Royal Docks Business School

MKM227 Postgraduate Dissertation

Student Number: U1252727

	Comments	Max Mark	Actual Mark
Introduction *Identification of a valid topic, research question and objectives framed to Masters Level standard with academic rationale developed, clear industry contextualisation of the research topic*	**Supervisor Comments:**	10%	
	2nd marker Comments:		
Critical Literature Review *Depth and breadth of literature search, engagement with seminal authors and papers, evidence of a critical approach toward the scholarly literature*	**Supervisor Comments:**	25%	
	2nd marker Comments:		
Research Methodology *Evaluation of research philosophies and*	**Supervisor Comments:**	15%	

perspectives. Justification of methodological approach, sampling strategy, data analysis and reliability and validity measures as applicable	**2nd marker Comments:**		
Data Analysis and Interpretation *Evidence of rigor in data analysis and interpretation procedures, identification of key patterns and themes in the research data, integration of academic theory into explanation of findings*	**Supervisor Comments:**	**35%**	
	2nd marker Comments:		
Conclusions and Recommendations *Research question and objectives addressed with implications to theoretical and managerial concepts considered. Recommendations provided for theory, practice and future research*	**Supervisor Comments:**	**10%**	
	2nd marker Comments:		
Organisation, presentation and references. *Well structured and ordered dissertation*	**Supervisor Comments:**	**5%**	

<table>
<tr><td>with correct use of grammar and syntax. In-text citation and bibliography conforming to “Cite Them Right”</td><td>2nd marker Comments:</td><td></td><td></td></tr>
<tr><td rowspan="2">Total</td><td>First Marker Total</td><td rowspan="2">100%</td><td></td></tr>
<tr><td>Second Marker Total</td><td></td></tr>
<tr><td colspan="2">Supervisor General Comments:</td><td colspan="2" rowspan="2">Agreed Mark:</td></tr>
<tr><td colspan="2">2nd Marker General Comments:</td></tr>
</table>

Supervisor’s Name: ……………………………………….. Signature: …………………………

2nd Marker’s Name: ………………………………………. Signature: …………………………

Forecasting Foreign Exchange Rates with the use of Artificial Neural Networks/Learning Machines and comparison with Traditional Concepts and Linear Models

A dissertation submitted in partial fulfilment of the requirements of the Royal Docks Business School, University of East London for the degree of **[Insert Full Programme Title Here]**

September 2014

Approx. 13,000

I declare that no material contained in the thesis has been used in any other submission for an academic award

Student Number: U1252727 Date: 09/09/2014

Dissertation Deposit Agreement

Libraries and Learning Services at UEL is compiling a collection of dissertations identified by academic staff as being of high quality. These dissertations will be included on ROAR the UEL Institutional Repository as examples for other students following the same courses in the future, and as a showcase of the best student work produced at UEL.
This Agreement details the permission we seek from you as the author to make your dissertation available. It allows UEL to add it to ROAR and make it available to others. You can choose whether you only want the dissertation seen by other students and staff at UEL ("Closed Access") or by everyone worldwide ("Open Access").

I DECLARE AS FOLLOWS:

- That I am the author and owner of the copyright in the Work and grant the University of East London a licence to make available the Work in digitised format through the Institutional Repository for the purposes of non-commercial research, private study, criticism, review and news reporting, illustration for teaching, and/or other educational purposes in electronic or print form
- That if my dissertation does include any substantial subsidiary material owned by third-party copyright holders, I have sought and obtained permission to include it in any version of my Work available in digital format via a stand-alone device or a communications network and that this permission encompasses the rights that I have granted to the University of East London.
- That I grant a non-exclusive licence to the University of East London and the user of the Work through this agreement. I retain all rights in the Work including my moral right to be identified as the author.
- That I agree for a relevant academic to nominate my Work for adding to ROAR if it meets their criteria for inclusion, but understand that only a few dissertations are selected.
- That if the repository administrators encounter problems with any digital file I supply, the administrators may change the format of the file. I also agree that the Institutional Repository administrators may, without changing content, migrate the Work to any medium or format for the purpose of future preservation and accessibility.

- That I have exercised reasonable care to ensure that the Work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

- That I understand that the University of East London does not have any obligation to take legal action on behalf of myself, or other rights holders, in the event of infringement of intellectual property rights, breach of contract or of any other right, in the Work.

I FURTHER DECLARE:

- That I can choose to declare my Work “Open Access”, available to anyone worldwide using ROAR without barriers and that files will also be available to automated agents, and may be searched and copied by text mining and plagiarism detection software.
- That if I do not choose the Open Access option, the Work will only be available for use by accredited UEL staff and students for a limited period of time.

Dissertation Details

Field Name	**Details to complete**
Title of thesis *Full title, including any subtitle*	**Forecasting Foreign Exchange Rates with the use of Artificial Neural Networks/Learning Machines and comparison with Traditional Concepts and Linear Models**
Author *Separate the surname (family name) from the forenames, given names or initials with a comma, e.g. Smith, Andrew J.*	Alrajhi, Omar Hamad
Supervisor(s)/advisor *Format as for author.*	Dr. Karathanasopoulos, Andreas
Author Affiliation *Name of school where you were based*	Royal Docks Business School U.E.L.
Qualification name *E.g. MA, MSc, MRes, PGDip*	MSc
Course Title *The title of the course e.g.*	MSc Finance & Risk
Date of Dissertation *Date submitted in format: YYYY-MM*	2014-09
Do you want to make the dissertation Open Access (on the public web) or Closed Access (for UEL users only)?	Open ■ Closed □

By returning this form electronically from a recognised UEL email address or UEL network system, I grant UEL the deposit agreement detailed above. I understand inclusion on and removal from ROAR is at UEL's discretion.

Name: Omar Hamad Alrajhi

Signature: .... Date: 09/09/2014

Royal Docklands Business School

UEL

Forecasting Foreign Exchange Rates with the use of Artificial Neural Networks/Learning Machines and comparison with Traditional Concepts and Linear Models

Post-Graduate Disseration

Due: 09/09/2014

Contents:

Abstract

Acknowledgement

1.0 Introduction

2.0 Literature Review

2.1 Artificial Neural Networks Historical Story

2.2 High-Order Neural Networks

2.3 Multi-Layer Perceptrons

2.4 Support-Vector Regression and Support-Vector Machines

2.5 Genetic Expression Programming

2.6 Hybrid Genetic Algorithm Support Vector Regression Models

2.7 Cons of Neural Networks

2.8 Theoretical Background

3.0 Methodology

3.1 Statistical Evaluation

4.0 Analysis

4.1 Null Hypothesis Test

4.2 Base Models

4.3 Random walk

4.4 Buy & Hold

5.0 Linear Models

5.1 Moving Average Convergence Divergence

5.2 Auto-Regressive Moving Average

6.0 Artificial Neural Networks

6.1 High-Order Neural Networks

6.2 Multi-Layer Perceptron

6.3 Gene Expression Programming and Genetic Algorithms

6.4 Statistical Evaluation

7.0 Conclusion

8.0 Recommendations

Sources

Appendix

Abstract

The prediction of Foreign Exchange has been an ever-going learning process. The development of methods of prediction has come a long way, from the beginning where the though there was no ability to predict the future, and behavior is an unpredictable entity to the development of simple statistical linear models that has come a long way to todays technology world where computers and their computational powers have made it possible for Artificial Intelligence to be born.

This paper will be going through previous studies on these Neural Networks to forecast the EUR/USD, GBP/USD and USD/JPY to test and review their ability to forecast one day ahead.

Acknowledgement:

I would like to start my thanks to my parents Hamad Nasser Alrajhi and Ruth Rebecca Alrajhi for their kind and loving support during my studies over the course of my life. I would like to thank Dr. Andreas Karathanasopoulos for his supportive help as my supervisor during my dissertation period and that he is the one who inspired me to work on this paper. I would like to thank everyone staff and friends that overwhelming kindness.

1.0 Introduction

The primordial foundation of any economy is money, with which we all know that it is a medium of exchange for goods and services. Given this as a fact, when an economy is vibrant, there would be more movement of in transactions where banks, businesses and individuals all try to strive to make a profit and vise-versa when it is on a slow down. This is a key component of the growth of an economy. With this in mind, there are as we all know other countries have an economy that interact with each other in exports, imports and of course the movement of capital to invest in a foreign market.

With this been said, this is where Foreign Exchange comes in. There are markets where banks act as an intermediary to help facilitate these transactions.

This market interests me due to the high volatility making it highly risky over the expanse of time. This inspired me to look at the tools in which practitioners use to reduce their risk.

A tool used to do this, was derived from the concept of prediction, which gave birth to forecasting. With this tool, help better manage their future outlook of exchange rate outlook.

The prediction of Foreign Exchange rates is crucial for Central Banks, Financial institutions, and organizations whether they're a business or a nonprofit entity or just simply individuals all whom engage in the space of transnational transactions. This can be a business that makes regular purchase of raw materials from another country requiring them to exchange their home currency for foreign currency. Or just simply a student living in a foreign country would undergo the same process. But as for businesses, larger the transactions are made at a regular biases, the cost of transactions increase and not to forget the fact that Exchange Rates fluctuate.

That's where the prediction of this fluctuation is vital for strategic decisions. In the sight of this, over time, professionals gradually developed the field of forecasting to help with their predictions.

The foreign Exchange is the most lucrative market due to its high volumes of daily trades. Approximately $5 trillion U.S. dollars is traded on a daily basis; the dissect brake down into 1.5 trillion dollars of daily spot trades making it extremely surrounded in volatility making it a challenge to predict.

Over time, statistics and econometric regressions made it possible to find patterns that help the prediction possible. Statistics helped the early stages of development by using

Moving Averages (MA) to help indicate a direction that the currency is heading towards.

Then later including Simple Moving Averages (SMAs) and Exponential Moving Averages (EMAs) to give early signals of a change of directive heading (Either getting stronger or getting weaker) of the currency. And so on statisticians came with Auto Regressive Moving Average (ARMA), Auto Regressive Integrated Moving Averages (ARIMA), Moving Average Convergence-Divergence (MACD) and other linear models, which all these are linear models. And now, highly advanced models which I'll be covering in this paper and comparing them with the previous models are Artificial Neural Networks (ANNs) and their advances and their evolution into different logic integrated systems which branch off into Higher-Order Neural Networks (HONNs), Support Vector Machines (SVM) and Multi-Layer Perceptions (MLP).

The objective of this paper is to run the simple linear models and compare them to the advanced non-linear (NNs) to see whether they're able to forecast the EUR/USD, GBP/USD. Then would be measure based on the main financial performance indicator Sharpe Ratio. And then calculate the Value-At-Risk (VaR) to measure the amount of risk.

To note, the process at which this study will be done are the returns of two currencies EUR/USD and GBP/USD and not the actual Rates of these currencies. Based on previous scholars, returns are normally distributed and that would assist the process of running the linear models forecasts.

Later in the literature review, there is great skepticism on the ability for these neural networks to actually predict Exchange Rates. The reason behind this is that there are a lot more to exchange rate fluctuations than simply historical information can actually say. An economy is a deep market, with this said, factors such as Interest Rates, Consumer Price Index (CPI), Inflation, Unemployment and political policies can have a great toll on unforeseen changes in exchange rates.

Another thing to add is the concept of the Efficient Market Hypothesis (EMH), conveying that if the markets were to mirror all information available, there would not be any room for investors or traders to make a profit because the prices would fine-tune so swiftly which would off set any arbitrage opportunity.

