T.R.
GEBZE INSTITUTE OF TECHNOLOGY
INSTITUTE OF SOCIAL SCIENCES

A COMPARATIVE ANALYSIS OF CLASSICAL MODELS AND ARTIFICIAL INTELLIGENCE BASED MODELS IN FORECASTING EXCHANGE RATES

ALİ FEHİM CEBECİ
MASTER'S THESIS
DEPARTMENT OF ECONOMICS

GEBZE
2014

T.R.
GEBZE INSTITUTE OF TECHNOLOGY
INSTITUTE OF SOCIAL SCIENCES

A COMPARATIVE ANALYSIS OF CLASSICAL MODELS AND ARTIFICIAL INTELLIGENCE BASED MODELS IN FORECASTING EXCHANGE RATES

ALİ FEHİM CEBECİ
MASTER'S THESIS
DEPARTMENT OF ECONOMICS

THESIS SUPERVISOR
PROF. DR. HÜSEYİN İNCE

GEBZE
2014

YÜKSEK LİSANS TEZİ JÜRİ ONAY SAYFASI

G.Y.T.E. Sosyal Bilimler Enstitüsü Yönetim Kurulu'nun 23/06/2014 tarih ve 2014/20 sayılı kararıyla oluşturulan jüri tarafından 21/07/2014 tarihinde tez savunma sınavı yapılan Ali Fehim CEBECİ'nin tez çalışması İktisat Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir.

JÜRİ

ÜYE
(TEZ DANIŞMANI) : Prof. Dr. Hüseyin İNCE

ÜYE : Prof. Dr. Salih Zeki İMAMOĞLU

ÜYE : Yrd. Doç. Dr. Mesut KARAKAŞ

ONAY

G.Y.T.E. Sosyal Bilimler Enstitüsü Yönetim Kurulu'nun ………/………/20… tarih ve ……………/………… sayılı kararı.

İMZA/MÜHÜR

ÖZET

Mesee ve Rogoff (1983) tarafından da belirtildiği gibi ekonomi temelli ekonometrik yöntemler ve zaman serileri yöntemleri gibi klasik yöntemler, döviz kuru tahmininde; döviz kurunun öngörülemez bir rassal süreç izlediğini belirten rassal yürüyüş modeline nazaran kayda değer bir başarı sergileyememişlerdir. Özellikle son 20 yılda, bilgi ve bilgisayar teknolojilerinde gözlenen hızlı gelişmeler sayesinde yapay zeka temelli yöntemler döviz kuru tahmininde umut vadeden yöntemler olarak yaygın bir şekilde kullanılmaya başlanmıştır. Bu çalışmanın amacı üç farklı Türk Lirası döviz kuru (TL/USD, TL/EUR, TL/GBP) için 2004-Ocak ve 2013-Aralık ayı arası 120 gözlemden oluşan aylık verileri kullanarak, klasik yöntemler ve yapay zeka temelli yöntemlerin tahmin performanslarının karşılaştırılmasıdır. İlgili yöntemlerin; bir adım öte, gözlem içi ve gözlem dışı tahmin performansları; kök ortalama kare hatası istatistiği ve Diebold-Mariano test yoluyla karşılaştırmalı olarak incelenmiştir. Gerçekleştirilen analiz neticesinde; yapay zeka temelli yöntemlerin gözlem-içi tahmin performansı açısından, klasik yöntemlere göre kayda değer bir şekilde daha başarılı olduğunu görülmüştür. Öte yandan klasik yöntemler, gözlem-dışı tahmin performansı açısından analize konu olan bir çok durumda yapay zeka temelli yöntemlere göre istatistiki olarak anlamlı bir şekilde daha başarılı olmuştur. Bunlara ek olarak analizde kullanılan yedi farklı model gözlem-dışı tahmin performansı açısından incelenen yirmi bir durumdan onunda, rassal yürüyüş modelinden istatistiki olarak anlamlı şekilde başarılı bir tahmin performansı sergilemiştir. Rassal yürüyüş modeli ise hiçbir durumda söz konusu modellerden istatistiki olarak anlamlı bir şekilde daha başarılı bir gözlem-dışı tahmin performansı sergileyememiştir.

Anahtar Kelimler: Döviz kurları, zaman serileri tahmini, yapay sinir ağları

SUMMARY

As emphasized by Meese and Rogoff (1983) in their seminal paper; economic fundamentals based models and time series models (i.e. classical models) are unable to significantly outperform a random walk model, implying that exchange rates behave in a purely random and unpredictable manner. During the last decades with the rapid advancements in computer and information technologies artificial intelligence based models came into use in forecasting exchange rates as a promising forecasting tool. This paper aims to investigate predictive accuracy of classical models and artificial intelligence based models in forecasting three different Turkish Lira exchange rates (TL/USD, TL/EUR, TL/GBP), using monthly period data from January 2004 to December 2013 with 120 observations. One step ahead, in-sample and out-of-sample forecasting accuracy of each model analyzed comparatively by utilizing root mean squared error statistics and Diebold-Mariano test. Analysis results proposed that; artificial intelligence based models performed significantly better for in-sample forecasts. Besides, for out-of-sample forecasts; classical models performed statistically significantly better than artificial intelligence based models in most instances. Furthermore, in ten out of twenty-one instances all seven forecasting models in consideration (both classical models and artificial intelligence based models) were capable of beating random walk model; while random walk model was not capable of statistically significantly beating any of these seven models in consideration.

Keywords: Exchange rates, time series forecasting, artificial neural networks

ACKNOWLEDGEMENTS

I would like to thank to my supervisor Prof. Dr. Hüseyin İnce for his valuable comments, advices and helpful guidance throughout the research.

I would also thank to my parents and family for their patience, tolerance and kind encouragements.

TABLE OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

Symbols and Abbreviations		Description
Δ	:	Difference Operator
$\sum$	:	Summation Operator
exp	:	Exponential Function
ε, u	:	Error Terms
$\beta, \alpha, \phi, \rho$	:	Coefficients
R	:	Correlation Coefficient
ARIMA	:	Autoregressive Integrated Moving Average
ARCH	:	Autoregressive Conditional Heteroscedasticity
GARCH	:	Generalized Autoregressive Conditional Heteroscedasticity
VAR	:	Vector Autoregressive
ECM	:	Error Correction Model
ANN	:	Artificial Neural Network
MLFFNN	:	Multilayer Feedforward Neural Network
RNN	:	Recurrent Neural Network
PNN	:	Probabilistic Neural Network
SVR	:	Support Vector Regression
LSSVR	:	Least Squares Support Vector Regression
RW	:	Random Walk

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

After the collapse of the Bretton-Woods system in 1973, fixed exchange rate regime obliged by the system disappeared and countries became vulnerable to sudden and drastic movements in exchange rate levels. In today's overwhelmingly globalized, economically integrated world, exchange rate between currencies have become increasingly crucial and have profound impacts on all levels of the economy. Because of such reasons, accurate and reliable forecasting of exchange rates emerged as an important necessity.

During the last decades exerting exchange rate movements, developing a model for forecasting exchange rates have been a dynamic and active research area in the literature. Many models have been proposed in economics literature, attempting to unveil exchange rate dynamics. These models are termed as economic fundamentals based models since they associate exchange rate behavior with macroeconomic variables. In addition to these standard econometric models with macroeconomic fundamentals, univariate and multivariate pure time series models are employed to exchange rate series for forecasting purposes. Meese and Rogoff (1983) in their seminal paper by conducting a comparative analysis of economic fundamentals based models and time series models (i.e. classical models) presented that these models are unable to significantly outperform a random walk model which implies that exchange rates behave in a purely random and unpredictable manner. This phenomenon is called in the literature as Meese-Rogoff puzzle. Subsequent, sophisticated empirical studies were also unable to strongly justify a solution to Mesee-Rogoff puzzle and prove that classical models are capable of beating naive random walk model.

Such dismal performance of classical models is mainly attributed to unrealistic assumptions of these models like; linear dependency between independent and dependent variables and assumptions regarding the data generating process that does not comply with real world situations. During the last two decades with the rapid advancements in computer and information technologies artificial intelligence based models came into use in forecasting exchange rates. Artificial intelligence based models don't involve any assumptions about data generating process. They are capable of dealing with nonlinear relationships. They extract the relationship between variables by learning from data and base their forecasts on these nonlinear approximations. These distinctive features made them a promising and attractive forecasting tool.

This paper conducts a comparative analysis of classical models such as; autoregressive moving average (ARIMA) models, autoregressive conditional heteroscedasticity (ARCH) models, vector autoregressive (VAR) models, error correction models (ECM) and artificial intelligence based models such as; multilayer feedforward neural networks (MLFFNN), recurrent neural networks (RNN), support vector regression (SVR) in forecasting Turkish Lira/U.S. Dollar (TL/USD), Turkish Lira/Euro (TL/EUR), Turkish Lira/British Pound (TL/GBP) exchange rates by using monthly observations from January 2004 to December 2013 of all three exchange rates, monetary aggregates of relevant countries and one year interbank borrowing rate of each currency.

The rest of this study is organized as follows; section 2 gives theoretical foundations of economic fundamentals based models of exchange rate determination (purchasing power parity, the balance of payments flow approach, the monetary approach, the portfolio balance approach); section 3 describes the econometric, (ARIMA models, ARCH models, VAR models, ECM models) and artificial intelligence based models (MLFFNN, RNN, PNN, SVR) that are commonly used in financial time series forecasting. Then in section 4 these models are employed to Turkish Lira exchange rates in order to compare forecasting performance of each model and to assess their predictive ability. Concluding remarks derived through analyses are provided in the final section.

2. ECONOMIC FUNDAMENTALS BASED MODELS OF EXCHANGE RATE DETERMINATION

2.1. Purchasing Power Parity

Origins of purchasing power parity date back to sixteenth century workings of the scholars of Salamanca school, Spain. Purchasing power parity, in modern context, emerged as a response to the financial issues brought by the World War I. During the war, many countries suspended or abandoned the gold standard in which exchange rate between currencies reflected their relative gold values. This raised the problem of how to set exchange rates without disturbing too much the prices and government finances (Rogoff, 1996). Swedish economist Gustav Cassel in his work “The World’s Monetary Problems” suggested use of purchasing power parity as a tool to solve the problem. He stated that in absence of the restrictions to international trade, exchange rate between two countries’ currencies adjust to general price level of the each country. Hence, exchange rates are determined by the purchasing powers of the currencies and purchasing power parities represent the true equilibrium of the exchange rates (Cassel, 1921, pp. 36-38). Relationship between exchange rates and price levels can be explained with two different approaches called; absolute and relative purchasing power parity. They are both based on the law of one price, which states that under the assumptions of (i) relevant commodities are tradable, (ii) there are no restrictions to trade, (iii) there are no transaction costs, (iv) relevant commodities are perfectly homogeneous, these commodities should trade at a same price across countries when converted in a single currency. In the absolute version, exchange rate between two currencies is determined by the ratio of price levels of these countries. Relative version is based on absolute version but the only difference is that it uses percentage changes rather than absolute differences. According to relative version of purchasing power parity; percentage change in exchange rate of currencies is determined by the difference of percentage change in price levels of relevant countries (Rosenberg, 1996, pp. 12, 14).

Validity of purchasing power parity has long been focus of debate. Frenkel (1978, 1981), in his two highly influential papers regarding the phenomenon found that purchasing power parity relation was generally valid for the period between the 1920s to 1970 but the relation nearly collapsed for the 1970s. Frenkel attributes this to the uncertain character of the 1970s and sharp, frequent changes of expectations resulting from real shocks like the oil crisis, supply shocks and shortages. Studies such as Adler and Lehmann (1983), Taylor (1988) argued that exchange rate behavior characterized by a random walk process rather than the behavior implied by purchasing power parity doctrine. Recently, research on the subject has questioned whether exchange rates follow a random walk process or they gravitate along some mean value and follow a mean reverting process. If exchange rates follow a mean reverting process along the path implied by purchasing power parity doctrine, then it can be traced that purchasing power parity is valid at least as a long run phenomenon. Jorion and Sweeney (1996) and Augustine et al. (2010) derived results that exchange rates tend to revert a long run equilibrium mean value while Baum et al. (1999) and Cushman (2008) derived results against long run equilibrium, mean reverting process.

2.2. The Balance of Payments Flow Approach

Balance of payments flow approach basically focuses on the role of international trade in determination of exchange rates. It was first introduced by the works of economists; Joan Robinson, Frizt Machlup and Gottfried Haberler (Rosenberg, 1996; pp. 69). Joan Robinson stated in "Essays in the Theory of Employment" for the theory of foreign exchange that the exchange rate is determined by supply and demand of home currency in terms of foreign currency. The demand for a foreign currency in terms of home currency (or the supply of home currency made available to the exchange market) mainly arises from the need to pay for goods and services purchased from foreigners and in order to make loans or purchase securities abroad. Demand for foreign currency arises from the domestic demand for foreign goods, services and financial instruments. In a similar vein, supply of foreign currency arises from the foreign demand for domestic goods, services and financial instruments (Robinson, 1947). Hence, if the international trade is balanced, i.e. domestic demand for foreign goods, services and financial instruments equals foreign demand for domestic goods, services and financial instruments then the supply and demand of foreign currency should be in equilibrium as well. An imbalanced trade or disequilibrium gives rise to excess supply or excess demand for foreign exchange. In this case exchange rate adjusts until the trade imbalance is vanished and the equilibrium is restored. For instance if the exchange rate is at a level corresponding to a disequilibrium, domestic currency is overvalued where demand for foreign currency exceeds supply of foreign currency, domestic country would be running a trade deficit and foreign country would be running a trade surplus. Excess demand for foreign currency drives the exchange rate up, domestic currency depreciates and the foreign currency appreciates, until the equilibrium is restored. The amount and the duration of the adjustment depends on the both domestic and foreign countries' elasticities of demand for exports and imports (Rosenberg, 1996, pp. 68-71).

2.3. The Monetary Approach

Monetary approach to exchange rate determination relates impacts of monetary policy to exchange rates; under the assumptions that; (i) money demand is a stable function of level of income and interest rates, (ii) purchasing power parity relation holds, (iii) arbitrage conditions implied by uncovered interest parity relation, i.e. interest yield of a domestic bond equals the interest yield on a foreign bond adjusted for the expected change in the exchange rate, holds each point in time. (Bilson, 1978, pp. 49). Its basic version; flexible price monetary model was first introduced by Frenkel (1976). Frenkel stated foundations of the model as; exchange rate, being relative price of two currencies, could be stated in terms of supply and demand for these currencies since the equilibrium exchange rate is attained when existing stock of these currencies are willingly held (Frenkel, 1976, pp. 201). Monetary approach, in its basic form, can be represented by following equation:

$$e = (m - m^*) - b_1(y - y^*) + b_2\left(\dot{m}^e - \dot{m}^{e^*}\right) \tag{2.1}$$

Where m and m^* are the log form of supply of money both in the domestic and foreign country respectively; b_1 and b_2 are income and interest rate elasticities of money demand which are assumed to be equal in domestic and foreign country; y

and y^* are the log form of level of income in domestic and foreign country respectively; $\dot{m}^e$ and $\dot{m}^{e^*}$ are expected future monetary growth rates in domestic and foreign country respectively. Transmission mechanism implied by the model can be summarized as follows: (i) a change in the money supply transforms to a change in the price level as proposed by the quantity theory of money and resulting change in price level transforms to a change in exchange rate through purchasing power parity relation (ii) in monetary model interest rate differentials are assumed to reflect differences in inflation rates which in turn reflects the differences in expected future monetary growth rates; if domestic country experiences a rise in domestic interest rate, that is sufficient to result a rise in interest rate differential, this means that inflationary expectations have risen in domestic country which imply that expected future monetary growth rate has relatively risen and this gives rise to depreciation of the domestic currency. Hence; in monetary model value of the domestic currency moves in the opposite direction of interest rate differential as opposed to the way popularly thought. Over time, various versions of the model were developed. These alternative versions termed in the literature as; sticky price version, real interest rate differential version, equilibrium real exchange rate version and exchange market pressure version (Rosenberg, 1996, pp. 140-143). Variations of the monetary model can be summarized by using following equation:

$$s = a_0 + a_1(m - m^*) + a_2(y - y^*) + a_3(r_s - r_s^*) + a_4\left(\pi^e - \pi^{e^*}\right) + a_5\left(\overline{TB} - \overline{TB}^*\right) + u \qquad (2.2)$$

where s is the logarithm of spot exchange rate between domestic and foreign country currencies; $m - m^*$ is the logarithm of domestic and foreign country money supplies in differential form, $y - y^*$ is the ratio of logarithm of domestic and foreign country real income levels, $r_s - r_s^*$ is the short term interest differential between domestic and foreign countries, π^e and π^{e^*} are the excpected future inflation rates in domestic and foreign countries respectively, $\overline{TB}$ and $\overline{TB}^*$ are the cumulated foreign trade balances of domestic and foreign countries and u is the error term. All variations of monetary model assume that $a_1 = 1$. The Frenkel-Bilson monetary model restricts $a_3 > 0$ and by assuming that the purchasing power parity holds, sets $a_4 = a_5 = 0$. The Dornbush-Frankel model by restricting $a_3 < 0$ and $a_5 = 0$ allows for deviations from purchasing power parity relation. Hooper-Morton monetary model by including the long run effect of trade balance on exchange rates extends the Dornbush-Frankel monetary model. None of the coefficients is set to zero in this model and they are restricted as $a_2 < 0$, $a_3 < 0$, $a_5 < 0$ (Meese and Rogoff, 1983; Kim and Mo, 1995).

