International Journal of
Finance
(IJF)

Autoregressive Neural Network EURO STOXX 50 Forecasting Model Based on Principal Component Stock Selection

Ahmad Abu Alrub, Tahir Abu Awwad and Emad Al-Saadi

International Journal of Finance
ISSN 2520-0852 (Online)
Vol. 6, Issue No. 2, pp 71 - 81, 2021
www.carijournals.org

Autoregressive Neural Network EURO STOXX 50 Forecasting Model Based on Principal Component Stock Selection.

[1*] Ahmad Abu Alrub
Assist. Prof. Dr. in Department of Accounting and Finance
Palestine Polytechnic University, Hebron P.O. Box 198, Palestine
[*1]Corresponding Author's E-mail: aabualrub@ppu.edu

[2*]Tahir Abu Awwad
Post Graduate Student: School of Computing
National College of Ireland
E-mail: tahirawwad@gmail.com

[3*]Emad Al-Saadi
PhD Student: School of Law and Economics
Istanbul Sabahattin Zaim University
E-mail: imad.sadi@islamicbank.ps

ABSTRACT

Purpose: The given study looks into forecast accuracy of a traditional ARIMA model while comparing it to Autoregressive Neural Network (AR-NN) model for 984 trading days on EURO STOXX 50 Index.

Methodology: A hybrid model is constructed by combining ARIMA model and feed-forward neural network model aiming to attain linear and non-linear price fluctuations. The study also incorporates the investigation of component stock prices of the index, that can be selected to improve the predictability of the hybrid model.

Findings:The reached ARIMA (1,1,3) model showed higher scores than AR-NN model however integrating selected exogenous stock prices from the index components gave much notable accuracy results. The selected exogenous stocks were extracted after conducting PCA and model scores were compared via MAPE and RMSE.

Unique contribution to theory, practice and policy: The major contribution of this work is to provide the researcher and fnancial analyst a systematic approach for development of intelligent methodology to forecast stock market. This paper also presents the outlines of proposed work with the aim to enhance the performance of existing techniques. Therefore, Empirical analysis is employed along with a hybrid model based on a feed-forward Neural Network. Lesser error is attained on the test set of Index stock price by comparing the performance of ARIMA and AR-NN while forecasting. Hence, The components of extracted Index stock price like exogenous features are added to make an influence from the AR-NN model.

Keywords: *ARIMA*, *Neural Networks*, *Time series*

1.0 INTRODUCTION

Equity securities or shares on a granular level represent an ownership of a company and on a more broad level, a combination of stocks (companies) represent the economic prosperity driver of a country. Stock markets represent a rich hub where these stocks can be traded for various resons as raising funds,merges and acquasistions, entry to new markets or merly seezing opertunity of price changes. Stock markets fill in the gap of a vital position in an economy's structure hence some financial actions in the market may highly effect the stability of a whole nation (Lin et al,. 2012).

Gaining a high profit from stock investments is a main advantage of stock investments over other financial instruments, however this possible rate of return comes with a rate of uncertainity and risk to the investor. Investors tend to forecast the direction or price fluctuation of a certain stock they intend to buy-in before they actually own it. According to Tay & Cao (2001), stock price forecasting is a challenging task for professional analysts. Investors have tackled the process of stock price predections in four main approaches. (1) fundamental analysis; where the future price of a companies stock is predicted from the companies past performance. Some metrics used in this method are companies turnover,annual reports, profit & loss reports etc. (2) Technical analysis approach relies more on the past stock price behaviour itself and attempt to extract patterns and shapes from price fluctuations. (3) statistical econometric approach rely on statistical theory and math to explain price changes. (4) the most recent approach is soft computing where it relys on the low cost of high computational power to solve complicated models and discover patterns in the data (Lam 2004).