This being said, Investors in the "real world" have managed to make profits over the expanse of time of trading. With this a fact, only proves that the Efficient Market Hypothesis (EMH) is well off broken by arbitrage. With this, left scalars to say that there is a delay in which price reacts to information that had been announced leaving room for investors to capitalize on.

Linear models have the disadvantage of being only valid for stationary data making it weak to analyze time series data; In contrast to non-linear models, they are more affective in running algorithms on time series making the forecast closer to accuracy. In the light of this, this paper will be using linear models in which Auto Regressive Moving Average (ARMA), Moving Average Convergence-Divergence (MACD) and with in mind the Random Walk Theory to compare them with the efficiency of non-linear learning machines to assist with our forecasting models.

In the late 20th century, traditional Artificial Neural Networks have morphed into the development of learning machines to solve issues with categorization of information that was inspired by the early Support Vector Machines (SVMs). This system has been used widely in day-to-day traders showing better results in narrowing down the error of the forecast.

2.0 LITRATURE REVIREW

The key to a stable economy or a market within it is its ability to be somewhat stable. This is where forecasting comes in to help maintain that satiability, making the adaption to changes much easier. As for the forecasting of exchange rates prolongs one of the hardest to be precise accuracy. (NEED SOURSE)

As proven predecessor forecasting models, MACD, ARMA and the Naïve Strategy to work but their ability to analyze non-linear time series data are not efficient compared to Artificial Neural Networks (ANNs). Previous studies had been done by Ghiassi et al.

(2005) to exhibit this fact, by forecasting both ARMA models and ANNs and comparing their performance.

Ever since the random walk theory existed investors and scholars of the past have always challenged it. Said in Eugene F. Fama (1965) to what extent past history of price of an asset can make a meaningful prediction to contribute to the forecast; concluding that it is impossible to beat the market by which the Efficient Market Hypothesis came to be derived. Nevertheless if models were to beat the random walk benchmark, it would challenge the Efficient Market Hypothesis (EMH). Messe & Rogoff (1983) suggested that the random walk outperforms the majority of linear models. Looking at the Autoregressive Integrated Moving Average (ARIMA), which been developed by Karemera and Kim (2006) was able to outperform the random walk theory; as they were on a monthly basis, it was not a profitable method that would be adopted by traders.

In the light of all this, Kondratenko et al (2003) mentions that Artificial Neural Networks have the advantage of analyzing non-stationary time series data in compared to traditional linear models. Making it superior to forecast foreign exchange markets, as they are non- stationary.

The mire importance of forecasting for practitioners whether they're for economists or business investors has driven the development of these models to advanced levels, resulting in a wide scope of preferred models that could be implemented. The range of these newer Artificial Neural Networks can factor in a multiple array of inputs, including vital economic indicators for example among these are Interest Rates, Unemployment Rates and Inflation, making it adaptable models that can take in account these factors as inputs. (Semprinis et al, 2013) The Generalized Auto Regressive Heteroscedasticity Model (GARCH) basically brought forth the ability of the analysis of data that is non-consistent in variations in data as well as it cogitated the development of learning machines derived from algorithmic models of biological functions. And so, Artificial Neural Networks have come to existence.

2.1 Artificial Neural Networks' Historical Story

The idea of Artificial Neural Networks (ANNs) originated from the concept of the human brains' learning process, from this, mathematical models to simulate the biological nervous system had been developed (Abraham 2005). And for this to be possible, McClulloch and Pitts (1943) built the first logical calculus model, which was regarded as the interception of two fields of research. The first is the theory of finite-state machines. And the other is the field of Artificial Neural Networks, from there; ANN models have evolved into the new advanced models we know now.

With the contribution of Donald Hebb (1949) he proclaimed the learning machine to the field of Neural Networks. This is by making a connection between psychology and physiology rendering it into neuropsychology and neuroanatomy making Neural Networks adaptable in the sense of unsupervised learning; and yet old used till today.

As for the perceptron as mentioned before, Frank Rosenblatt (1958) created this method that is based on pattern recognition algorithms which is now called 'Multi-Layer perceptron (MLP)'. The system 'MLP' had been criticized due to its lack of its ability of making conditional logic Minksky and Papert (1969). Furthermore, not to forget that fact that computers were lacking superior computing powers that were needed to algorithmize complex conditions.

In the late 70s, a paper had been done on how to solve this Exclusive-Or problem by the modification of the back-propagation algorithm by instead of having a single-layer perceptron model but adding other layers Werbos (1975). And then, a big pause accrued due to the fact as mentioned before, the lack of computational power at its time.

Moving into the 90s, with the advancement of technology and its capability to computer complex algorithms, sophisticated statistical models can be now ran without limitations. This rendered the capability of learning machines to find the relationship within the data set making it a strong form of ANNs. As a result, by using historical information of prices to forecast the future variations has never been better (Lawrence and Andriola 1992). This advance in ANNs have shown better performance results rather than previous linear regression models in forecasting the Arbitrage Pricing Theory (APT) (Refenes, 1994), including Yoon (1993) in forecasting bond ratings and stock prices. In addition ANNs have shown a positive performing attribute in its adaption to non-linear characteristics of foreign exchange rate prediction (Franses and Grievensen, 1998).

As Rime el al (2010) mentioned in his analysis paper in forecasting foreign exchange rates. Forecasting "one day ahead" was done due to the fact that it is implementable because of the data is available, and that practitioners prefer to undergo high frequency trading on a day-to-day basis.

Now to highlight the functionality of a Neural Network, data is processed in three stages: The integration stage, in which all input variables (x_i) will be weighted with parameters ($w_{i,j}$). Second stage, is the non-linear stage at which all the weighted inputs are transformed into a non-linear function according to the relevance of the network desired. Third stage, and which is the final stage, is the propagation process where outputs are transmitted to the corresponding neurons.

The equation to this process should look like this:

$$y = f\left[\sum_{j=0}^{n-1} w_{i,j}\, x_i - \theta_i\right]$$

Where as:

$y =$ Output

$w_{i,j}$ = Adaptive Parameters (Weights)

x_i = Input Values

$f(x) =$ Non-Linear Function

$\theta_i = The\ Threshold$

Based on previous scholars in the field of forecasting; previous information of price movement whether they are stationary or non-stationary, Artificial Neural Networks have established their place in the world of prediction. They have developed over history, starting simple linear regression models to non-linear learning machines. Till this day ANNs are being developed to place a better accuracy in their forecasting abilities. And yet till today, the two worlds of academia and real world traders clash when it comes to trading strategies in which should be implemented.

2.1 High-Order Neural Networks:

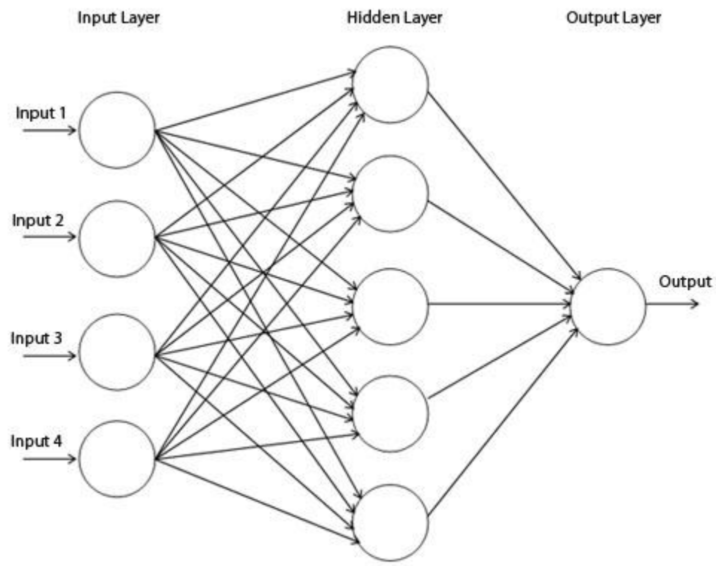
Over the wide spread of different types of Artificial Neural Networks (ANNs) is the High-Order Neural Network (HONNs) which was introduced in the late 80s by Giles and Maxwell. Well known of its aptitude of impressive computational, storage and learning capabilities (Gilers and Maxwell, 1987). The reason behind this is that the order structure can be customized to the order structure of a problem. Therefore, previous knowledge such as geometric invariances can be encoded in the high-Order Neural Network. This type of Neural Network attempts to adaptably generate useful discriminate functions, thereafter led to the study of Threshold Logic. Given this, another name to High-Order Neural Networks is High-Order Threshold Logic Unit (HOTLU). Criticism by Minsky and Papent when they concluded that High-Order Threshold Logic Units (HOTLU) were impractical, the reason behind this, is that the combinational explosion of high order term also including that First-Order TLUs were too limited to have any interest.

The structure at which the network is assembled (HONNs) similar to single layer perceptrons inputs goes through hidden layers before becoming outputs. A. Ismail (2001) pointed out a key difference in this network, which are the way inputs are combined.

2.2 Multi-Layer Perceptrons:

Another type of Artificial Neural Network (ANN) is the Multi-Layer Perceptron (MLP). Rosenblatt introduced the perceptron in 1957 initiating the Single-Layer perceptron and thereon the Multi-layer perceptron took off. Another word for this type of network is the feed-forward neural network, given that data is forwarded through the network to give a possible output.

These networks can be visualized in the diagram below:



To shine some light from the diagram above, the NN process undergoes three steps, the first; data is inputted into the network in a manor that would be recognized by training of the network can be done and would go into the input layer; later in the paper this would be explained. The second is the weighting (as each line has a weight connector) as the transit of values in the input layer into the hidden layer (shown as lines connecting each neuron to each hidden layer) they under go the learning process narrowing down patters in the data. The third is the final step is values will be weighted again to render the one output value. To simplify, I will be only one layer transit (From the hidden layer to the output) as the first process from input layer to hidden layer should be the same but with a longer equation. The equation to this would simulate as the following:

$$y = a(H_1 * w_1) + (H_2 * w_2) + (H_3 * w_3) + (H_4 * w_4) + (H_5 * w_5)$$

Where as:

$y =$ The Multi-Layer Output Value

$a(x) =$ The Hyperbolic tangent/Sigmoid Function

$H_n =$ Hidden Layer

$w_n =$ The Weight

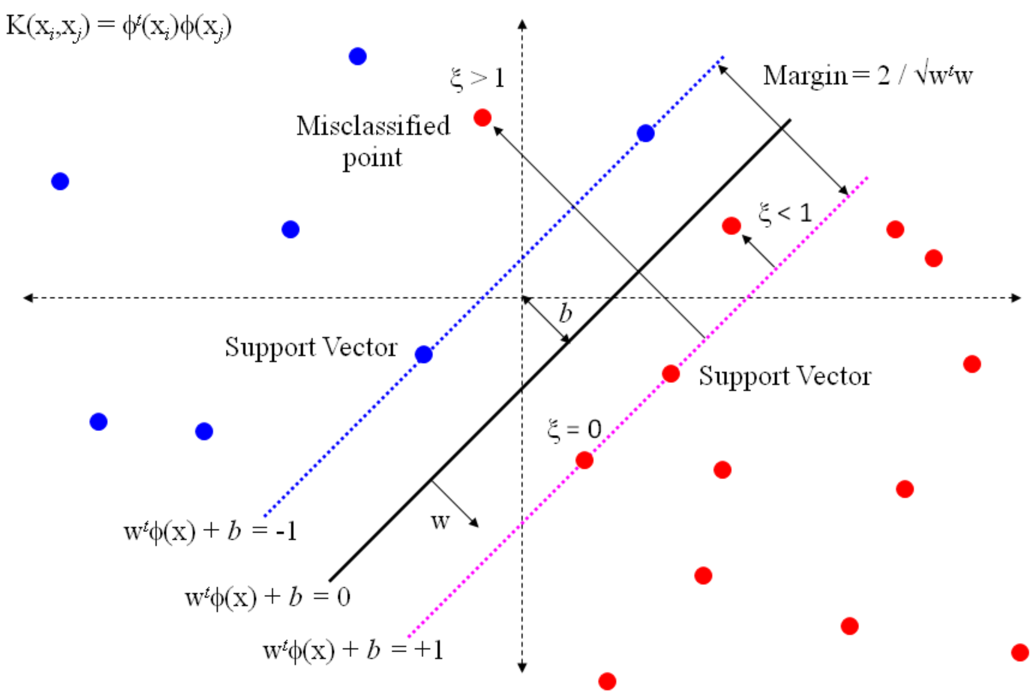
Basically the inputs are fed through a multiple layers to help with the learning part before the networks are trained.