There exists a huge literature regarding empirical tests of the various versions of the monetary model and its predictive ability. Studies following early successful implementation of monetary model by Frenkel (1976) and Bilson (1978) produced unsatisfactory results. Meese and Rogoff (1983) in their distinctive paper found that a random walk model outperforms monetary model in every forecasting horizons. Recent studies claimed that; adjustment of exchange rates to level implied by the monetary models is governed by a nonlinear adjustment process. Theoretical models are fundamentally sound but the linear methods used for estimating these models were the major cause of dismal empirical performance until then (Killian and Taylor, 2003). Taylor and Peel (2000), Wu and Chen (2001), Kim et al. (2010), Junttila and Korhonen (2011), Beckmann (2013) presented that in contrast to studies based on linear methods, predictability of the monetary model increases when the nonlinear adjustment process is taken into account.

2.4. The Portfolio Balance Approach

Similar to the monetary approach, the portfolio balance approach links determination of exchange rates with interaction of supply and demand for financial assets but in a much broader sense. In portfolio balance approach exchange rates are determined not only by relative money supplies but also by relative bond supplies. Isard (1995) gives a brief, coherent background of its theoretical framework.

According to portfolio balance approach the menu of assets that a resident of a country can allocate his/her wealth is as follows: (i) a non-interest bearing asset; reserve money of central bank (M), (ii) domestic bonds that are issued by government and held by domestic private sector asset holders (B), (iii) foreign bonds that are issued by foreign governments and held by domestic private sector asset holders (F).

$$W = M + B + eF \tag{2.3}$$

Net wealth of a domestic asset holder equals to sum of the amount of reserve money held, value of domestic government bonds and value of foreign government bonds in denominated in domestic currency, i.e. its value is multiplied by the exchange rate (e) (see equation 2.3). The way monetary policy, fiscal policy and current account imbalances affect exchange rate in portfolio balance approach can be exemplified as follows; (i) If central bank of the domestic country pursues a expansionary monetary policy through an open market purchase; with the resulting swap between reserve money (M) and domestic bonds (B) domestic wealth remains constant. Such a monetary expansion, leave domestic asset holders with excess supply of reserve money and an equivalent excess demand for domestic bonds. These forces simultaneously drive the domestic interest rate downward and domestic currency depreciates as domestic asset holders attempt to substitute from domestic bonds to foreign bonds. (ii) If the government pursues an expansionary fiscal policy which is financed through issuing government bonds; resulting rise in the supply of domestic bonds (B) induces a rise in domestic wealth (W). While it is obvious that a rise in domestic interest rate is necessary to eliminate excess demand for money, net effect of the policy on exchange rates is ambiguous. Rise in wealth of the domestic asset holders (resulting from rise in the domestic bonds) will leave them with an excess supply for domestic bonds and an excess demand for remaining two components of wealth; reserve money and foreign bonds. While excess demand for foreign bonds drives the value of domestic currency downwards; the rise in domestic interest rate drives the value of domestic currency upwards by inducing domestic asset holders to substitute foreign bonds to domestic bonds. Net effect on exchange rate depends on which force will dominate. (iii) Assuming that a country runs a current account surplus and consequently foreign bonds that are held by the domestic asset holders rise. Such a rise in foreign bonds leaves domestic asset holders with an excess supply of foreign bonds and an excess demand for domestic bonds and reserve money. Excess demand for domestic bonds can only be offset if domestic interest rates decline while excess demand for reserve money can only be offset if the domestic interest rates rise. Since it is impossible for interest rates to rise and decline at the same time; the only way that the portfolio balance can be ensured is that the exchange rate (e) to decline by the same proportion as the rise in foreign bonds (F) and hence holding domestic currency value of foreign bonds (eF) constant (Rosenberg, 1996, pp. 188-197; Enders, 1977)

3. EXCHANGE RATE FORECASTING MODELS

3.1. Econometric Models

3.1.1. Autoregressive Integrated Moving Average (ARIMA) Models

Structure of an ARIMA (p, d, q) model which represents a combination of a autoregressive model (AR) of order; p, moving average model (MA) of order; q, integrated of order d, was first introduced by Box and Jenkins (1976). Box and Jenkins, compiled and integrated existing knowledge on time series forecasting and delivered a coherent, unique; three stage iterative process for model identification, parameter estimation and diagnostic checking. This iterative procedure is regarded in the literature as Box-Jenkins methodology. During the last half a century ARIMA models have been frequently used in many areas of time series forecasting. In an ARIMA (p, d, q) process, the future value of a variable is assumed to be a linear combination of its past values and past error terms which has the following form:

$$\begin{aligned} y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} \\ +\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \end{aligned} \tag{3.1}$$

where y_t is the value of the variable at time t, ε_t is random error term at time t which is independent and identically distributed with mean 0 and variance σ^2, θ_i and ϕ_j are coefficients, p and q are autoregressive and moving average polynomials respectively (Hamilton, 1994, pp. 58). For instance, an ARIMA (1, 0, 1) process can be represented as follows:

$$y_t = \theta_0 + \phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \tag{3.2}$$

It is a prerequisite to warrant that the data sample is stationary; in order to apply Box-Jenkins methodology and to estimate coefficients of the series. There are two main methods for detecting nonstationarity: (i) Graphical inspection: visual representation of stationary series against time, exhibit a smooth looking pattern, evenly distributed around a constant mean with no trend or seasonal deviation. It is not easy to judge whether the series are stationary or not by visual inspection alone. A more useful tool is to analyze correlogram which represent behavior of autocorrelation and partial autocorrelation functions of the series. A white noise stationary process has autocorrelation and partial autocorrelation value which is not statistically significantly different from zero for arbitrarily chosen lag values. Hence stationary time series should represent a correlogram feature consistent with this. (ii) Unit root tests: in equation 3.3 y_t is the value of the variable at time t and ε_t is random error term; if ρ=1 series become a pure random walk process which is a characteristic nonstationary, unit root process. Unit root tests, question whether ρ coefficient is statistically significantly different from 1 or not. If ρ=1, or in difference form if δ=1- ρ =0 (see equation 3.4), then series are said to be nonstationary.

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{3.3}$$

$$\Delta y_t = \delta y_{t-1} + \varepsilon_t \tag{3.4}$$

Unit root tests developed by Dickey and Fuller (1979) and Phillips and Perron (1988) are widely used in the literature for testing stationarity of time series (Johnston and Dinardo, 1997).

In order to estimate an ARIMA (p, d, q) model, one needs to determine whether the series follow; a pure AR (p) process and if so what is the value of p, a pure MA (q) process and if so what is the value of q or an ARIMA process that is combination of an AR (p) process and MA (q) process. Box-Jenkins methodology identifies a three steps process that gives an answer to such questions. First step is identification. This step aims to find appropriate values of p, d and q for possible candidate models with the aid of autocorrelation functions and partial auto correlation functions of the series. (See table 3.1 for typical patterns of processes). Second step is estimation. Having identified possible candidates in the first step, in this step these alternative models are estimated. Third and the last step is diagnostic checking. The model that fits data best among alternatives is determined in this step. Statistical significance of coefficients of each estimation are questioned, estimates are compared with criterions such as Akaike's information criterion and Schwarz's information criterion; autocorrelation functions and partial autocorrelation functions of residuals from each alternative regression is examined with the aid of Ljung-Box and Box-Pierce Q statistics to determine if they follow a white noise process (Gujarati, 2009, pp. 777).

Table 3.1: Theoretical Correlation Patterns of AR and MA processes

Type of Model	Typical Pattern of ACF	Typical Pattern of PAC
AR (p)	Decays exponentially or with damped sine wave pattern or both	Significant spikes through lags p
MA (q)	Significant spikes through lags q	Declines exponentially
ARMA (p, q)	Exponential decay	Exponential decay

Source: Gujarati, 2009, pp. 781

Meese and Rogoff (1983) presented that a bivariate autoregressive model is unable to beat a random walk model in all forecasting horizons from one to twelve months; for monthly point-sample data of Dollar/Pound, Dollar/Mark, Dollar/Yen and trade weighted Dollar exchange rates, regardless of whether root mean squared error or mean absolute error methodology used as a comparison criteria. Fernandez-Rodriguez et al. (1999) applied non-parametric nearest neighbor forecasting technique and an ARIMA (1,1,0) model to currencies of nine countries participated in exchange rate mechanism of European Monetary System. Data set consist of daily spot exchange rates between the Deutsche mark and Belgian franc (BFR), the Danish Crown (DKR), the Portuguese Escudo (ESC), the French Franc (FF), the Dutch Guilder (HFL), the Irish Pound (IRL), the Italian Lira (LIT), the Spanish Peseta (PTA), the Pound Sterling (UKL) for the period from 1st January 1978 to 31st December 1994. Forecasting performances are compared by Theil’s U statistic and Diebold-Mariano test; ARIMA model offers lower U statistics only for three out nine cases and Dielbold-Mariano test results suggest that ARIMA model outperforms nearest neighbor predictor only for the Pound Sterling (UKL) exchange rate. Botha and Pretorius (2009) compared univariate models and multivariate models in forecasting South African Rand, U.S. Dollar exchange rate for one step ahead, two quarters and four quarters ahead horizons with mean absolute deviation criterion.

Results shown that while an ARIMA (1, 1, 1) model performed poorly for one step ahead forecast, its performance is improved as forecasting horizon is extended to two quarters and four quarters. McCrae et al. (2002) analyzed one to forty days ahead forecasts of ARIMA models in comparison with cointegration based error correction models (ECM) using an ARIMA (2, 1, 0) model for Japanese Yen, an ARIMA (4, 1, 0) model for Malaysian Ringgit, an ARIMA (1, 1, 2) model for Philippines Peso, an ARIMA (0, 1, 5) model for Thai Baht and an ARIMA (0, 1, 1) model for Singapore Dollar. Data sample consisted of log daily US Dollar exchange rates of five currencies from 1 January 1985 to 28 February 1997. They proposed that forecasting accuracy of ARIMA models as regard to error correction models depend on exchange rate series under consideration and forecast horizon. Individual currency analysis revealed that; over short horizons, ARIMA models provide more accurate forecasts for moving average terms of order higher than one and ARIMA models dominated error correction models at all horizons for only Thai Baht exchange rate.

3.1.2. Autoregressive Conditional Heteroskedasticity (ARCH) Models

Financial time series such as; exchange rates, stock prices, inflation often exhibit a random walk process in the level form and they also represent wide swings and volatility in difference form; suggesting that variance of these series is not constant and varies over time. Autoregressive conditional heteroskedasticity (ARCH) model which is developed by Engle (1982) gives a procedure for modeling this conditional variance in time series forecasting (Gujarati, 2009, pp. 791).

Within ARCH representation square of random error term (ε_t) is described as itself following an AR (p) process:

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + w_t \qquad (3.5)$$

where w_t is a white noise process with zero mean and constant variance. Conditional variance of this random error term (ε_t) then depends on lagged squared error terms (Hamilton, 1994, pp. 658, 659).

$$Var(\varepsilon_t|\varepsilon_{t-1}\varepsilon_{t-2}..\varepsilon_{t-p}) = \sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \cdots + \alpha_p\varepsilon_{t-p}^2 \tag{3.6}$$

Procedure for testing whether the autoregressive conditional heteroskedasticity is present or not can be summarized as follows: (i) Data set (Y_t) is transformed to return series; log mean adjusted relative change (X_t) for instance (see equation 3.7);

$$\begin{gathered} Y_t^* = \log Y_t \\ dY_t^* = Y_t^* - Y_{t-1}^* = relative\ change\ in\ Y_t \\ d\bar{Y}_t^* = mean\ of\ dY_t^* \\ X_t = dY_t^* - d\bar{Y}_t^* \end{gathered} \tag{3.7}$$

(ii) Following AR (p) model of volatility is estimated with appropriate lag values:

$$X_t^2 = \beta_0 + \beta_1\, X_{t-1}^2 + \beta_2\, X_{t-2}^2 + \cdots + \beta_p\, X_{t-p}^2 + \varepsilon_t \tag{3.8}$$

(iii) Residuals obtained from original regression ($\hat{\varepsilon}_t$), i.e. residuals obtained from estimation of equation 3.8 are regressed in the following form by ordinary least squares (OLS) estimation technique:

$$\hat{\varepsilon}_t{}^2 = \hat{\alpha}_0 + \hat{\alpha}_1\, \hat{\varepsilon}_{t-1}^2 + \hat{\alpha}_2\, \hat{\varepsilon}_{t-2}^2 + \cdots + \hat{\alpha}_p\, \hat{\varepsilon}_{t-p}^2 \tag{3.9}$$

(iv) Null hypothesis:

$$H_0\!: \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0 \tag{3.10}$$

is tested for joint significance of $\hat{\alpha}_1$, $\hat{\alpha}_2$,…, $\hat{\alpha}_p$ by usual F test. If these coefficients ($\hat{\alpha}_1$, $\hat{\alpha}_2$,…, $\hat{\alpha}_p$) are statistically significantly different from zero, i.e. if null hypothesis is rejected, then one can conclude that error terms do not have constant conditional variance and they follow conditional variation as implied by ARCH model. When the ARCH effect is present, applying generalized least squares (GLS) estimation technique instead of using ordinary least squares (OLS) can be a remedial measure (Johnston and Dinardo, 1997, pp. 196).

ARCH models gained enormous attention after the introduction by Engle (1982) and became widely used by researchers. Many variations of the model were developed. Among them one of the most frequently used is the generalized autoregressive conditional heteroskedasticity (GARCH) model which is developed by Bollerslev (1986). Within the GARCH presentation conditional variance is determined by not just lagged values squared error terms but also by lagged conditional variances (see equation 3.11) (Gujarati, 2009, pp. 796).

$$Var\left(\varepsilon_t \middle| \varepsilon_{t-1}\varepsilon_{t-2}\varepsilon_{t-1}\dots\varepsilon_{t-p}\right) = \sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \cdots + \alpha_p\varepsilon_{t-p}^2 + \gamma_1\sigma_{t-1}^2 + \cdots + \gamma_q\sigma_{t-q}^2 \tag{3.11}$$

West and Cho (1995) compared out-of-sample forecasting performance of GARCH models with autoregressive models and a nonparametric Gaussian kernel model. Data sample consist of five different (Canada, France, Germany, Japan, United Kingdom) bilateral weekly U.S. Dollar exchange rates from 14 March 1973 to 20 September 1989 with 863 observations. They concluded that; while GARCH models perform better than others for one-week horizon it is difficult to choose one model over another for twelve- and twenty-four-week-ahead forecasts. Vilasuso (2002) evaluated forecasting accuracy of GARCH model and its two variations; fractionally integrated GARCH (FIGARCH) model and integrated GARCH (IGARCH) model according to mean absolute error (MAE) and mean square error (MSE) criterion by using six different daily spot U.S. Dollar exchange rates within the period from 13 March 1979 to 31 December 1997. FIGARCH model dominated other models at all 1-, 5-, and 10-day forecast horizons with both mean absolute error (MAE) and mean square error (MSE) criterion. McMillan and Speight (2004) used alternative volatility measure suggested by Andersen and Bollerslev (1998) based on the cumulative squared returns from intra-day data instead of daily squared returns. Comparative analysis provided that simple GARCH outperforms the exponential smoothing and moving average models for almost all of the seventeen different exchange rates by means of a variety of error statistics: mean error, mean absolute error and root mean squared error. Chortareas et al. (2011) employed the intraday GARCH and FIGARCH models to 15 min data on returns of four Euro exchange rates with 509.472 observations from the period between January 4th, 2000, and

October 31st, 2004 and concluded that intraday FIGARCH outperformed traditional GARCH, AR and MA models.

3.1.3. Vector Autoregressive (VAR) Models

In a univariate time series model such as ARIMA model a variable is modeled depending merely on its own lagged terms. VAR models are used for modeling multivariate time series for which each variable is a linear function of past values of itself and past values of other explanatory variables. The model is termed as "autoregressive" because it consist past, lagged values of dependent variable and "vector" because it deals with a vector of two or more variables. A VAR (p) process with k different variables is as follows:

$$\boldsymbol{y_t} = \boldsymbol{m} + \boldsymbol{\varphi_1} y_{t-1} + \boldsymbol{\varphi_2} y_{t-2} + \cdots + \boldsymbol{\varphi_p} y_{t-p} + \boldsymbol{\varepsilon_t} \tag{3.12}$$

where $\boldsymbol{y_t}$ is a k x 1 vector of dependent variables, $\boldsymbol{\varphi_i}$ are k x k matrices of coefficients, $\boldsymbol{m}$ is a k x 1 vector of constants and $\boldsymbol{\varepsilon_t}$ is k x 1 vector of white noise error terms. For instance a simple VAR (1) model with two variables can be represented as follows:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} + \begin{bmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \tag{3.13}$$

And it can be written out in following explicit equation system (Johnston and Dinardo, 1997, pp. 287-288):

$$\begin{aligned} y_{1t} &= m_1 + \varphi_{11} y_{1,t-1} + \varphi_{12} y_{2,t-1} + \varepsilon_{1t} \\ y_{2t} &= m_2 + \varphi_{21} y_{1,t-1} + \varphi_{22} y_{2,t-1} + \varepsilon_{2t} \end{aligned} \tag{3.14}$$

Procedure for estimating a simple VAR model can be summarized as follows: since VAR model proposes a multiple equations system that consider joint behavior of several variables with all variables are endogenously determined by past values of itself and remaining variables in consideration, it is best suited with variables of bilateral causation. Bilateral causation can be examined with tests such as Granger causality test. VAR model can be specified in stationary form. If it is so; stationarity of the series can be ascertained by transforming the data set and can be ensured through unit root tests. Appropriate degree of lagged values that will be included in the model can be determined by Akaike's and Schwarz's information criterion. For the purpose of forecasting, all equations including lagged values of each variable can be estimated by usual ordinary least squares (OLS) estimation technique (Gujarati, 2009, pp. 784-788).