To get the better of perfect forecasting in stock market prices or at least reaching a prediction price as close as possible to the actual price. Practitioners and Academic researchers have put many efforts to develop a perfect model for varying market conditions. The main issues with the ongoing variation are the unexpected turn of events, unexpected changes, the unexpected direction of the country's economy, or the challenges regarding the policies of a company and the absence of information (Daniel & Moskowitz 2016). One of the modern theories of stock market named Efficient Market Hypothesis (EMH) states that if all the information regarding stock price is available, the stock market is said to be in an efficient state. In an efficient stock market, attaining extra returns from the price fluctuations is not accessible (Brown, 2020).

The EMH theory is contradicted through traditional economic techniques that propose a higher level of accuracy for forecasting the short period of time for the future by a model called 'Auto-Regressive Moving Average ARIMA' model. For a single time series, ARIMA is a stochastic model that is developed on merging the models of Auto-regressive (AR) and Moving-Average (MA) in a single model (Musa & Joshua 2020). This method is then generalized to get time series of multiple models in the Vector Auto-Regression (VAR) and the Multivariate ARIMA. Assuming the linearity of proposed models during observations and the result of short term-based predictions are the two main limitations of the implementation of new models (Siami-Namini et al. 2018). Machine learning and Neural Network techniques are the important aspects of the limitations of new models along with a computer power source and easy requirements (Mehtab & Sen 2020). The interpretability of the results of a model can be diminished due to data transformations and complex models' development in linear common models and they become responsive to a certain dataset and values rather than the general access. General forecast models need to be established by deep learning and machine learning techniques which can predict the long term forecasts for the future with minimum error and higher accuracy rate of real values in stock prices.

This study evaluates the accuracy level of making predictions through the Auto-Regressive Integrated Moving Average (ARIMA) technique and by an Auto-Regressive Neural Network (AR-NN). ARIMA is a univariate time series forecasting model based on conventional methods while the AR-NN is based on a hybrid model for forecasting. In the next step, exogenous features are added into the AR-NN for better implementation. These attributes of the proposed models are established through the model of Principal Component Analysis (PCA) of the components of Euro Stock 50 Index. The main contributions to models are:

- To investigate the efficiency of traditional stock price forecasting, empirical analysis is employed along with a hybrid model based on a feed-forward Neural Network.
- Lesser error is attained on the test set of Index stock price by comparing the performance of ARIMA and AR-NN while forecasting.
- The components of extracted Index stock price like exogenous features are added to make an influence from the AR-NN model.

The literature on the stock price forecasting is described in section two of this paper. Section three is based on the mathematical approaches, an analysis conducted, and in section four of this paper, the developed algorithms of AR-NN, ARIMA, and PCA are described. In the final part, results and conclusions are provided.

2.0 LITERATURE REVIEW

Several macro-economic factors affect the fluctuations of the stock market and it also includes the economic situation of the country or a company, bank rate, commodity rate, currency exchange rate, price of gold, movement of stock market expectations of investors policies made by organizations, the psychology of investors, etc. (Biso & Dash2015; Haleh et al. 2011). Many other techniques have been developed in the past few years to forecast the stock market and to suggest some intelligence-based decision-making systems. To predict the stock price time series, the most commonly employed techniques are statistical methods and software-based computer approaches (Wang et al. 2011). The future stock prices are predicted based on past values through traditional modes of forecasting like exponential smoothing (ES), autoregressive moving average (ARMA), autoregressive conditional heteroskedasticity (ARCH), autoregressive integrated moving average (ARIMA), and generalized autoregressive conditional heteroskedasticity (GARCH) (Box et al. 2011). The proposed systems are developed on the basis of financial time series under study which is generated from the linear process and forecasts the future series of the proposed model (Kumar & Murugan 2013). The series of data of stock time is non-linear, very noisy, dynamic, complex, chaotic, and non-parametric (Si & Yin 2013; Adebiyi et al. 2012). Older techniques of statistical analysis cannot be employed to investigate the non-static and complex nature of the stock market. The performance has been tested for the combination of the regression model and model classification in the study of Mehtab and Sen (2020). The performance of the model proposed was investigated in their study employing the correlation, Mean/RMSE, and mismatched cases. The analysis made by Sen

(2020) investigated the outperformance of the LSTM model among all the other performance accuracy measures, some mismatched cases, Mean/RMSE, and correlation.