Previously, research had been done on the foreign exchange market using the Multi-Layer Perceptron, as they were not prosperous. Castiglione (2001) forecasted three major indexes in the United States stock market using the Multi-Layer Perceptron Neural Network and had satisfying results, having an accuracy of nearly 50%. Maybe this is because he had not used stock market announcements as classified inputs to help with the networks training process to increase its accuracy. By using binary signal systems would help with output accuracy as suggested by Tino et al (2001) as he forecasted the United Kingdom's stock market. And yet found it unsuccessful, making his conclusion that forecasting financial markets using Neural Networks was not proficient in the learning process of historical data.

A previous study had been done by on the major currency pairs Yao et al (2000) using Multi-Layer Perceptrons (MLP). Two inputs were used in the neural network; exchange rates of these currency pairs, which were time lagged, and the moving average values at different time frames. The exchange rate values were collected on a weekly frame, with this the fact; the forecast will predict the week to come. Results shown an accurate forecast, (Kondratenko et al. 2003) the fact that it most traders in the practical world prefer a forecast on a daily basis rather than a weekly approach.

2.3 Support Vector Machines and Support Vector Regressions:

Vladimir N. Vapnik invented the original Support Vector Machine algorithm in 1995, in which classifying data is the common task done by these machines. This is done by classifying values are either true of false. As elaborated by Zinzalian et al. (2009) the determinant of a price rise or fall would be encoded as a simple binary classification problem, that is, a 1 for a rise and 0 as a fall. To visualize this, below is a diagram to portray this function:



As we can see in the diagram above, a process of the creation of decisions plane where the split of data space into two planes by construction a support vector ranging between 1 and -1, this called the margin zone. The SVM approximates the function in the following form:

$$f(x) = w * \emptyset(x) + b$$

Where:

$\emptyset(x) =$ The high dimensional feature spaces mapping of the nonlinear input space x

$w =$ Estimated coefficient of regularized risk function

$b =$ Estimated coefficient of regularized risk function

According to Cao and Francis (2003) Support Vector Machines were based on Structure Risk Minimization (SRM), which strives to a goal of minimizing the upper bound of the generalization error (in consideration of the sum of training errors and confidence intervals) this was introduced by Vapnik's-Insensitive loss function. This induced the principle what differs it from Neural Networks that commonly use Empirical Risk Minimization (ERM) that only minimizes the training error. This function can be visualized as the following:

$$Minimize\ \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

Whereas:

ξ_i & ξ_i^* = Slack Variables

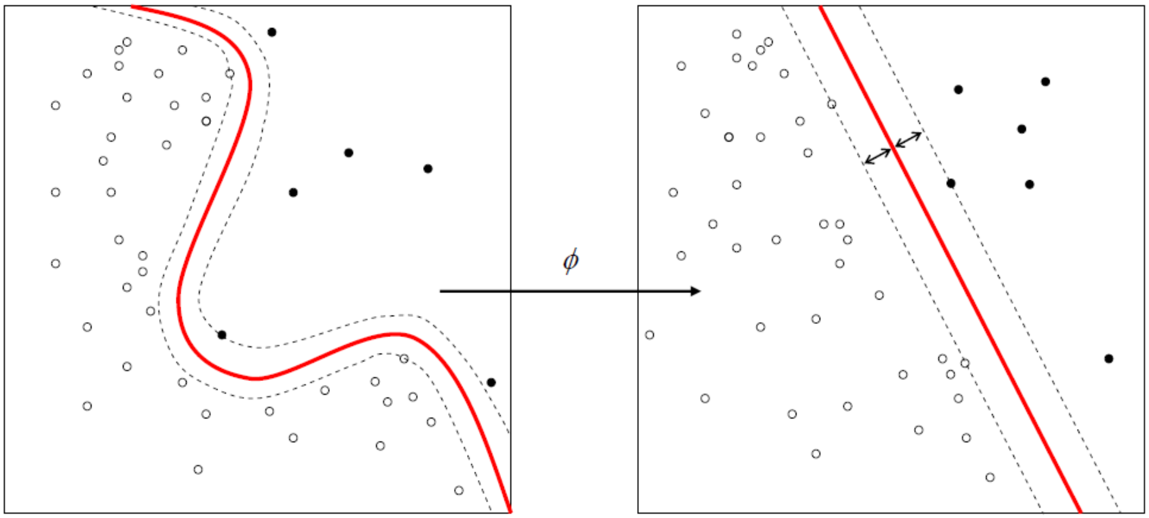
C = Cost Parameter

$$w = \sum_{i=1}^{n} (a_i - a_i^*) x_i$$

The highlight benefits of Support Vector Machines mentioned in Cao and Francis (2003) is that they achieve high generalization performance, as traditional Neural Networks only implement the Empirical Risk Minimization (ERM). Also, the training process is equivalent to solving a linear quadratic programming problem, whereas Other Neural Networks training requires nonlinear optimization.

One disadvantage of SVMs mentioned by Cao and Francis (2003) is that the training requires a lot of computational time to solve large sample sizes. Bahramy et al (2013) mentioned the fact that there were mixed results have been found when forecasting foreign exchange markets due to the non- linearity of exchange rates.

SVMs are able to forecast exchange rates according to Ince and Trafalis (2005) and Brandl et al (2009) paper. And P.Dharnidaharka (2012) goes into detail that linear classifier models do not cut it when solving more complex problems in non-stationary realms as like foreign exchange markets. A more complex model structure to help elaborate a more effective when classifying these types of data sets is seen below; which the transformation into a higher dimensional space making the function feasible:



The transformation of the function into a higher dimensional structure making the curve a complex mapping method using Kernels mathematical functions to arrange and map objects in the universe being studied.

For this to take place the Kernel function is introduced the function looks as the following:

$$f(x) = \sum_{i=1}^{n} (a_i - a_i^*)\, K(x_i, x) + b$$

There are, a variety of Kernel functions in the Support-Vector Regression, here are some of them:

The Linear Kernel Function: $K(x_i, x) = \; x_i^T x$

Polynomial Kernel Function: $K(x_i, x) = (x_i^T x + 1)^{\,d}$

Sigmoid Kernel Function: (Multi-Layer Perceptron): $K(x_i, x) = \tanh(yx_1^T x + r)$

Radial Basis Kernel Function: $K(x_i, x) = \frac{\exp(-||x - x_i||_2^2)}{\sigma^2}$

Kernel is a function that regulates inputs/outputs of a network. Sermpinis et al. (2013) stressed the importance of the selection of Kernel parameters for the success and precision of Support-Vector Machines.

In 2007 studies had been done on Support Vector Regressions (SVRs) and results shown a promising future compared to traditional Neural Networks (Ulrich et al, 2007). As well as Tay and Cao (2001) they have used SVMs to forecast time series and their numerical findings found to point out that SVMs outperform Multi-Layer back-propagation neural network.

Later studies led the development of SVRs into Hybrid Support Vector Machines (HSVMs). Pai and Hong and Lin and Chen (2005) elaborated in their paper that one particular model can not apprehend all data patterns easily, and therefore, proposed the Hybrid Support Vector Machine (HSVMs). Where this made exploitation of the uniqueness strength of linear and nonlinear SVMs models at forecasting foreign exchange rates. Moreover, parameters of both linear and nonlinear SVM models are determined later by Genetic Algorithms (GAs). Results of these proposed models show an outperforming other approaches previously made.

Tan & Wang (2004) and Bahramy et al. (2013) zeroed in the success of Support-Vector Machines with their presented structure and Kernel selection. With this fact, the primer of the joining Genetic Algorithms with Support Vector Machines (GA-SVMs) made it possible to heighten its Kernel function and feature. In the height of this, to confirm the assertions made previously on Genetic Algorithm Support Vector Machines (GA-SVMs), equivalent variables will be used as other NN models will be used in the directive to retain uniformity in the analysis.

Also, Lee et al (2004) proposed a multi-category Support Vector Machine was fashioned on the well-established binary SVM that showed good results. With this inspiration, Liu & Shen (2006) developed a Multi-category $\Psi - \text{learning}$ system that a $\Psi - \text{loss}$ function is included because SVMs lack the data reduction. Multi-category systems reduce the number of support vectors giving more solutions to a problem at hand.

To help understand more about Kernel parameters and feature selection as mentioned before those are featured in Genetic Algorithm Support Vector Machines (GA-SVMs), the key competence that it contributes to these learning machines is that they help in the training process impacts on the classification accuracy. In other words, those " Kernel parameters and feature selection" make it possible to optimize subclasses in the constraint conditions.

2.4 Gene Expression Programming:

Gene expression programming is a genotype/phenotype genetic algorithm that is linear and ramified. This type of programming is a new technique in the creation of computer programs. These genetic programs use as mentioned previously genetic algorithms, in which the use of character linear chromosomes comprised of gene structure (head and a tail). These chromosome and act as a genome and can be subjected to adjustments by means of transmutation, rearrangement, root rearrangement, gene transposition and other adaptive measures (Ferreira, 2002). The main impartial job of a genetic algorithm is to search and naturally select a solution in the vast and complex universe it's attempting to analyze, even if a mathematical analysis isn't available. As mentioned, a Chromosome contains genes, which are considered as optimization parameters. All of these iteration/generation(s) over goes a fitness function that evaluates the chromosomes accuracy of coinciding solution set; of which the "fitness" chromosome is elected to exist. The criterion that is blueprinted in the logic, when met, will determine when the cycles will terminate. Holland (1995) goes on to mention an advantage of Genetic Algorithms that they can comprehend large search universes with out getting confined in regional solution optima.

