMSE). Actually NMSE and MAE are the parameters used to evaluate how close the prediction results are to the eventual outcomes MAE, and NMSE parameter are calculated as follows

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\text{Actual } V_i - \text{Predicted } V_i| \tag{10}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{\text{Actual } V_i - \text{Predicted } V_i}{\text{Actual } V_i}\right| \tag{11}$$

$$NMSE = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{(\text{Actual } V_i - \text{Predicted } V_i)^2}{\sigma^2 \text{Actual } V_i}\right) \tag{12}$$

iii. By examining accuracy of the predicted model for the out of sample data with equation (13)

$$\text{Accuray power} = 100\% - (\text{Prediction Error})\% \tag{13}$$

4. Results and discussion

The estimation of anomaly for the day of the week in KSE 100 Index was designed by using secondary data through ARIMA and ANN models by Microsoft excel and R.

The secondary data of KSE 100 Index for daily closing price index from 01-01-1990 to 31-12-2019 was obtained from yahoo finance. This study found Wednesday third trading day of the week significant. Wednesday from each week of the daily closing price index separated from the whole data. In total 1456 observations of Wednesday were collected in which 1441 observations were kept for training sample and last 15 samples for testing sample.

4.2 Days of the week effect in KSE 100 Index

The linear regression model used for the day of the week effect purpose. The descriptive statistics for residuals of the linear regression model are illustrated in table 1. First trading day Monday found negative and not significant while Wednesday third trading day was valued significant (Table 1). Coefficient of correlation between trading days and returns were observed. Monday found with negative correlation and Wednesday with positive correlation. All other trading days with poor correlation (Table 1). Similar findings were also reported by Gharaibeh and Hammadi (2013); Berument and Kiymaz (2001) ; Aly, Mehdian and Perry (2004).

Table 1 Day of the week effect analysis

Descriptive statistics of residuals & linear regression model						
Min	Q1	Median	Q3	Max		
-5.6948	-0.2762	0.0076	0.2953	5.4715		
Linear regression model						
Days	Estimates	Std. Error	t-value	Pr (>|t|)		
Monday	-0.03276	0.03361	-0.975	0.3298		
Tuesday	0.05200	0.03364	1.546	0.1222		
Wednesday	0.08196	0.03361	2.438	0.0148 *		
Thursday	0.04491	0.03397	1.322	0.1862		
Friday	0.05029	0.03489	1.441	0.1496		
Correlation between daily returns and days						
Coefficients	Estimated returns	Monday	Tuesday	Wednesday	Thursday	Friday
Returns	1					
Monday	-0.056	1				
Tuesday	0.0127	-0.251	1			
Wednesday	0.0370	-0.2524	-0.2512	1		
Thursday	0.0067	-0.2391	-0.2379	-0.2392	1	
Friday	0.0097	-0.2119	-0.2109	-0.2120	-0.2008	1

Significant at level: 0.01 ‘*’

4.3 ARIMA model

The normality of the returns series was checked with Q-Q plot (Figure 1). The returns series obtained stationary from logarithmic first difference of the closing price index. The correlogram of the stationary series shown mean is constant (Figure 2a). The descriptive statistics of the stationary series is reported in Table 2. The kurtosis was noted more than 3 for the stationary series. Augmented dickey Fuller test was used to conform the returns series is stationary (Table 2). The different ARIMA orders were evaluated. ARIMA order (2, 1, 3) with drift was found appropriate with the criteria of smallest value of AIC, BIC, and maximum value of Log-likelihood function. The performance of different ARIMA orders were checked with the smallest values of MAE and MAPE. Box-Ljung test conformed selected ARIMA order is suitable for prediction (Table 3). The data was predicted with the selected ARIMA order (2, 1, 3) for 15 step ahead (Figure 3a). The coefficient of correlation was observed 88.71% for out of sample data (table 4) and shown in figure 3b. The power of accuracy for the out of sample data was estimated 88.765% (Table 4). The findings of the current study are equivalent to Adebiyi and Adewumi (2014); Prapanna, Labani and Saptarsi (2014).

