Exchange Rates Prediction via Deep Learning and Machine Learning: A Literature Survey on Currency Forecasting

Oyiwona Ajumi[1], Abhishek Kaushik[2]

School of Business and Law, Dublin Business School

[2]ADAPT Centre, School of Computing, Dublin City University, Dublin 9, Ireland

Abstract: ***The main objective of this article is to present an up-to-date review on time series forecasting techniques. This paper provides a thorough critique on how forecasting of exchange rates in the FOREX market using Machine Learning methods differ from Deep Learning methods. Empirical research on exchange rate forecasting using Deep-learning methods for emerging economies in the Middle East and Africa is almost non-existent. In any case, exchange rate forecasting is not an easy task. There is no single forecasting method that is superior for obtaining accurate exchange rate all the time and for different exchange rate. This review willprovide guidance to analysts for future research in this application area.***

Keywords: Forecasting, Exchange Rate, Machine Learning, Time series

1. Introduction

The significance of the Foreign Exchange (FOREX) market wasn‘t generally as solid as it is today. When the value of a currency was tied to physical commodity, it was impossible to benefit from currency trading. Since the breakdown of the Bretton Woods system of fixed exchange rates in 1971 – 1973, and the implementation of floating exchange rate system that allowed currency rates to be determined by its demand and supply, the topic of exchange rates predictability has been discussed and analyzed in many studies [1] .Wang Peijie defines a FOREX market as a market where a convertible currency is exchanged for another convertible currency or other convertible currencies [2] .Accurately Forecasting the exchange rate is a critical objective of all international investors and policy makers [3] . Market supply and demand impacts exchange rate movement daily, imposing risks on participants in the FOREX market. Therefore, it is imperative to accurately forecast exchange rates, which would allow investors and businesses to make quick, confident and effective decisions. This technical approach to analyzing the forex market is based on chart analysis, which generates results by evaluating the recurring patterns in graphs of exchange rate movements. The success of this approach depends on the forecaster’s ability to discover patterns that repeat themselves [4] .

There is reasonable skepticism in the ability to make money by predicting price changes in any given market. This skepticism mirrors the efficient market hypothesis according to which markets fully integrate all the available information and prices fully adjust immediately once new information becomes available. In essence predicting the market movement is useless because the market is fully efficient. However forecasting is possible due to the market’s delayed reaction to new information [4] . Applying fundamental as well as technical approaches, shows some evidence that exchange rate movements may be predictable for a longer time horizon [5] . Other researchers challenge the idea that economic fundamentals help predict currency values and attribute changes in exchange rates to random luck [6] . They claim that the performance of a random walk model is just as good as the performance of a model based on economic fundamental variables.

Another concern often raised is the validity of using past price data to predict the future. One view posits that the future is a logical extension of the past. The only type of data one has to go to for analysis is past data. We can only estimate the future by projecting past experiences into that future [4] . That being said, time series forecasting relies on a single stream of data to evaluate the past i.e. to account for trend, volatility or seasonality and to forecast the future. Any major disruptions in the past data need to be explained to avoid potentially replicating the disruption in the future forecast.

Researchers have used a number of linear and non-linear techniques for currency predictions with the utmost degree of reliability. Some of the most often used techniques are: Naive models, Autoregressive Moving Average model (ARMA), Exponential smoothing, and Autoregressive Conditional Heteroskedasticity models (ARCH) and its variations (GARCH, EGARCH etc.), which model volatility. In the past few years a substantial amount of research works have been carried out towards the application of artificial neural networks (ANN) for time series modelling and forecasting. The excellent feature of ANNs, when applied to time series forecasting problems is their inherent capability of non-linear modelling, without any presumption about the statistical distribution followed by the observations. The most common and popular among them are the multi-layer perceptron (MLP) which is characterized by a single hidden layer Feed Forward Network [7] .Support vector machines (SVM), which were initially proposed to solve classification problems, are now being applied to time series.Least-square SVM (LS-SVM) and Dynamic Least-square SVM (LS-SVM) are two popular SVM models for time series forecasting.

Financial series and especially exchange rates are dominated by factors that time-series analysis and statistics are unable to capture in a single model. Based on this, ensembles and hybrid methods have been proposed in order to achieve a prediction accuracy that exceeds any individual model, which in turn minimises market investment risk.

This paper provides a critical survey of recent literature on Machine Learning and Deep Learning for exchange rate forecasting taken from reputed journals and books.

One of the objectives of this paper is to provide guidance to analysts on exploring the existing literature in this area as well as provide a dependable overview of established findings that can be helpful in future research. In addition, forecasters at Central banks, financial organizations and private traders will be keen on knowing which models and methodologies successfully predict exchange rates.

The article is organized as follows. Section 2address the efficient market hypothesis, and the reasoning and measures taken to accurately forecast exchange rates: Section 3 reviews the literature on Machine Learning and Deep Learning models that have been used to successfully forecast exchange rate, Section 4 examines the characteristics and methodology of the models used in literature, their assumptions and limitations: Section 5 overviews the forecast evaluation methods and Section 6 concludes the paper.

2. Exchange rate is unpredictable: Random walk theory

The current consensus is that the random walk is explained by the efficient market hypothesis (EMH). The EMH developed by Eugene Fama states that financial markets are efficient and that prices already reflect all known information concerning a stock or other security and that prices rapidly adjust to any new information [4] .In a nutshell, it claims that price movement is random and price changes are "serially independent" and a simple "buy and hold" market strategy is best. While there seems little doubt that a certain amount of randomness or "noise" does exist in all markets, it's just unrealistic to believe that all price movement is random [4] .

The strongest argument against the random movement supporters is price anticipation. One can argue that all participants (the market) know exactly where prices should move considering all publicly available information. For example, following a breaking news, a market participant may have bought or sold stock to increase/lower the price, forcing the move to the current price. Excluding anticipation, the apparent random movement of prices depends on the frequency of the data used and the time interval. When a longer time horizon is used and the data is averaged to increase smoothing process, the trending characteristics will change and smooth out the irregular daily movements and results in noticeable correlations between successive prices. In the long run, most future prices find a level of equilibrium and, over time show characteristics to mean reverting(return to a local average price) [8] .

One of the first researchers to look into this topic was Mussa [9] . In his research, he determined that the exchange rate is largely unpredictable after analyzing the behavior of the United States dollar exchange rates against major currencies. He concluded that the spot exchange rate of the U.S. dollar approximately follows a random-walk process under condition that the exchange rate is not controlled by interventions. Meese and Rogoff arrived at the same conclusion in their 1983 paper. The authors collected monthly, seasonally unadjusted exchange rate data on U.S. dollar against British pound, Japanese yen and German mark from 1973 to 1981. Splitting the datasets and comparing the out-of-sample forecast of time series models with the random walk without drift, they determined that the random walk model outperforms the other models at horizons up to a year. Although, the model failed to yield the minimum forecast errors when the forecast horizon was extended to periods beyond twelve months [6] .