To help understand how the genetic mutation and crossover evolutionary procedure is done, Montana (2002) elegantly describes how the program operates by simplifying the procedure in visualized parse trees as illustrated below in the first diagram is the base followed by the crossover mutation:

Diagram (#)

$if\ x > 0\ then$

$x \coloneqq (x + 3) * (4 - y)$

$else$

$y \coloneqq x + y - 1$

$end\ if$

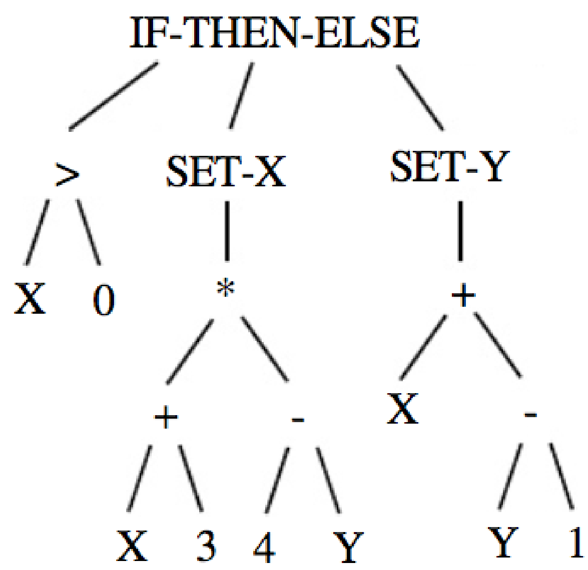
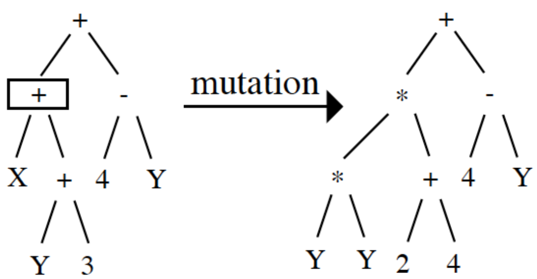
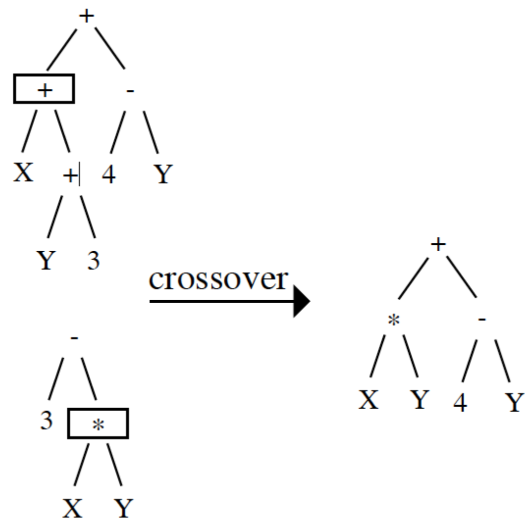


Diagram (#)





The crossover and mutation is suggested by Montana (2002), which is an extension to Koza (1992) model. Where Koza's model is shown in the diagram (1), and whereas Montana's model (the extension) in diagram (2)

The procedure that the crossover undergoes 3 steps:

1. A Random selection of a node within each respective tree as a crossover point.
2. Proceeding to take the subtree rooted at a selected node in a second parent and then use it to replace the subtree rooted the selected node in the first parent to generate a child (and the option of reversion to obtain a second child).
3. The use of the child if its maximum depth is less than or equal to Max-Tree-Depth.

(Source: Montana, 2002)

A very similar procedure is done for the mutation part in 3 steps; different in the creation of a new tree rather than a child tree. Procedure shown below:

1. A random selection of nodes within the respected tree parent tree as a mutation point.
2. A new tree is generated constrained to the parameters set by the program (similar to the process before).
3. The newly generated tree goes through the process of validation with respect to the compliance of the program. THEN, if conditions met, the tree is considered.

(Source: Montana, 2002)

Allen and Karjalainen (1999) inaugurated a similar model using a Genetic Algorithm to undergo supervised learning rules to commence technical trades in the S&P 500 realm. Saying Allen and Karjalainen (1999) that "the rules do not earn excess return over a simple buy-and-hold strategy" this is due to after traction costs and the trading rules undertake positions when only returns are positive and daily volatility is low, and when returns are negative and volatility is high, trades are not commenced.

2.5 Hybrid GA-SVR Models:

This is a method of which more than one model is used to coincide its focus on resolving a desired goal. With the goal of this paper in mind, the use of more than one model will help optimize the trading ability in forecasting foreign exchange rates. The hybridity of Genetic Algorithms and Support Vector Regressions would optimum this.

In consideration of past literature, Support Vector Regressions achieve a better process of classification in compared to Support Vector Machines. This is due to the fact that when generation of end of day forecasts, SVMs produce only a binary output; the requirement of scrupulous forecast desires in trading options is needed, that only SVRs can offer.

2.6 Cons of Artificial Neural Networks:

Given what had been covered in the literature review, all findings from previous scholars and researchers indicate that Artificial Neural Networks tend to perform well with consideration of it's application of what sort of data being analyzed. However, in this section, I will be going to focus on the cons of ANNs. The axiom that they lack the qualitative aspect in the input entry due to the fact that only historical prices variation is only taken into account. There are more to the reasons behind currency price movement, like for example news reports, interest rate announcement, Inflation and any unemployment which effects the disposable income of the privet sector, with these factors from both sides of the economic spectrum which are playing a tug of war with the currencies appreciation and depreciation. With this in mind, Yoon (1991) proposed made a respectful effort to converging qualitative and quantitative data fed into the input layer perceptron's. Other attempts had been made by other research scholars to achieve this same method of analysis (Parra and Agell, NEED DATE).

Another disadvantage of artificial neural networks, is that the weights of connectors are difficult to interpret and estimate as it been given the term "black box" problem. Benitez, Castro and Requena (1997) Go and speak of ANNs are devices working as black boxes, as they capture "hidden layers" and find the relationship between inputs and outputs. Due to the fact of the complexity of nonlinear functions are set up, the network makes it hard to interpret the given networks weights. This is where Back-propagation comes in, when the networks adjusts it's weights from the output back to the input to find where and which weight connectors are effecting the Mean Square Errors. This method of weight correctification is one of the most simplest ways and most common led to efficient models; this simplicity led to slow and non-maximized result system. This is where Genetic Algorithms (GAs) come in to boost the process of weight rearrangement.

2.7 Trading Application:

As mentioned before, in this paper, I will be conducting a comparison between traditional simple statistical models as benchmarks (MACD; Naïve strategy; Random Walk and ARMA) with the proposed Artificial Neural Networks (ANNs) and to find whether they outperform based on the yielding Rate of Return and Mean Square Errors of each model.

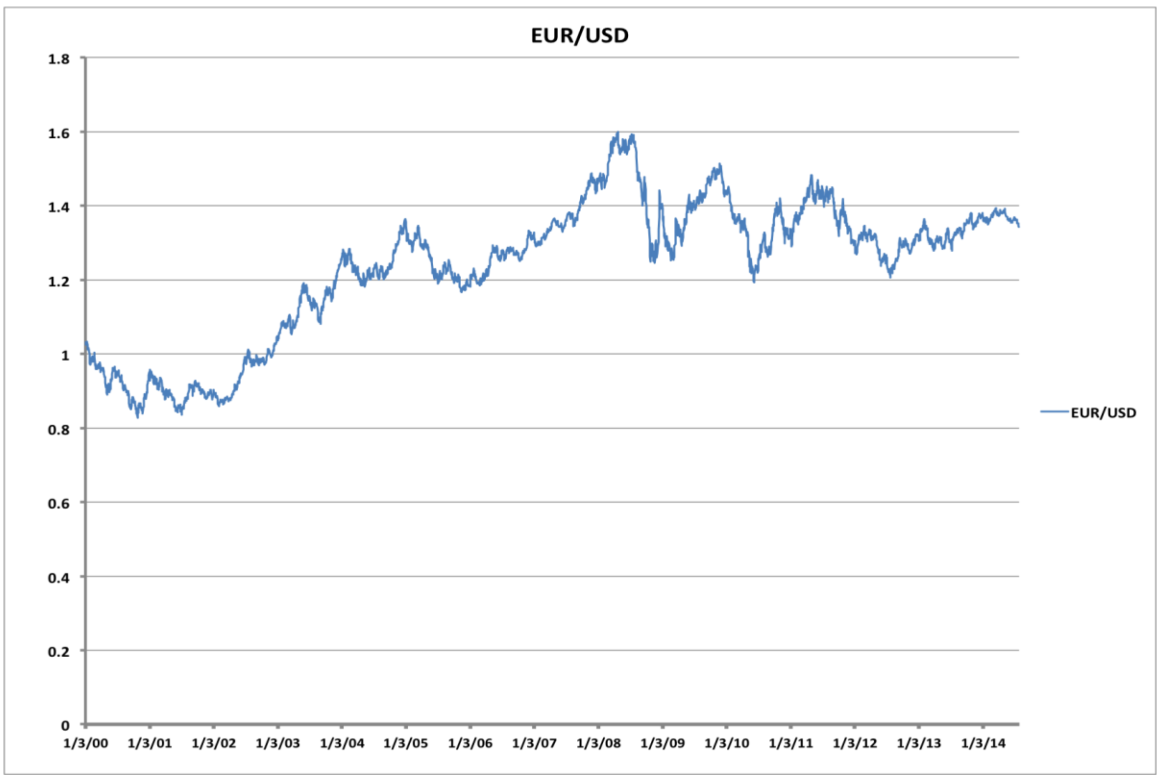
As seen in the literature review, that Artificial Neural Networks yield promising returns. Alexander (1964) contradicts these high yielding returns, saying that after considering transaction costs of trades made, the promising returns of ANNs diminish making them no better compared to benchmarks. Others have confirmed this finding down history yielding the same results (Blume, 1966; Fama, 1970). A different approach was angled towards this problem, (Reay, 2002; Allen, 1990; Neely, 2003) constructed a system that is founded on trading rules to over come the issue. Realizing they came to the same dead end.