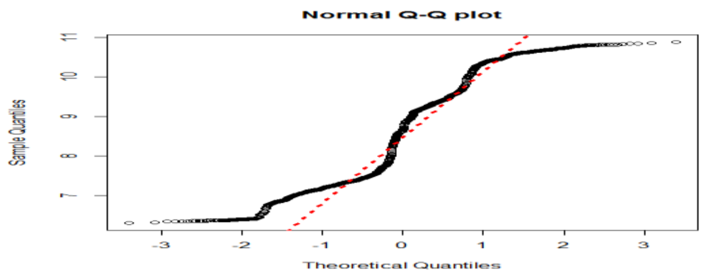

Figure 1 Normal Q-Q Plot

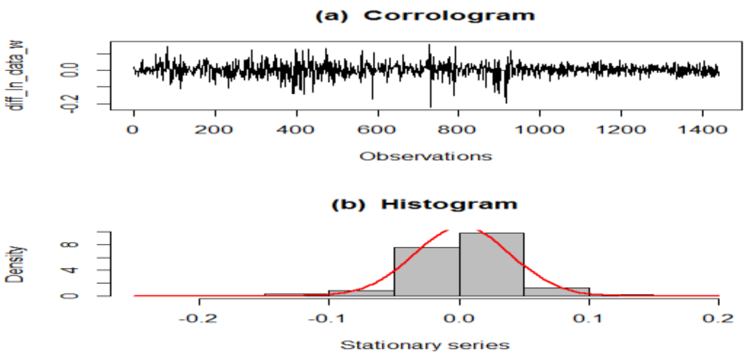

Figure 2 (a) Corrologram of stationary series, (b) Histogram of stationary series

Table 2 Descriptive statistics, Augmented Ducky-Fuller test & Selected ARIMA order

Descriptive Series	Mean	S.D	Skewness	Kurtosis
Stationary series	0.00275	0.0368	-0.582	4.0168

https://cibg.org.au/ P-ISSN: 2204-1990; E-ISSN: 1323-6903

Augmented Ducky-Fuller test						
Series	Estimate	Lag order		Augmented Dickey-Fuller		p-value
Stationary series		11		-11.558		0.01
ARIMA order (2,1,3)						
Coefficients	ar_1	ar_2	ma_1	ma_2	ma_3	drift
Values	0.5253	-0.5135	-0.4133	0.5663	0.0774	0.0027
s.e	0.2146	0.1451	0.2151	0.1250	0.0411	0.0012

Table 3 selection criteria of ARIMA order (2, 1, 3) & Box-Ljung test

Log likelihood	AIC	BIC	MAE	MAPE
2736.9	-5459.81	-5422.9	0.0256	0.3135
Box-Ljung test				
df	5	10	15	20
x-squared	0.8642	2.2085	11.416	16.571
p-value	0.9728	0.9945	0.7226	0.6806

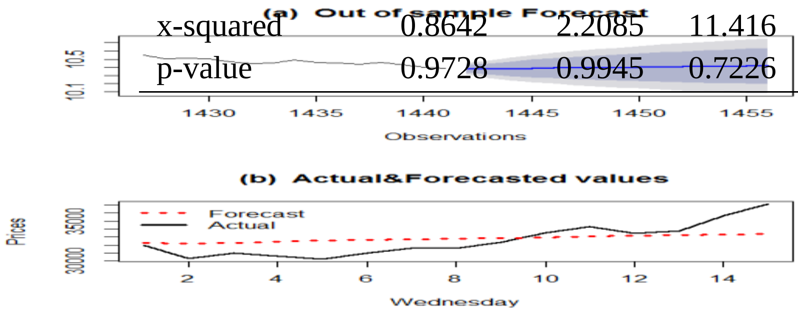

Figure 3 (a) Out of sample forecasted values, (b) Actual and Forecasted values

Table 4 Out of sample performance & correlation

Performance measurement			Correlation			Power of accuracy
			Correlation	Actual	Forecast	
MAE	MAPE	NMSE	Actual	1.0000	0.8871	88.765%
1385.82	0.0424	0.6561	Forecasted	0.8871	1.0000	

4.4 ANN model

In ANN model building 5 lags were used from the estimation of the day of the week effect in KSE 100 Index. The designed ANN model derived 11671 steps and produced 0.060921 errors as shown in figure 4. The next 15 steps ahead were predicted (Figure 5a). The performance of the derived ANN model for out of sample data between the actual and forecasted values were estimated with MAE, MAPE and NMSE reported in Table 5. The coefficient of determination R^2 was noted 98% shown in (Figure 5b). The coefficient of correlation for out of the sample data was found 92.4% (Table 5). Power of Accuracy was noted 97.18% (Table 5). The observations of the current study was established similar results with previous studies such as Ju-Jie, et al. (2012); Saiful and Yoshiki (2011); Mayankkumar and Sunil (2014); Kumar and Murugan (2013).