Liu et al. (1994) used different VAR specifications such as full vector autoregressive (FVAR), mixed vector autoregressive (MVAR), Bayesian vector autoregressive (BVAR) specification in order to forecast three U.S. Dollar exchange rates by monetary asset model of exchange rate determination and compared forecasting outcomes of each VAR specification. Out of sample forecast for the period from January 1983 to December 1989 suggested that VAR presentation may be a reliable alternative for exchange rate forecasting purposes and restricted VAR specifications provide much more accurate forecasts than unrestricted specifications. Joseph (2001) examined forecasting performance of VAR models in non-stationary, stationary and error-correction forms for seven U.S. Dollar exchange rates in daily, weekly and monthly horizons. Data set spanned the period 24 October 1983 to 12 May 1997 for daily bid-ask prices of seven currencies from the London FX market. Results proposed that non-stationary specification of VAR dominates its stationary and error correction specifications.

3.1.4. Error Correction Models (ECM)

Time series analysis often involves dealing with non-stationary series. Hence, transformation methods such as; differencing or seasonal adjustment applied to series in order to ascertain stationarity. Series that are not stationary in level form but becomes stationary after differencing say d times are called integrated of order d and presented as; I(d).

Linear combination of integrated series generally produces integrated series with highest order of the combined series. But if dependent variable (y_t) and explanatory variable (x_t) are both integrated then the error term (ε_t) which is linear combination of them (see equations 3.15 and 3.16); will be integrated of higher order of both too. For instance if y_t and x_t are I (1) then resulting error term (ε_t) is expected to be I(1).

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{3.15}$$

$$\varepsilon_t = y_t - \beta_0 + \beta_1 x_t \tag{3.16}$$

If series (y_t and x_t) share a common trend and drift together; then resulting error terms from regressing these series may be stationary, even if y_t.and x_t are non-stationary series. In such cases both series are said to be cointegrated. In economic terms two variables are considered to be cointegrated if there exist a long-run equilibrium or relationship between them. β_1 in equation 3.16 is called cointegrating parameter and for multivariate presentations, vector of coefficient parameters is called cointegrating vector.

Error correction mechanism is first introduced by Sargan (1964). Given that dependent variable is cointegrated with explanatory variable; error correction model, ties short run behavior of dependent variable with its long run equilibrium value by treating error term (ε_t in equation 3.16) as equilibrium error and reveals short run relation between dependent and explanatory variables (Greene, 2003, pp. 649-653).

Procedure for building a simple error correction model with a single regressor and regressand both integrated of order one is summarized as follows: Initially dependent variable is estimated through usual ordinary least squares technique to get equation 3.17. A trend variable (t) may be included to the regression equation if necessary, i.e. if residuals obtained from the regression follow a trend stationary process.

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_t + \hat{\beta}_2 t \tag{3.17}$$

Residuals of the regression equation are obtained by subtracting actual values (y_t) from the estimated values ($\hat{y}_t$) for each observation (see equation 3.18).

$$\hat{\varepsilon}_t = y_t - \hat{y}_t = y_t - \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_t \tag{3.18}$$

After that, stationartiy of the residuals is tested by performing a cointegration test such as augmented Engle-Granger test. According to test results if it is conferred that residuals are stationary then it is concluded that variables are cointegrated and they have a long-term relationship. Finally, following error correction model is regressed and significance of the coefficients are tested:

$$\Delta\hat{y}_t = \alpha_0 + \alpha_1 \Delta\hat{x}_t + \alpha_2 \hat{\varepsilon}_{t-1} + \nu_t \tag{3.19}$$

where $\Delta\hat{y}_t$ is the differenced form of dependent variable, $\Delta\hat{x}_t$ is the differenced form of explanatory variable, $\hat{\varepsilon}_{t-1}$ is lagged value of error term from equation 3.17, ν_t is white noise error term and α_2 is error correction coefficient which indicates how much of the discrepancy between long term and short term values corrected in one period (Gujarati, 2009, pp. 762-765).

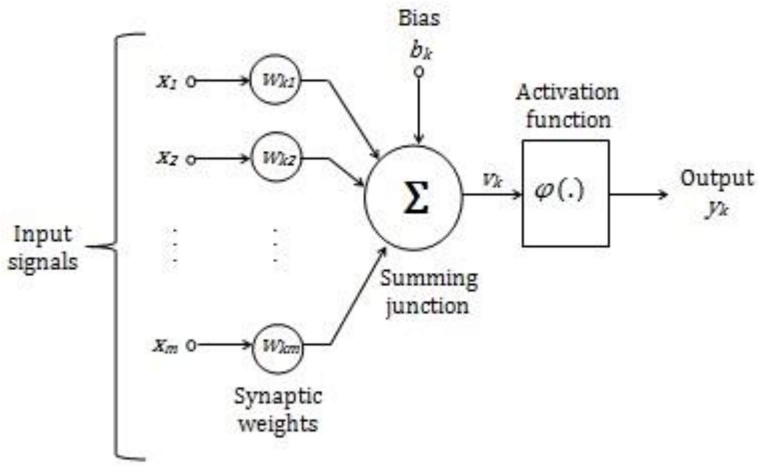
Kim and Mo (1995) applied error correction model to variations of monetary model (the Frenkel-Bilson monetary model, the Dornbusch-Frankel monetary model and the Hooper-Morton monetary model) in order to forecast U.S. Dollar/Deutsche Mark exchange rate. They concluded that while a random walk model dominates in the short run, all three structural monetary models strongly dominated random walk model in terms of long horizon forecasting accuracy. Fritsche and Wallace (1997) compared purchasing power parity model in error correction specification with a random walk model for U.S. Dollar exchange rate of four different countries' (UK, Germany, Canada and Japan) currencies. Estimations are generated by quarterly data from the period between 1974:Q1 to 1992:Q4 and the period between 1993:Q1 to 1994:Q4 used for forecasting. Purchasing power parity model in error correction specification was unable to beat a random walk model for currencies of UK and Germany but it performed better than random walk model for the currencies of Canada and Japan. Van Aarle et al. (2000) estimated the sticky-price monetary model with vector error correction model (VECM) to forecast four different Euro exchange rates (U.S. Dollar, British Pound, Japan and Swiss Franc). Data set consisted of monthly average data from January 1980 to February 1999. Results showed that a monetary model in vector error correction form is capable of beating the naive random walk model and its forecasts outperformed forecasts of random walk model for all currencies except Swiss Franc. Chen and Leung (2003) introduced a new method called Bayesian vector error correction model (BVECM). This method is based on enhancing accuracy of a Bayesian vector autoregressive model by enabling it to capture information from the long-run uncovered interest-parity (UIP) relationship. Regression results for three different exchange rates (Australia Dollar/U.S. Dollar, Japanese Yen/U.S. Dollar, and Korea Won/U.S. Dollar) demonstrated that out-of-sample forecasting performance of proposed Bayesian vector error correction model is better than Bayesian vector autoregressive models and random walk models. Gharleghi and Nor (2012) used relative price monetary model (RPMM) in vector error correction (VEC) specification with the aim of forecasting Malaysian Ringgit/U.S. Dollar exchange rate by using monthly data set from January 1986 to September 2010. They concluded that; long term dynamics suggested by monetary model is valid according to Johansen-Juselius cointegration approach and a vector error correction based monetary model outperformed the random walk model.

3.2. Artificial Intelligence Based Models

Artificial intelligence based models; artificial neural networks, which have wide range of application area in time series forecasting were inspired by functioning of brain and nerve systems of living organisms (Luk et al., 2001). According to Haykin (1999) artificial intelligence era can assumed to be start with the seminal work of McCulloch and Pitts (1943). As a psychiatrist and mathematician; McCulloch and Pitts described logical calculus of neural networks and represented a formal model of artificial neuron that mimics biological neurons. Minsky (1961) in his work titled “Steps Towards Artificial Intelligence” defined the term neural networks and brought artificial neural networks as a theory of computation (Haykin, 1999, pp. 38).

Artificial neural networks resemble the learning process of the brain by interconnected units which serve as artificial neurons (Hinton, 1992). Artificial neurons of artificial neural networks composed of three main elements: (i) Synapses or connecting links which weights the input signals. (ii) An adder which sums input signals weighted by synapses. (iii) An activation function which limit outcome of the neuron by transforming the inputs. Figure 3.1 represents a typical neuron of an artificial neural network.

Figure 3.1: Typical Model of an Artificial Neuron



Source: Haykin, 1999, pp. 13

Mathematical structure of a neuron can be demonstrated by following equations:

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \tag{3.20}$$

$$y_k = \varphi(u_k + b_k) \tag{3.21}$$

where x_1, x_2, …, x_m are input signals w_{k1}, w_{k2}, …, w_{km} are synaptic weights, u_k is the output of adder, b_k is the bias, $\varphi(.)$ is activation function and y_k is the output signal of the neuron. Three basic types of activation function can be identified: (i) Threshold function, represented in equation 3.22

$$\varphi(v) = \begin{cases} 1 & if\ v \geq 0 \\ 0 & if\ v < 0 \end{cases} \tag{3.22}$$

(ii) Piecewise-linear function represented in equation 3.23

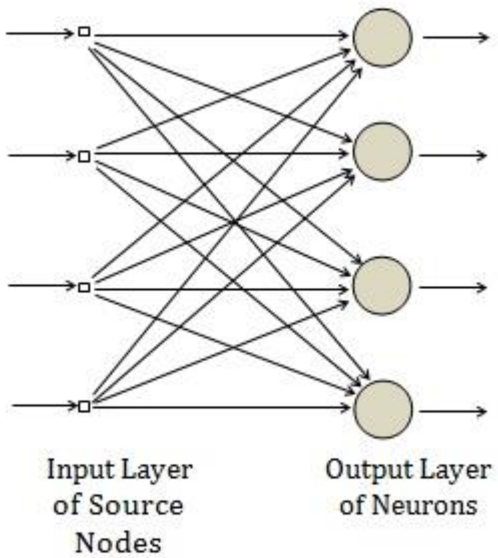
$$\varphi(v) = \begin{cases} 1, & v \geq +1/2 \\ v, & 1/2 > v > -1/2 \\ 0, & v \leq -1/2 \end{cases} \tag{3.23}$$

(iii) Sigmoid function represented in equation 3.24 by a logistic function where a is the slope parameter. (Haykin, 1999, pp. 10-14)

$$\varphi(v) = \frac{1}{1 + \exp(-av)} \tag{3.24}$$

In general a typical layered artificial neural network structure is composed of connected, simple knowledge processing elements called; nodes or neurons. Each node is connected to others by synapses, all of which is associated with a relevant weight factor (Palmer et al., 2006). Nodes with similar characteristics are organized in the form of layers. A layer is a group of nodes that broadcasts information to other connected layers and external environment which has no interconnections. Figure 3.2 illustrates the simplest form of single layer feedforward (in which source nodes are transmitted to output nodes but not vice versa) neural network.

Figure 3.2: A Simple Single Layer Feedforward Neural Network



Source: Haykin, 1999, pp. 21

First layer in the figure, connecting the input nodes, called input layer and the last layer consisting the output nodes is called output layer. The number of nodes in the input layer and output layer depends on the number input parameters and output parameters respectively (Luk et al., 2001).

Sophisticated empirical studies for nearly last half century proved that classical econometric models and time series models for forecasting exchange rates were not sufficient to reliably beat a naive random walk model which implies that exchange rates behave in a purely random and unpredictable manner (Ni and Yin, 2009). Classical models assume that the future value of a variable is linearly dependent on its own past values and past values of some other theoretical determinant variables with a random, white noise error term. These models are also based on the assumption that time series are generated by stationary processes. Disappointing performance of classical models was mainly attributed to these unrealistic assumptions accompanied by these models about the nature of data that does not comply with real world situations (Khashei et al., 2009). Autoregressive conditional heteroskedasticity (ARCH) and generalized autoregressive conditional heteroskedasticity models are developed by Engle (1982) and Bollerslev (1986) in order to overcome the difficulty of modeling non-constant variance feature of time series. These models are proven to be more effective in exchange rate forecasting than linear models. But while these models performed well for particular situations they were unable to serve as a convincing tool for modeling different type of situations (Ince and Trafalis, 2006).

Using artificial neural networks in time series forecasting have become a common practice among researchers during the last decades. Distinctive features of artificial neural networks that made them attractive in time series forecasting are as follows: (i) Artificial neural networks do not rely on assumptions about the nature of the data, which is the case for classical models, instead they are data-driven techniques. Artificial neural networks are nonparametric techniques in the sense that; they can learn from experience incorporated within the data and extract functional relationships with no need for theoretical framework (ii) Artificial neural networks can reliably reveal characteristics of sample population by learning form sample data even if the sample data consist noisy information. (iii) Artificial neural networks are not subject to limitations from functional forms. They are universal functional approximators that can identify and approximate any continuous function. (iv)

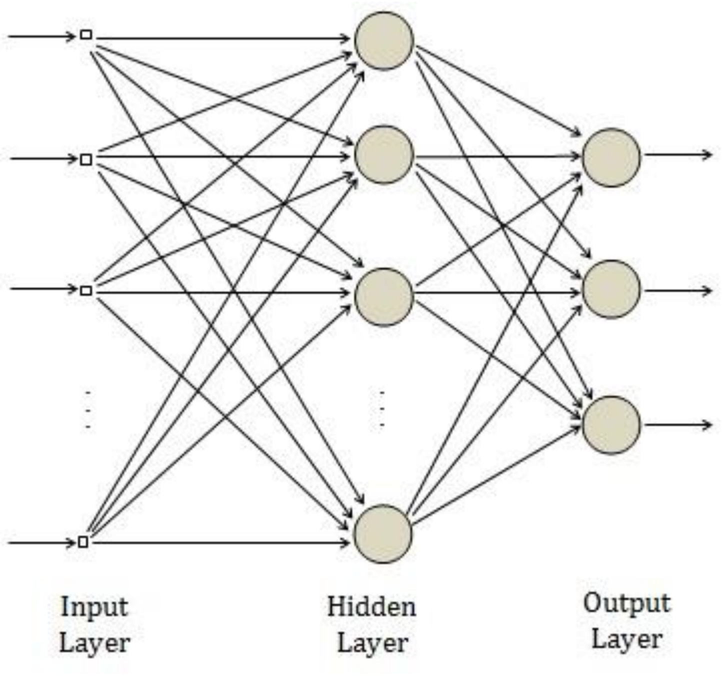
Artificial neural networks don't involve any assumptions about data generating process. They are capable of dealing with nonlinear relationships. They extract the relationship between variables by learning from data and base their forecasts on these nonlinear approximations (Zhang et. al., 1998).

Charytoniuk and Chen (2000) identified process for developing artificial neural networks for forecasting purposes which involves a five steps procedure: (i) Selection of input variables: A neural network can deal with correlated variables but the higher the number of correlated variables used as input; necessitates the larger the architecture of network and longer training time. Hence, autocorrelation and partial autocorrelation patterns of relevant variables can be analyzed and also principal component analysis techniques can be applied in order to avoid including unnecessary variables as inputs. (ii) Design of neural network structure: determining optimal network type and number of layers is crucial for forecasting performance of the neural network. A trial error procedure or applying an algorithm like pruning algorithm may be useful. (iii) Extraction of training data: Since the measurement value of the each variable may differ, all data should be scaled within the same range via normalization techniques. Artificial neural networks require data sets that are representative of the population but not in the way that the statistical methods involve. In statistics, a sample is accepted to be representative of population only if it shares the distribution of the population. Representative in neural networks procedure means that; all variables in training data are represented equally, i.e. they have uniform distribution regardless of the population distribution. (iv) Training of the designed neural networks: training a neural network aims to minimize error, i.e. difference between desired output and actual output, by adjusting weights of each unit. An optimization algorithm, such as; back-propagation algorithm, may be resorted. (v) Validation of trained neural networks: Performance of the neural network should be validated for a different data set. This can be done by randomly selecting a part of training data set and taking them aside for testing to see if the errors are still acceptable.

3.2.1. Multilayer Feedforward Neural Networks (MLFFNN)

Multilayer feedforward neural networks are extended structures of the single layer feedforward neural networks. They are termed as feedforward in the sense that information is processed in one direction; from input to output and multilayer in the sense they consist one or more hidden layers between the input and output layers (Haykin, 1999, pp. 21-22). In the one hidden layer case as represented in figure 3.3; input layer broadcasts to hidden layer by processing inputs through its nodes and after that the output layer use the outputs of hidden layer as inputs. Hence hidden layers are not directly connected to neural network's inputs and outputs. Introduction of a hidden layer to the network aims to enhance the learning ability of the neural. network and to enable it to deal with functions of greater complexity (Hammerstrom, 1993).