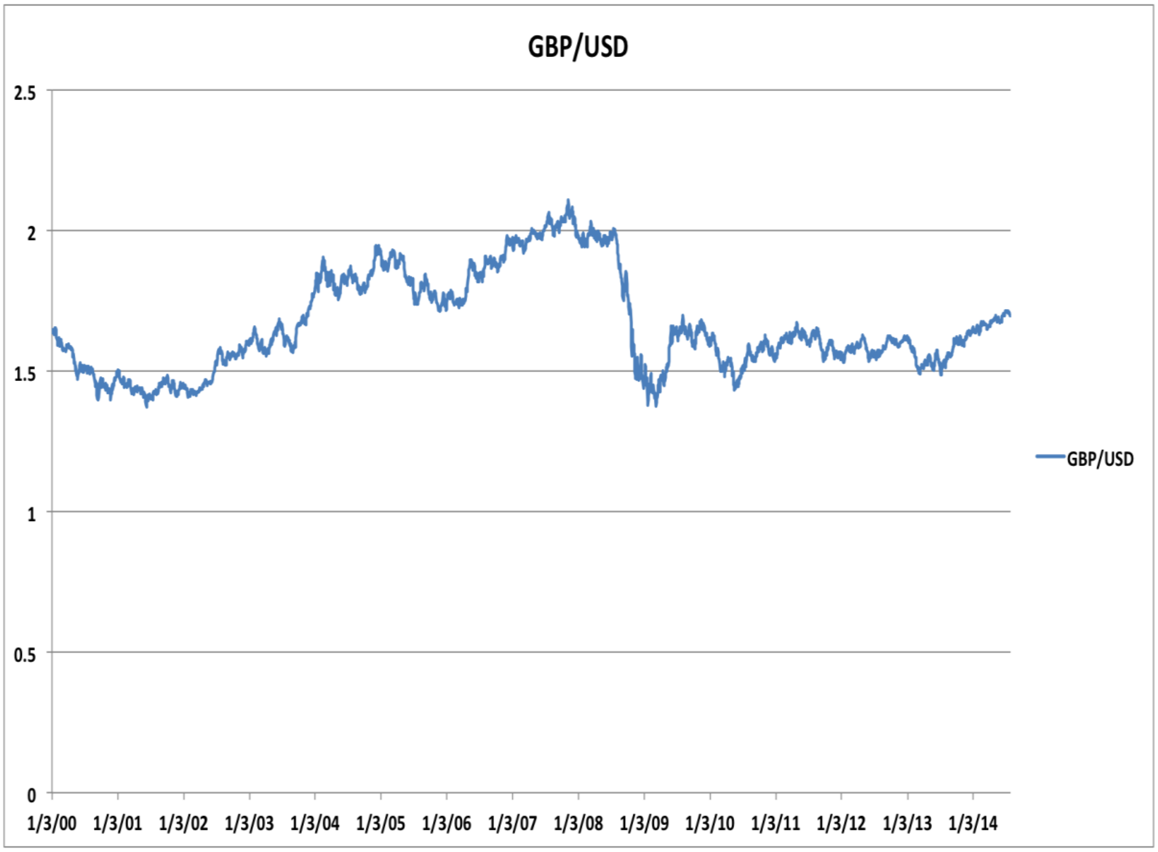
2.8 Theoretical background

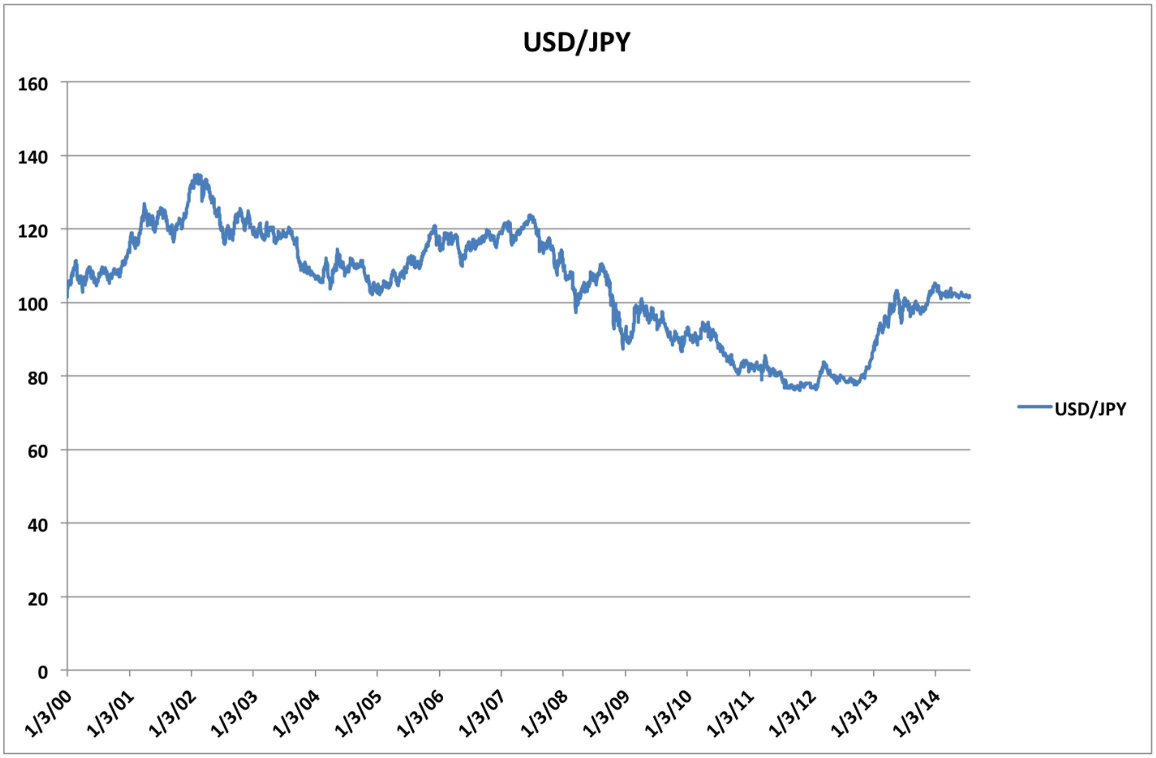
As I had stated this paper of the importance of Exchange Rate forecasting, daily changes of Exchange Rates can't be overlooked. The consideration of flow of money between countries is an ever-lasting event; moreover, slight changes in the rates can make losses or gains for entities with boundaries of an economy. Being the fundamental object of medium of exchange, large volumes are traded every day making currency to have its well know characteristic of volatility.

The data that will be acquired for this dissertation is vastly available from a variety financial online websites. 'Yahoo finance, Oanda.com and Bloomberg network' have all the worlds' exchange rates available for extraction to a worksheet. This will be the method of collection of historical data for this paper.

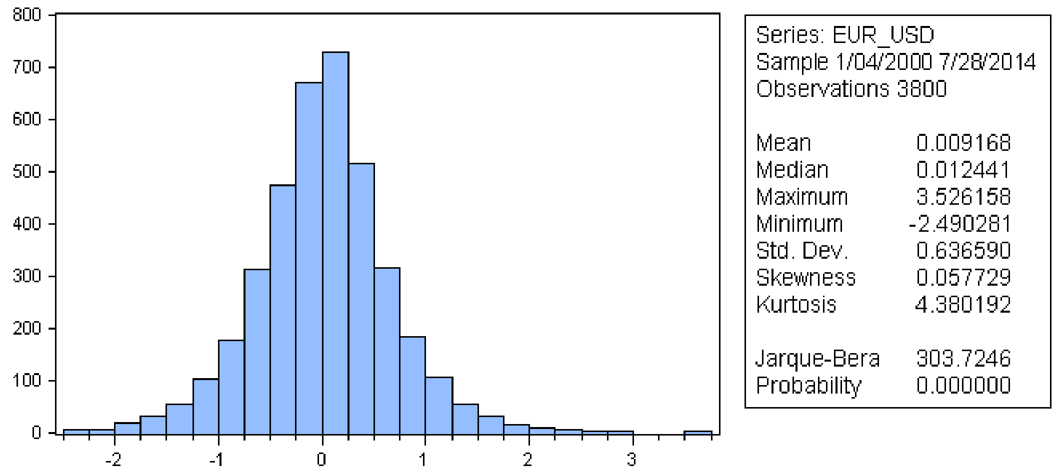
The period of which GBP/USD, EUR/USD (ADD AS NEEDED) from the period 01/01/2001 to 31/01/2014 that has been extracted, giving (NUMBER OF SAMPELS) per currency, this is shown in the table below. Then after the In-Sample and Out-Sample will be divided to tailor the data for the Neural Networks.

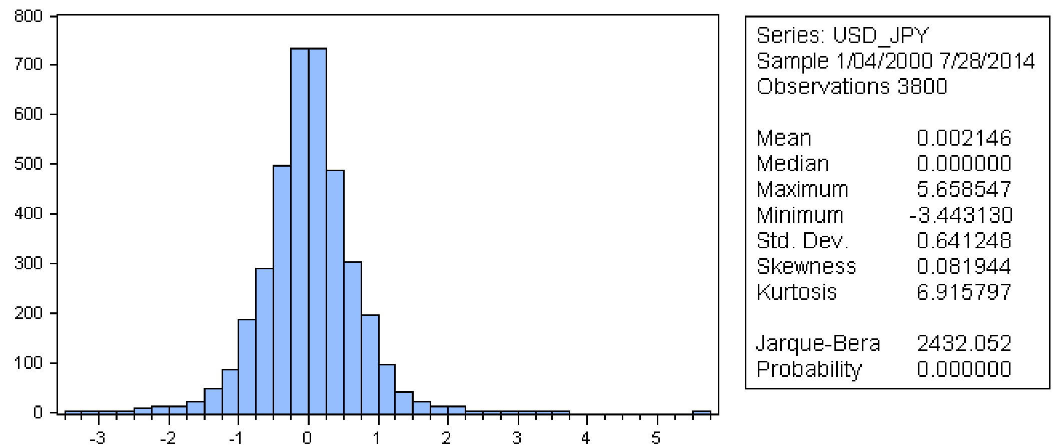


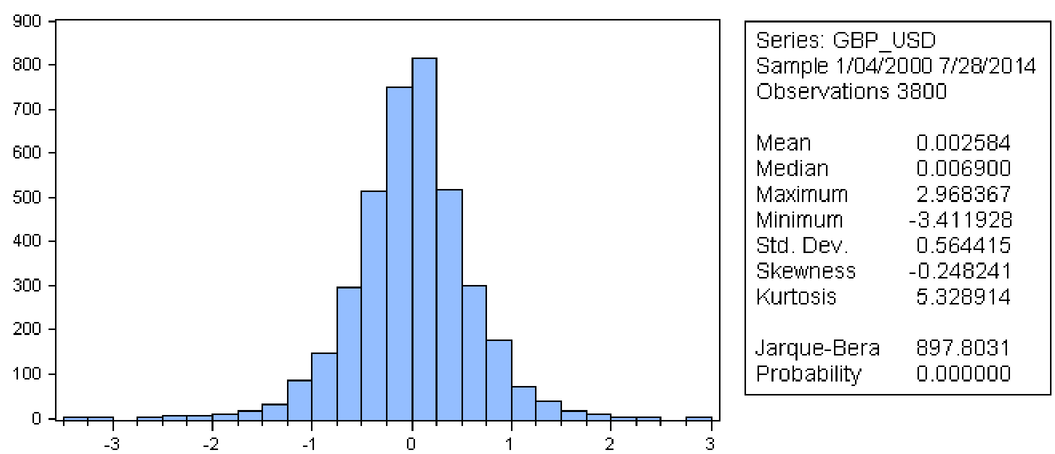

GBP/USD
2.5
2
1.5
1
0.5
0
1/3/00
1/3/01
1/3/02
1/3/03
1/3/04
1/3/05
1/3/06
1/3/07
1/3/08
1/3/09
1/3/10
1/3/11
1/3/12
1/3/13
1/3/14
GBP/USD



After the collection of all the data of daily prices of "will be studied" exchange rates, the calculation of the rate of return of each period 'daily' to help train the Neural Network to forecast the Rate of Return of one day ahead. After the conversion of prices into returns, below will visualize the returns on a histogram distribution (ADD TO THIS):


Series: EUR_USD
Sample 1/04/2000 7/28/2014
Observations 3800
Mean 0.009168
Median 0.012441
Maximum 3.526158
Minimum -2.490281
Std. Dev. 0.636590
Skewness 0.057729
Kurtosis 4.380192
Jarque-Bera 303.7246
Probability 0.000000


Series: USD_JPY
Sample 1/04/2000 7/28/2014
Observations 3800
Mean 0.002146
Median 0.000000
Maximum 5.658547
Minimum -3.443130
Std. Dev. 0.641248
Skewness 0.081944
Kurtosis 6.915797
Jarque-Bera 2432.052
Probability 0.000000


Series: GBP_USD
Sample 1/04/2000 7/28/2014
Observations 3800
Mean 0.002584
Median 0.006900
Maximum 2.968367
Minimum -3.411928
Std. Dev. 0.564415
Skewness -0.248241
Kurtosis 5.328914
Jarque-Bera 897.8031
Probability 0.000000

3.0 METHODOLOGY

The method of which is data will be extracted would be from the Bloomberg data bank of the currencies that will be analyzed. The current number of observations are (____) these are the ends of day values of each currency. Then after, they will be separated into two groups; In-Sample and Out-Sample as shown before in the section before. The actual foreign exchange rate won't be forecasted, instead, the rates of returns will be: $R(t) = (P_t - P_{t-1})/P_{t-1}$. Thereafter, The Null Hypothesis (Ho) will be tested to establish the footing whether the data is stationary or non-stationary. In the case whether if the data is non-stationary would indicate that there is no correlation what so ever, resulting in that the data will follow the random walk theory. Moreover, other forecasting models will be undependable on the long run.

