



University  
of Glasgow

Gordon, Ross (2019) *Applications of artificial neural networks in financial market forecasting*. PhD thesis

<https://theses.gla.ac.uk/74373/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)

# Applications of Artificial Neural Networks in Financial Market Forecasting

Ross Gordon

Submitted in fulfilment of the requirements  
for the Degree of Doctor of Philosophy

University of Glasgow

Adam Smith Business School

College of Social Sciences

March 2019

# Abstract

This thesis evaluates the utility of Artificial Neural Networks (ANNs) applied to financial market and macroeconomic forecasting. In application, ANNs are evaluated in comparison to traditional forecasting models to evaluate if their nonlinear and adaptive properties yield superior forecasting performance in terms of robustness and accuracy. Furthermore, as ANNs are data-driven models, an emphasis is placed on the data collection stage by compiling extensive candidate input variable pools, a task frequently underperformed by prior research. In evaluating their performance, ANNs are applied to the domains of: exchange rate forecasting, volatility forecasting, and macroeconomic forecasting.

Regarding exchange rate forecasting, ANNs are applied to forecast the daily logarithmic returns of the EUR/USD over a short-term forecast horizon of one period. Initially, the analytic method of Technical Analysis (TA) and its sub-section of technical indicators are utilised to compile an extensive candidate input variable pool featuring standard and advanced technical indicators measuring all *technical* aspects of the EUR/USD time series. The candidate input variable pool is then subjected to a two-stage Input Variable Selection (IVS) process, producing an informative subset of technical indicators to serve as inputs to the ANNs. A collection of ANNs is then trained and tested on the EUR/USD time series data with their performance evaluated over a 5-year sample period (2012 to 2016), reserving the last two years for out of sample testing. A Moving Average Convergence Divergence (MACD) model serves as a benchmark with the in-sample and out-of-sample empirical results demonstrating the MACD is a superior forecasting model across most forecast evaluation metrics.

For volatility forecasting, ANNs are applied to forecast the volatility of the Nikkei 225 Index over a short-term forecast horizon of one period. Initially, an extensive candidate input variable pool is compiled consisting of implied volatility models and historical volatility models. The candidate input variable pool is then subjected to a two-stage IVS process. A collection of ANNs is then trained and tested on the Nikkei 225 Index time series data with their performance evaluated over a 4-year sample period (2014 to 2017), reserving the last year for out-of-sample testing. A GARCH (1,1) model serves as a benchmark with the out-of-sample empirical results finding the GARCH (1,1) model to be the superior volatility forecasting model.

The research concludes with ANNs applied to macroeconomic forecasting, where ANNs are applied to forecast the monthly per cent-change in U.S. civilian unemployment and the quarterly per cent-change in U.S. Gross Domestic Product (GDP). For both studies, an extensive candidate input variable pool is compiled using relevant macroeconomic indicator data sourced from the Federal Bank of St Louis. The candidate input variable pools are then subjected to a two-stage IVS process. A collection of ANNs is trained and tested on the U.S. unemployment time series data (UNEMPLOY) and U.S. GDP time series data. The sample periods are (1972 to 2017) and (1960 to 2016) respectively, reserving the last 20% of data for out of sample testing. In both studies, the performance of the ANNs are benchmarked against a Support Vector Regression (SVR) model and a Naïve forecast. In both studies, the ANNs outperform the SVR benchmark model.

The empirical results demonstrate that ANNs are superior forecasting models in the domain of macroeconomic forecasting, with the Modular Neural Network performing notably well. However, the empirical results question the utility of ANNs in the domains of exchange rate forecasting and volatility forecasting. A MACD model outperforms ANNs in exchange rate forecasting both in-sample and out-of-sample, and a GARCH (1,1) model outperforms ANNs in volatility forecasting.

# Table of contents

<b>Abstract</b>	1
<b>Contents Table</b>	3
<b>List of Tables</b>	8
<b>List of Figures</b>	10
<b>Acknowledgements</b>	12
<b>Declaration</b>	13
<b>List of Abbreviations</b>	14
<b>1. Introduction</b>	
1.1 Background & Motivation	16
1.2 Outline & Contribution	19
<b>2. Financial Forecasting and Trading Strategies: A Review</b>	
2.1.1 Technical Analysis Overview	23
2.1.2 Dow Theory	24
2.1.3 Criticisms of Dow Theory	28
2.1.4 Technical Analysis & Fundamental Analysis	29
2.1.5 Technical Analysis: Application & Criticism	29
2.1.6 Random Walk Theory	31
2.1.7 Efficient Market Hypothesis	34
2.1.6.1 Criticism of the Efficient Market Hypothesis	34
2.1.7 Behavioural Finance	34
2.1.8 Adaptive Market Hypothesis	34
2.2 Trading Strategies	37
2.2.1 The Buy & Hold Strategy	37
2.2.2 Technical Trading Strategies	37
2.2.2.1 Moving Averages	38
2.2.2.2 Moving Average Literature Review	42
2.2.2.3 Oscillators	43
2.2.2.4 Oscillators Literature Review	46
2.2.2.5 Technical Analysis & Computational Intelligence	48