Although the conclusion drawn by Meese-Rogoff has been resilient throughout the years, some researchers dispute the claim, stating that it is indeed possible to outperform the naïve random walk model using models that account for econometric problems. Such models can explain a small amount of variation in the exchange rate [10] . However, Giorgianni [11] questioned the inference procedures of the results of these studies, showing that the forecasting ability of exchange rates is critically dependent on the assumptions of the data generating process and that co-integration between fundamentals and exchange rates did not help in long horizon predictability. A different approach was taken by Kilian and Taylor [12] , in which they accounted for the nonlinearity between exchange rates and fundamentals. The authors applied exponential smoothing AR technique to the dataset and found strong evidence of predictability at longer horizons.

The results discouraged others from analyzing this area of economics for a while. Many studies about the predictability of exchange rates came up with mixed results. On one hand, they showed that it is very difficult to beat the random walk at short horizons, e.g. up to one year but there was evidence of predictability at longer horizons. Later on in the 2000s, some studies have been able to prove that some models could outperform the random walk model, at least in specific conditions and for short-horizon predictability. It seems doubtful that statistical evidence will ever totally prove or disprove the Random Walk Theory. However, the idea that markets are random is totally rejected by the technical community. If the markets were truly random, no forecasting technique would work.

3. Exchange Rate is Predictable

The idea that the FOREX market is not so efficient is supported by Campbell and MacKinley [13] in which they state that "Recent econometric advances and empirical evidence seem to suggest that financial asset returns are predictable to some degree".

3.1 Forecasting Via Machine Learning

It's been established that although it is clearly difficult to accurately forecast exchange rates, it is not impossible. In time series forecasting, past observations are collected and analysed to develop a suitable mathematical model that captures the underlying data generating process for the series. Time series models can have many forms and represent different stochastic processes. There are two widely used linear time series models viz. Autoregressive (AR) and Moving Average (MA). Combining these two, the ARMA and the Autoregressive Integrated Moving Average(ARIMA) have been proven effective in various studies [14] . The Box Jenkins methodology for finding the hyper parameters of the ARIMA model refers to the iterative approach of model identification, parameter estimation and diagnostic checking [15] . Because of its popularity, the ARIMA model has been used as a benchmark to evaluate some new modeling approaches [16] . Result of the study by Newaz [17] in which he used Box-Jenkins methodology on monthly exchange rate data of Indian rupee for building an ARIMA, exponential smoothing, naïve 1 and naïve 2 models to forecast the exchange rate showed that the ARIMA model provides a better forecast of exchange rates than the other models. He came to this conclusion by comparing the Mean Absolute Error (MAE), Median Absolute Error (MEAE), Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE) and Root Mean Square Error (RMSE) of the predictions. Maria and Eva [18] investigated the behaviour of daily exchange rates of the Romanian Leu against the most important currencies in terms of international trade, namely the Euro, United States dollar, British pound, Japanese yen, Chinese yuan and the Russian ruble. Simple Exponential Smoothing, Double Exponential Smoothing, the Simple Holt-Winters, the Additive Holt-Winters and the ARIMA were generated and compared with each other. All the results indicated the appreciation of the Romanian Leu against the other currencies. The authors concluded that the exponential smoothing techniques outperform the ARIMA models in forecasting the evolution of the exchange rate because of the speed with which they adapt to the smallest changes in market conditions. The authors also observed that the ARIMA models were more effective in rendering the medium-term trend. Paraschos Maniatis [19] attempted to model the exchange rate between Euro and U.S. dollar using ARIMA as well as Exponential smoothing techniques. The authors failed to validate the assumptions of the linear models; hence, they were unable to obtain reliable predictions.

One question that arises is how the fluctuations in exchange rate might affect the financial market players and the investors as well as the economy as a whole. In an attempt to identify a model forecast for the Nigerian naira against U.S dollar using the Box Jenkins approach, Onasanya and Adeniji [20] analyzed the exchange rate data from January 1994 to December 2011. An out-of-sample forecast of 12 months was made and results indicated that the naira would continue to depreciate against the US dollar, which was has been the case. Based on the results, the authors recommended an incorporation of fiscal policies, and devaluation methods to stabilize the naira by eliminating over dependence of the nation on imports. Ghalayini [21] attempted to explain the value of the dollar/euro exchange rate which, based on its linear relationship was estimated by a simple ARIMA model. However, the limited computational power of the ARIMA model failed to model the currency pair volatility, therefore, motivating the author to perform a multivariate co-integration analysis. Further research was conducted by Olatunji and Bello [22] on monthly exchange rate data in Nigeria. They compared the performance of ARIMA and ARMA based on the MAE and RMSE of the predictions and concluded that the ARIMA model performed best.

Another traditionally used linear time series model that incorporate multivariate systems is the vector autoregressive model (VAR) and the vector error correction (VEC) model. As the random walk without drift has proven to be a very competitive model in forecasting exchange rates, Carriero*et al.* [23] developed a Bayesian Vector Autoregression (BVAR) with a Normal-inverted Wishart prior, imposing a-priori a univariate driftless random-walk representation, but allowing the data to speak about the relevance of other available information. The proposed BVAR model was used to forecast a panel of 33 exchange rates vis-a-vis the US Dollar finding that it could lead to gains at all forecast horizons, including the very short ones where the random walk forecast is typically extremely hard to outperform. In contrast, Cuaresma and Hlouskova [24] compared the accuracy of VAR, restricted vector autoregressive (RVAR), BVAR, VEC and Bayesian vector error correction (BVEC) models in forecasting the exchange rates for five Central and Eastern European currencies (Czech Koruna, Hungarian Forint, Polish Zloty, Slovak Koruna and Slovenian Tolar) against the Euro and the US dollar. Majority of the models failed to beat the random walk, which presents further evidence on the difficulties associated with complex time series models. In forecasting, and even in economics, multivariate models are not necessarily better than univariate ones. While multivariate models are convenient in modelling interesting interdependencies and achieve a better fit within a given sample, it is often found that univariate methods outperform multivariate methods out of sample [25] .