The consistency of data is empirical for prediction; stationary data have these characteristics over time. In other words, the mean-variance follows a consistent trend this fact points out of the stationarity of the data. This is a fact says that in short time frames data tend to be non-stationary, whereas data on larger time frames tend to be more stationary. The test of the Null Hypothesis will be done to help confirm the state at which the data is. By using Eviews (Econometric software) in which the root test would be available to analyze. To give a glimpse to how this is done, the data is measured through the Dickey Fuller GLS leading to the root that would be tested for the trend and intercept. Results of this test will point a confirmation of the (Ho). There are 4 times of non-stationary (random walk); Pure Random Walk, Random walk with a drift, deterministic trend and Random Walk with drift and deterministic trend.

So, with all this in consideration, non-stationary data is the random walk theory, because of the fact that the data can't be predicted. In this case, data needs to undergo a transformational process by differentiation. Whether if the data were Random Walk with a drift or a pure Random Walk, the difference between periods would be calculated ($Y_{(t-1)} - Y_t$). As for the Deterministic trend, the process of detrending must be completed, where $Y_t - \beta_t$ Giving the formula $Y_t - \beta_t = \propto + \varepsilon_t$.

After completing the process of transformation of the data form non-stationary to stationary status, stage of dividing the total data set into In-Sample and Out-Sample sets for it to be ready for the modeling process.

The first part of the analysis will be by computing the benchmarks for each currency using 2 main strategies as mentioned: The Random Walk and Buy and Hold. This will be the main paralleling indicator to the basic learning machines and basic linear models. The method of computation of the mentioned strategies is done through a calculation of excel with the concept if the tomorrows expected return is higher than the rate of the return, a buy order position would be considered, and if tomorrows expected return is lower than the todays then a sell order position would be considered, and if there's no change from the expected and todays return a hold position would be considered.

Once this is done, Mean Average Percentage Errors, the Root Mean Square and the Sharpe Ratio would unfold helping to consider the Risk/Return ratio giving the actual value of trades that are undertaken.

This strategy process is done solely through Excel software. After finding the expected returns, a position is taken based on this strategy. If the predicted value is a negative a short position is taken, and if the predicted value is a positive, a long position will be taken, and finally if a zero comes up, a hold position will be considered. A final annualized return is summed for each year, and then summed again at the end to see the actual return over the years to make it comparable.

As for the MACD and ARMA, they're models that can be customized, making the average movable. In other words, the period length can be adjusted meaning that the average is based on how many previous periods (Example: a ARMA 10 would consider ever 10 lapsing consecutive days).

Basically ARMA are two terms coincided, the Autoregressive term $X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-1} + \varepsilon_t$ and the moving average term $X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$, Once joined together, the ARMA term comes to be:

$$X_t = x + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-1}$$

Where:

θ & φ = Parameters

c = Constant

ε_t = The error term at time t

The ARMA and MACD models are preformed using Eviews software, where running the Least-Square Method would give an approximation of a solution of historical data. Where the square residuals are minimized to reduce in respects of the data to a minimum. This process makes a "best fit" line between data variables and the residual.

Once completed, a forecasted return is calculated for it to be compared to the returns of other forecasting models with consideration of the error terms and risk/reward ratio.

Now moving on to the MLP and HONN models, they are constructed in a similar way in terms of data preparation. The concept of Neural Networks uses a string of past lags as input values. This will be used in this paper when forecasting using these networks, 9 lags have been selected all with a 1-day difference: this is shown below:

OUTPUT	INPUTS								
-0.41166									
1.57903	-0.41166								
0.57785	1.57903	-0.41166							
0.76064	0.57785	1.57903	-0.41166						
-0.77096	0.76064	0.57785	1.57903	-0.41166					
0.04047	-0.77096	0.76064	0.57785	1.57903	-0.41166				
0.04854	0.04047	-0.77096	0.76064	0.57785	1.57903	-0.41166			
0.33961	0.04854	0.04047	-0.77096	0.76064	0.57785	1.57903	-0.41166		
-0.74140	0.33961	0.04854	0.04047	-0.77096	0.76064	0.57785	1.57903	-0.41166	
-0.37347	-0.74140	0.33961	0.04854	0.04047	-0.77096	0.76064	0.57785	1.57903	-0.41166

This sequence of data structuring helps the neural network to learn from changes in historical data variation. Each lag gives a reference to the output from different lengths in previous periods performance.

With this system of arrangement, data is as mentioned, is separated into In-Samples and Out-Samples, this is crucial for the process of training of the network. This process uses in-samples lagged inputs to target the outputs as an actual value prediction. This process is called “Pattern Recognition” where only historical data is considered as inputs to predict the future.

This technique is then again done for the other sample ‘out-samples’, this process is like a reconciliation process, where based on the in-samples actual values to be forecasted are compared to the actual values, this activity, reviles the errors of forecasts. Once this is done, the neural network undergoes testing of forecasted values and then a validation stage. These steps are done for both the HONNs and MLPs and all are done using a software called ‘Matlab’

Once the software runs the neural network, graphs the one-day rate of return forecast as well as the errors of the forecast for it to be ready for comparisons.

The Support-Vector Regression as well as the Genetic algorithms are the last two neural networks that will be done in this paper, these models will be again done in Matlab, the frame work will be the same. The core superiority of these models is that they can capture asymmetries and nonlinearities in past data, this is to optimize given forecasted values. As for the Genetic Algorithms, they’re capable of capturing and analyzing large sums of input data. As been read in previous literature, the use of economic indicators as inputs such as interest rates and unemployment can be used as other indicators of exchange rate fluctuations.

3.1 STATISTICAL EVALUATION:

In this section, I will be going over the statistical indicators of performance. These indicators help the process of evaluating the risk/reward part, to help signal a comparison indication to which model out performs the other, and then summarizing the ranking of which model is the best. This section will be divided into two parts, Risk and Reward

Risk

- *Root Mean Squared Error (RMSE)*

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{Y}_t - Y_t)^2}$$

Where:

$n =$ Number of data sets.

$\hat{Y}_t =$ The Forecasted Value at time t

$Y_t =$ The Actual Value at time t

RMSE is a deviation type of error indicator; it measures the differences between forecast and the actual values of observation. As mentioned in methodology, Neural Networks make a calculation of the residuals in the in-samples and when it goes through the out-sample data the term becomes prediction errors. This is the indicator to display the accuracy of the forecast, so the larger the error is, the more inaccurate the forecast is, and vise-versa.

This type of risk measurement is compatible with various forecasting models, but does not accept different type of data sets. Under the umbrella of the unbiased compliance, the RMSE would be equivalent to the standard deviation, and in a perfect fit situation RMSE would be equal to zero.

- *Mean Absolute Percentage Error*

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\frac{\hat{Y}_t - Y_t}{Y_t}$$

Where:

$n =$ Number of data sets.

$\hat{Y}_t =$ The Forecasted Value at time t

$Y_t =$ The Actual Value at time t

This risk measure is normally used as a trend estimation. Same as the RMSE, it finds the difference between the forecasted and actual values and displays it in a percentage term.

- Sharpe Ratio:

This is an indicator to the risk/reward of an asset known as the return to variability ratio. It one of the pronounced evaluation tools among the investment world.

$$Sharpe\ Ratio = \frac{Return_{(asset)}}{Standard\ Deviation_{(asset)}}$$

RETURN:

Is also known as the rate of return, where it helps to evaluate whether if there's a profit or loss to an investment.

$$Annulized\ Return = \frac{\sum_{t=1}^{n}(\frac{P_t - P_{t-1}}{P_{t-1}} * 100)}{n} * 252$$

Where:

n = Number of Traded Days

P_t = Price at Time t

The annualized return is an indicator to how well the asset or portfolio has done over a given period.

4.0 THE ANALYSIS

4.1Null Hypothesis Test:

This part is the process of which to test the data we have whether the data sets we have is stationary or not. If the case that Null Hypothesis is rejected, the processes data transformation of mean and standard deviation into stationary sets as had been mentioned before in the methodology part is required for linear statistical models in order to be forecasted. This test of the Null Hypothesis is done in Eviews software, which the root-testing tool is available

4.1 Base Model:

The main idea of this paper is to compare linear models, Neural Networks, Learning Machines the benchmark models (The Random Walk, at which the random walk is the market behavior. This is done as mentioned before is to compare the forecasted one-day ahead profit/loss, the errors of forecast and statistical Sharpe Ratio.

4.2 The Random Walk:

In 1973 Burton Malkiel published a book called "A Random Walk down Wall Street" and it attracted attention to the randomness of the exchange market. Meaning, that the exchange rate changes are unpredictable and it is not possible to predict the markets movements. Basically it says that tomorrow's prices are depicted by yesterdays price with an error term. This is seen below in a formula state:

$$Y_t = \mu + Y_{t-1} + \varepsilon_t$$

The interest revolved around this concept, is that it's been known to beat most classical statistical models. With this being said, it should be a good benchmark to compare all other forecasting models to see whether they beat the market or not.

Below are results of using this concept model:

	EUR/USD	GBP/USD	USD/JPY
Annualized Return	29.30%	24.90%	26.03%
RMSE	0.0552	0.0653	0.0552
MAPE	92.23%	88.23%	76.23%
Sharpe Ratio	0.77	0.71	0.72

As previous scholars had realized it, this model shows a promising result pattern. Including the fact that the RMSE and MAPE are at a tolerable level of error. The deterrent to the use of this concept in the real world is that security selections with the appropriate weights are completely random. This fact makes it a not a real implicational option in the real world of investment.

4.4 Buy and Hold:

This is done buy making a one-time purchase of an asset rather than liquidating every year once with a repurchase. This concept is popular within the stock market domain of investors; signaling the idea that stocks tend to be always on an uptrend in the long run. This makes the investor resilient to volatility in business cycles; day traders on the other hand tend to endure more volatility by the fact that they commence in daily trades taking advantage to the movements of the market.

Below are the values to be considered for this concept:

	EUR/USD	GBP/USD	USD/JPY
Annualized Return	0.90%	0.77%	1.02%
RMSE	0.20	0.111	0.102
MAPE	880.06%	760.77%	554.34%
Sharpe Ratio	0.1	0.05	0.12

In the Foreign Exchange market is it a bit different from the Stock Market, where most currencies do not follow a constant uptrend or down trend on the long run. Results show a low rate of return and a horrifying percentage of error. This is understandable, due to the fact the foundation to this model is that previous prices is the determinant to the forecasted price; which is never the case, making the percentage error really high.

5.0 Linear Models:

As mentioned before, linear models that will be looking into in this paper: Moving Average Convergence Divergence (MACD) and the Autoregressive Moving Average (ARMA)

These models are used today on daily base trading within the practical world. They signal for the trader of a trend forming, making their decisions based on the crossing of moving averages.

5.1 Auto-Regressive Moving Average:

ARMA is pronounced for its ability to be modified and configured (Periods of trade) meaning when a converging point on moving average with the price happens (a trend his forming) a decision of a long or short position is considered based on this.

Results show a well preforming method of trading yielding well off returns for in and out of samples. Below are the results to this model:

	EUR/USD	GBP/USD	USD/JPY
Annualized Return	22.44%	19.23%	12.22%
RMSE	0.075	0.068	0.055
MAPE	43.39%	48.39%	48.39%
Sharpe Ratio	0.87	0.89	0.92

IF WANT ADD SMALL PART

5.2 Moving Average Convergence Divergence (MACD):

MACD is a method of which order decisions whether to go short or long are made based on the convergence of two moving averages of different periods. Similar to the

Auto-Regressive Moving Average (ARMA), this method is an indicator to a change in trend direction. As shown below in the diagram taken from MetaTrader's platform (A FX trading Platform):

As you can see in the diagram above, the Moving Average Convergence Divergence (MACD) is as mentioned before that trading decisions are made on the basis of when a given Moving Average period converges with a different period Moving Average.