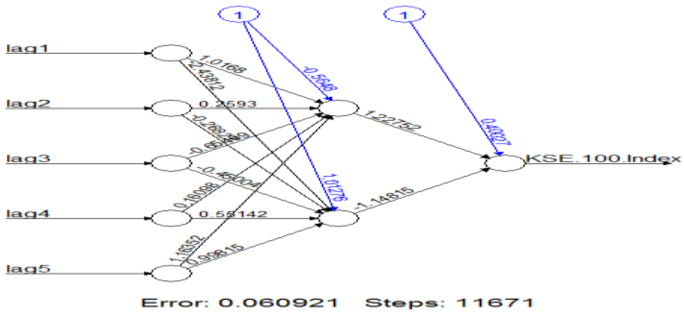

Figure 4 neural network model with lags

Table 5 Out of sample performance & correlation

performance measurement			Correlation			Power of accuracy
			Correlation	Actual	Forecast	
MAE	MAPE	NMSE	Actual	1.000	0.924	97.18%
908.848	0.0277	0.2701	Forecasted	0.924	1.000	

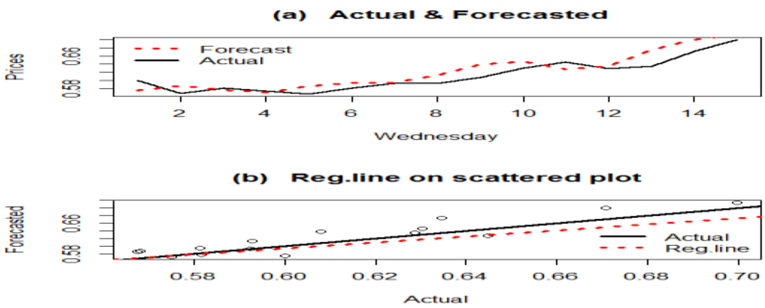

Figure 5 (a) Actual and forecasted prices (b) Forecasted regression line on scattered plot

4.5 Comparison of ANN and ARIMA model

The comparative study revealed that two models performed better (Figure 6). ANN model displayed quite low errors as compared to ARIMA model (Table 6). The study observed out of sample forecasting pattern of ARIMA model is directional which shows linear pattern and ANN model is towards value prediction (Figure 6). Similar results were also reported by Pieleanu (2016) ; Adebiyi, Adewumi and Ayo (2014) ARIMA model prediction for non-linear data.

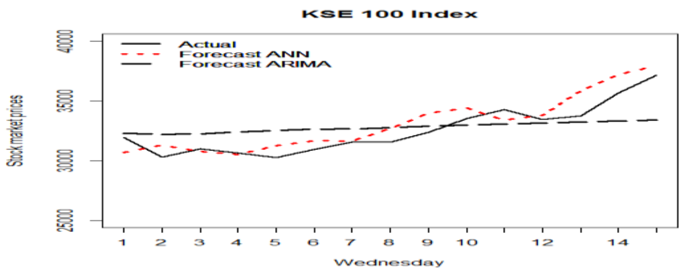

Figure 6 Actual and Forecasted values

Table 6 Out of sample error measurement

Model Error	ANN	ARIMA
MAE	908.848	1385.82
MAPE	0.0277	0.0424
NMSE	0.2701	0.6561

5. Conclusion

This study concluded that Wednesday the third trading day of week was significant for day of the week. 15 steps ahead Wednesday were predicted with quite low errors. The coefficient of correlation was observed 0.8871 by ARIMA model and 0.924 by ANN model. The power of accuracy was found 88.765% and 97.18% by ARIMA and ANN models respectively. The comparative study of two model determined ANN model has performed better than the ARIMA model with smallest prediction error.

References

Adebiyi, A. A., C. K. Ayo, M. O. Adebiyi, and S. O. Otokiti. 2012. "An improved stock price prediction using hybrid market indicators." *African Journal of Computing & ICT* 5 (5): 124-135.

Adebiyi, A. Ayodele, and Aderemi O. Adewumi. 2014. "Stock price prediction using the ARIMA model." *2014 UKSim-AMSS 16th International conference on computer modelling and simulation* . 105-111.

Adebiyi, Ayodele Ariyo, Aderemi Oluyinka Adewumi, and Charles Korede Ayo . 2014. "Comparision of ARIMA and Artifical neural networks models for stock price prediction." *Journal of applied mathematics* (Hindawi Publishing Corporation) 2014: 1-7. http://dx.doi.org/10.1155/2014/614342.

Aly, Hassan , Seyed Mehdian , and Mark J. Perry. 2004. "An analysis of the day -of-the-week effects in the egyptian stock market." *International Journal of Business* 9 (3): 301-308.