Linear models have drawn much attention due to their relative simplicity in understanding and implementation. However, many practical time series show non-linear patterns [14] . Non-linear models are appropriate for predicting volatility changes in economic and financial time series [26] .Volatility is considered a measure of risk, and investors want a premium for investing in risky assets. Exchange rate volatility is a major challenge facing development of an economy, making planning more problematic and investment, more risky.In modelling volatility, popular and frequently applied models include the ARCH and Generalized Autoregressive Conditional Heteroskedasticity (GARCH).ARIMA is best suited for stationary exchange rate and ARCH and GARCH are suited for modelling shocks to the series. Pahlavani and Roshan [27] attempted to compare the forecasting performance of the ARIMA model and hybrid ARIMA-GARCH, ARIMA-EGARCH Models by using daily data of the Iran's exchange rate against the U.S. Dollar (IRR/USD). The results indicated that in terms of the lowest RMSE, MAE and TIC criteria, the best model was the ARIMA, then EGARCH. Bello [28] investigated the volatility of the daily Dollar/Naira exchange rate using different variations of the GARCH model for the period of January 1999 to April 2013. Standard GARCH models assume that positive and negative error terms have a symmetric effect on the volatility. In other words, good and bad news have the same effect on the volatility in this model. The GARCH extensions were applied due to their capability of modeling leverage effects as noted by [29] . Analysing the MAE, RMAE, MAPE and Theil inequality Coefficient (TIC), the results showed that TS-GARCH model, with its specification on conditional standard deviation instead of conditional variance, provided the most accurate forecasts for future Naira-dollar exchange rate volatility.

Recently, the Support Vector Machine (SVM) method, which was first suggested by Wu & Vapnik [30] has emerged as a new and powerful technique for constructing non-linear empirical regression models in time series forecasting. SVMs estimate the regression using a set of linear functions that are defined in a high dimensional space using Vapnik's ε-insensitive loss function, a risk function consisting of the empirical error and a regularization principle [31] . Lixia & Wenjing [32] proposed a technique based on least squares support vector machine (LS-SVM) to predict the FOREX market. Grid search was applied to automatically determine the model's hyper-parameter during the forecasting process. The experiment was carried out on four kinds of daily exchange rate records comprising data from 2003 to 2007. By testing the out-of-sample proportion of the data, and evaluating the RMSE, MAE and MAPE, the authors concluded that LS-SVM is a feasible and valid approach to forecasting exchange rate time series. However, the performance of the model wasn't compared with other benchmark models, making the experiment incomplete.

Hybrid ensemble models, which combine the output of several different models, have also been proposed to increase forecast accuracy. The study by Fu [33] combined empirical mode decomposition with SVR (EMD-SVR) on an ensemble based framework. EMD decomposes a nonlinear and non-stationary time series into a sum of intrinsic mode function (IMF) components with individual intrinsic time scale properties. The effectiveness of the model was demonstrated on the euro (EUR) /Chinese renminbi (RMB) exchange rate. The result showed that EMD-SVR has a strong forecasting ability and is remarkably superior to normal SVR. Another author, Lin [34] applied a hybrid EMD-LSSVR (least squares support vector regression) forecasting model to FOREX rate forecasting. LSSVR is constructed to forecast the IMF obtained after EMD decomposition, and the residual values individually, and then all these forecasted values are aggregated to produce the final forecasted value. The result of the analysis showed that the proposed EMD-LSSVR model outperformed the EMD-ARIMA as well as the LSSVR and ARIMA models without decomposition. Sermpinis *et al* [35] introduced a hybrid Rolling Genetic Algorithm-Support Vector Regression (RG-SVR) model for optimal parameter selection and feature subset combination. RG-SVR genetically searches over a feature space (pool of individual forecasts) and then combines the optimal feature subsets for each exchange rate. Adopting a sliding window approach, the algorithm was applied to EUR/USD, EUR/GBP and EUR/JPY exchange rates for forecasting and trading decisions. The proposed model was benchmarked against non-genetically optimized SVR and SVM models, and it exhibited predominance over the other models. Further research was conducted by Plakandaras *et al.* [36] , in which they combined signal processing to machine learning methodologies by introducing a hybrid Ensemble Empirical Mode Decomposition (EEMD), Multivariate Adaptive Regression Splines (MARS) and Support Vector Regression (SVR).After the decomposition of the original exchange rate series with EEMD, MARS selects the most informative variables included in the initial dataset using a forward and backward recursive partitioning strategy. The selected variables are fed into two distinctive SVR models for forecasting each component separately one period ahead with the summation providing exchange rate forecast. The proposed model was applied to monthly and daily Euro (EUR) /United States Dollar (USD), USD/Japanese JPY (JPY), Australian Dollar (AUD) /Norwegian Krone (NOK), New Zealand Dollar (NZD) /Brazilian Real (BRL) and South African Rand (ZAR) /Philippine Peso (PHP) exchange rates. The authors determined that the proposed model is data driven and relies on minimum initial assumptions to make accurate predictions.

3.2 Forecasting Via Deep Learning

Over the last decade, brain-inspired deep learning techniques, originally introduced by Hinton [37] has been shown to be a robust and efficient method in a variety of application domains. These include face detection, speech recognition, document categorization and natural language processing. Deep learning networks are effective not only for pattern recognition tasks, but also for prediction problems of sequential data. Some recent studies have used deep learning for exchange rate predictions. The classical methods used for time series prediction like Box-Jenkins ARIMA assumes that there is a linear relationship between inputs and outputs. However, Neural Networks have the advantage of approximating nonlinear functions. Yu *et al.* [38] proposed an intelligent system framework integrating forex forecasting and trading decision. An advanced intelligent decision support system (DSS) incorporating a back-propagation neural network-based forex forecasting subsystem and Web-based forex trading decision support subsystem was developed based on this framework. To assess the performance of the developed DSS, daily exchange rate data from 1990 to 2003 was trained and tested using the proposed model. The performance of four classical methods were compared to that of the model and evaluated on the RMSE. The results indicated superiority of the proposed model. Azad and Mahsin [39] developed a feed forward multilayer neural network using back propagation learning algorithm to predict the monthly average exchange rates of Bangladesh. They concluded that Neural Networks could approximate any continuous function and had better predictability than ARIMA model, evaluated by means of MAE, RMSE and MAPE.

Hybrid models have become popular in the field of financial forecasting in recent years since studies theoretically proved that a combination of multiple models could produce better results. Sermpinis *et al.* [40] once again investigated the performance of a hybrid neural network architecture comprising Particle Swarm Optimization and Adaptive Radial Basis Function (ARBF-PSO), a time varying leverage trading strategy based on Glosten, Jagannathan and Runkle (GJR) volatility forecasts and a neural network fitness function for FOREX rate forecasting. The forecast performance was evaluated on EUR/USD, EUR/GBP and EUR/JPY ECB exchange rates over the period of January 1999-March 2011 using the last 2 years for out-of-sample testing. The ARBF-PSO results were benchmarked with a Nearest Neighbours algorithm (k-NN), ARMA, a moving average convergence/divergence model (MACD) plus a naïve strategy. The ARBF-PSO architecture outperformed all other models in terms of statistical accuracy and trading efficiency for the three exchange rates. Falat, Marcek, and Durisova [41] proposed a new hybrid neural network, which is a combination of the standard Radial Basis Function (RBF) neural network, a genetic algorithm, and a moving average. The genetic algorithm was used to optimize the parameters of the neural network and the simple moving average was used to model the error of the RBF network in order to enhance the prediction outputs of the model. To determine the forecasting efficiency, they performed a comparative statistical one-day-ahead out-of-sample analysis of the tested model with autoregressive models and the standard neural network on USD/CAD exchange rate. The authors concluded that their suggested hybrid neural network is able to produce more accurate forecasts than the standard models and can be helpful in eliminating the risk of making a bad decision in forex trading. The recent success of deep networks is partially attributable to their ability to learn abstract features from raw data [42] .