Results to this method is shown below:

	EUR/USD	GBP/USD	USD/JPY
Annualized Return	16.23%	20.43%	18.46%
RMSE	0.0088	0.0120	0.023
MAPE	47.33%	48.33%	48.23%
Sharpe Ratio	0.57	0.87	0.66

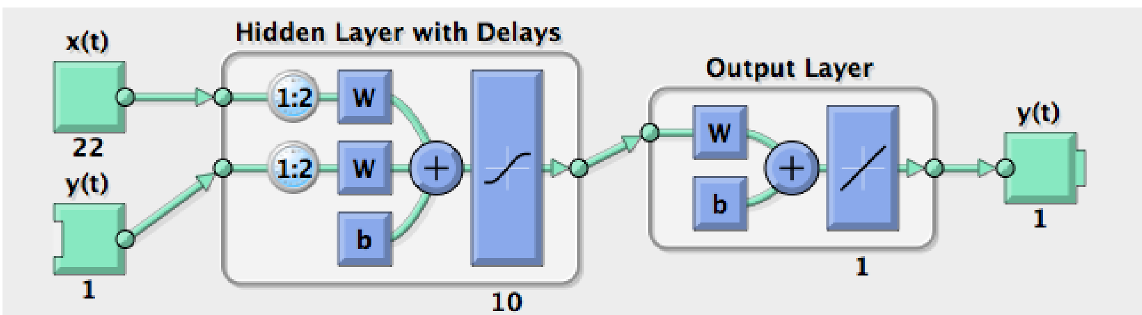
Results show the MACD model show a weak Sharpe Ratio. The best from the three pair was the GBP/USD, giving some promising results. As for the rest, the rates of return and Sharpe Ratio aren't as the greatest.

As literature has commented on the MACD, they mentioned that it's not practical in the real world application due to its inconsistent linearity and sensitivity. Contradicting this literature, practitioners tend to use it in high frequency trading on a daily basis as they find it as a good supportive indicator of a forming trend within the Foreign Exchange market.

6.0 Artificial Neural Networks (ANNs):

6.1 High-Order Neural Networks (HONNs)

High-Order Neural Networks as presented in the literature review are node-based network, where inputs in this case consist of nine-lagged rate of return values. These input nodes get weighted as they go through the hidden layers. After this part of the journey, information is then transmitted again with a given weight in a non-linear manner as an activation function as was mentioned in the literature review in detail.



Results for this forecast of one-day for all the pairs ahead is below:

	EUR/USD	GBP/USD	USD/JPY
Annualized Return	34.78%	22.66%	21.67%
RMSE	0.025	0.25	0.025
MAPE	46.55%	47.55%	47.55%
Sharpe Ratio	1.1	0.98	0.78

Results show that HONNs is a well performing model; as mentioned in previous literature, large input samples as was done here in this model contributes to a lowered forecasted values and increasing the errors. As suggested, no more than 4 inputs should in include in this model. Whereas in this paper there was 9 lagged inputs with a sample size of 3800 end of day closing rate of returns.

6.2 Multi-Layer Perceptron:

The multi-Layer Perceptron is quite similar to the High-Order Neural Network in certain aspects and has small differentiation as from MLPs; all this had been mentioned in the literature review part and methodology. Same as all the other models, this has been done using Matlab and the results are shown below:

	EUR/USD	GBP/USD	USD/JPY
Annualized Return	35.76%	29.55%	30.25%
RMSE	0.025	0.017	0.0124
MAPE	44.23%	46.23%	46.21%
Sharpe Ratio	1.30	0.87	0.77

Results show an acceptable mean of forecasting using this model. The MLP is a simple setup making back-propagation well enough to adjust for errors and optimizing the forecasted returns.

The returns in all the currency pairs show promising returns with the accuracy pleasing. The Root Mean Square Error is at a minimal. Literature suggestion this model is well established in the use in practical cases.

6.3 Gene Expressions and Genetic Programming:

As mentioned in the Literature review, GAs is well renounced as good problem sovling algorithms (the process of evolution). The main difference how these networks work from the classic Artificial Neural Network is that they go through a process of natural selection of data in order to find the optimal solution to a problem when going though the stages of training. They have seen to level off of the benchmarks that were set as

been seen in previous work. Results are shown below:

	EUR/USD	GBP/USD	USD/JPY
Annualized Return	34.05%	36.44%	30.49%
RMSE	0.050	0.044	0.48
MAPE	49.26%	49.30%	49.24%
Sharpe Ratio	0.67	0.457	0.89

With these results, it is showing the highest forecasted rate of return of all other models with acceptable Sharpe ratios.

6.4 Support-Vector Machines/Support-Vector Regressions:

This model has displayed a lot of interest in the field of financial forecasting, as it been widely studied in the academic world. Literature suggests that Learning Machines such as this type has recently attracted a lot of eyes. This is maybe because of the required amount of computational power required by these Learning Machines, and now, in this day and age, supercomputers are now available to run complex and dynamic algorithms.

This type of model is later combined with other previous models making a rolling hybrid model expecting more success in this field. Previous literature suggests this combination to make the network better at creating featured subset selections. Gene Expressions are epic in the field of program selection, with this being combined, only makes this and later to come models more efficient and modeling forecasts.

Results are show of this model:

	EUR/USD	GBP/USD	USD/JPY
Annualized Return	65.05%	49%	44.09%
RMSE	0.0190	0.0190	0.0185
MAPE	49.22%	48.22%	48.22%
Sharpe Ratio	1.87	1.45	1.09

With these results, clearly show a well performing forecast in terms of Sharpe Ratio making it to 1 -1.87. Returns are well off high with low errors in the forecast. Making this model the best performing one in consideration of errors.

		EUR/USD	GBP/USD	JPY/USD
Random Walk	Annualized Return	29.30%	24.90%	26.03%
	RMSE	0.0552	0.0653	0.0552
	MAPE	92.23%	88.23%	76.23%
	Sharpe Ratio	0.77	0.71	0.72
Buy & Hold	Annualized Return	0.90%	0.77%	1.02%
	RMSE	0.20	0.111	0.102
	MAPE	880.06%	760.77%	554.34%
	Sharpe Ratio	0.1	0.05	0.12
ARMA	Annualized Return	22.44%	19.23%	12.22%
	RMSE	0.075	0.068	0.055
	MAPE	43.39%	48.39%	48.39%
	Sharpe Ratio	0.87	0.89	0.92
MACD	Annualized Return	16.23%	20.43%	18.46%
	RMSE	0.0088	0.0120	0.023
	MAPE	47.33%	48.33%	48.23%
	Sharpe Ratio	0.57	0.87	0.66
HONNs	Annualized Return	34.78%	22.66%	21.67%
	RMSE	0.025	0.25	0.025
	MAPE	46.55%	47.55%	47.55%
	Sharpe Ratio	1.1	0.98	0.78
MLP	Annualized Return	35.76%	29.55%	30.25%
	RMSE	0.025	0.017	0.0124
	MAPE	44.23%	46.23%	46.21%
	Sharpe Ratio	1.30	0.87	0.77
Gene Expression	Annualized Return	34.05%	36.44%	30.49%
	RMSE	0.050	0.044	0.48
	MAPE	49.26%	49.30%	49.24%
	Sharpe Ratio	0.67	0.457	0.89
SVM-SVR	Annualized Return	65.05%	49%	44.09%
	RMSE	0.0190	0.0190	0.0185
	MAPE	49.22%	48.22%	48.22%
	Sharpe Ratio	1.87	1.45	1.09

By looking at the table above of all the forecasted values and their errors, we can conclude some have done better than others. Some with high Root Mean Average Percentage Error, and others had a lower MAPE; and the same goes with the Annualized Returns.

Neural networks and Random walk showed the best performing models in terms of probability and lowest risk. And for ARMA and MACD showed low profitability and higher risk in terms of the Root Mean Average Error making them not a practical method or model to be used in real world application.

In light of what has been done in this paper, Neural Networks have a bright future in financial forecasting; Learning Machines have made great progress in the Gene Algorithms Support-Vector Regressions (GA-SVR), they have shown the best preforming models to come.

7.0 Conclusion:

The objective of this paper is to determine which model is the best model in forecasting Foreign Exchange Rates whether it be the simple "Random Walk", "Buy and Hold" or the use of linear models, or if Artificial Neural Networks and their variation and advances. The process to identify this, as pervious literature suggested, the time scale to which the forecast had been done is "one-day ahead" forecast. The collection of all data using the "Bloomberg" system to extract all the data prices from 2000-2014 period and then converted into returns as suggested as well by previous literature.

Linear models are well know to be used more in technical trading, where day traders use many linear indicators to help them decide quickly whether a trend is forming in high-frequency trading. As in this paper, software base modeling was used to make these predictions. ARMA and MACD did fairly well when predicting the chosen currency pairs despite their high errors, they preformed well against the common model the Buy and Hold strategy. On the other hand, they did not preform well against the Random Walk model as it was mentioned so previously by other scalars.

As for the Multi-Layer Perceptron, it preformed as I expected. Previous literature pointed out this fact that this model known for its performance showing promising returns and with a low Error term. This is lined with the HONNs model; returns and error terms were fairly acceptable.

Ending with the Genetic Algorithm (Gene Expression and Genetic Programming) that is basically Genetic Algorithm Support Vector Machines, they have outperformed all previous models in terms of Rate of Return and with the lowest Error terms. To top it off, it showed as well as a well and a healthy Sharpe Ratio. Proving this, confirms that Learning Machines are the new frontier in comparisons, and should be the foundation to which new research should be lunched off from. As seen in literature, This model is being developed into Hybrid Genetic Algorithm Support Vector Regressions and are showing a promising future with the use of Classification and superior back-propagation methods.

8.0 Recommendations:

The objective of this paper was to compare various forecasting models and to find and confirm previous literature that had been done in this field. Finding that literature has been confirmed my beliefs in this field that they do work, explicitly Artificial Neural Networks against other forecasting models. To be more exact to where the direction of these Neural Networks, Learning Machines has found a place in the forecasting universe. Making them these Learning Machines superior in the accuracy of forecasts.

I believe that given the fact the Foreign Exchange market is highly volatile, I my self, conduct FX trades for the past years, but not with Neural Networks. But sole on the raw method of technical trading, where trades are done in a high-frequency manor. And so, I do still have belief that linear models do still have their place in the prediction world do the fact that the on 5 minute - 15 minute charts are ever changing in the speed at which trends formulate. With this being said, linear model trend catching techniques tend to be helpful with the use of course the wide verity of other linear models like for example: Relative Strength Indicator (RSI), Commodity Channel Index or other types of oscillators. They tend to help and have been till today used by ForeX traders.

By which in the hype of the development of Artificial Neural Networks, I have come to believe that these systems are epic when forecasting on the longer run, and be a good tool in risk management. I also believe that the GA-SVR is the new frontier in Learning Machines and so they would be best to be used in the managing Hedge Funds and Mutual Funds; as those portfolios are extremely large and well diversified, I would assume that the input sets are enormous making the computational power vital, and where the use of supercomputers would be forerunner in this voyage of perfection.