Berument, Hakan, and Halil Kiymaz. 2001. "The Day of the Week Effect on stock Market Volatility." *Journal of Economic and Finance* 25 (2): 181-193.

Box, G., and G. Jenkins. 1970. *Time series analysis ,Forecasting and Control.* San Francisco,CA:Holden-Day.

Emin AVCI. 2007. "Forecasting daily and seasonal returns of the ISE-100 Index with neural network models." *Dogus University Dergisi* 8 (2): 128-142.

Falinouss, Pegah. 2007. *Stock trend prediction using news articles: a text mining approach.*

Farooq Rasheed, and Muhammad Arshad. November 14, 2009. "The significance of financial literacy." *Proceedings 2nd CBRC, Lahore, Pakistan.*

Gharaibeh, Ahmad M.O., and Fatima Ismail Hammadi. 2013. "The Day of the Week Anomaly in Bahrain's Stock Market." *International Management Review* 9 (2): 60-68.

Haroon, Mohammad Arshad. 2012. "Testing the week form efficiency of Karachi Stock Exchange." *Pakistan Journal of Commerece and Social Sciences* 6 (2): 297-307.

Ju-Jie Wang , Jian-Zhou Whang, Zhe-George Zhang, and Shu-Po Guo. 2012. "Stock index forecasting based on a hybrid model." *Omega* 40: 758-766.

Ko, Wang, Yuming Li, and John Erickson. 1997. "A new look at the Monday effect." *The Journal of Finance* LII (5): 2171-2186.

Kumar , D. Ashok, and S. Murugan. 2013. "Performance Analysis of Indian Stock Market Index using Neural Network Time Series Model." *Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME) .* 21-22.

Kurt, Hornik, Stinchcombe Maxwell, and White Halbert. 1990. "Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks." *Neural networks* 3 (5): 551-560.

Mayankkumar, B Patel, and R Yalamalle Sunil. 2014. "Stock price prediction using artifical neural network." *International journal of innovative research in science and technology* 6 (3): 13755-13762.

Olatunji, Sunday Olusanya, Mohammad Saad Al-Ahmadi , Moustafa Elashafei, and Yaser Ahmed Fallatah. 2011. "Saudi Arabia Stock Prices Forecasting Using Artifical Neural Networks." *In Fourth International Conference on the Applications of Digital Information and Web Technologies.* 81-86, August.

Pieleanu, Florin Dan. 2016. "Predicting the evolution of bet index, using an ARIMA model." *Journal of information systems & operations management* 10 (1): 151-162.

Prapanna, Mondal, Shit Labani, and Goswami Saptarsi. 2014. "Study of effectiveness of time series modelling (ARIMA) in forecasting stock prices." *International journal of computer science,Engineering and applications* 4 (2): 13-29.

Rossi, Matteo, and Ardi Gunardi. 2018. "Efficient market hypothesis and stock market anomalies : Emprical evidence in four european countries." *The Journal of applied business research* 34 (1): 183-192.

Saiful Anwar, and Yoshiki Mikami. 2011. "Comparing accuracy performance of ANN, MLR, and GARCH model in predicting time deposit return of Islamuc Bank." *International journal of trade ,economic and finance* 2 (1): 44-50.

Selvan Simon, and Arun Raoot. 2012. "Accuracy driven artifical neural networks in stock market prediction." *International Journal on soft Computing* 3 (2).

Setty, D. V., T. m. Rangaswamy, and K. N. Subramanya. 2010. "A review on data mining applications to the performance of stock marketing." *International Journal of Computer Applications* 3 (1): 33-43.

Shagufta Parveen, and Muhammad Ayub Siddiqui. 2018. "Anchoring heuristic effect and overconfidence bias in investors: A case of Pakistan stock exchange." *Abasyn Journal of Social Sciences* 11 (2).

Wijaya, Yohanes Budiman, S. Kom, and Togar Alam Napitupulu. 2010. "Stock price prediction: comparision of Arima and artifical neural network Methods." *2010 Second International Conference on Advances in Computing ,Control,and Telecommunication Technologies.* IEEE computer Society. 176-179.

YAO, Jing Tao, and Chew Lim TAN. 2001. "Guidelines for Financial Forecasting with Neural Networks." *Proceedings of international Conference on Neural information Processing* 34: 14-18.

Zhang , G. Peter. 2003. "Time series forecasting using a hybrid ARIMA and neural network model." *Neurocomputing* 50: 159-175.