Siami-Namini et al,. (2018) studied the traditional statistical ARIMA that can still replace the LSTM deep learning method. The study included monthly information of 6 financial stock prices and 6 months of economic variables. The results of the study showed an outrun the developed LSTM model as compared to the ARIMA. The performance of the forecast has been questioned in the study of Wen et al. (2020) from CNN, PCA, MLP, LSTM, and MA.

The results obtained are confined with RMSE in the accuracy forecasted along with the PCS technique while comparing to the other proposed models. Wang et al. (2019) studied the performance of the PCA-DNN model over the PCA-ANN model for the classification of variations in stock prices. Sixty economic variables were employed for the prediction of the direction of the SPDR S&P 500 ETF Index. The purpose of this research work is to review the current models employed to find solutions to the stock market prices and their forecasting. The study will be useful to guide a research scholar, investor, and analyst to build a smart stock market by employing forecasting technology.

3.0 ANALYSIS TECHNIQUE

A brief description is provided here to elaborate on the methodology to attain results and an explanation is provided for the mathematical theory of PCA, AR-NN, ANN, and ARIMA.

3.1 Autoregressive Integrated Moving Average

It is a model composed of composite time series as indicated by the acronym ARIMA (p, d, q) by combining the Moving Average (MA) and Autoregression (AR). The I in this shows the integration and defines the level through which a series varies to reach the point of stationarity. The Autoregression model is based on the regression of lagging observation and depends upon the $(y_{t-1}, i = 1,2, .. p)$ and as an unrestrained model in addition to the already available models.

$$y_t = c + \sum_{i=1}^{p} \emptyset_i y_{t-1} + \epsilon_t \qquad (1)$$

The y_t describes the response at the given time t to satisfy the stationarity. Here c is a constant, an autocorrelation coefficient between y_t and y_{t-i} is $\emptyset_i$. The term showing error ϵ_t should fulfill the requirement of Gaussian noise series and along with a zero average and with a variance of σ^2. The model of Moving Average (MA) describes that the observation is dependent on time t with the error term of residual values and with lagged observation based on the number of moving averages (q). The expected value of y_t demonstrated by μ, the θ_i terms demonstrate the weights inserted to the present and previous stochastic terms.

$$y_t = \mu + \sum_{i=0}^{q} \theta_i \epsilon_{t-1} \qquad (2)$$

By combining equations (1),(2) an ARIMA model is formed to the order (p, q) where $\theta_i, \emptyset_i \neq 0$ and $\sigma_\epsilon^2 \neq 0$

$$y_t = c + \sum_{i=1}^{p} \emptyset_i y_{t-i} + \epsilon_t + \sum_{i=0}^{q} \theta_i \epsilon_{t-i} \qquad (3)$$

3.2 Artificial Neural Network

An Artificial Neural Network (ANN) is an information-driven numerical technique depending upon a non-direct enactment method. The versatility of this strategy and its casual forecast gives it the upside of catching patterns in value fluctuations, the capacity to acknowledge the high dimensionality in input includes just as producing a higher chance of anticipated results (White, 1998). Considering the high ups and downs in economic business sectors and differing highlights merged with price changes, ANN makes the proper way for financial specialists to take advantage of available opportunities and moving resources in the most encouraging business sector for the forecasted period (Wood and Dasgupta, 1996). In the given research, the engineering of the suggested model of ANN is a full feed forward 3-layer neural system. The first layer is composed of the hubs of slacked index costs, the concealed layer, and the yield layer 1is composed of gestures of the normal Index cost. The lagged prices are demonstrated by the mathematical relationship of $(y_{t-1}, y_{t-2}, \ldots . y_{t-n})$. ω is a vector of all parameters and f is a function determined by the network structure and weights (Musa & Joshua 2020).