I also believe that having a Learning Machine that can actually comprehend historical financial disasters and being able to know what were the leading components of that disaster and then regress them and factor it into the network of evolutions. In other words, by factoring in all the main affecters that affect a given price that the target to be forecasted and finding their regression using the GARCH models of regression, to realize by how much they tend to be affected when changes in the dependable. And then adjust the forecast when the external factor changes.

Sources:

Acimovic, E., Ghabour, R. and Karlsson, E. (2012). Is the Swedish stock market efficient?.

Allen, F. and Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. *Journal of financial Economics*, 51(2), pp.245--271.

Allen, F. and Karjalainen, R. (1995). Using genetic algorithms to find technical trading rules. *Rodney L. White Center for Financial Research Working Paper*, pp.20--95.

Bandini, S., Bonomi, A. and Vizzari, G. (2009). Simulation supporting the design of self-organizing ambient intelligent systems. pp.2082--2086.

Brock, W., Lakonishok, J. and LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance*, 47(5), pp.1731--1764.

Cheung, Y. (1993). Long memory in foreign-exchange rates. *Journal of Business \& Economic Statistics*, 11(1), pp.93--101. Cheung, Y. (1993). Long memory in foreign-exchange rates. *Journal of Business \& Economic Statistics*, 11(1), pp.93--101.

Clark, G. (1950). The organization of behavior: A neuropsychological theory. DO Hebb. John Wiley And Sons, Inc., New York, 1949, 335 pages, 19 illustrations, 288 references. $4.00. *Wiley Online Library*.

Cheung, Y. (1993). Long memory in foreign-exchange rates. *Journal of Business \& Economic Statistics*, 11(1), pp.93--101.

Clark, G. (1950). The organization of behavior: A neuropsychological theory. DO Hebb. John Wiley And Sons, Inc., New York, 1949, 335 pages, 19 illustrations, 288 references. $4.00. *Wiley Online Library*.

Dase, R. and Pawar, D. (2010). Application of Artificial Neural Network for stock market predictions: A review of literature. *International Journal of Machine Intelligence*, 2(2), pp.14--17.

Duan, K., Keerthi, S. and Poo, A. (2003). Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing*, 51, pp.41--59.

Dunis, C., Laws, J. and Karathanassopoulos, A. (2011). Modelling and trading the Greek stock market with mixed neural network models. *Applied Financial Economics*, 21(23), pp.1793--1808.

Dupernex, S. (2007). Why Might Share Prices Follow A Random Walk?. *Student Economic Review*, 21(1), pp.167--179.

Ehrhardt, G. (2013). The Not-so-Random Drunkard's Walk. *Journal of Statistics Education*, 21(2).

Emanuel, D. (1980). A Note on Filter Rules and Stock-Market Trading in New Zealand*. *Economic Record*, 56(155), pp.378--381.

F., B. and S. F., C. (2013). Forecasting foreign exchange rates using Support Vector Regression. IEEE Conference, pp.34-41.

Fama, E. (1976). Efficient capital markets: reply. *Journal of Finance*, pp.143--145.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work*. *The journal of Finance*, 25(2), pp.383--417.

Fama, E. and Blume, M. (1966). Filter rules and stock-market trading. *Journal of business*, pp.226--241.

Fesler, J. (1948). Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization. By Herbert A. Simon.(New York: The Macmillan Company. 1947. Pp. xvi, 259. $4.00.). *The Journal of Politics*, 10(01), pp.187--189.

Franses, P. and van Griensven, K. (1998). Forecasting exchange rates using neural networks for technical trading rules. *Studies in Nonlinear Dynamics \& Econometrics*, 2(4).

Giles, C. and Maxwell, T. (1987). Learning, invariance, and generalization in high-order neural networks. *Applied optics*, 26(23), pp.4972--4978.

Glen Donaldson, R. and Kamstra, M. (1999). Neural network forecast combining with interaction effects. *Journal of the Franklin Institute*, 336(2), pp.227--236.

Gunawardana, K. (2012). Predicting stock price performance: a neural network approach. 4.
Harrald, P. and Kamstra, M. (1997). Evolving artificial neural networks to combine financial forecasts. *Evolutionary Computation, IEEE Transactions on*, 1(1), pp.40--52.

Holland, J., (1995). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*.

Huang, C. and Wang, C. (2006). A GA-based feature selection and parameters optimizationfor support vector machines. *Expert Systems with applications*, 31(2), pp.231--240.

J., L. and P., A. (1992). Three-step method evaluates neural networks for your application. pp.93-110.

Kai, F. and Wenhua, X. (1997). Training neural network with genetic algorithms for forecasting the stock price index. 1, pp.401--403.

Kai, F. and Wenhua, X. (1997). Training neural network with genetic algorithms for forecasting the stock price index. 1, pp.401--403.

Kanungo, R. (2004). Genetic Algorithms: Genesis of Stock Evaluation. *Economics WPA Working Paper*, (0404007).

Kanungo, R. (2004). Genetic Algorithms: Genesis of Stock Evaluation. *Economics WPA Working Paper*, (0404007).

Karemera, D. and Kim, B. (2006). Assessing the forecasting accuracy of alternative nominal exchange rate models: The case of long memory. *Journal of Forecasting*, 25(5), pp.369--380.
Karemera, D. and Kim, B. (2006). Assessing the forecasting accuracy of alternative nominal exchange rate models: The case of long memory. *Journal of Forecasting*, 25(5), pp.369--380.

Knowles, A., Hussein, A., Deredy, W., Lisboa, P. and Dunis, C. (2005). Higher-order neural networks with Bayesian confidence measure for prediction of EUR/USD exchange rate. *Artificial Higher Order Neural networks for Economics and Business, ed. by M. Zhang (Idea Group, Hershey, 2009)*, pp.48--59.

Knowles, A., Hussein, A., Deredy, W., Lisboa, P. and Dunis, C. (2005). Higher-order neural networks with Bayesian confidence measure for prediction of EUR/USD exchange rate. *Artificial Higher Order Neural networks for Economics and Business, ed. by M. Zhang (Idea Group, Hershey, 2009)*, pp.48--59.

Koza, J. (1992). *Genetic programming*. 1st ed. Cambridge, Mass.: MIT Press.

Koza, J., Bennett III, F. and Stiffelman, O. (1999). Genetic programming as a Darwinian invention machine. *Springer*, pp.93--108.

Kwon, Y. and Moon, B. (2007). A hybrid neurogenetic approach for stock forecasting. *Neural Networks, IEEE Transactions on*, 18(3), pp.851--864.

Kwon, Y. and Moon, B. (2007). A hybrid neurogenetic approach for stock forecasting. *Neural Networks, IEEE Transactions on*, 18(3), pp.851--864.

Kwon, Y. and Moon, B. (2007). A hybrid neurogenetic approach for stock forecasting. *Neural Networks, IEEE Transactions on*, 18(3), pp.851--864.

Lee, Y., Lin, Y. and Wahba, G. (2004). Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data. *Journal of the American Statistical Association*, 99(465), pp.67--81.

Lee, Y., Lin, Y. and Wahba, G. (2004). Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data. *Journal of the American Statistical Association*, 99(465), pp.67--81.

Lee, Y., Lin, Y. and Wahba, G. (2004). Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data. *Journal of the American Statistical Association*, 99(465), pp.67--81.

Leen, T., Tesauro, G. and Touretzky, D. (1995). *Advances in neural information processing systems 7*. 1st ed. Cambridge, Mass.: MIT Press.

Lo, A. (2000). Finance: A selective survey. *Journal of the American Statistical Association*, 95(450), pp.629--635.

Malkiel, B. and Yang, M. (2004). *Man bu Hua'er jie*. 1st ed. Taibei Shi: Tian xia yuan jian chu ban gu fen you xian gong si.

McCulloch, W. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), pp.115--133.

Meese, R. and Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample?. *Journal of international economics*, 14(1), pp.3--24.

Mendelsohn, L. (2012). *Trend Forecasting with Intermarket Analysis*. 1st ed. Hoboken: Wiley.

Nicholas Refenes, A., Zapranis, A. and Francis, G. (1994). Stock performance modeling using neural networks: a comparative study with regression models. *Neural Networks*, 7(2), pp.375--388.

Papadimitriou, T., Gogas, P. and Plakandaras, V. (2013). Forecasting daily and monthly exchange rates with machine learning techniques.

Refenes, P. (1995). *Neural networks in the capital markets*. 1st ed. Chichester: Wiley.

S. S., A. (1964). The Random Character of Stock Market Prices. *MIT Press*.

Sampson, J. (1976). Adaptation in Natural and Artificial Systems (John H. Holland). *SIAM Review*, 18(3), pp.529--530.

Sermpinis, G., Stasinakis, C., Theofilatos, K. and Karathanasopoulos, A. (2013). *Modeling, Forecasting and Trading the EUR Exchange Rates with Hybrid Rolling Genetic Algorithms – Support Vector Regression Forecast Combinations*. 1st ed. London.

Sermpinis, G., Theofilatos, K., Karathanasopoulos, A., Georgopoulos, E. and Dunis, C. (2013). Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and Particle Swarm Optimization. *European Journal of Operational Research*, 225(3), pp.528--540.

Spector, L. (2007). *Automatic quantum computer programming*. 1st ed. [S.l.]: Springer.

Spitzer, F. (2001). *Principles of random walk*. 1st ed. New York: Springer.

Srivastava, S. and Tripathi, K. (n.d.). Artificial Neural Network and Non-Linear Regression: A Comparative Study.

Timmermann, A. and Granger, C. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting*, 20(1), pp.15--27.

Trafalis, T. and Ince, H. (2000). Support vector machine for regression and applications to financial forecasting. 6, pp.6348--6348.

Ullrich, C., Seese, D. and Chalup, S. (2007). Foreign exchange trading with support vector machines. *Springer*, pp.539--546.

Ullrich, C., Seese, D. and Chalup, S. (2007). Foreign exchange trading with support vector machines. *Springer*, pp.539--546.

Vaisla, K. and Bhatt, A. (2010). An analysis of the performance of artificial neural network technique for stock market forecasting. *International Journal on Computer Science and Engineering*, 2(06), pp.2104--2109.

X., S. and L., L. (2006). Multicategory E-Learning. *Journal of the American Statistical Association*, (101(474), pp.500-509.

Xiru, Z. (1994). Non-linear intraday prediction models for intraday. *International Journal of Intelligent Systems in Accounting and Finance Management*, (54(3), pp.234-252.

Yao, J. and Tan, C. (2000). A case study on using neural networks to perform technical forecasting of forex. *Neurocomputing*, 34(1), pp.79--98.

Yao, J. and Tan, C. (2000). A case study on using neural networks to perform technical forecasting of forex. *Neurocomputing*, 34(1), pp.79--98.

Yao, X. and Liu, Y. (1997). A new evolutionary system for evolving artificial neural networks. *Neural Networks, IEEE Transactions on*, 8(3), pp.694--713.

Yoon, Y. and Swales, G. (1991). Predicting stock price performance: A neural network approach. 4, pp.156--162.

Yoon, Y., Swales Jr, G. and Margavio, T. (1993). A comparison of discriminant analysis versus artificial neural networks. *Journal of the Operational Research Society*, pp.51--60.

Zhang, M., Xu, S. and Fulcher, J. (2002). Neuron-adaptive higher order neural-network models for automated financial data modeling. *Neural Networks, IEEE Transactions on*, 13(1), pp.188--204.

ZINZALIAN, A., SIMEONOV, D. and ROBEVA, E. (2009). FX FORECASTING WITH HYBRID SUPPORT VECTOR MACHINES.