### **3. Forecasting Model**

3.1	Artificial Neural Networks	50
3.1.1	Multilayer Perceptron	50
3.1.2	Generalised Feed Forward	52
3.1.3	Modular Neural Network	53
3.1.4	Recurrent Neural Network	54
3.1.5	Radial Basis Function	56
3.1.6	Probabilistic Neural Network	57
3.1.7	Co-Active Neuro-Fuzzy Inference System (CANFIS)	58
3.1.8	Forecast Combination	62
3.2	Benchmark Models	62
3.2.1	Moving Average Convergence Divergence	62
3.2.2	Naïve Forecast	63
3.2.3	Support Vector Regression	63
3.2.4	GARCH(1,1)	63

### **4. Artificial Intelligence in Trading: Forecasting the EUR/USD Exchange Rate with Artificial Neural Networks**

4.1	Introduction	63
4.2	Motivation	64
4.3	EUR/USD Time Series Data	65
4.3.1	EUR/USD Time Series Data (Closing Price)	65
4.3.2	EUR/USD Time Series Data (Log Returns)	66
4.4	Input Variable Selection	68
4.4.1	Filter Methods	68
4.4.2	Stepwise Multivariate Linear Regression	68
4.5	Methodology	70
4.5.1	Artificial Neural Networks	70
4.5.2	Number of Hidden Layers	71
4.5.3	Transfer Function	71
4.5.4	Learning Rule	71
4.5.5	Processing Elements	71
4.5.6	Epochs	72
4.5.7	Weights	72
4.5.8	Data Segregation	72

4.6	Forecast Evaluation	73
4.6.1	In-Sample Forecast Evaluation	74
4.6.2	Out-of-Sample Forecast Evaluation	75
4.6.3	Point Forecast	77
4.6.4	Forecast Evaluation under Increased Volatility	77
4.7	Trading Performance	78
4.7.1	In-Sample Trading Performance	79
4.7.2	Out-of-Sample Trading Performance	80
4.8	Conclusion	81

## **5. Volatility Modelling & Forecasting: A Nikkei 225 Index Example**

5.1	Introduction	82
5.2	Motivation	85
5.3	Nikkei 225 Index Time Series Data	86
5.3.1	Nikkei 225 Index Time Series Data (Closing Price)	86
5.3.2	Nikkei 225 Index Time Series Data (Absolute Log>Returns)	87
5.4	Input Variable Selection	89
5.4.1	Filter Methods	89
5.4.2	Multivariate Linear Regression (Stepwise)	89
5.5	Methodology	91
5.5.1	Artificial Neural Networks	91
5.5.2	Number of Hidden Layers	91
5.5.3	Transfer Function	92
5.5.4	Learning Rule	92
5.5.5	Processing Elements	92
5.5.6	Epochs	93
5.5.7	Weights	93
5.5.8	Data Segregation	93
5.5.9	Benchmark Model: GARCH(1,1)	93
5.6	Forecast Evaluation	94
5.6.1	In-Sample Forecast Evaluation	94
5.6.2	Out-of-Sample Forecast Evaluation	95
5.7	Conclusion	96

## **6. Macroeconomic Forecasting with Artificial Neural Networks: A U.S. Unemployment Example**

6.1	Introduction	98
6.2	Motivation	100
6.3	U.S. Unemployment Time Series Data	100
6.3.1	U.S. Unemployment Time Series Data (Persons)	101
6.3.2	U.S. Unemployment Time Series Data (Percent Change)	102
6.4	Input Variable Selection	104
6.4.1	Filter Methods	104
6.4.2	LASSO	104
6.5	Methodology	105
6.5.1	Artificial Neural Networks	106
6.5.2	Number of Hidden Layers	106
6.5.3	Transfer Function	106
6.5.4	Learning Rule	107
6.5.5	Processing Elements	107
6.5.6	Epochs	107
6.5.7	Weights	107
6.5.8	Data Segregation	107
6.6	Forecast Evaluation	108
6.6.1	In-Sample Forecast Evaluation	109
6.6.2	Out-of-Sample Forecast Evaluation	110
6.6.3	Point Forecast	110
6.7	Conclusion	111