4. Forecasting Models

Time-series models predict on the assumption that the future is a function of the past [43] .

4.1 Naive Model

4.1.1 Simple random walk model

Naïve (or Random Walk) Model is the most cost-effective forecasting model, and provides a benchmark against which more sophisticated models can be compared. This forecasting method is only suitable for time series data. A random walk is defined as a process where the current value of a variable is composed of the past value plus an error term defined as a white noise (a normal variable with zero mean and variance one). Algebraically a random walk is represented as follows:

$$y_t = y_{t-1} + \epsilon_t \qquad (1)$$

The implication of a process of this type is that the best prediction of y for next period is the current value, or in other words the process does not allow one to predict the change $y_t - y_{t-1}$. That is, the change of y is absolutely random. It can be shown that the mean of a random walk process is constant but its variance is not. Therefore a random walk process is non-stationary, and its variance increases with t. In practice, the presence of a random walk process makes the forecast process very simple since all the future values of y_{t+s} for $s > 0$, is simply y_t.

4.1.2 Random walk model with drift

A drift acts like a trend, and the process has the following form:

$$y_t = y_{t-1} + a + \epsilon_t \qquad (2)$$

For $a > 0$ the process will show an upward trend. This process shows both a deterministic trend and a stochastic trend. Using the general solution for the previous process,

$$y_t = y_0 + at + \sum_{t=1}^{n} \epsilon_t \qquad (3)$$

Where $y_0 + at$ is the deterministic trend and $\sum_{t=1}^{n} \epsilon_t$ is the stochastic trend.

The relevance of the random walk model is that many economic time series follow a pattern that resembles a trend model. Furthermore, if two time series are independent random walk processes, then the relationship between the two does not have an economic meaning.

4.2 The Autoregressive Moving Average (ARMA) Models

An *ARMA (p, q)* model is a combination of *AR (p)* and *MA (q)* models and is suitable for univariate time series modeling. In an *AR (p)* model the future value of a variable is assumed to be a linear combination of *p* past observations and a random error together with a constant term. A moving average*(MA)* model is a linear regression of the current observation of the time series against the random shocks of one or more prior observations. Autoregressive *(AR)* and moving average *(MA)* models can be effectively combined together to form a general and useful class of time series models, known as the *ARMA* models. Mathematically an *ARMA (p, q)* model is represented as:

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \epsilon_t$$
$$= c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \epsilon_t \qquad (4)$$

Here y_t and ϵ_t are respectively the actual value and random error (or random shock) at time period t, φ_i ($i = 1, 2 \ldots p$) are model parameters and c is a constant. The integer constant p is known as the order of the model. Sometimes the constant term is omitted for simplicity.

4.2.1 Autoregressive Integrated Moving Average (ARIMA) Models

In theory, these are the most general class of models for forecasting a time series. ARIMA models are applied in some cases where data show evidence of non-stationarity, (*A stationary series has no trend, its variations around its mean have a constant amplitude, and it wiggles in a consistent fashion),* where differentiation can be applied (corresponding to the "integrated" part of the model. This can be in conjunction with nonlinear transformations such as logging or deflating (if necessary). An *ARIMA* model can be viewed as a "filter" that tries to separate the signal from the noise, and the signal is then extrapolated into the future to obtain forecasts.

The ARIMA forecasting equation for a stationary time series is a *linear* (i.e., regression-type) equation in which the predictors consist of *lags of the dependent* variable and/or *lags of the forecast errors*. Mathematically:

$$\phi(L)(1-L)^d y_t = \theta(L)\epsilon_t, i.e \quad (5)$$

$$\left(1-\sum_{i=1}^{p}\varphi_i L^i\right)(1-L)^d y_t = (1+\sum_{j=1}^{q}\theta_j L^j)\epsilon_t \quad (6)$$

- Here, p, d and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model respectively.
- The integer d controls the level of differencing. Generally d=1 is enough in most cases. When d=0, then it reduces to an ARMA (p, q) model.

A useful generalization of ARIMA models is the Autoregressive Fractionally Integrated Moving Average (ARFIMA) model, which allows non-integer values of the differencing parameter d. ARFIMA has useful application in modeling time series with long memory The ARIMA model is for non-seasonal non-stationary data. Box and Jenkins have generalized this model to deal with seasonality. Their proposed model is known as the **Seasonal ARIMA (SARIMA) model**. In this model seasonal differencing of appropriate order is used to remove non-stationarity from the series.

4.3 Box-Jenkins Methodology

Statisticians George Box and Gwilym Jenkins developed a practical approach to building an ARIMA model, which best fit to a given time series and also satisfy the parsimony principle thattells us to choose the simplest scientific explanation that fits the evidence [29] .The Box-Jenkins methodology does not assume any particular pattern in the historical data of the series to be forecasted. Rather, it uses a three step iterative approach of *model identification*, *parameter estimation* and *diagnostic checking* to determine the best parsimonious model from a general class of ARIMA models.A useful device for initially assessing the values for p and q are the autocorrelation function (ACF) and partial autocorrelation function (PACF). The Box-Jenkins forecast method is schematically shown:

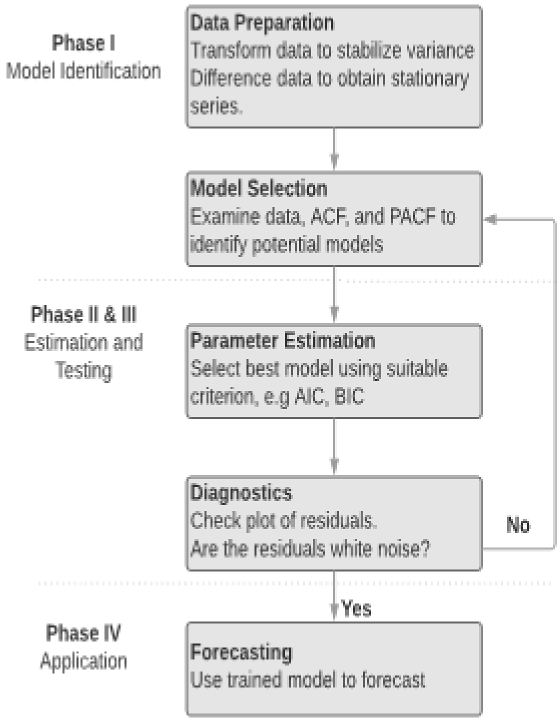

Figure 1: Box-Jenkins Methodology of identifying, fitting, checking, and using integrated autoregressive, moving average (ARIMA) time series models.