$$y_t = f(y_{t-1}, y_{t-2}, \ldots . y_{t-n}, \omega) + \epsilon_t \quad (4)$$

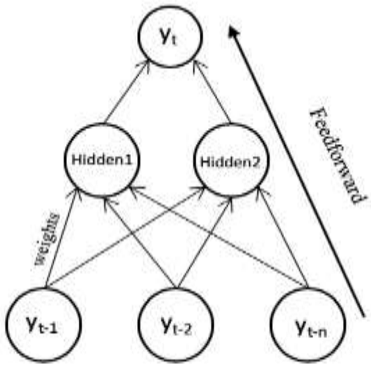


Figure1: Neural Network Architecture

3.3 Hybrid Model

The idea of the hybrid model is to overcome the shortcomings of developing a linear or non-linear approach on an individual basis. The models based on statistics are represented along with the Machine learning Neural Network systems (Wang et al. 2019). Hybrid models can be employed to research empirical models to make predictions about performance models (Zhong & Enke 2017; White 1988). The hybrid methodology devised by Zhang shows that the expected price is determined by linear or non-linear models. The ARIMA model was employed to determine the linear attributes of price changes by L_t . the remaining non-linear attributes are determined by the residual model that is analyzed by ANN.

$$Y_t = L_t + N_t \quad (5)$$

4.0 PERFORMANCE MEASURES

To measure and compare the power of prediction and performance of the given study and its models, the outcomes are determined by Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The term (A_t) is employed to show the actual index price and the (F_t) is used for forecasted index price and the N point out the number of observations of daily price variations.

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{A_t - F_t}{A_t}\right| \times 100\% \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(A_t - F_t)^2} \quad (7)$$

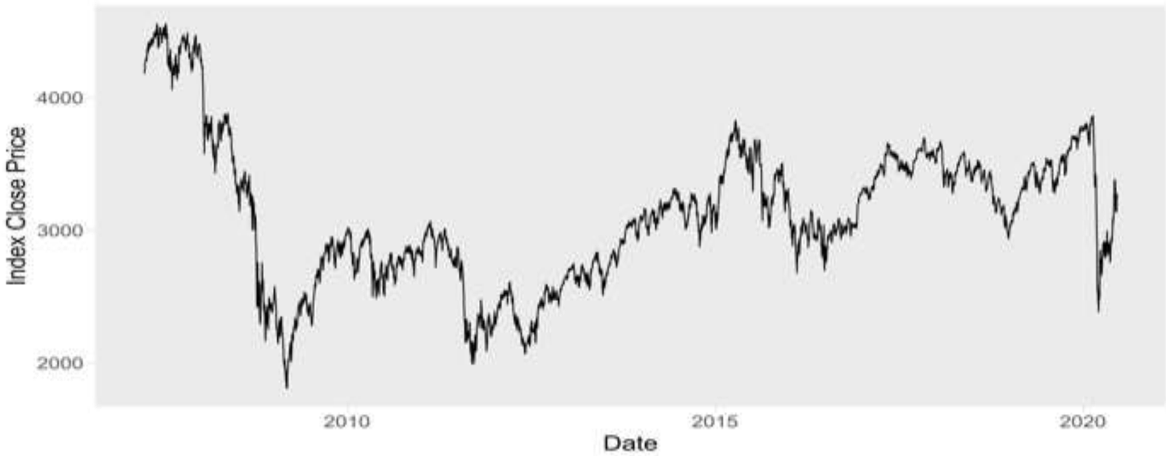
International Journal of Finance
ISSN 2520-0852 (Online)
Vol. 6, Issue No. 2, pp 71 - 81, 2021 www.carijournals.org

5.0 EMPIRICAL ANALYSIS

This section provides the information of steps taken that led to the experimental results of this study. The developed models forecasted the index values between 20th June 2016 to 19th June 2020 for EURO STOXX 50 values. The largest organizations in Eurozone are represented by the STOXX 50 index which is taken as a blue-chip index. The free float capitalization methodology is employed to study and develop a weighted average for 50 component companies.