Figure 3.3: A Multilayer Feedforward Neural Network



Source: Haykin, 1999, pp. 159

In order to make a neural network of a specified architecture to perform some specific task; one need to determine optimal level of strength of connections, i.e. synaptic weights, that minimizes the difference between predicted and actual levels of outputs. This procedure is called learning of neural networks or training. Training procedure requires to adjust weights until artificial neural network mimic the known behavior of the system to be modeled, as close as possible (Malinov et al., 2001). During the initial applications, training of artificial neural networks was based on an inefficient trail-error process that required large amount of effort to implement and produced unsatisfactory outcomes which caused artificial neural networks to receive little attention until 1970’s (Rumelhart et al, 1994). During the mid-1970’s Paul J. Werbos introduced an efficient procedure for training a multilayer feedforward neural network. This procedure is now known as back-propagation algorithm and was largely ignored and its usefulness was not well appreciated until the end of 1980’s. David E. Rumelhart, Ronald J. Williams from University of California and Geoffrey E. Hinton were popularized the algorithm by demonstrating that it could teach the hidden units efficiently with a fair performance (Hinton, 1992).

Multilayer feedforward neural networks have been successfully applied to solve some difficult and complex problems by training them with back-propagation algorithm. Back-propagation algorithm basically consists of two stages: a forward pass and a backward pass. During the forward pass; input layer broadcasts an activity pattern (vector of inputs) to all hidden nodes (nodes on the hidden layers) and hence its effect propagates layer by layer through the output node. Each output node then generates actual results by calculating weighted sum and passing this through the activation (transfer) function. During the forward pass synaptic weights of each node are all fixed. The second stage, backward pass, is based on adjusting synaptic weights in accordance with an error correction rule. During the backward pass; actual outcome of the neural network is subtracted from the target outcome to form output errors. Synaptic weights are adjusted to minimize output errors by propagating these output errors backward through opposite direction of the synaptic connections of the layers of the neural network. The name, “back-propagation” is derived from this back-propagation nature of the procedure (Haykin, 1999, pp. 156).

To be more specific the way that the back-propagation algorithm works, can be explained as follows: (i) Explaining back-propagation algorithm necessitate representing outputs of a neural network in mathematical terms. Assuming that node j is a typical node in output layer and node i is a typical node in the previous layer.

An output node broadcast the outputs of the neural network in two stages. First; it computes total weighted sum of inputs using the formula in equation 3.25.

$$x_j = \sum_i y_i w_{ij} \tag{3.25}$$

where y_i is the output of the *i*th node in the previous layer and w_{ij} is the fixed synaptic weight between *i*th and *j*th nodes. Then, each node on the output layer calculates the output of the neural network by processing weighted sums with activation function. In general, back-propagation algorithm entails the use of sigmoid type activation function.

$$y_j = \frac{1}{1 + e^{-x_j}} \tag{3.26}$$

(ii) Once the outputs of all nodes on the output layer have been determined then total error (E) is calculated which is defined as:

$$E = \frac{1}{2} \sum_j (y_j - d_j)^2 \tag{3.27}$$

where y_j is the output of the *j*th node in the output layer and d_j is the corresponding actual output of the *j*th node. Steps represented so far, constitute forward pass of back-propagation algorithm and remaining steps will represent backward pass of the algorithm. (iii) Error derivative (EA) is computed to see by how much the error changes as the output of the neural network (y_j) is changed:

$$EA_j = \frac{\partial E}{\partial y_j} = y_j - d_j \tag{3.28}$$

(iv) Chain rule can be applied to see by how much the error changes as the output of the previous layer (x_j) changes (EI):

$$EI_j = \frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial x_j} \tag{3.29}$$

Differentiating equation 3.26 with respect to x_j and replacing it in equation 3.29 gives:

$$EI_j = \frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial x_j} = EA_j\, y_j(1 - y_j) \tag{3.30}$$

(v) The speed with which the error changes as a synaptic weight on the connection of an output (w_{ij}) changes (EW) can be computed by:

$$EW_{ij} = \frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial x_i}\frac{\partial x_j}{\partial w_{ij}} = EI_j\, y_i \tag{3.31}$$

Equation 3.31 implies that the quantity EW equals to the quantity EI, which is calculated on the previous step, multiplied by the output of the node on the previous level, i.e. output of the node from which the connection emanates. (vi) Since when the output of a node in the previous layer (*i*th layer) changed it simultaneously affects outputs of all nodes that are connected to it; this last step involves computing overall effect of these changes on the error (EA) by adding all these separate effects on output nodes:

$$EA_i = \frac{\partial E}{\partial y_i} = \sum_j \frac{\partial E}{\partial x_j}\frac{\partial x_j}{\partial y_i} = \sum_j EI_j w_{ij} \tag{3.32}$$

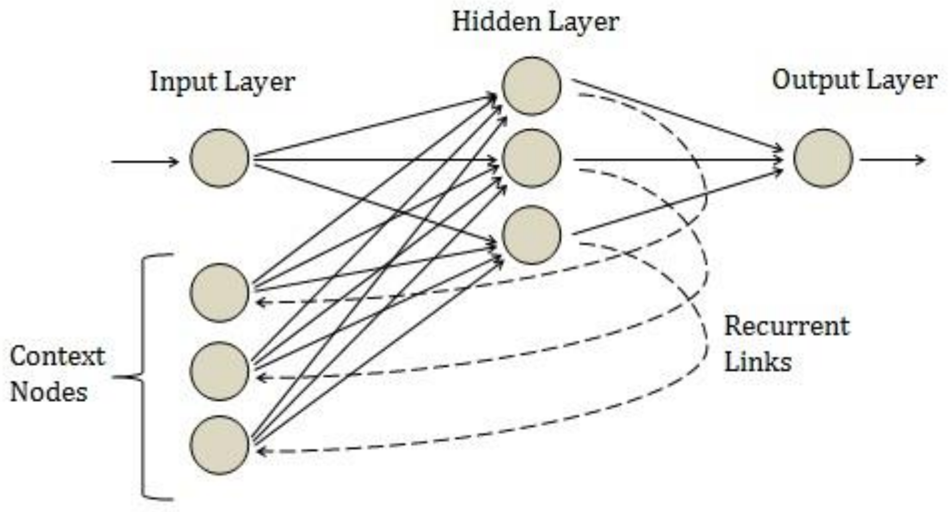
Once the EA is computed this procedure can be repeated for successively earlier layers as desired (Hinton, 1992; Rumelhart et al., 1986).

Wu (1995) compared one step ahead and six step ahead forecasting performance of ARIMA and multilayer feedforward neural network (MLFNN) models using monthly average exchange rates between U.S. Dollar and New Taiwan Dollar; from January 1979 to December 1992. First 162 observations, from January 1979 to December 1992, were used for regression and model building and remaining six observations were used for performance evaluation. Empirical analysis proposed that MLFNN model perform better than an ARIMA model for both one step ahead and six step ahead forecasts. Verkooijen (1996) used ordinary least squares (OLS) technique and artificial neural networks (ANN) for regressing structural models of exchange rate determination and compared forecasting performance of OLS and ANN regressions. Data sample consisted of 228 monthly observations of U.S. Dollar/Deutsche Mark exchange rate from January 1974 to January 1993. Using rolling regression technique and starting with first 100 observations, compared forecasting performance of OLS and ANN regressions of three different structural models for one month, six months, twelve months, twenty four months and thirty six months ahead forecasts. According to root mean square error criteria while being closely followed by OLS regressions, ANN regressions always performed better than OLS regressions and also ANN regressions were much better in percentage of correctly predicted signs. Zhang and Hu (1998) examined the effects of; training sample sizes and network structures (i.e. number of inputs and number of hidden nodes) on out of sample forecasting performance of multilayer feedforward neural networks. Data set is composed of daily British Pound/U.S. Dollar exchange rates from the beginning of 1976 to end of 1993. MLFNN structures used in the study consisted of; ten levels of input nodes (from 1 to 10), five levels of hidden nodes (4, 8, 12, 16, 20), two different training periods (from 1976 to 1992 and from 1988 to 1992). It is derived that the impact of number of hidden nodes on forecasting performance is higher than the impact of number of hidden nodes and also; forecasting errors diminish with larger training sets. Panda and Narasimhan (2007) conducted a comparative analysis of multilayer feedforward neural networks with random walk models and linear autoregressive models using 496 weekly observations of Indian Rupee/U.S. Dollar exchange rates from January 6, 1994 to July 10, 2003. Initial 350 observations are kept for training and remaining 146 observations are used for forecast comparison. Using six different forecasting evaluation criteria they concluded that neural network model is capable of beating random walk and linear autoregressive models for both out-of-sample and in-sample forecasting.

3.2.2. Recurrent Neural Networks (RNN)

Recurrent neural networks differ from feedforward neural networks by including at least one loop back either from output or an intermediate layer to input layer. Recurrent neural networks involve the use of outputs of hidden, intermediate, layers or the errors derived from the actual outcomes of neural network as inputs to next period depending on whether the outcomes of the hidden layers or the output layer loop back (Dunis and Huang, 2002). In addition to input nodes and hidden nodes of a feedforward neural network, recurrent neural networks' input layer or hidden layers include context nodes. These context nodes store the outputs of hidden layers or the errors of the neural network for the current period and broadcast them to next period. Recurrent neural network structure introduced by Elman (1990), uses outputs of all hidden neurons as inputs to next period thus having context nodes as many as the hidden nodes (More and Deo, 2003). Figure 3.4 depicts an Elman type recurrent neural network with one input node, one output node, three hidden nodes and hence three context nodes.

Figure 3.4: A Typical Elman Recurrent Neural Network

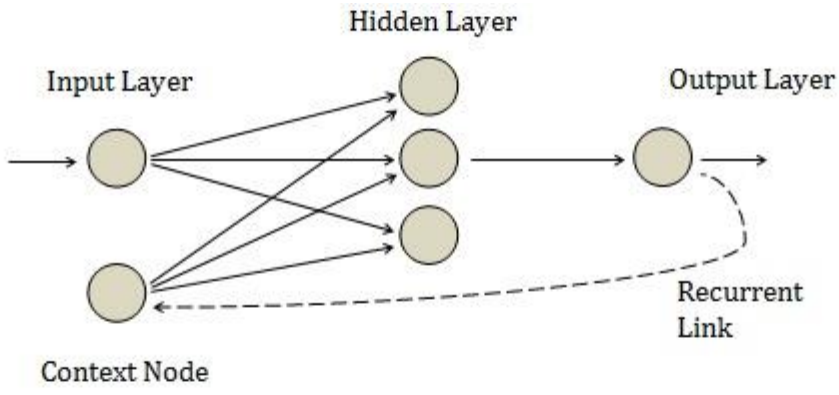


Source: Modified from Marra and Morabito, 2005

Another widely used recurrent network structure is Jordan type recurrent neural network, introduced by Jordan (1989). This neural network structure loops the output errors of the neural network back as an input; through the context nodes in the input layer (Yasdi, 1999). Figure 3.5 depicts a Jordan type recurrent neural network with one output node and hence one context node.

Recurrent neural networks ensure a richer dynamic structure than feedforward neural networks by enabling use of processed data of the previous period as an additional input to current period. This distinguishing structure allows them to perform better when dealing with noisy data and extracting features of hidden states for outputs that depend on a number of outputs of previous periods (Saad et al., 1998). Besides recurrent neural networks has a major drawback with respect to their feedforward counterparts; that they require more complex architectures with larger number of connections. This feature of recurrent neural networks has been subject to criticism concerning the lack of transparency (Dunis and Huang, 2002).

Figure 3.5: A Typical Jordan Recurrent Neural Network



Source: Modified from More and Deo, 2003

Kuan and Liu (1995) conducted a comparative analysis of random walk model and various structures of feedforward and recurrent neural networks in forecasting five different U.S. Dollar exchange rates (British Pound, Canadian Dollar, Deutsche Mark, Japanese Yen, Swiss Franc). Results provided limited evidence on neural networks as a forecasting tool. Neural networks while having significant ability to predict direction of movements of exchange rates, they were unable to strongly outperform random walk model (for only two currencies out of five). Tenti (1996) evaluated prediction accuracy of recurrent neural networks architecture using three different recurrent neural network structures. Analysis is based on profitability of two different trading strategies, associated with Deutsche Mark exchange rate forecasts of neural networks. Using various performance criteria based on forecasting accuracy and trade profitability, concluded that recurrent neural networks are useful and reliable tools for financial forecasting applications. Preminger and Franck (2007) investigated predictive accuracy of recurrent neural networks in comparison with random walk model and a linear autoregressive model. The models are applied to Japanese Yen and British Pound exchange rates against U.S. Dollar for one, three and six months of forecasting horizons. Monthly observations are obtained for the analysis, end of month prices of each currency, starting from January 1971 to October 2004 with 406 observations. 307 observations are used for estimation and remaining observations are left for out-of-sample forecasts. Each model is estimated by using moving regression technique in which for an h-step ahead forecast 307-h observations are used for estimation. Results of the analysis suggested that; forecasting performances vary for forecasting horizon and currency under consideration while novel estimation technique presented in the study which is robust to outliers improved forecasting performances of each model.

3.2.3. Probabilistic Neural Networks (PNN)

Probabilistic neural network is another artificial neural network architecture that fundamentally differs from multilayer feedforward neural network and recurrent neural network architectures by using Bayesian decision rule in order to classify observations (Parry et. al., 2011). With reference to workings of Parzen (1962) and Cacoullos (1966), Specht (1990) delivered that both Bayesian theory of conditional probability and the nonparametric estimation method for probability density functions introduced by Parzen (Bayes-Parzen classifier) could be applied to large number of independently and parallelly running simple processes within a multilayer neural network architecture (Hajmeer and Basheer, 2003).

Bayesian theory dictates that an observation vector $x = [x_1, x_2, \dots, x_m]$ from a population consisting of K different categories (1, 2, ..., i, ..., K) belongs to category i if equation 3.33 holds for all categories with $j \neq i$:

$$h_i c_i f_i(x) > h_j c_j f_j(x) \tag{3.33}$$

where h_i is the prior probability that the sample belongs to category i, c_i is the cost or loss associated with misclassifying a sample from category i and $f_i(x)$ is the probability density function of category i. Primary shortcoming of Bayesian classification is to estimate probability density function of each category when underlying probability densities are actually unknown. A usual practice is to assume normal (Gaussian) distribution. However this assumption produces high classification errors when actual distribution significantly deviates from normal distribution. Parzen (1962) proposed an estimator of probability density functions for univariate case and Cacoullos (1966) extended it to multivariate case. According to Cacoullos; joint probability density function of category i for a set of m observations ($x = [x_1, x_2, \dots, x_m]$) is as follows:

$$g_i(x_1, x_2, \dots, x_m) = \frac{1}{M\,\sigma_1\sigma_2 \dots \sigma_m} \sum_{i=1}^{m} W\left(\frac{x_1 - x_{1,i}}{\sigma_1}, \frac{x_2 - x_{1,i}}{\sigma_2}, \dots, \frac{x_n - x_{m,i}}{\sigma_m}\right) \qquad (3.34)$$

where W is weighting function, σ_1, σ_2, …, σ_n are the smoothing parameters (alternatively called window or kernel width), i.e. standard deviation of random variables around the mean, and M is the total number of training parameters in category i. Equation 3.34 becomes to following reduced form when all smoothing parameters are assumed to be equal (σ_1 = σ_2= …= σ_n = σ) and a Gaussian function replaced for W:

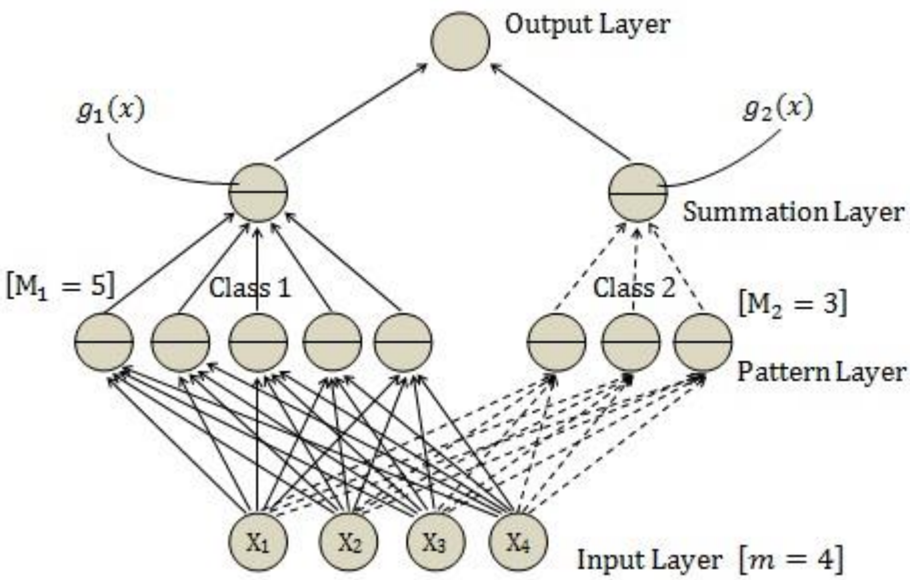
$$g_i(x) = \frac{1}{(2\pi)^{m/2} M \sigma^m} \sum_{i=1}^{M} exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \qquad (3.35)$$

where x is the input vector to be classified and x_i is the vector of training parameters in category i.

General structure of a probabilistic neural network is comprised of four layers: (i) an input layer, (ii) a pattern layer (first hidden layer), (iii) a summation layer (second hidden layer), (iv) an output layer. Figure 3.6 illustrates a simple probabilistic neural network with four inputs (m=4), hence four nodes on input layer, two population classes (K=2); Class 1 and Class 2; eight training observations with five belonging to Class 1 (M_1=5) and remaining three belonging to Class 2 (M_2=3).