This three-phase process as shown in Fig.1 is repeated several times until a satisfactory model is finally selected. The best model is selected based on following criteria; smallest AIC (Akaike's information criteria) or SBC (Schwarz's information criteria), a minimum value of the standard error of the residuals and white noise (random) residuals of the model (which shows that there is no significant pattern left in the ACFs of the residuals). This model can then be used for forecasting future values of the time series.

4.4 Exponential Smoothing

Exponential smoothing was proposed in the late 1950s and has motivated some of the most successful forecasting methods. "Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, the more recent the observation the higher the associated weight" [44] . Naturally, a pattern does not always occur exactly the same as it did previously. Thus, a smoothing constant is included to reduce the random fluctuations that may exist within a pattern. There are three main types of exponential smoothing time series forecasting methods.

Single Exponential Smoothing, SES for short, also called Simple Exponential Smoothing, is a time series forecasting method for univariate data without a trend or seasonality. It requires a single parameter, called *alpha* (∝), as shown in Eq.8.

$$S_t = \alpha\frac{y_t}{I_{t-L}} + (1-\alpha)(S_{t-1} - b_{t-1}) \quad (7)$$

where

- □ = the smoothing constant *(0 ≤ ∝≤1)*
- y is the observation
- S is the smoothed observation
- b is the trend factor
- I is the seasonal index

Double Exponential Smoothing adds support for trends in the univariate time series. In addition to the *alpha* parameter for controlling smoothing factor for the level, an additional smoothing factor is added to control the decay of the influence of the change in trend called *beta* ((β).

$$b_t = \beta(S_t - S_{t-1}) + (1-\beta)b_{t-1} \qquad (8)$$

Where β is a constant that is chosen with reference to α. Like α it can be chosen through the Levenberg–Marquardt algorithm.

Triple Exponential Smoothing is an extension that explicitly adds support for seasonality to the univariate time series. This method is sometimes called Holt-Winters Exponential Smoothing, named for two contributors to the method: Charles Holt and Peter Winters. In addition to the alpha and beta smoothing factors, a new parameter is added that controls the influence on the seasonal component.

$$l_t = \gamma\frac{y_t}{S_t} + (1-\gamma)l_{t-L} \qquad (9)$$

$$F_{t+m} = (S_t + mb_t)l_{t-L+m} \qquad (10)$$

Where F is the forecast at *m* periods ahead.

4.5 Vector Autoregressive Models

A Vector autoregressive model of N endogenous variables $y_t = (y_{1t} \dots y_{nt} \dots y_{Nt})$ of order p, VAR(p), is defined as:

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t \qquad (11)$$

WhereA_i is an $(N \times N)$ matrix of autoregressive coefficients for $i = 1, \dots, p$. The $(Nx1)$ vector u_t is a vector generalization of white noise:

$$E(u_t) = 0$$

$E(u_t, u_\tau') = \begin{cases}\Omega \\ 0\end{cases}$ for $t = \tau$, otherwise

with Ω an $(N \times N)$ symmetric positive definitive matrix.

Unrestricted VAR models often have poor forecasting results because of overfitting.

Thus, as to make the forecasting result close to the real situation, we can give some restrictions to the parameters either using the RVAR or BVAR methods.

4.6 Nonlinear Time Series Models

Financial data display sudden bursts in volatility from time to time and linear models aren't able to capture such behavior. In order to capture nonlinear behaviour several nonlinear models have been proposed.Campbell, Lo, & MacKinlay [13] provide a review of these models. According to them almost all non-linear time series can be divided into two branches: models non-linear in mean and others non-linear in variance (Heteroskedasticity). As an illustrative example, here we present two nonlinear time series models:

Nonlinear Moving Average (NMA) Model: This model is non-linear in mean but not variance.

$$y_t = \varepsilon_t + \alpha\varepsilon_{t-1}^2 \qquad (12)$$

$\varepsilon_t, \varepsilon_{t-1}^2$ are the white noise error terms

Engle's (1982) ARCH Model: This model is heteroskedastic, i.e. non- linear in variance, but linear in mean. This model has several other variations, like GARCH, EGARCH etc.

$$y_t = \varepsilon_t + \alpha\sqrt{\varepsilon_t^2} \qquad (13)$$

4.6.1 Generalised Autoregressive Conditional Heteroskedasticity Models

The GARCH model [45] is a generalisation of the Autoregressive Conditional Heteroskedasticity (ARCH) model introduced by Engle in 1982. Instead of being conditionally fixed over time, the variance of the ARCH models is modelled as being dependent on lags of past squared residuals. Bollerslev [45] extended ARCH models to produce GARCH models in which the variance also depends on lags of past variances. The GARCH model can be considered as an infinite order ARCH model. A basic GARCH(p, q) model consists of a mean equation and a variance equation that can be represented as follows:

$$Y_t = \sum_{i=1}^{K}\beta_i y_{t-i} + \varepsilon_{3t}, \varepsilon \sim N(0, h_t), \qquad (14)$$

$$h_t = \alpha_0 + \sum_{i=1}^{p}\gamma_i h_{t-i} + \sum_{i=1}^{q}\lambda_i\varepsilon_{3t-i}^2 \qquad (15)$$

This simple GARCH model has a shortcoming. This model also restricts the impact of shock to be independent of its sign, whereas there is evidence of an asymmetric response for financial markets. In the basic ARCH model only squared residuals enter the conditional variance equation. Therefore, the signs of the residual or shocks have no influence on conditional volatility.

4.7 Multi-Layer Perceptrons

The use of Neural Networks in financial forecasting is not new and several researchers have successfully applied them to the task of identifying patterns in time series or estimating the profitability of technical trading rules. The most widely used ANNs in forecasting problems are multi-layer perceptrons (MLPs), which use a single hidden layer feed forward network (FNN). The model is characterized by a network of three layers, viz. input, hidden and output layer, connected by acyclic links. There may be more than one hidden layer.