5.1 Data Preparation

The data employed for index values were attained from yahoo finance by "tidyquant" and "Batch Get Symbols" packages in R. The main interest of this study was to attain all the information on closing index prices available in the stock market. The number of observations in the final estimations is 3279 which represents the daily index numbers and their closing prices. The values of median, mean, maximum and minimum are \$3184, \$3184, \$4842 and \$1526 respectively.



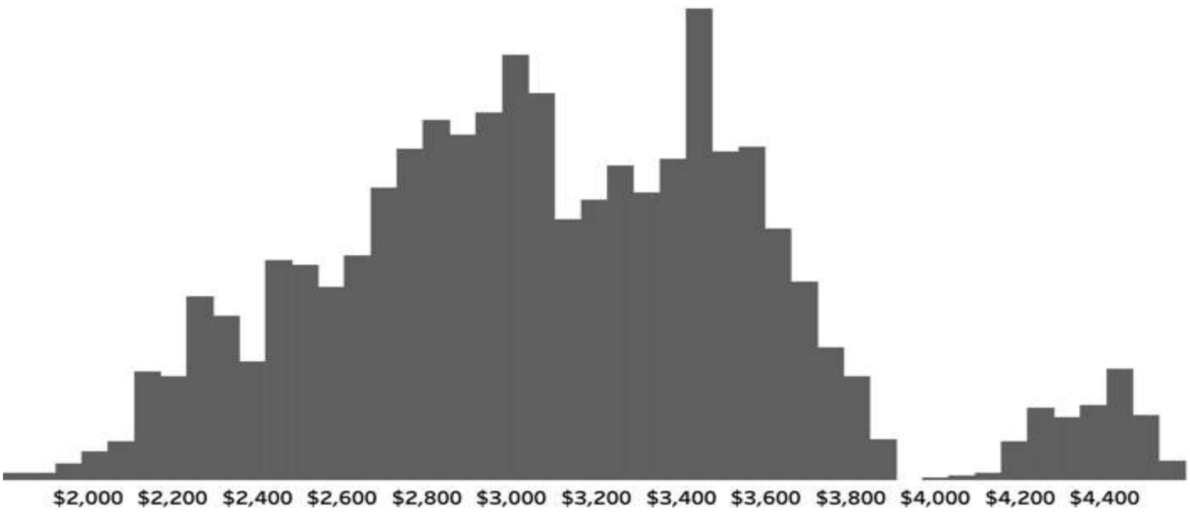


Figure2: Euro STOXX 50 yearly closing price \ closing price distribution

5.2 ARIMA Model

In this model, the steps taken are parametric tweaking, identification of model, and diagnostic confirmation (Faruk, 2010). The null hypothesis of the non-stationarity of the augmented Dickey–Fuller test (ADF) was rejected to show the first difference (d1) as stationary as p-value = 0.01. The order for this model was decided as per the models of PACF, ACF graphs and in the final attempt best fit model was selected based on the minimum Akaike Information Criterion (AIC). For the satisfaction of requirements of white noise, diagnostic checking was confirmed by fitted residual models as shown in figure 3. The most-suited model is ARIMA along with a minimum AIC of -12518.62.