Working process of a probabilistic neural network begins with input layer; each observation within the sample that is subject to classification represented in the input layer with a single node. Input layer broadcast inputs to pattern layer. Nodes in the pattern layer compute the distance between each input observation and the training pattern presented by the relevant pattern node. Pattern nodes then subject these differences to activation function (exponential part of the equation 3.35.) and broadcast its outputs to summation layer. Summation layer comprise of number of nodes equal to the number of classes. Each node sums the outputs of the pattern nodes of corresponding classes and subjects them to constant part of the equation 3.35. Finally output node compares the outcomes of summation nodes and yields the computed class with the probability that the sample will belong to that class (Specht, 1990; Hajmeer and Basheer, 2003; Chen et al., 2003; Gan et al., 2005; Parry et al., 2011).

Figure 3.6: A Simple Probabilistic Neural Network Structure



Source: Hajmeer and Basheer, 2003

3.2.4. Support Vector Regression

Vapnik (1995) introduced a novel neural network algorithm that originally used for pattern recognition practices called; support vector machines. Afterwards, Vapnik et al. (1996) presented the way that the support vector machines can be utilized to nonlinear regression applications by introducing ε-sensitive loss function and this procedure is called in the literature as support vector regression (Tay and Cao, 2001).

Support vector regression has distinctive features that made it attractive compared to other neural networks. Support vector regression follows structural risk minimization principle and seeks to minimize upper bound of the generalization error instead of minimizing prediction error (empirical risk minimization principle) followed by other neural networks (Chen, 2011). Use of structural risk minimization enhances the ability of support vector regression to generalize input output relationship derived during the training phase and enables it to avoid vulnerability of over-fitting problem of other neural networks which may result in poor prediction ability when tested for different data sets (Wu and Akbarov, 2011). Architecture and weights of a support vector regression can be determined easily and rapidly by solving a quadratic optimization problem through a standard programming algorithm while the same process for other neural networks necessitates a tough and time-consuming, trial-error procedure. This also means that the solution of support vector regression is unique, optimal and is not subject to risk of being stuck in local minima; which is the case for other neural networks (Lin et al., 2009; Tay and Cao, 2001).

Support vector regression performs a linear regression in high-dimensional feature space in order to estimate an unknown function $g(x)$ by mapping given noisy set of data points ($G = [(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)] = \{(x_i, y_i)\}_{i=1}^{N}$; where x_i is the input values, y_i is the actual values, N is the total number of data points) into this feature space through nonlinear mapping. Hence support vector regression transforms a nonlinear regression problem in low-dimensional space to a linear regression problem in high dimensional space by regressing a linear function ($f(x)$) through solving a convex optimization problem.

$$f(x) = \hat{y} = \langle \boldsymbol{w}, \boldsymbol{x} \rangle + b_0 , \quad \boldsymbol{w} \in \mathbb{R}^{\boldsymbol{d}}, b_0 \in \mathbb{R} \tag{3.36}$$

where; $\hat{y}$ is the estimated value of y with an error tolerance of ε, $\mathbb{R}^{\boldsymbol{d}}$ represents dimensional space of input variables (x_i)**,** $\mathbb{R}$ represents dimensional space of actual values (y_i), $\langle \boldsymbol{w}, \boldsymbol{x} \rangle$ represents dot product of two vectors in $\mathbb{R}^{\boldsymbol{d}}$, $\boldsymbol{w}$ and $\boldsymbol{x}$ are the weight factors to be estimated and the intercept term respectively and b_0 is the bias term. In estimating $\boldsymbol{w}$ support vector regression aims minimize norm of the $\boldsymbol{w}$ in order to ensure that the estimated function is as flat as possible. This estimation problem can be written as following convex optimization problem:

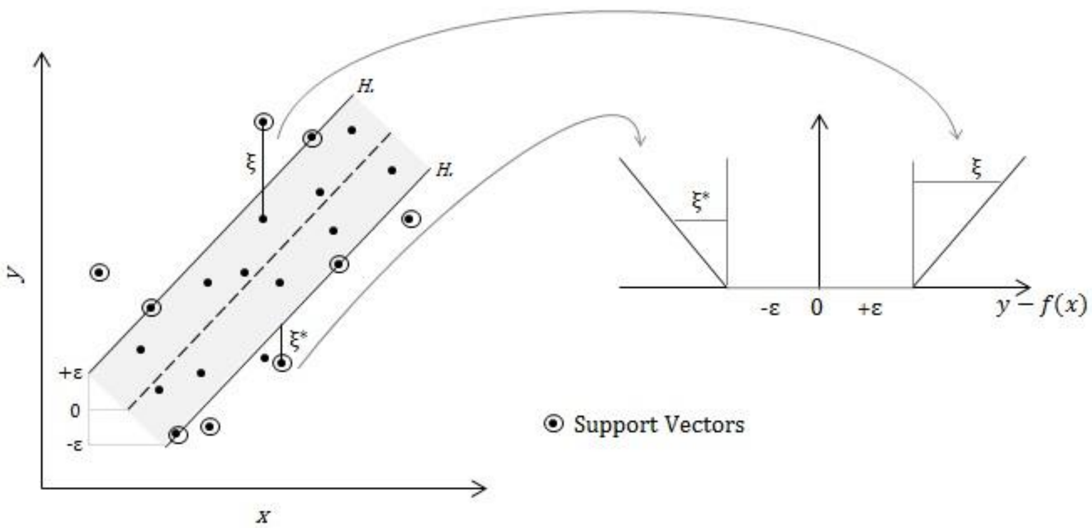
$$\begin{aligned} &\textit{minimize} && \frac{1}{2}\|\boldsymbol{w}\|^2 \\ &\textit{subject to} && \begin{cases} y_i - \langle \boldsymbol{w}, \boldsymbol{x_i} \rangle - b_0 \le \varepsilon, \\ \langle \boldsymbol{w}, \boldsymbol{x_i} \rangle + b_0 - y_i \le \varepsilon \end{cases} \end{aligned} \tag{3.37}$$

Equation 3.37 implicitly assumes that a function $f(x)$ exist which estimate all pairs of (x_i, y_i) with a precision error of ε, i.e. convex optimization is feasible. However, usually this is not the case and it is often required to allow some error larger than ε. In order to cope with such infeasible constraints of the optimization problem slack variables (ξ, ξ^*) can be introduced.

$$\varepsilon(x, y, f) = \max(0, |y - f(x)| - \varepsilon) \tag{3.38}$$

To be more concrete; ε-sensitive loss function of the support vector regression defined in the equation 3.38 omits the error if the difference between estimated value ($f(x)$) and actual value (y) is lower than ε ($|y - f(x)| < \varepsilon$) and otherwise equal it to: $|y - f(x)| - \varepsilon$.

Figure 3.7: The Soft Margin Loss Function



Source: Moura et al., 2011

Hence ε-sensitive loss function allows only error free tube of radius ε. In order to extend this radius, one may define following slack variables: $|y - f(x)| - \varepsilon = \xi$ for data points above the tube and $|y - f(x)| - \varepsilon = \xi^*$ for data points below the tube (soft margin loss function); as illustrated in figure 3.7.

With the introduction of slack variables optimization problem becomes:

$$
\begin{aligned}
&\mathit{minimize} && \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \\
&\mathit{subject\ to} && \begin{cases} y_i - \langle \boldsymbol{w}, \boldsymbol{x_i} \rangle - b_0 \le \varepsilon + \xi_i, \\ \langle \boldsymbol{w}, \boldsymbol{x_i} \rangle + b_0 - y_i \le \varepsilon + \xi_i^*, \\ \qquad \xi_i, \xi_i^* \ge 0 \end{cases}
\end{aligned}
\tag{3.39}
$$

Constant term C (which is constrained to be; $C > 0$) represents the trade-off between flatness of $f(x)$ and the amount with which the errors higher than ε are tolerated (i.e. up to what amount the radius of the error tube extended). In order to solve the primal optimization problem presented by equation 3.39, Lagrange function of its dual problem can be formulated by introducing some equalities provided by Kuhn-Tucker conditions:

$$
\begin{aligned}
&\mathit{maximize} && L_D(\alpha, \alpha^*) = -\frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j \rangle \\
& && -\varepsilon\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) \\
&\mathit{subject\ to} && \begin{cases} \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \\ 0 \le \alpha_i \le C \\ 0 \le \alpha_i^* \le C \end{cases}
\end{aligned}
\tag{3.40}
$$

where α and α^* are Lagrangian multipliers of the first two constraints of the primal optimization problem (equation 3.39). Solving for w yields:

$$
w^* = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)x_i \tag{3.41}
$$

And hence $f(x)$ equals to:

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \langle \boldsymbol{x_i}, \boldsymbol{x} \rangle + b \tag{3.42}$$

Support vector regression can deal with situations where $f(x)$ is nonlinear. This can be simply done by preprocessing input values (x_i) by mapping $\Phi: \mathbb{R}^{\boldsymbol{d}} \rightarrow \mathrm{F}^{\boldsymbol{d}}$ where $\mathrm{F}^{\boldsymbol{d}}$ is some d dimensional feature space. It is sufficient to determine: $k(\boldsymbol{x_i}, \boldsymbol{x}) = \langle \Phi(\boldsymbol{x_i}), \Phi(\boldsymbol{x}) \rangle$ for support regression to fulfill such mapping since it only involves dot products (see equation 3.42). Thus $f(x)$ becomes:

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)\, k(\boldsymbol{x_i}, \boldsymbol{x}) + b \tag{3.43}$$

where $k(\boldsymbol{x_i}, \boldsymbol{x})$is the so-called Kernel function that may take the following forms presented in equations 3.44, 3.45 and 3.46: linear, polynomial and radial basis (Mohandes, 2002; Kamruzzaman et al., 2003; Smola and Schölkopf, 2004; Ince and Trafalis, 2006; Wu and Akbarov, 2011; Moura et al., 2011, Shabri and Suhartono, 2012).

$$Linear: \quad \langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle \tag{3.44}$$

$$Polynomial: \quad \left(\langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle + r \right)^d \tag{3.45}$$

$$Radial\ Basis: \quad exp\left(\left\| x_i - x_j \right\|^2 / 2\sigma^2 \right) \tag{3.46}$$

where; σ^2, r and d are kernel parameters.

Since; as the dataset gets larger and larger, quadratic optimization problem involved in the training process of the above mentioned, support vector regression technique become a very complex issue and hence this technique entails time consuming computations when applied to large datasets. Least squares support vector regression technique (LSSVR) which is proposed by Suykens and Vandewalle (1999) simplifies the training process by offering a solution to system of linear equations through least squares technique instead of quadratic optimization problem involved in standard support vector regression technique.

In LSSVM; ε-sensitive loss function is replaced by least square of the error (e_i) as loss function. With this replacement, the optimization problem becomes:

$$
\begin{aligned}
&\textit{maximize} \qquad J(w,e) = \frac{1}{2}\,\langle \boldsymbol{w}, \boldsymbol{w}\rangle + \frac{1}{2} C \sum_{i=1}^{N} e_i^2 \\
&\textit{subject to} \qquad y_i = w^T \Phi(\boldsymbol{x_i}) - b_0 + e_i, \;\; i = 1,2,\dots,N
\end{aligned}
\tag{3.47}
$$

And the Lagrange function for the solution of the optimization problem can be constructed as:

$$
L(w,e,\alpha,b_0) = J(w,e) - \sum_{i=1}^{N} \alpha_i w^T \Phi(\boldsymbol{x_i}) + b_0 + e_i - y_i \tag{3.48}
$$

Karush-Kuhn-Tucker conditions imply that the solution of the Lagrange function can be achieved by partially differentiating the function with respect to parameters; w, b_0, e_i and α_i:

$$
\begin{aligned}
\frac{\partial L}{\partial w} &= 0 \;\Rightarrow\; w = \sum_{i=1}^{N} \alpha_i \Phi(\boldsymbol{x_i}) \\
\frac{\partial L}{\partial b_0} &= 0 \;\Rightarrow\; b_0 = \sum_{i=1}^{N} \alpha_i \\
\frac{\partial L}{\partial e_i} &= 0 \;\Rightarrow\; \alpha_i = C\, e_i \\
\frac{\partial L}{\partial \alpha_i} &= 0 \;\Rightarrow\; w^T \Phi(\boldsymbol{x_i}) + b_0 + e_i - y_i = 0
\end{aligned}
\tag{3.49}
$$

After eliminating e_i and w the solution is transformed to following system of linear equations:

$$
\begin{bmatrix} b_0 \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 & I_v^T \\ I_v & \Phi(\boldsymbol{x_i})\Phi(\boldsymbol{x_j}) + C^{-1} I \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{3.50}
$$

where; $y = [y_1, y_2, \dots y_N]^T$, $I_v = [1,1,\dots 1]^T$and $\alpha = [\alpha_1, \alpha_2, \dots \alpha_N]^T$ (Mak and Yang, 2007; Zhang and Liu, 2009; Zhou et al., 2011; Liao et al., 2011; Zhu and Wei, 2013).

Kamruzzaman et al. (2003) applied alternative SVR structures to forecast seven different exchange rates by using 565 weekly observations from January 1991 to July 2002. First 500 observations are used for training the model and the remaining observations are used for forecast comparison. Forecasts are based on differing SVR structures, in terms of kernel function types (linear, polynomial, radial basis and spline) for which the range of the regularization parameter (C) is set to (0,1 10^5). Results proposed that the SVR models with radial basis kernel function and polynomial kernel function produce lower prediction errors while varying the value of parameter C did not affect prediction performance on a consistent basis. Bahramy and Crone (2013) conducted a comparative analysis of alternative SVR structures with differing kernel functions and input variables; in forecasting direction of the Euro/USD exchange rate movements. Data set is comprised of 4810 daily observations from which the 50% is retained as the training set, 25% as the validation set and remaining 25% as the test set. Grid search algorithm is utilized to determine the optimal values of the free parameters (C, σ^2). Range of the parameter C is set to $\{2^{-12}$ to $2^{12}\}$ and range of the parameter σ^2 is set to $\{2^{-17}$ to $2^0\}$ both with step-sizes of 2^{-1}. Results of the analysis showed that the novel SVR model suggested in the study, which use technical indicators as inputs, has considerable predictive ability and performed better than the benchmark models.

4. EMPIRICAL RESULTS

In this section; classical models (ARIMA, GARCH, VAR, ECM) and artificial intelligence based models (MLFNN, RNN, SVR) are applied to forecast Turkish Lira/U.S. Dollar (TL/USD), Turkish Lira/Euro (TL/EUR), Turkish Lira/British Pound (TL/GBP) exchange rates. One step ahead, in-sample and out-of-sample forecasting accuracy of each model analyzed comparatively.

Following statistical measures are utilized throughout the analysis for model selection and forecast evaluation purposes: (i) Root Mean Squared Error (RMSE): it measures how well the forecast outcomes of a specific model fit the actual data, lower values mean better fit and a value of zero mean perfect fit. It can be calculated by following equation:

$$RMSE = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4.1}$$

where y_i is the actual value of the variable at time t, $\hat{y}_i$ is the fitted value of the variable at time t and n is the number of observations.

(ii) Correlation Coefficient (R): which measure the strength of linear dependency between actual and fitted values of a model. Higher values indicate strong linear relationship and lower values indicate weak linear relationship. It can be calculated by following equation:

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{4.2}$$

where; $\bar{y}$ is the mean of the actual values and $\bar{\hat{y}}$ is mean of the fitted values (Pai and Hong, 2007; Fang et al. 2009; Samsudin et al. 2011).

(iii) Diebold-Mariano Test: it aims to test forecasting accuracy of two competing models under the null hypothesis of equal forecast accuracy by using following test statistic (DM) which follows a standard normal distribution:

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{n}}} \tag{4.3}$$

where $\bar{d}$ is the mean of the loss differential (d_i) which can be defined in the form of absolute deviation of errors (e_{1i}, e_{2i}): $|e_{1i}| - |e_{2i}|$ or in the form of squared deviation of errors: $e_{1i}^2 - e_{2i}^2$; $\hat{f}_d(0)$ is a consistent estimator of spectral density of the loss differential function (d) with a frequency zero and n is the number of observations (Fernandez-Rodriguez et al., 1999; Chung, 2006; Preminger and Franck, 2007)

Applications of classical models are carried out by software package Eviews and applications of artificial intelligence based models are carried out by software package Matlab.

4.1. Data

Dataset is comprised of monthly period data from January 2004 to December 2013 with 120 observations. TL/USD, TL/EUR, TL/GBP exchange rate series, monetary aggregates (M3) of the relevant currencies and the one year interbank borrowing rate of the relevant currencies are utilized throughout the analysis.

First 96 observations (80% of the dataset) are used for estimation and model specification purposes and remaining 24 observations (20% of the dataset) are retained for out-of-sample forecast comparison. Monthly exchange rate series are generated from daily closing rate of the first working day of each month. Variables and sources are reported in Table 4.1.