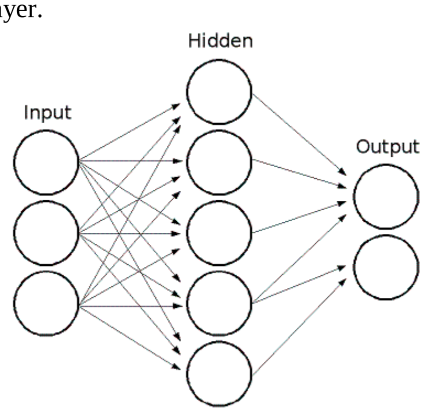

Figure 2: A three-layer perceptron with three inputs, two outputs, and the hidden layer of five neurons

All the neurons in MLP are similar. Each of them has several input links (it takes the output values from several neurons in the previous layer as input) and several output links (it passes the response to several neurons in the next layer). The values retrieved from the previous layer are summed up with certain weights, individual for each neuron, plus the bias term. The sum is transformed using the activation function f that may also be different for different neurons.

The output of the model is computed using the following mathematical expression

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y_{t-1}\right) + \varepsilon_t, \forall t \qquad (16)$$

In Eq. (17),

- $y_{t-i} (i = 1, 2, \ldots, p)$ are the p inputs and y_t is the output.
- The integers p, q are the number of input and hidden nodes respectively.
- $\alpha_j (j = 0, 1, 2, \ldots, q)$ and $\beta_{ij} (i = 1, 2, \ldots, p; j = 0, 1, 2, \ldots, q)$ are the connection weights and
- ε_t is the random shock; α_0 and β_{0j} are the bias terms.

Usually, the logistic sigmoid function $g(x) = \frac{1}{1+e^{-x}}$ is applied as the nonlinear activation function. Other activation functions, such as linear, hyperbolic tangent, Gaussian, etc. can also be used.

The feed forward ANN model in fact performs a non-linear functional mapping from the past observations of the time series to the future value, i.e. $y_t = f(y_{t-1}, y_{t-2}, \ldots\ldots y_{t-p}, w)$

Where w is a vector of all parameters and f is a function determined by the network structure and connection weights. To estimate the connection weights, non-linear least square procedures are used, which are based on the minimization of the error function.

$$F(\psi) = \sum_t e_t^2 = \sum_t (y_t - \hat{y}_t)^2 \qquad (17)$$

Ψ is the space of all connection weights in (17) above.

The optimization techniques used for minimizing the error function are referred to as *Learning Rules*. The best-known learning rule in literature is the *Backpropagation* or *Generalized Delta Rule*.MLP are notoriously good at detecting nonlinearities, but suffer from long training time and a very high number of alternatives as far as architectures and parameters go.

4.8 Radial Basis Function (RBF) Network

The RBF neural network is very different from other types of neural networks. It consists only of three layers, the input layer, hidden layer, and the output layer (Fig.3). Each node in the hidden layer calculates the distance between itself and the input and then a radial basis function changes the result. The final result of the radial basis function will be the output of the hidden layer and multiplies to a weight and is then fed to the output layer of RBF. In the output node all the inputs from the hidden layer are summed together and it yields the final output of the RBF neural network.

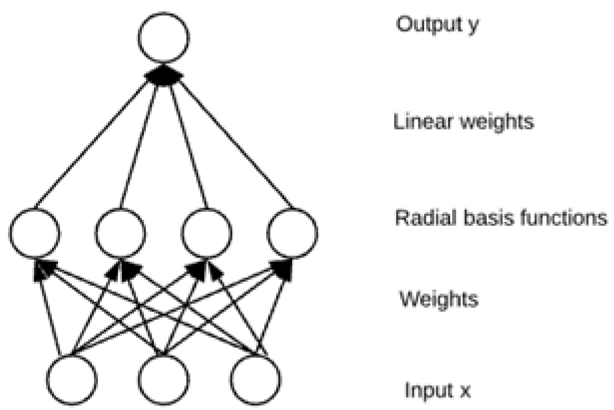

Figure 2: RBF architecture

The calculation procedure of an RBF neural network is presented below:

$$y_t = \alpha_0 + \sum_{i=1}^{n} w_i \phi(d_i) \qquad (18)$$

- w_i is the weight of the edge i which transfers the result of the hidden layer to the output layer.
- α_0 bias is a value which is added to the output at the end as an independent weight value. The function ϕ is the radial basis function and d_i is the distance between the input and the centre of the node in the hidden layer.

For calculating the distance, Euclidian distance can be used [46] .

$$d_i = \sqrt{\sum_{j=1}^{n} (x_j - c_{ij})^2} \qquad (19)$$

A desired network model should produce reasonably small error not only on within sample (training) data but also on out of sample (test) data [47] . Due to this reason immense care is required while choosing the number of input and hidden neurons. However, it is a difficult task as there is no theoretical guidance available for the selection of these parameters.A network with fewer weights is less complex than one with more weights. It is known that the “simplest hypothesis/model is least likely to overfit”. A network that uses the least number of weights and biases to achieve a given mapping is least likely to overfit the data and is most likely to generalize well on the unseen data [47] .

4.9 Support Vector Machines

Recently, a new statistical learning theory, viz. the Support Vector Machine (SVM) has been receiving increasing attention for classification and forecasting. Initially SVMs were designed to solve pattern classification problems, such as optimal character recognition, face identification and text classification, etc. But soon they found wide applications in other domains, such as function approximation, regression estimation and time series prediction problems.SVMs originate from Vapnik’s statistical learning theory [30] and is based on the *Structural Risk Minimization (SRM)* principle. The objective of SVM is to find a decision rule with good generalization ability through selecting some particular subset of training data, called support vectors.

Their major advantage over Neural Networks is that they formulate the regression problem as a quadratic optimization problem. SVMs perform by nonlinearly mapping the input data into a high dimensional feature space by means of a kernel function and then do the linear regression in the transformed space. The whole process results in nonlinear regression in the low-dimensional space [48] .

Given a dataset $G = \{(x_i,y_i)\}_{i=1}^{N}$ independently and randomly generated from some unknown function $g(x)$, a function f that approximates $g(x)$, based on the knowledge of G is as follows:

$$f(x) = \sum_i \omega_i \phi_i(x) + b \quad (20)$$

where $\phi_i(x)$ are the features, coefficients ω_i and b can be estimated by minimizing the regularized risk equation.

$$R(C) = C\frac{1}{N}\sum_{i=1}^{N} L(d_i, y_i)\frac{1}{2} + \|\omega\|^2 \quad (21)$$

$$R(C) = C\frac{1}{N}\sum_{i=1}^{N} L(d_i, y_i)\frac{1}{2} + \|\omega\|^2 \quad (21)$$

$$L(d,y) = \begin{cases} |d-y| - \varepsilon \, |d-y| \geq \varepsilon \\ 0 \ other, \end{cases} \quad (22)$$

In **SVM regression**, Vapnik proposed the use of ε -insensitive loss function, ε being known a priori.C and ε are prescribed parameters. C is called the regularization constant while ε is referred to as the regularization constant. $L(d,y)$ is the intensive loss function and the term $C\frac{1}{N}\sum_{i=1}^{N} L(d_i,y_i)$ is the empirical error while the $\frac{1}{2} + \|\omega\|^2$ indicates the flatness of the function.