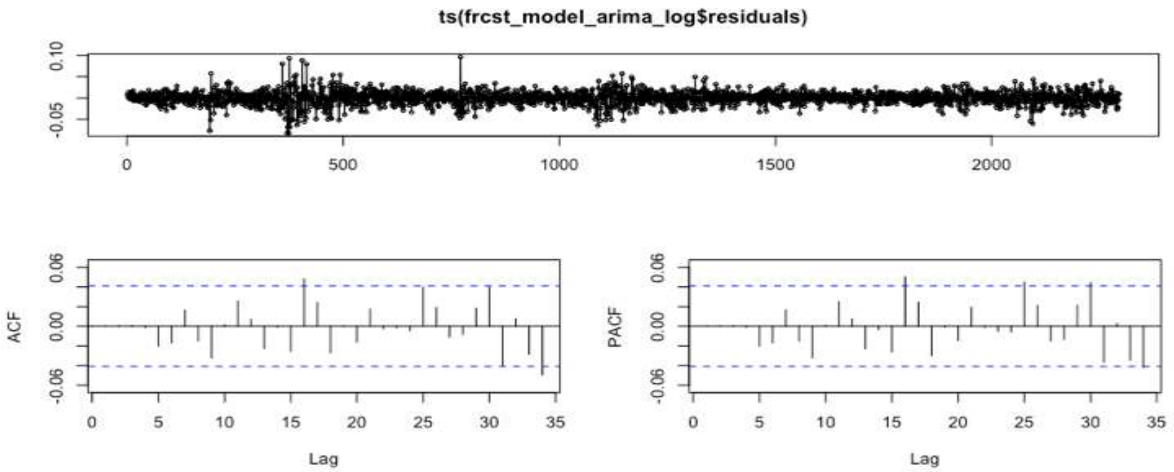


Figure 3: ARIMA(1,1,3) white noise residuals

5.3 Hybrid Model

A statistical software called 'R' is employed in the package of NNFOR for forecasting the index price values with the help of a multi-layer Neural network. A set of twenty repetitions are used for the training of the network. Random weights were employed as starting an experiment in every repetition. To enhance the hidden layer, given weights are connected to 5 nodes in the hidden layer, but it produced no benefit for the network. In the input layer, the optimal ANN has 2 nodes which represent the 1st and 4th lag prices. The inputs to the system are lagged prices and ANN for the creation of an Autoregressive neural network.

5.4 Principal Componenet Analysis

The main examination is directed on the 50 independent segment weights of the EURO STOXX 50 Index. As indicated by the Eigenvalues show the 56% of variety in the list is reasonable by (PC1) and also 24.7% constantly head segment PC2. In this examination we studied the supplies of (PC1) costs and attribute them as outside regressors into the past got ARNNET model. From the segments of (PC1), the stocks having Eigenvalues above estimation of 1 are picked as under the 1st method which has no clarification power and contains 24 out of 29 values of the stock market.

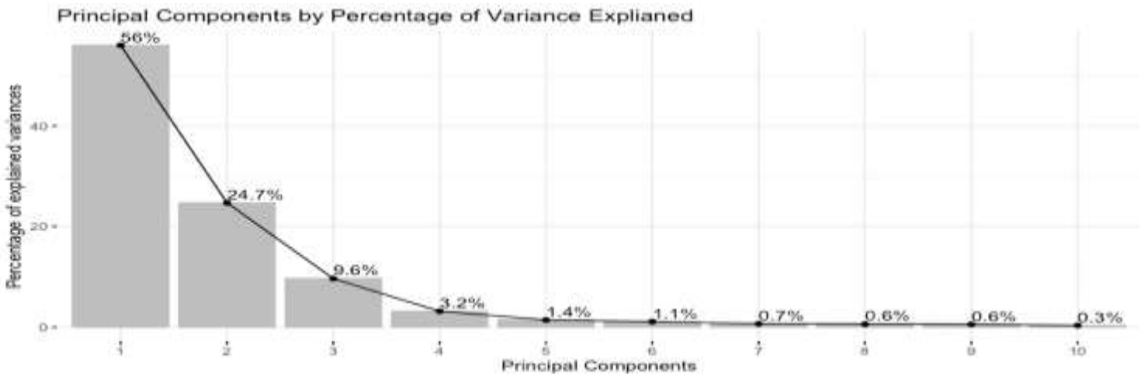


Figure 4: Principal Components by Percentage of Variance Explained.