Table 4.1: Variables and Sources

Variables	Sources
TL/USD Exchange rate	Central Bank of the Republic of Turkey
TL/EUR Exchange Rate	Central Bank of the Republic of Turkey
TL/GBP Exchange Rate	Central Bank of the Republic of Turkey
TL One Year Interbank Borrowing Rate	The Banks Association of Turkey
EUR One Year Interbank Borrowing Rate	European Banking Federation
TL Monetary Aggregates (M3)	Central Bank of the Republic of Turkey
USD Monetary Aggregates (M3)	Federal Reserve Bank of St. Louis
EUR Monetary Aggregates (M3)	European Central Bank
GBP Monetary Aggregates (M3)	Bank of England

4.2. ARIMA Models

Box-Jenkins methodology implemented in order to determine ARIMA specification of all three exchange rate series. Initially unit root tests (Philips-Perron and Augmented Dickey Fuller tests) are applied to each serie in order to asses stationarity. If it is conferred that the serie is nonstationary in level form; it is differenced until stationarity is ensured. After that, patterns of autocorrelation functions and partial autocorrelation functions are examined to determine the ARIMA specification.

Table 4.2: Augmented Dickey-Fuller Unit Root Test Results of Exchange Rate Series

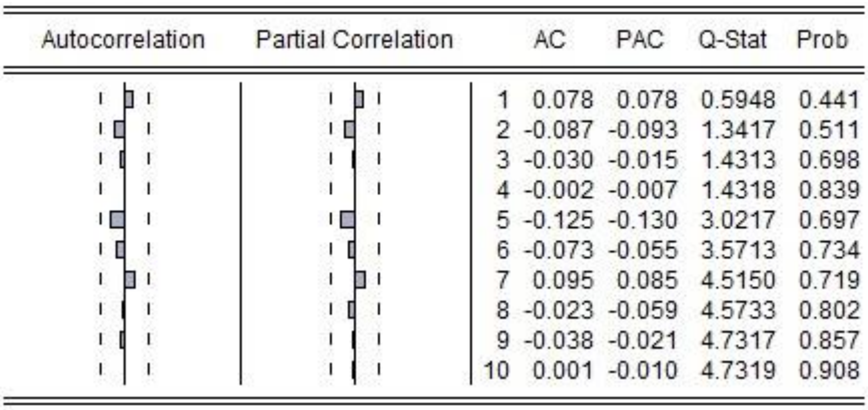
	TL/USD		TL/EUR		TL/GBP	
	Level	Difference	Level	Difference	Level	Difference
Test Statistic	-2,10	-8,85	-2,80	-8,94	-2,55	-8,68
Critical Value	-3,46	-3,46	-3,46	-3,46	-2,89	-2,89
P-Value	0,54	0,00	0,20	0,00	0,11	0,00

Table 4.3: Phillips-Perron Unit Root Test Results of Exchange Rate Series

	TL/USD		TL/EUR		TL/GBP	
	Level	Difference	Level	Difference	Level	Difference
Test Statistic	-2,25	-8,82	-2,98	-8,94	-2,49	-9,68
Critical Value	-3,46	-3,46	-3,46	-3,46	-2,89	-2,89
P-Value	0,45	0,00	0,14	0,00	0,12	0,00

Augmented Dickey-Fuller and Phillips-Perron unit root test results for each of three series are presented in tables 4.2 and 4.3 both in level and first difference forms. Since including trend and intercept terms to regression; reduce degrees of freedom and power of the test (ie., increase the probability of failing to reject a false null hypothesis); three different regression specifications are estimated consecutively when applying unit root tests; (i) none (with no trend and intercept term), (ii) intercept (with intercept term), (iii) trend and intercept (including both trend and intercept terms). Trend and intercept specification for TL/USD and TL/EUR series and intercept specification for TL/GBP series are seen to be adequate. Unit root test results, proposed that the series are nonstationary in level form but they become stationary in first difference form; for a 1% significance level. Hence all three series are integrated of order one; I (1).

Figure 4.1: ACF and PACF Pattern of TL/USD Exchange Rate



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.078	0.078	0.5948	0.441
		2	-0.087	-0.093	1.3417	0.511
		3	-0.030	-0.015	1.4313	0.698
		4	-0.002	-0.007	1.4318	0.839
		5	-0.125	-0.130	3.0217	0.697
		6	-0.073	-0.055	3.5713	0.734
		7	0.095	0.085	4.5150	0.719
		8	-0.023	-0.059	4.5733	0.802
		9	-0.038	-0.021	4.7317	0.857
		10	0.001	-0.010	4.7319	0.908

Figure 4.2: ACF and PACF Pattern of TL/EUR Exchange Rate

Lag	AC	PAC	Q-Stat	Prob
1	0.079	0.079	0.6195	0.431
2	-0.118	-0.125	1.9995	0.368
3	-0.056	-0.036	2.3125	0.510
4	-0.049	-0.057	2.5575	0.634
5	-0.051	-0.054	2.8208	0.728
6	-0.006	-0.013	2.8242	0.831
7	-0.000	-0.017	2.8242	0.901
8	-0.041	-0.051	2.9983	0.934
9	0.015	0.014	3.0216	0.963
10	-0.084	-0.106	3.7943	0.956

Figure 4.3: ACF and PACF Pattern of TL/GBP Exchange Rate

Lag	AC	PAC	Q-Stat	Prob
1	0.007	0.007	0.0051	0.943
2	-0.255	-0.255	6.4565	0.040
3	-0.038	-0.036	6.6005	0.086
4	0.069	0.004	7.0824	0.132
5	-0.064	-0.088	7.4962	0.186
6	-0.007	0.012	7.5007	0.277
7	0.216	0.196	12.395	0.088
8	0.035	0.032	12.525	0.129
9	-0.118	-0.014	14.021	0.122
10	-0.109	-0.087	15.315	0.121

Autocorrelation functions (ACF) and partial autocorrelation functions (PACF) of the series are presented in figures 4.1, 4.2 and 4.3. As can be seen from the figures, living aside the TL/GBP exchange rate, autocorrelation functions and partial auto correlation functions of exchange rate series do not follow typical patterns of autoregressive and moving average processes; hence autocorrelation functions and partial autocorrelation functions of the series do not reveal the ARIMA specifications of the series clearly. For each of the series 24 different ARIMA specification is estimated and among these an ARIMA (2, 1, 2) model for TL/USD exchange rate, an ARIMA (2, 1, 1) model for TL/EUR exchange rate and an ARIMA (1, 1, 1) model for TL/GBP exchange rate proven to be the best in terms of statistical significance of the coefficients, specification criterion (Akaike and Schwarz information criterion) and residual diagnostics. ARIMA model regression results are summarized in tables

4.4, 4.5 and 4.6 where e_{t-k} represent lagged values of exchange rate series (AR terms), ε_{t-k} represent lagged values of error terms (MA terms) and β_0.is the constant term.

Table 4.4: ARIMA Model Regression Results of TL/USD Exchange Rate

Variable	Coefficient	Std. Error	t-Statistic	P-Value
e_{t-1}	0,506	0,035	14,418	0,000
e_{t-2}	-0,919	0,034	-26,763	0,000
ε_{t-1}	-0,511	0,025	-20,159	0,000
ε_{t-2}	0,955	0,017	56,617	0,000

Table 4.5: ARIMA Model Regression Results of TL/EUR Exchange Rate

Variable	Coefficient	Std. Error	t-Statistic	P-Value
β_0	-0,004	0,001	-3,946	0,000
e_{t-1}	0,984	0,105	9,410	0,000
e_{t-2}	-0,158	0,107	-1,479	0,143
ε_{t-1}	-0,985	0,017	-57,410	0,000

Table 4.6: ARIMA Model Regression Results of TL/GBP Exchange Rate

Variable	Coefficient	Std. Error	t-Statistic	P-Value
e_{t-1}	-0,863	0,058	-14,909	0,000
ε_{t-1}	0,979	0,014	68,217	0,000

4.3. ARCH Models

When applying ARCH models to exchange rate series, returns are calculated by the logarithm-difference of the exchange rate series. Monthly return series (r_t) are presented in equation 4.4 where E_t is the exchange rate between two currencies and log is the natural logarithm of the relevant data.

$$r_t = \log\left(\frac{E_t}{E_{t-1}}\right) = \log E_t - \log E_{t-1} \qquad (4.4)$$

Descriptive statistics of the return series are summarized in table 4.7. Jarque-Berra statistic for testing whether the return series follow a normal distribution shows that a null hypothesis of normal distribution can be strongly rejected. Besides each of the series are skewed, kurtosis values largely exceed three (which is the value corresponding to normal distribution) and standard deviations are significantly different from zero. Combining all these; it can be traced that exchange rate series are volatile and they are suitable for ARCH or GARCH presentation.

Table 4.7: Descriptive Statistics of the Exchange Rate Return Series

Exchange Rate	Sample Size	Mean (%)	Standard Deviation	Skewness	Kurtosis	Jarque-Berra Test (P-Value)
TL/USD	95	-0,293	0,045	-2,092	11,075	0,000
TL/EUR	95	-0,351	0,040	-1,429	7,905	0,000
TL/GBP	95	-0,151	0,041	-1,513	8,834	0,000

In order to model volatility, initially ARIMA specification of the each return series is determined. Estimation results for various ARIMA specifications of each return series proposed that an ARIMA (2, 1, 2) specification for TL/USD return series, an ARIMA (3, 1, 2) model for TL/EUR and an ARIMA (1, 1, 1) model for TL/GBP fit the data best.

Table 4.8: ARCH Model Regression Results of TL/USD Exchange Rate

Variable	Coefficient	Std. Error	z-Statistic	P-Value
β_0	0,002	0,000	14,891	0,000
ε_{t-1}^2	-0,033	0,015	-2,260	0,024

Table 4.9: GARCH Model Regression Results of TL/EUR Exchange Rate

Variable	Coefficient	Std. Error	z-Statistic	P-Value
ε_{t-1}^2	0,398	0,257	1,548	0,122
σ_{t-1}^2	0,461	0,236	1,953	0,051

Table 4.10: GARCH Model Regression Results of TL/GBP Exchange Rate

Variable	Coefficient	Std. Error	z-Statistic	P-Value
ε_{t-1}^{2}	-0,046	0,010	-4,811	0,000
σ_{t-1}^{2}	1,081	0,043	25,407	0,000

Afterwards, square of residuals from these regressions are used as an estimator to conditional variance (volatility) of the series. A number of alternative ARCH/GARCH specifications are estimated. Estimation results showed that the volatility of TL/EUR and TL/GBP exchange rates are best identified by a GARCH (1, 1) model and the volatility of TL/USD exchange rate by an ARCH (1) model. Regression results are summarized in tables 4.8, 4.9 and 4.10 where; ε_{t-1}^{2} represent ARCH term and σ_{t-1}^{2} represent GARCH term and β_0 represent the constant term.

4.4. VAR Models

In literature, empirical studies regarding the forecasting performance of VAR models such as; Bikker (1998) and Joseph (2001) proposed that a VAR model performs better in non-stationary form than in differenced, stationary, form when applied to cointegrated variables with same degree of integration. In this respect; candidate determinant variables that are suggested by economic fundamentals based models and previous empirical studies such as; Hoque and Latif (1993), Botha and Pretorius (2009) are examined in order to specify determinant variables of TL/USD, TL/EUR, TL/GBP; Turkish Lira exchange rates (e).

$$r = \log R - \log R^{*} \tag{4.5}$$

$$m = \log M - \log M^{*} \tag{4.6}$$

After examining a number of variables with subject to Granger causality tests, cointegration relationships and unit root structure; monetary differential (m) which is the log difference of domestic country's money supply (M) and foreign country's money supply (M^{*}) and interest rate differential (r) which is the difference of one year interbank borrowing rates of domestic currency (R) and foreign currency (R^*) are picked as determinant variables among alternatives (see equation 4.5 and 4.6).

Table: 4.11: Unit Root Test Results of Determinant Variables

Variable	Specification	TL/USD		TL/EUR		TL/GBP	
		Level	Difference	Level	Difference	Level	Difference
m	Trend and Intercept	0,768	0,000	0,742	0,000	0,889	0,000
r	Intercept	0,152	0,000	0,271	0,000	0,271	0,000

Augmented Dickey-Fuller unit root test results with P-Values corresponding to each entry are presented in table 4.11. Results suggest that; both variables are non-stationary in level form and become stationary in first difference form with same degree of integration; I (1) for each of three exchange rates.

Table 4.12: Cointegration Test Results of Determinant Variables

Dependent Variable	Independent Variable	TL/USD (P-Value)	TL/EUR (P-Value)	TL/GBP (P-Value)
e	m	0,012	0,003	0,008
e	r	0,041	0,003	0,012
e	m, r	0,015	0,001	0,000
r	e	0,043	0,011	0,018
r	m	0,033	0,011	0,041
r	e, m	0,042	0,010	0,018
m	r	0,072*	0,087*	0,184*
m	e	0,006	0,050	0,057
m	r, e	0,004	0,042	0,045

Table 4.12 summarizes Engle-Granger cointegration test results for determinant variables. Results reveal that except three results, that are highlighted by asterisk in the table, all of the determinant variables are cointegrated at about 5% significance level for each of the three different Turkish Lira exchange rates. It is conferred that a VAR (2) model is the best specification for each of the three Turkish Lira exchange rates according to residual diagnostics and specification criterion such as Akaike and Schwarz information criteria. VAR model regression results are presented in tables 4.13, 4.14 and 4.15.

Table 4.13: VAR Model Regression Equation of TL/USD Exchange Rate

Independent Variables	e_{t-1}	e_{t-2}	m_{t-1}	m_{t-2}	r_{t-1}	r_{t-2}	β_0
Coefficient	1,065	-0,105	-2,033	1,960	0,0002	-0,003	-0,182
Standard Error	0,098	0,101	0,298	0,294	0,004	0,004	0,068
t-statistic	10,863	-1,032	-6,829	6,679	0,063	-0,807	-2,671

Table 4.14: VAR Model Regression Equation of TL/EUR Exchange Rate

Independent Variables	e_{t-1}	e_{t-2}	m_{t-1}	m_{t-2}	r_{t-1}	r_{t-2}	β_0
Coefficient	0,990	-0,139	-1,930	1,807	-0,0008	-0,003	-0,413
Standard Error	0,082	0,080	0,201	0,206	0,003	0,003	0,135
t-statistic	12,099	-1,745	-9,607	8,757	-0,300	-1,165	-3,055

Table 4.15: VAR Model Regression Equation of TL/GBP Exchange Rate

Independent Variables	e_{t-1}	e_{t-2}	m_{t-1}	m_{t-2}	r_{t-1}	r_{t-2}	β_0
Coefficient	0,766	0,039	-0,603	0,514	-0,009	0,006	-0,278
Standard Error	0,128	0,127	0,245	0,250	0,004	0,004	0,115
t-statistic	6,000	0,309	-2,459	2,059	-2,111	1,378	-2,419

4.5. ECM Models

Cointegration relations and unit root structures of exchange rate series and determinant variables were examined in previous sections. Since all variables are cointegrated and integrated of same order I (1); following cointegration equation can be estimated by ordinary least squares technique for each of the exchange rate series:

$$\hat{e}_t = \hat{\beta}_0 + \hat{\beta}_1 \widehat{m}_t + \hat{\beta}_1 \hat{r}_t + \hat{\beta}_2 t + \hat{\varepsilon}_t \qquad (4.7)$$

where $\hat{e}_t$, $\widehat{m}_t$, $\hat{r}_t$, t and $\hat{\varepsilon}_t$ are estimated values of exchange rate at time t, monetary differential at time t, interest rate differential at time t, trend term and residuals or errors from regression, respectively. Cointegration equation regression results for all three are presented in tables 4.16, 4.17 and 4.18.

Table 4.16: Cointegration Regression Equation of TL/USD Exchange Rate

Variable	Coefficient	Std. Error	t-Statistic	P-Value
m	0,823	0,152	5,403	0,000
r	-0,004	0,002	-1,695	0,093
β_0	2,698	0,512	5,273	0,000
t	-0,011	0,001	-8,411	0,000

Table 4.17: Cointegration Regression Equation of TL/EUR Exchange Rate

Variable	Coefficient	Std. Error	t-Statistic	P-Value
m	-0,549	0,028	-19,443	0,000
r	-0,013	0,002	-7,318	0,000
β_0	-2,178	0,067	-32,497	0,000

Table 4.18: Cointegration Regression Equation of TL/GBP Exchange Rate

Variable	Coefficient	Std. Error	t-Statistic	P-Value
m	-0,913	0,086	-10,655	0,000
r	-0,012	0,002	-7,228	0,000
β_0	-2,401	0,179	-13,414	0,000
t	0,004	0,001	5,421	0,000

Residuals or error terms ($\hat{\varepsilon}_t$) from cointegration regressions of each exchange rate series are derived through following equation:

$$\hat{\varepsilon}_t = \hat{e}_t - \hat{\beta}_0 - \hat{\beta}_1\widehat{m}_t - \hat{\beta}_1\hat{r}_t - \hat{\beta}_2 t \tag{4.8}$$

Unit root test results of the residual series proved that the series are stationary and consecutively these residual series are used for estimating following error correction model:

$$\Delta e_t = \alpha_0 + \alpha_1 \Delta m_t + \alpha_1 \Delta r_t + \hat{\varepsilon}_{t-1} + \nu_t \tag{4.9}$$

where $\hat{\varepsilon}_{t-1}$ represents the lagged values of the residuals that are derived from equation 4.8 and ν_t is the error term of error correction model equation. Error correction model regression results for each of the exchange rate series are delivered in tables 4.19, 4.20 and 4.21.