Using the slack variables ζ and ζ^* the optimization problem solved by SVM under appropriate constraints depends on the finite number of parameters and has the following form.

$$f(x_i, \alpha_i, \alpha_i)^* = \sum_{i=1}^{N} (\alpha_i - \alpha_i)^* K(x, x_i) + b \quad (23)$$

$K(x, x_i)$ is the kernel function.

To avoid the explicit computation of the non-linear mapping φ(**x**), the associated kernel function must satisfy the *Mercer's Condition.* Below are some well-known kernels used in SVM literature:

The Linear Kernel

$$K(x,y) = x^T y \quad (24)$$

The Polynomial Kernel

$$K(x,y) = (1 + x^T y)^2 \quad (25)$$

The Radial Basis Function (RBF) Kernel

$$K(x,y) = \exp(-\|x - y\|^2 / 2\sigma^2) \quad (26)$$

The Neural Network Kernel

$$K(x,y) = \tanh a x^T y + b \quad (27)$$

where a, b are constants

$$W(\alpha, \alpha^*) = \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{N} (\alpha_i + \alpha_i^*) + \frac{1}{2}\sum_{i,j=1} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) \quad (28)$$

Equation (28) defines the optimization problem. The constraints $0 \leq \alpha_i^*, a_j \leq C, \sum_{i=1}^{N}(\alpha_i^* - \alpha_i) = 0$ .Only a number of coefficients satisfy $(\alpha_i^* - \alpha_i) \neq 0$ and the associated points to those coefficients are called support vectors.An LS-SVM formulation employs the equality constraints and a sum-squared error (SSE) cost function, instead of quadratic program in traditional SVM.

Another important characteristic of SVM is that here the training process is equivalent to solving a linearly constrained quadratic programming problem. So, contrary to other networks' training, the SVM solution is always unique and globally optimal. However a major disadvantage of SVM is that when the training size is large, it requires an enormous amount of computation, which increases the time complexity of the solution

5. Forecast Evaluation Methods

Many accuracy measures have been proposed to evaluate the performance of forecasting models. Error measurement statistics play a critical role in estimating forecast accuracy, monitoring for outliers, and benchmarking the forecasting process. The accuracy of forecasts can only be determined by considering how well a model performs on new data that was not used when estimating the model i.e. splitting the time series into train and test sets. It is well known that no single measure is superior to all others. Commonly used measures include:

5.1Scale dependent measures

The measures are useful in comparing forecasting methods on the same set of data, having the same scale. Commonly used scale dependent measures are Mean Absolute Error (MAE), Mean Squared Error (MSE) and RMSE [49] :

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \quad (29)$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \quad (30)$$

$$RMSE = \sqrt{MSE} \quad (31)$$

5.2 Percentage-based measures

To be scale-independent, a common approach is to use percentage errors based on observation values. Two example measures based on percentage errors are MAPE and sMAPE defined as:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|e_t|}{|Y_t|} \quad (32)$$

$$sMAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{2.|e_t|}{|Y_t| + |\square_t|} \quad (33)$$

5.3 Relative-based measures

Another approach for accuracy measures to be scale-independent is to use relative errors based on the errors produced by a benchmark method (e.g. the naïve method).

The most commonly used are Mean Relative Absolute Error (MRAE) [49] .

$$MRAE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{e_t}{e_t^*}\right| \quad (34)$$

6. Conclusion

This article critically analyses some recent research on exchange rate prediction and the various methodologies that have been applied to disprove the efficient market hypothesis. It is clear that Research on forecasting exchange rates to date has mainly focused on advanced economies, the big players in the FOREX market. The significance of exchange rate forecasts cannot be understated. Accurate prediction of exchange rates is critical for devising robust monetary policies. It can be a tool for correcting internal and external imbalances as well as an instrument of improving the efficiency of resource allocation. For developing economies, prediction performance can outline exchange rate instability. This constitutes a serious hindrance to business planning and growth and if detected early, fiscal policies like adopting an exchange rate band can be implemented in order to minimise volatility. Research shows that SVM, Deep learning methods such as Feedforward Neural Network and hybrid ensembles produce higher predictive accuracy than traditional time series models. Future research should be carried out to assess the performance of these models, however, no single forecasting model consistently stands out as the best when assessed by different criteria and on different currency pairs and decisions made based on the models predictions should be used with caution.