6.0 Results and Discussion

In the final part of this report, forecasted results are discussed. The total training time was 984 days to forecast the implemented results. The univariate model for forecasting employed was ARIMA which denotes the 1st order correlation. The proposed model provided better efficiency as it provides the RMSE of 0.1809 as compared to the training set of 0.015766. The model of the neural network presented the RMSE value of 0.2479 and 0.015760 for the testing set by the hybrid auto-regressive model. The reported stocks for the principal component produced the RMSE of 0.120173 for the test set. On this basis, the null hypothesis is rejected and other hypotheses for research are accepted which state the individual stock prices of the EURO STOXX 50 index that can affect the predicting capability of the AR-NN model. The RMSE test set has been provided in table 1 to forecast comparisons. The table 3 describes the nodes of the input layer which are described along with the lag values for every stock price.

Table1: Test Set Forecast Performance Metrics

Model	RMSE	MAPE (%)
ARIMA(1,1,3)	**0.18093**	**2.0434934**
NNETAR(2,5,1)	**0.24790**	**2.8530023**
PC1_NNETAR (36,5,1)	**0.12017**	**1.2919216**

Table 2: Training set Forecast Performance Metrics

MODEL	**RMSE**	**MAPE (%)**
ARIMA (1,1,3)	0.01576697	0.1408455
NNETAR (2,5,1)	0.01576032	0.1404156
PC1_NNETAR (36,5,1)	0.01293683	0.1224619

The power of predictability was represented by previous forecast values in this study and the proposed models compared the log models. To achieve better visibility for the models, the prediction is made for actual index prices. Figure 5 below shows:

Figure 5: Euro STOXX 50 Index test set price forecast, model comparison.

The prediction method was found to be effective during the experiments and analysis made in the study. This has provided some practical and theoretical values to learn the foreign stock market in deep applications. The following areas can be further investigated based on present work done. The scale of the network can be enhanced, network structure can be improved, the mechanism of attention can be improved and more complex structures can be further investigated. It has been found that the stock prices contain typical attributes of non-linear dynamic systems which are very complicated in terms of changes. The trend of stock prices cannot be predicted by using the analysis method, so there is a need for some non-linear method for prediction of stock prices with better efficiency to meet the demands of the market.

7.0 CONCLUSION AND POLICY IMPLICATION

Predictability of the traditional ARIMA model is investigated along with a hybrid auto-regressive neural network AR-NN model. EURO STOXX 50 Index is forecasted for a total time span of 984 trading days via both models. Linear and no-linear fluctuations of the price are aimed to be captured by the hybrid AR-NN model. This study takes an extra mile in the analysis by Integrating a PCA on the AR-NN model. Individual stock components of the index are accordingly selected as exogenous variables injected to the AR-NN model.

This study's findings contradicted the findings of previous research by (Musa & Joshua 2020; Areekul et al,. 2009) as the reported ARIMA (1,1,3) model showed better accuracy rates than the AR-NN model in comparision of RMSE accuracy prediction measure. However after injecting the selected exogenous stock prices into the hybrid model, forecast accuracy rates improved significantly reflecting an RMSE 0.0129 and MAPE 0.122. The study's findings proposes a more accurate method of forcasting which policy makers and investors may seek to decrease uncertainity and risks of stock price future fluctuations.

Major findings of this study are; (1) the accuracy of forecasting in the stock market can be enhanced through identifying suitable pre-processing models and feature selection techniques as the PCA. (2) the individual stock price's of an index components significantly improve the forecast accuracy rates of the index price itself. (3) the hybrid auto-regressive neural network

(AR-NN) model is able to give better forecast over ARIMA model, after the selected exogenous variables integration to the model.

8.0 RECOMMENDATIONS

We end by providing a recommendation on how this work can be taken a step further. Although hybrid auto-regressive neural network AR-NN model considered in this paper outperform traditional econometric models in forecasting EURO STOXX 50 Index is forecasted , we believe that the forecasting accuracy of Integrating a PCA on the AR-NN model. Individual stock components of the index are accordingly selected as exogenous variables injected to the AR-NN model. can still be improved. We performed a grid search to determine the optimum number of hidden nodes and training required and performed some experimentation to find the optimum set of input variables. However, ideally, a Integrating a PCA on the AR-NN mode for a particular task has to be optimised over the entire parameter space of the learning rate, momentum rate, number of hidden layers and nodes, combination of input variables and activation functions. In that respect, Integrating a PCA on the AR-NN mode combined with genetic algorithm optimization techniques can potentially be used for building more accurate models for forecasting EURO STOXX 50 Index and is recommended for future research.