Table 4.19: ECM Regression Equation of TL/USD Exchange Rate

Variable	Coefficient	Std. Error	t-Statistic	P-Value
Δm_t	-0,258	0,251	-1,029	0,306
Δr_t	-0,018	0,003	-5,511	0,000
$\hat{\varepsilon}_{t-1}$	-0,057	0,049	1,170	0,245

Table 4.20: ECM Regression Equation of TL/EUR Exchange Rate

Variable	Coefficient	Std. Error	t-Statistic	P-Value
Δm_t	-0,206	0,208	-0,989	0,325
Δr_t	-0,013	0,003	-4,371	0,000
$\hat{\varepsilon}_{t-1}$	-0,212	0,063	-3,379	0,001

Table 4.21: ECM Regression Equation of TL/GBP Exchange Rate

Variable	Coefficient	Std. Error	t-Statistic	P-Value
Δm_t	-0,102	0,174	-0,586	0,559
Δr_t	-0,017	0,003	-5,936	0,000
$\hat{\varepsilon}_{t-1}$	-0,275	0,065	-4,241	0,000

Regression results imply that; roughly 6% of error (deviation from long run equilibrium) is corrected in one period (a month) for TL/USD exchange rate series, %21 percent is corrected for TL/EUR exchange rate series and %27 percent is corrected for TL/GBP series.

4.6. MLFFNN Models

While being so crucial for the performance of a MLFFNN there is no universally accepted methodology to determine optimal number of hidden nodes of a MLFFNN. Training, in-sample part of the dataset can be used to determine most accurate and precise neural network structure for forecasting exchange rates by using a trial-error process (Zhang et al., 2001; Palmer et al., 2006).

Number of input nodes of the MLFFNN is determined by outcomes of the VAR and cointegration analyses that are carried out in previous sections. MLFFNN structures each with six input nodes in input layer that represent variables: e_{t-1}, e_{t-2}, m_{t-1}, m_{t-2}, r_{t-1}, r_{t-2}, one hidden layer with 2, 4, 6, 8, 10, 12 hidden nodes and an output layer with one output node that represents one-step-ahead predicted value of the relevant exchange rate series, are estimated.

Optimization of weights is achieved by Levenberg-Marquardt back-propagation algorithm, a sigmoid transfer function utilized in hidden layer and linear transfer function utilized in output layer. Among these 6 different MLFFNN structures, the one that produces lowest in-sample RMSE value and correlation coefficient (R) value is selected and out-of-sample predictions are based on these MLFFNN structures.

Regression results indicated that a MLFFNN structure with 6 hidden nodes for USD and EUR exchange rates and 10 hidden nodes for GBP series came up to be fit the data best by revealing lowest in-sample RMSE values and highest R values. RMSE and R values for each exchange rate series are summarized in table 4.22.

Table 4.22: Optimal MLFFNN Structures

Exchange Rate	Number of Hidden Nodes	In-Sample RMSE	R
TL/USD	6	0,0209	0,9895
TL/EUR	6	0,0198	0,9903
TL/GBP	10	0,0271	0,9510

4.7. RNN Models

A recurrent neural network structure in which the output errors are looped back to input layer is estimated by using inputs; e_{t-1}, e_{t-2}, m_{t-1}, m_{t-2}, r_{t-1}, r_{t-2} with 2, 4, 5, 6, 8, 10, 12 hidden nodes on hidden layer. As in MLFFNN application, optimization of weights is achieved by Levenberg-Marquardt back-propagation algorithm, a sigmoid transfer function utilized in hidden layer and linear transfer function utilized in output layer.

$$y(e_t) = f(e_{t-1}, e_{t-2}, m_{t-1}, m_{t-2}, r_{t-1}, r_{t-2}, u_{t-1}, u_{t-2}) \qquad (4.10)$$

Functional form of the recurrent neural network is presented in equation 4.10 where u represent the output error. Among these 7 different RNN structures, the one that produces lowest in-sample RMSE value and R value is selected and out-of-sample predictions are based on these RNN structures.

Regression results indicated that a RNN structure with 10 hidden nodes for TL/USD exchange rate and 12 hidden nodes for TL/EUR and TL/GBP exchange rates came up to be fit the data best by revealing lowest in-sample RMSE values and highest R values. RMSE and R values for each exchange rate series are summarized in table 4.23.

Table 4.23: Optimal RNN Structures

Exchange Rate	Number of Hidden Nodes	In-Sample RMSE	R
TL/USD	10	0,0248	0,9842
TL/EUR	12	0,0172	0,9925
TL/GBP	12	0,0218	0,9705

4.8. SVR Models

Least squares support vector regression (LSSVR) methodology is employed to forecast each of the exchange rate series. Optimal selection of kernel function (linear, polynomial or radial basis function) and free parameters such as; C and σ^2 is crucial for the forecasting accuracy and the stability of the LSSVM model. However, there exist no analytical, universal way to determine kernel function and parameters optimally (Kamurazzaman et al., 2003; Ince and Trafalis, 2005; Samsudin et al., 2011; Shabri and Suhartono, 2012). Based on VAR and cointegration analysis (as in MLFFNN and RNN applications) input set: e_{t-1}, e_{t-2}, m_{t-1}, m_{t-2}, r_{t-1}, r_{t-2} is employed. Kernel function is arbitrarily chosen to be radial basis function. A 10-fold cross-validation with grid search algorithm for each of the predetermined differing ranges of the parameters is employed in order to determine optimal parameters and LSSVM structure.

For TL/USD and TL/GBP exchange rates; range of the parameter C is fixed to range (0,01 10000); lower bound of the range of the parameter σ^2 is fixed to 1, upper bound of the parameter is alternated between from 5 to 100 with a step size of 5. For TL/EUR exchange rate; range of the parameter C is fixed to range (0,01 10000); lower bound of the range of the parameter σ^2 is fixed to 0,01, upper bound of the parameter is alternated between from 100 to 2000 with a step size of 100. Therefore a 10-fold cross-validation is employed 20 times for each of the exchange rate series.

Table 4.24: Optimal SVR Structures

	C	σ^2	RMSE	R
TL/USD	1192,1	152,5758	0,0325	0,9506
TL/EUR	2329	718,6	0,0246	0,9559
TL/GBP	8,1796	57,0765	0,0368	0,7414

After determining LSSVM structure with optimal parameters through cross-validation for each range of the parameters; in-sample RMSE and R values are calculated consecutively and out-of-sample predictions are based on the LSSVR structure that produces best in-sample fit, i.e. the one that produces lowest in-sample RMSE value and highest in-sample R value. Values of the parameters, in-sample RMSE and R values are presented in table 4.24.

4.9. Forecast Evaluation

Root mean squared error (RMSE) statistic corresponding to in-sample and out-of-sample TL/USD, TL/EUR and TL/GBP exchange rate forecasts of all seven models and their ranks are tabulated in tables 4.25, 4.26, 4.27.

Table 4.25: RMSE Statistics for TL/USD Exchange Rate

Model	In-Sample RMSE	Rank	Out-of-Sample RMSE	Rank
ARIMA (2, 1, 2)	0,0433	6	0,0261	2
GARCH (1, 0)	0,0445	7	0,0265	3
VAR (2)	0,0342	4	0,0420	7
ECM	0,0380	5	0,0221	1
MLFFNN (6-6-1)	0,0209	1	0,0305	4
RNN (6-10-1)	0,0248	2	0,0318	5
SVR (1192,1 152,6)	0,0325	3	0,0363	6

Regarding the in-sample forecasts of TL/USD exchange rate; among classical models VAR model performed best and among artificial intelligence based models MLFFNN model performed best. Furthermore; MLFFNN model produced lowest in-sample RMSE value among all seven models. For out-of-sample forecasts; among classical models ECM performed best and among artificial intelligence based models MLFFNN model performed best. Among all seven models; ECM produced significantly lower out-of-sample RMSE than remaining six models.

Table 4.26: RMSE Statistics for TL/EUR Exchange Rate

Model	In-Sample RMSE	Rank	Out-of-Sample RMSE	Rank
ARIMA (2, 1, 1)	0,0384	6	0,0226	2
GARCH (1, 1)	0,0391	7	0,0235	3
VAR (2)	0,0258	4	0,0289	7
ECM	0,0344	5	0,0199	1
MLFFNN (6-6-1)	0,0198	2	0,0279	4
RNN (6-12-1)	0,0172	1	0,0283	5
SVR (2329 718,6)	0,0246	3	0,0288	6

For the in-sample forecasts of TL/EUR exchange rates; among classical models VAR model performed best and among artificial intelligence based models RNN model performed best. Besides; lowest in-sample RMSE value among all seven models is achieved by RNN model. Regarding out-of-sample forecasts; among classical models ECM performed best and among artificial intelligence based models MLFFNN model performed best. As with the out-of-sample forecast results for TL/USD exchange rate ECM performed significantly better than remaining models. Another salient feature of the out-of-sample forecasts of TL/EUR exchange rate is that; its ranks are exactly the same as the out-of-sample forecasts of the TL/USD exchange rate.

Table 4.27: RMSE Statistics for TL/GBP Exchange Rate

Model	In-Sample RMSE	Rank	Out-of-Sample RMSE	Rank
ARIMA (1, 1, 1)	0,0403	6	0,0249	2
GARCH (1, 1)	0,0409	7	0,0257	3
VAR (2)	0,0374	5	0,0318	6
ECM	0,0305	3	0,0233	1
MLFFNN (6-10-1)	0,0271	2	0,0297	5
RNN (6-12-1)	0,0218	1	0,0370	7
SVR (8,18 57,1)	0,0368	4	0,0284	4

In terms of in-sample forecasting ability of TL/GBP exchange rate; ECM performed best among classical models and among artificial intelligence based models RNN model performed best. In addition, lowest in-sample RMSE value among all seven models is achieved by RNN model. Regarding out-of-sample forecasts; among classical models ECM performed best and among artificial intelligence based models SVR model performed best. As in TL/USD and TL/EUR exchange rates ECM produced lowest out-of-sample RMSE among all seven models.

Combining all these results regarding the in-sample and out-of sample forecasting performances of seven different models for TL/USD, TL/EUR, TL/GBP exchange rates; it can be traced that; in-sample forecasting performance of artificial intelligence based models is significantly better than classical models. Besides, classical models produced lower out-of-sample RMSE values than artificial intelligence based models in most instances. Among all seven models; ECM performed best in out-of-sample forecasts of all three exchange rates.

Table 4.28: Diebold-Mariano Test Results of TL/USD Exchange Rate

Model	DMTEST1		DMTEST2		DMTEST3	
	MSE Ratio	DM Stat.	MSE Ratio	DM Stat.	MSE Ratio	DM Stat.
ARIMA	0,7323	-1,4916	0,6736	-2,0400**	0,5170	-2,7941**
GARCH	0,7549	-1,3125	0,6944	-1,7830	0,5329	-2,5404**
VAR	1,8963	4,5960*	1,7444	2,2737*	1,3387	1,5066
ECM	0,5250	-2,8050**	0,4830	-2,6561**	0,3707	-3,3083**

Table 4.29: Diebold-Mariano Test Results of TL/EUR Exchange Rate

Model	DMTEST1		DMTEST2		DMTEST3	
	MSE Ratio	DM Stat.	MSE Ratio	DM Stat.	MSE Ratio	DM Stat.
ARIMA	0,6562	-1,2760	0,6377	-1,6015	0,6158	-1,4038
GARCH	0,7095	-1,1178	0,6895	-1,3708	0,6658	-1,3159
VAR	1,0730	0,7995	1,0429	0,3784	1,0070	0,0681
ECM	0,5087	-1,9091	0,4945	-2,2613**	0,4774	-2,0865**

Table 4.30: Diebold-Mariano Test Results of TL/GBP Exchange Rate

Model	DMTEST1		DMTEST2		DMTEST3	
	MSE Ratio	DM Stat.	MSE Ratio	DM Stat.	MSE Ratio	DM Stat.
ARIMA	0,7029	-1,5038	0,4529	-2,0729**	0,7687	-1,3763
GARCH	0,7488	-1,3331	0,4825	-2,0140**	0,8189	-1,2089
VAR	1,1464	0,7638	0,7387	-1,1833	1,2538	1,8119
ECM	0,6155	-1,5988	0,3966	-2,2605**	0,6731	-1,2660

In order to assess; statistical significance of differences between out-of-sample forecasts of classical models and artificial intelligence based models, i.e. to test whether the RMSE differences among models are statistically significant; Diebold-Mariano (DM) test is employed for all three exchange rates. MSE ratios (ratio of MSE of classical model to MSE of benchmark model), calculated Diebold-Mariano test statistics, for each of the three exchange rates are presented in tables 4.28, 4.29 and 4.30. DMTEST1, DMTEST2 and DMTEST3 panels represent corresponding outcomes when MLFFNN model is chosen to be as benchmark, RNN model is chosen to be as benchmark and SVR model is chosen to be as benchmark respectively. In tables, test results suggesting that; for a %5 significance level benchmark model statistically significantly outperform classical model are highlighted by an asterisk at the corresponding model's Diebold-Mariano test statistic entry and test results suggesting that for a %5 significance level the classical model statistically significantly outperform benchmark model are highlighted by two asterisks at the corresponding model's Diebold-Mariano test statistic entry.

As can be seen from tables; among thirteen statistically significant results, i.e. Diebold-Mariano test results that suggest rejection of the null hypothesis of equal forecast accuracy, only two outcomes (for TL/USD exchange rate, VAR model when MLFFNN model is the benchmark and again for TL/USD exchange rate, VAR model when RNN model is the benchmark) are in favor of artificial intelligence based models. For TL/USD exchange rate; classical models statistically significantly outperformed artificial intelligence based models in six instances (for ECM when MLFFNN model is the benchmark; for ARIMA model and ECM when RNN model is the benchmark; for ARIMA, GARCH and ECM models when SVR model is the benchmark). For TL/EUR exchange rate; classical models statistically significantly outperformed artificial intelligence based models in two instances (for ECM when RNN and SVR models are the benchmarks). For TL/GBP exchange rate classical models statistically significantly outperformed artificial intelligence based models in three instances (for ARIMA, GARCH and ECM models when RNN model is the benchmark).

To sum up; according to DM test results that are represented in tables 4.28, 4.29 and 4.30 for three different exchange rates, among thirty-six instances, thirteen are statistically significant and among these statistically significant results eleven are in favor of classical models and only two are in favor of artificial intelligence based models.

Table 4.31: Diebold-Mariano Test Results for RW as Benchmark Model

Model	TL/USD		TL/EUR		TL/GBP	
	MSE Ratio	DM Stat.	MSE Ratio	DM Stat.	MSE Ratio	DM Stat.
ARIMA	0,4215	-3,4426**	0,5869	-1,4187	0,5061	-2,3352**
GARCH	0,4345	-3,3055**	0,6346	-1,3818	0,5392	-2,2826**
VAR	1,0916	0,5127	0,9597	-0,1570	0,8255	-1,0695
ECM	0,3022	-3,6333**	0,4551	-1,9677**	0,4432	-2,1360**
MLFFNN	0,5756	-2,6842**	0,8945	-0,5492	0,7201	-1,2355
RNN	0,6258	-2,4456**	0,9203	-0,2735	1,1176	0,8013
SVR	0,8154	-1,3958	0,9531	-0,2177	0,6584	-2,2163**

In addition to these findings; following ongoing debates in exchange rate forecasting literature related with finding a model that is capable of beating random walk model; forecasts of all seven models in consideration are compared with the forecasts of a pure random walk model on the basis of Diebold-Mariano test. Test results, which are represented in table 4.31, suggest that; models in consideration are capable of beating random walk model in ten out of twenty-one instances which are highlighted by two asterisks in the corresponding Diebold-Mariano test entry while random walk model is not capable of statistically significantly beating any of the models in consideration.

5. CONCLUDING REMARKS

This paper investigated predictive ability of several classical models and artificial intelligence based models in forecasting; TL/USD, TL/EUR and TL/GBP exchange rates. Both in-sample and out-of-sample forecasts of each model are evaluated by root mean squared error statistic and statistical significance of the forecast accuracy differences among the models are assessed by Diebold-Mariano test.

Empirical analysis aimed to contribute the exchange rate forecasting literature in a number of ways. First, several classical, pure time series models and artificial intelligence based models are employed to three different Turkish Lira exchange rates (TL/USD, TL/EUR and TL/GBP) and their predictive ability is analyzed comparatively. Second, candidate determinant variables suggested by the theories and previous empirical studies are examined for Turkish Lira exchange rates through causality and cointegration analyses. Third, after identifying determinant variables of Turkish Lira exchange rates, two models are estimated (a model in VAR representation and a model ECM representation) using these variables and their forecasting accuracy is compared by those of artificial intelligence based models. Fourth; predictive ability of all of the seven models in consideration (ARIMA, GARCH, VAR, ECM, MLFFNN, RNN, SVR) is investigated by comparing their forecasts with forecasts from a pure random walk model.

Comparative analysis of root mean squared error (RMSE) statistic corresponding to in-sample and out-of-sample TL/USD, TL/EUR, TL/GBP exchange rate forecasts of all models proposed that; artificial intelligence based models performed significantly better than classical models for in-sample forecasts. Besides, classical models produced lower out-of-sample RMSE values than artificial intelligence based models in most instances. Among all seven models; ECM model performed best in out-of-sample forecasts for all three exchange rates.

According to Diebold-Mariano test results for out-of-sample forecasts of classical and artificial intelligence based models, among thirteen statistically significant results only two of them were in favor of artificial intelligence based models and remaining eleven results were in favor of classical models. Forecasting models in consideration were capable of beating random walk model in ten out of twenty-one instances while random walk model was not capable of statistically significantly beating any of the models in consideration.