References

[1] R. A. DeFusco, *Quantitative investment analysis*. 2007.
[2] Wang Peijie, "The Economics of Foreign Exchange and Global Finance, " 2009.
[3] M. R. Rosenberg and D. Folkerts-Landau, "The Deutsche Bank Guide to Exchange Rate Determination, " *DB Res.*, 2002.
[4] J. J. Murphy, "Technical Analysis Of The Financial Markets, " *Pennsylvania Dental Journal*. 1999.
[5] M. Osinska, "Quality of the Forecasts of Currencies Exchange Rates Volatilities, " *Int. Res. J. Financ. Econ.*, 2010.
[6] R. A. Meese and K. Rogoff, "Empirical exchange rate models of the seventies. Do they fit out of sample?, " *J. Int. Econ.*, 1983.
[7] P. G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model, " *Neurocomputing*. 2003.
[8] P. Kaufman, *Trading Systems and Methods*. 2005.
[9] M. Mussa, "Empirical regularities in the behavior of exchange rates and theories of the foreign exchange market, " *Carnegie-Rochester Confer. Ser. Public Policy*, 1979.
[10] N. C. Mark, "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability, " *Am. Econ. Rev.*, vol. 85, no. 1, pp. 201–218, 1995.
[11] J. Berkowitz and L. Giorgianni, "Long-horizon exchange rate predictability?, " *Rev. Econ. Stat.*, 2001.
[12] L. Kilian and M. P. Taylor, "Why is it difficult to beat the random walk forecast of exchange rates?, " *J. Int. Econ.*, 2003.
[13] J. Y. Campbell, A. W. Lo, and A. C. MacKinlay, "Event-Study Analysis, " in *The Econometrics of Financial Markets*, 1997.
[14] R. Adhikari and R. K. Agrawal, "An Introductory Study on Time Series Modeling and Forecasting, " 2013.
[15] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: Forecasting and control: Fourth edition*. 2013.
[16] H. B. Hwarng and H. T. Ang, "A simple neural network for ARMA(p, q) time series, " *Omega*, 2001.
[17] M. K. Newaz, "COMPARING THE PERFORMANCE OF TIME SERIES MODELS FOR FORECASTING EXCHANGE RATE, " 2008.
[18] D. Eva and F. C. Maria, "Exchange-Rates Forecasting: Exponential Smoothing Techniques and Arima Models, " *Ann. Univ. Oradea Econ. Sci.*, 2011.
[19] P. Maniatis, "Forecasting the exchange rate between euro and USD: Probabilistic approach versus ARIMA and exponential smoothing techniques, " *J. Appl. Bus. Res.*, vol. 28, no. 2, pp. 171–192, 2012.
[20] O. K. Onasanya and * Adeniji, "FORECASTING OF EXCHANGE RATE BETWEEN NAIRA AND US DOLLAR USING TIME DOMAIN MODEL, " 2013.
[21] L. Ghalayini, "Modeling and Forecasting the US Dollar/Euro Exchange Rate, " *Int. J. Econ. Financ.*, vol. 6, no. 1, pp. 194–207, 2013.
[22] O. M. Olatunji and A. Bello, "A Suitable Model for the Forecast of Exchange Rate in Nigeria (Nigerian Naira versus US Dollar), " *Int. J. Sci. Res. 2015*, vol. 4, no. 5, pp. 2669–2676, 2015.
[23] A. Carriero, G. Kapetanios, and M. Marcellino, "Forecasting exchange rates with a large Bayesian VAR, " *Int. J. Forecast.*, 2009.
[24] J. C. Cuaresma and J. Hlouskova, "Forecasting exchange rates in transition economies: A comparison of multivariate time series models, " *Empir. Econ.*, 2004.
[25] R. Kunst, "Multivariate Forecasting Methods, " *Building*, pp. 28–39, 2000.
[26] T. Teräsvirta, "An introduction to univariate GARCH models, " 2015.
[27] M. Pahlavani and R. Roshan, "The Comparison among ARIMA and hybrid ARIMA-GARCH Models in Forecasting the Exchange Rate of Iran, " *Int. J. Bus. Dev. Stud.*, 2015.
[28] M. Y., T. M, and A. Bello, "Forecasting of Exchange Rate Volatility between Naira and US Dollar Using GARCH Models, " *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 4, no. 7, pp. 369–381, 2014.
[29] F. Black, "Studies in stock price volatility changes, " *Proceedings of the 1976 American Statistical Association, Business and Economical Statistics Section*. 1976.
[30] Y. Wu and V. N. Vapnik, "Statistical Learning Theory, " *Technometrics*, 1999.
[31] L. J. Cao and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting, " *IEEE Trans. Neural Networks*, 2003.
[32] L. Lixia and W. Wenjing, "Exchange rates forecasting

with least squares support vector machine, ” in *Proceedings - International Conference on Computer Science and Software Engineering, CSSE 2008*, 2008.
[33] C. Fu, “Forecasting exchange rate with EMD-based Support Vector Regression, ” in *2010 International Conference on Management and Service Science, MASS 2010*, 2010.
[34] C. S. Lin, S. H. Chiu, and T. Y. Lin, “Empirical mode decomposition-based least squares support vector regression for foreign exchange rate forecasting, ” *Econ. Model.*, 2012.
[35] G. Sermpinis, C. Stasinakis, K. Theofilatos, and A. Karathanasopoulos, “Modeling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms - Support vector regression forecast combinations, ” *Eur. J. Oper. Res.*, 2015.
[36] V. Plakandaras, T. Papadimitriou, and P. Gogas, “Forecasting Daily and Monthly Exchange Rates with Machine Learning Techniques, ” *J. Forecast.*, 2015.
[37] G. E. Hinton, S. Osindero, and Y. W. Teh, “A fast learning algorithm for deep belief nets, ” *Neural Comput.*, 2006.
[38] L. Yu, S. Wang, K. K. Lai, and W. W. Huang, “Developing and assessing an intelligent forex rolling forecasting and trading decision support system for online e-Service, ” *Int. J. Intell. Syst.*, 2007.
[39] A. K. Azad and M. Mahsin, “Forecasting Exchange Rates of Bangladesh using ANN and ARIMA Models: A Comparative Study, ” *Int. J. Adv. Eng. Sci. Technol.*, 2011.
[40] G. Sermpinis, K. Theofilatos, A. Karathanasopoulos, E. F. Georgopoulos, and C. Dunis, “Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and Particle Swarm Optimization, ” *Eur. J. Oper. Res.*, 2013.
[41] L. Falat, D. Marcek, and M. Durisova, “Intelligent Soft Computing on Forex: Exchange Rates Forecasting with Hybrid Radial Basis Neural Network, ” *Sci. World J.*, 2016.
[42] S. Galeshchuk and S. Mukherjee, “Deep networks for predicting direction of change in foreign exchange rates, ” *Intell. Syst. Accounting, Financ. Manag.*, 2017.
[43] Jay Heizer and Barry Render, “Principles of Operations Management, ” *Genes & genetic systems*. 2006.
[44] J. Hyndman, Rob and G. Athanasopoulos, *Forecasting: Principles and Practice: Notes*. 2014.
[45] T. Bollerslev, “Bollerslev, Generalized Autoregressive Conditional Heteroskedasticity.pdf, ” vol. 31, pp. 307–327, 1986.
[46] A. Kia, M. Fathian, and M. Gholamian, “Using MLP and RBF Neural Networks to Improve the Prediction of Exchange Rate Time Series with ARIMA, ” *Int. J. Inf. …*, vol. 2, no. 4, pp. 2–5, 2012.
[47] J. Kamruzzaman and R. A. Sarker, “ANN-Based Forecasting of Foreign Currency Exchange Rates, ” *Neural Inf. Process. Rev.*, vol. 3, no. 2, pp. 49–58, 2004.
[48] J. Kamruzzaman, R. A. Sarker, and I. Ahmad, “SVM based models for predicting foreign currency exchange rates, ” in *Third IEEE International Conference on Data Mining*, 2003.
[49] C. Chen, J. Twycross, and J. M. Garibaldi, “A new accuracy measure based on bounded relative error for time series forecasting, ” *PLoS One*, 2017.

Author Profile

Oyiwona Ajumi is currently in the final phase of completing her Masters degree in Data Analytics from Dublin Business School, Ireland. She received a Bachelor of Science in Computer Engineering from Saint Petersburg State University (2016), Russia. Research interests include theory and applications of Data mining, Deep Learning and Machine Learning.

Abhishek is currently a lecturer in Dublin Business School. He is doing his research in ADAPT Research Lab (DCU), Ireland. His research interests include theory and applications of Information Retrieval Data mining, Deep Learning, and Machine Learning.