References

Adebiyi, A. A., Ayo, C. K., Adebiyi, M., & Otokiti, S. O. (2012). Stock price prediction using neural network with hybridized market indicators. Journal of Emerging Trends in Computing and Information Sciences, 3(1)

Areekul, P., Senjyu, T., Toyama, H., & Yona, A. (2009). Notice of violation of ieee publication principles: A hybrid arima and neural network model for short-term price forecasting in deregulated market. IEEE Transactions on Power Systems, 25(1), 524-530

Bisoi, R., & Dash, P. K. (2015). Prediction of financial time series and its volatility using a hybrid dynamic neural network trained by sliding mode algorithm and differential evolution. International Journal of Information and Decision Sciences, 7(2), 166-191.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. Journal of Financial economics, 122(2), 221-247.

Faruk, D. Ö. (2010). A hybrid neural network and ARIMA model for water quality time series prediction. Engineering applications of artificial intelligence, 23(4), 586-594.

Haleh, H., Moghaddam, B. A., & Ebrahimijam, S. (2011). A new approach to forecasting stock price with EKF data fusion. International Journal of Trade, Economics and Finance, 2(2), 109.

Kumar, D. A., & Murugan, S. (2013, February). Performance analysis of Indian stock market index using neural network time series model. In 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (pp. 72-78). IEEE.

Lam, M. (2004). Neural network techniques for financial performance prediction: integrating fundamental and technical analysis. Decision support systems, 37(4), 567-581.

Lin, C. S., Chiu, S. H., & Lin, T. Y. (2012). Empirical mode decomposition–based least squares support vector regression for foreign exchange rate forecasting. Economic Modelling, 29(6), 2583-2590.

Mehtab, S., & Sen, J. (2020). A time series analysis-based stock price prediction using machine learning and deep learning models. International Journal of Business Forecasting and Marketing Intelligence, 6(4), 272-335.

Musa, Y., & Joshua, S. (2020). Analysis of ARIMA-artificial neural network hybrid model in forecasting of stock market returns. Asian Journal of Probability and Statistics, 42-53.

Si, Y. W., & Yin, J. (2013). OBST-based segmentation approach to financial time series. Engineering Applications of Artificial Intelligence, 26(10), 2581-2596.

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018, December). A comparison of ARIMA and LSTM in forecasting time series. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 1394-1401). IEEE.

Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. omega, 29(4), 309-317.

Valentini, V. (2020). Efficient market hypothesis: diving into the cryptocurrency world.

Wang, J. Z., Wang, J. J., Zhang, Z. G., & Guo, S. P. (2011). Forecasting stock indices with back propagation neural network. Expert Systems with Applications, 38(11), 14346-14355.

Wang, X., Dai, W., & Yan, Y. (2021). Numerical analysis of a new conservative scheme for the 2D generalized Rosenau-RLW equation. Applicable Analysis, 100(12), 2564-2580.

Wen, Y., Lin, P., & Nie, X. (2020, March). Research of stock price prediction based on PCA-LSTM model. In IOP Conference Series: Materials Science and Engineering (Vol. 790, No. 1, p. 012109). IOP Publishing.

White, H. (1988, July). Economic prediction using neural networks: The case of IBM daily stock returns. In ICNN (Vol. 2, pp. 451-458).

Wood, D., & Dasgupta, B. (1996). Classifying trend movements in the MSCI USA capital market index—A comparison of regression, arima and neural network methods. Computers & Operations Research, 23(6), 611-622.

Zhong, X., & Enke, D. (2017). A comprehensive cluster and classification mining procedure for daily stock market return forecasting. Neurocomputing, 267, 152-168.