Such dismal performance of artificial intelligence based models may be due to relatively low frequency data (monthly) employed for the analysis. Hence, conducting a comparative analysis for forecasting Turkish Lira exchange rates by making use of higher frequency data, daily for instance, and also proposing a hybrid model by combining classical models and artificial intelligence based models or enhancing accuracy of the structures of the artificial intelligence based models through optimization algorithms remain as future research subjects.

REFERENCES

Adler, M., & Lehmann, B. (1983). Deviations from purchasing power parity in the long run. *The Journal of Finance,* 37(5), 1471-1487.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review,* 39,: 885–905.

Augustine, C. A., Malindretos, J., & Kiseok N. (2010). Cointegration, dynamic structure, and the validity of purchasing power parity in African countries. *International Review of Economics and Finance,* 19, 755-768.

Bahramy, F., & Crone, S. F. (2013). Forecasting Foreign Exchange Rates using Support Vector Regression. *Proceedings of the 2013 Ieee Conference on Computational Intelligence for Financial Engineering & Economics (Cifer),* 34-41.

Baum, C. F., Barkoulas, J. T., & Caglayan, M. (1999). Long memory or structural breaks: can either explain nonstationary real exchange rates under the current float. *Journal of International Financial Markets Institutions & Money,* 9, 359-376.

Beckmann, J. (2013). Nonlinear Exchange Rate Adjustment and the Monetary Model. *Review of International Economics*, 21(4), 654-670.

Bikker, J. A. (1998). Inflation forecasting for aggregates of the EU-7 and EU-14 with Bayesian VAR models. *Journal of Forecasting*, 17(2), 147-165.

Bilson, J. F. O. (1978). The monetary approach to the exchange rate: some empirical evidence. *IMF Staff Papers*, 25, 48-75.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327.

Botha, I., & Pretorius, M. (2009). Forecasting the exchange rate in South Africa: A comparative analysis challenging the random walk model. *African Journal of Business Management*, 3(9), 486-494.

Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis; forecasting and control. San Francisco, Holden-Day

Cacoullos T., (1966). Estimation of a multivariate density. *Annals of Institute of Statistical Mathematics*, 18(2), 179-189.

Cassel, G. (1921). The world's monetary problems. E. P. Dutton and Co., New York.

Charytoniuk, W., & Chen, M. S. (2000). Neural network design for short-term load forecasting. Drpt2000: International Conference on Electric Utility

Deregulation and Restructuring and Power Technologies, Proceedings, 554-561.

Chen, K. Y. (2011). Combining linear and nonlinear model in forecasting tourism demand. *Expert Systems with Applications*, 38(8), 10368-10376.

Chen, A. S., & Leung, M. T. (2003). A Bayesian vector error correction model for forecasting exchange rates. *Computers & Operations Research*, 30(6), 887-900.

Chen, A. S., Leung, M. T., & Daouk, H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. *Computers & Operations Research*, 30(6), 901-923.

Chortareas, G., Jiang, Y., & Nankervis, J. C. (2011). Forecasting exchange rate volatility using high-frequency data: Is the euro different? *International Journal of Forecasting*, 27(4), 1089-1107.

Chung, S. K. (2006). The out-of-sample forecasts of nonlinear long-memory models of the real exchange rate. *International Journal of Finance & Economics*, 11(4), 355-370.

Cushman, D. O. (2008). Long-run PPP in a system context: No favorable evidence after all for the U.S., Germany, and Japan. *Journal of International Financial Markets, Instutions & Money*, 18, 413-424.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time-Series with a Unit Root. J*ournal of the American Statistical Association*, 74(366), 427-431.

Dunis, C. L., & Huang, X. H. (2002). Forecasting and trading currency volatility: An application of recurrent neural regression and model combination. *Journal of Forecasting*, 21(5), 317-354.

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, 14(2), 179-211.

Enders, W. (1977). Portfolio Balance and Exchange-Rate Stability. *Journal of Money Credit and Banking*, 9(3), 491-499.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom Inflation. *Econometrica*, 50(1), 987-1007.

Fang, S. F., Wang, M. P., & Song, M. (2009). An approach for the aging process optimization of Al-Zn-Mg-Cu series alloys. *Materials & Design*, 30(7), 2460-2467.

Fernandez-Rodriguez, F., Sosvilla-Rivero, S., & Andrada-Felix, J. (1999). Exchange-rate forecasts with simultaneous nearest-neighbour methods: evidence from the EMS. *International Journal of Forecasting*, 15(4), 383-392.

Frenkel, J. A. (1976). A monetary approach to the exchange rate: doctrinal aspects and empirical evidence. *Scandinavian Journal of Economics,* 78, 200-224.

Frenkel, J. A. (1978). Purchasing power parity doctrinal perspective and evidence from the 1920s. *Journal of International Economics*, 8(2), 169-191.

Frenkel, J. A. (1981). The collapse of purchasing power parities during the 1970's. *European Economic Review*, 16, 217-237.

Fritsche, C. P., & Wallace, M. (1997). Forecasting the exchange rate PPP versus a random walk. *Economics Letters*, 54(1), 69-74.

Gan, C., Limsombunchai, V., Clemes, M., & Weng, A. (2005). Consumer Choice Prediction: Artificial Neural Networks versus Logistic Model. Modsim 2005: *International Congress on Modelling and Simulation: Advances and Applications for Management and Decision Making*, 51-57.

Gharleghi, B., & Nor, A. H. S. M. (2012). Is Monetary Variable a Determinant in the Ringgit-Dollar Exchange Rates Model?: A Cointegration Approach. *Sains Malaysiana*, 41(9), 1163-1169.

Greene, W. H. (2003). Econometric analysis (5th ed.). Upper Saddle River, N.J.: Prentice Hall.

Gujarati, D. N., & Porter, D. C. (2009). Basic econometrics (5th ed.). Boston: McGraw-Hill Irwin

Hajmeer, M., & Basheer, I. (2003). Comparison of logistic regression and neural network-based classifiers for bacterial growth. *Food Microbiology*, 20(1), 43

Hamilton, J. D. (1994). Time series analysis. Princeton, N.J.: Princeton University Press

Hammerstrom, D. (1993). Neural Networks at Work. *Ieee Spectrum*, 30(6), 26-32.

Haykin, S. S. (1999). Neural networks : a comprehensive foundation (2nd ed.). Upper Saddle River, N.J.: Prentice Hall.

Hinton, G. E. (1992). How Neural Networks Learn from Experience. *Scientific American*, 267(3), 145-151.

Hoque, A., & Latif, A. (1993). Forecasting Exchange-Rate for the Australian Dollar Vis-a-Vis the United-States Dollar Using Multivariate Time-Series Models. *Applied Economics*, 25(3), 403-407

Ince, H., & Trafalis, T. B. (2006). A hybrid model for exchange rate prediction. *Decision Support Systems*, 42(2), 1054-1062.

Isard, P. (1995). Exchange rate economics. Cambridge England ; New York: Cambridge University Press.

Johnston, J., & Dinardo, J. (1997). Econometric methods (4th ed.). New York: McGraw-Hill.

Jordon, M.I. (1989). Serial order: A parallel distributed processing. In J. L. Elman and D. E. Rumelhart , editors, Advance in connectionist theory: Speech, Elbaum.

Jorion, P., & Sweeney, R. J. (1996). Mean reversion in real exchange rates: evidence and implications for forecasting. *Journal of International Money and Finance*, 15(4), 535-550.

Joseph, N. L. (2001). Model specification and forecasting foreign exchange rates with vector autoregressions. *Journal of Forecasting*, 20(7), 451-484.

Junttila, J., & Korhonen, M. (2011). Nonlinearity and time-variation in the monetary model of exchange rates. *Journal of Macroeconomics*, 33(2), 288-302.

Kamruzzaman, D., Sarker, R. A., & Ahmad, I. A. (2003). SVM based models for predicting foreign currency exchange rates. *Third Ieee International Conference on Data Mining, Proceedings*, 557-560.

Khashei, M., Bijari, M., & Ardali, G. A. R. (2009). Improvement of Auto-Regressive Integrated Moving Average models using Fuzzy logic and Artificial Neural Networks (ANNs). *Neurocomputing*, 72(4-6), 956-967.

Kilian, L., & Taylor, M. P. (2003). Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, 60(1), 85-107.

Kim, B. H., Min, H. G., & Moh, Y. K. (2010). Nonlinear dynamics in exchange rate deviations from the monetary fundamentals: An empirical study. *Economic Modelling*, 27(5), 1167-1177.

Kim, B. J. C., & Mo, S. W. (1995). Cointegration and the Long-Run Forecast of Exchange-Rates. *Economics Letters*, 48(3-4), 353-359

Kuan, C. M., & Liu, T. (1995). Forecasting Exchange-Rates Using Feedforward and Recurrent Neural Networks. *Journal of Applied Econometrics*, 10(4), 347-364

Liao, R. J., Zheng, H. B., Grzybowski, S., & Yang, L. J. (2011). Particle swarm optimization-least squares support vector regression based forecasting model on dissolved gases in oil-filled power transformers. *Electric Power Systems Research*, 81(12), 2074-2080.

Lin, G. F., Chen, G. R., Wu, M. C., & Chou, Y. C. (2009). Effective forecasting of hourly typhoon rainfall using support vector machines. *Water Resources Research*, 45.

Liu, T. R., Gerlow, M. E., & Irwin, S. H. (1994). The Performance of Alternative Var Models in Forecasting Exchange-Rates. *International Journal of Forecasting*, 10(3), 419-433.

Luk, K. C., Ball, J. E., & Sharma, A. (2001). An application of artificial neural networks for rainfall forecasting. *Mathematical and Computer Modelling*, 33(6-7), 683-693.

Mak, K. L., & Yang, D. H. (2007). Forecasting Hong Kong's container throughput with approximate least squares support vector machines. *World Congress on Engineering* 2007, Vols 1 and 2, 7-12.

Malinov, S., Sha, W., & McKeown, J. J. (2001). Modelling the correlation between processing parameters and properties in titanium alloys using artificial neural network. *Computational Materials Science*, 21(3), 375-394.

Marra, S., & Morabito, F. C. (2005). A new technique for solar activity forecasting using recurrent Elman networks. *Enformatika*, Vol 7: Iec 2005 Proceedings, 68-73.

Mcculloch, W. S., & Pitts, W. (1943). *Bulletin of Mathematical Biophysics*, Vol 5, 115-133,.

McCrae, M., Lin, Y. X., Pavlik, D., & Gulati, C. M. (2002). Can cointegration-based forecasting outperform univariate models? An application to Asian exchange rates. *Journal of Forecasting*, 21(5), 355-380.

McMillan, D. G., & Speight, A. E. H. (2004). Daily volatility forecasts: Reassessing performance of GARCH models. *Journal of Forecasting*, 23(6), 449-460.

Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: do they fit out of sample? *Journal of International Economics*, 14, 3-24.

Minsky, M. L. (1961). Steps Towards Artificial Intelligence. *Proceedings of the Institute of Radio Engineers*, 49, 8-30.

Mohandes, M. (2002). Support vector machines for short-term electrical load forecasting. *International Journal of Energy Research*, 26(4), 335-345.

More, A., & Deo, M. C. (2003). Forecasting wind with neural networks. *Marine Structures*, 16(1), 35-49.

Moura, M. D., Zio, E., Lins, I. D., & Droguett, E. (2011). Failure and reliability prediction by support vector machines regression of time series data. *Reliability Engineering & System Safety*, 96(11), 1527-1534.

Ni, H., & Yin, H. J. (2009). Exchange rate prediction using hybrid neural networks and trading indicators. *Neurocomputing*, 72(13-15), 2815-2823.

Pai, P. F., & Hong, W. C. (2007). A recurrent support vector regression model in rainfall forecasting. *Hydrological Processes*, 21(6), 819-827.

Palmer, A., Montano, J. J., & Sese, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management*, 27(5), 781-790.

Panda, C., & Narasimhan, V. (2007). Forecasting exchange rate better with artificial neural network. *Journal of Policy Modeling*, 29(2), 227-236.

Parry, M. E., Cao, Q., & Song, M. (2011). Forecasting new product adoption with probabilistic neural networks. *Journal of Product Innovation Management*, 28, 78-88.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33, 1065-1076.

Phillips, P. C. B., & Perron, P. (1988). Testing for a Unit-Root in Time-Series Regression. *Biometrika*, 75(2), 335-346.

Preminger, A., & Franck, R. (2007). Forecasting exchange rates: A robust regression approach. *International Journal of Forecasting*, 23(1), 71-84.

Robinson, J. (1947). The foreign exchanges. *Essays in the theory of employment,* A. R. Mowbray & Co. Limited, London.

Rogoff, K. (1996). The purchasing power parity puzzle. *Journal of Economic Literature*, 34(2), 647-668.

Rosenberg, M. R. (1996). Currency forecasting: methods and models for predicting exchange rates. Times Mirror Higher Education Group Inc., Chicago.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning Representations by Back-Propagating Errors. *Nature*, 323(6088), 533-536.

Rumelhart, D. E., Widrow, B., & Lehr, M. A. (1994). The Basic Ideas in Neural Networks. *Communications of the Acm*, 37(3), 87-92.

Saad, E. W., Prokhorov, D. V., & Wunsch, D. C. (1998). Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. *Ieee Transactions on Neural Networks*, 9(6), 1456-1470.

Samsudin, R., Saad, P., & Shabri, A. (2011). River flow time series using least squares support vector machines. *Hydrology and Earth System Sciences*, 15(6), 1835-1852.

Sargan, J. D. (1964). Wages and prices in the United Kingdom: a study in econometric methodology in K. F. Wallis and D. F. Henry, eds., Quantitative Economics and Econometrics Analysis, Basis Blackwell, Oxford

Shabri, A., & Suhartono. (2012). Streamflow forecasting using least-squares support vector machines. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 57(7), 1275-1293.

Smola, A. J., & Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199-222.

Specht, D. F. (1990). Probabilistic Neural Networks. *Neural Networks*, 3(1), 109-118.

Suykens, J. A. K., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3), 293-300.

Tay, F. E. H., & Cao, L. J. (2001). Application of support vector machines in financial time series forecasting. *Omega-International Journal of Management Science*, 29(4), 309-317.

Taylor, M. P. (1988). An empirical examination of long run purchasing power parity using cointegration techniques. *Applied Economics*, 20, 1369-1381.

Taylor, M. P., & Peel, D. A. (2000). Nonlinear adjustment, long-run equilibrium and exchange rate fundamentals. *Journal of International Money and Finance*, 19(1), 33-53.

Tenti, P. (1996). Forecasting foreign exchange rates using recurrent neural networks. *Applied Artificial Intelligence*, 10(6), 567-581.

Vapnik, V. N. (1995). The nature of statistical learning theory. Springer-Verlag: New York.

Vapnik, V. N., Golowich, S. E., & Smola, A. J. (1996). Support vector method for function approximation, regression and estimation and signal processing. *Advance in Neural Information Processing Systems*, MIT Press: Cambridge, 281-287.

Van Aarle, B., Boss, M., & Hlouskova, J. (2000). Forecasting the euro exchange rate using vector error correction models. *Weltwirtschaftliches Archiv-Review of World Economics*, 136(2), 232-258.

Verkooijen, W. (1996). A neural network approach to long-run exchange rate prediction. *Computational Economics*, 9(1), 51-65.

Vilasuso, J. (2002). Forecasting exchange rate volatility. *Economics Letters*, 76(1), 59-64.

West, K. D., & Cho, D. (1995). The Predictive Ability of Several Models of Exchange-Rate Volatility. *Journal of Econometrics*, 69(2), 367-391.

Wu, B. (1995). Model-Free Forecasting for Nonlinear Time-Series (with Application to Exchange-Rates). *Computational Statistics & Data Analysis*, 19(4), 433-459.

Wu, S. M., & Akbarov, A. (2011). Support vector regression for warranty claim forecasting. *European Journal of Operational Research*, 213(1), 196-204.

Wu, J. L., & Chen, S. L. (2001). Nominal exchange-rate prediction: evidence from a nonlinear approach. *Journal of International Money and Finance*, 20(4), 521-532.

Yasdi, R. (1999). Prediction of road traffic using a neural network approach. *Neural Computing & Applications*, 8(2), 135-142.

Zhang, G., & Hu, M. Y. (1998). Neural network forecasting of the British pound US dollar exchange rate. *Omega-International Journal of Management Science,* 26(4), 495-506.

Zhang, Y., & Liu, Y. C. (2009). Traffic forecasting using least squares support vector machines. *Transportmetrica*, 5(3), 193-213.

Zhang, G. Q., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), 35-62.

Zhang, G. P., Patuwo, B. E., & Hu, M. Y. (2001). A simulation study of artificial neural networks for nonlinear time-series forecasting. *Computers & Operations Research*, 28(4), 381-396.

Zhou, J. Y., Shi, J., & Li, G. (2011). Fine tuning support vector machines for short-term wind speed forecasting. *Energy Conversion and Management*, 52(4), 1990-1998.

Zhu, B. Z., & Wei, Y. M. (2013). Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. *Omega-International Journal of Management Science*, 41(3), 517-524.

CURRICULUM VITAE

Ali Fehim Cebeci was born in İstanbul. He received his bachelor’s degree in economics from Marmara University in 2011. He worked as a research assistant at Department of Economics of Adnan Menderes University (Aydın-Turkey) and currently works as a research assistant at Department of Economics of Gebze Institute of Technology.