## **7. Macroeconomic Forecasting with Artificial Neural Networks: A U.S. GDP Example**

7.1	Introduction	122
7.2	Motivation	114
7.3	U.S. GDP Time Series Data	115
7.3.1	U.S. GDP Time Series Data (\$Bn)	116
7.3.2	U.S. GDP Time Series Data (Percent-change)	117
7.4	Input Variable Selection	119
7.4.1	Filter Methods	119
7.4.2	LASSO	119



7.5	Methodology	120
7.5.1	Artificial Neural Network	121
7.5.2	Number of Hidden Layers	121
7.5.3	Transfer Function	122
7.5.4	Learning Rule	122
7.5.5	Processing Elements	122
7.5.6	Epochs	122
7.5.7	Weights	122
7.5.8	Data Segregation	122
7.5.9	Benchmark Models	123
7.6	Forecast Evaluation	123
7.6.1	In-Sample Forecast Evaluation	124
7.6.2	Out-of-Sample Forecast Evaluation	125
7.6.3	Point Forecast	126
7.6.4	Forecast Evaluation During Crisis Period	126
7.7	Conclusion	127
<b>8.</b>	<b>General Conclusion</b>	<b>128</b>

# List of Tables

## Chapter 4

Table 4-1: EUR/USD Descriptive Statistics (Closing Prices)	66
Table 4-2: EUR/USD Descriptive Statistics (Log>Returns)	67
Table 4-3: Model Summary	69
Table 4-4: Model ANOVA	69
Table 4-5: Model Coefficients	70
Table 4-6: Input Variable Descriptions	72
Table 4-7: Data Segregation of Sample Period	74
Table 4-8: In-Sample Forecast Evaluation	75
Table 4-9: Out-of-Sample Forecast Evaluation	76
Table 4-10: In-Sample Trading Evaluation	77
Table 4-11. Out-of-Sample Trading Evaluation	78
Table 4-12: Point Forecasts	79

## Chapter 5

Table 5-1: Nikkei 225 Index, Descriptive Statistics (Closing Price)	87
Table 5-2: Nikkei 225 Index, Descriptive Statistics (Log>Returns)	88
Table 5-3: Model Summary	90
Table 5-4: Model ANOVA	90
Table 5-5: Model Coefficients	90
Table 5-6: Input Variable Descriptions	93
Table 5-7: Data Segregation of Sample Period	93
Table 5-8: In-Sample Forecast Evaluation	94
Table 5-9: Out-of-Sample Forecast Evaluation	95

## Chapter 6

Table 6-1: US Civilian Unemployment, Descriptive Statistics (%)	101
Table 6-2: U.S. Civilian Unemployment, Descriptive Statistics, (% $\Delta$ )	102
Table 6-3: Model Summary	107
Table 6-4: Model ANOVA	108
Table 6-5: Model Coefficients	108

Table 6-6: Input Variable Descriptions	109
Table 6-7: U.S. Civilian Unemployment, Data Segregation of Sample Period	109
Table 6-8: In-Sample Forecast Evaluation	109
Table6- 9: Out-of-Sample Forecast Evaluation	110

## **Chapter 7**

Table 7-1: U.S. GDP, Descriptive Statistics (\$Bn)	117
Table 7-2: U.S. GDP, Descriptive Statistics, (% $\Delta$ )	118
Table 7-3: Model Summary	123
Table 7-7: U.S. GDP, Data Segregation of Sample Period	125
Table 7-8: In-Sample Forecast Evaluation	126
Table 7- 9: Out-of-Sample Forecast Evaluation	126

## **Appendices**

Appendix. 1: Forecast Evaluation Metrics	130
Appendix. 2: Trading Evaluation Metrics	133
Appendix. 3: EUR/USD Summary Statistics (Closing Price)	134
Appendix. 4: EUR/USD Summary Statistics (Log>Returns)	138
Appendix. 5: Nikkei 225 Index Summary Statistics (Closing Price)	141
Appendix. 6: Nikkei 225 Index Summary Statistics (Absolute Log>Returns)	143
Appendix. 7: U.S. Civilian Unemployment Summary Statistics (%)	146
Appendix. 8: U.S. Civilian Unemployment Summary Statistics (% $\Delta$ )	148
Appendix. 9: U.S. GDP Summary Statistics (%)	150
Appendix. 10: U.S. GDP Summary Statistics (% $\Delta$ )	153

# List of Figures

## Chapter 2

Fig. 2-1: EUR/GBP Daily Chart, November 2016 to May 2017, SMA Application

Fig. 2-2: EUR/GBP Daily Chart, November 2016 to May 2017, SMA, EMA, WMA

## Chapter 3

Fig. 3-1: Feed-Forward Artificial Neural Network (One Hidden Layer)

Fig. 3-2: Generalized Feed-Forward Neural Network Featuring One Hidden Layer

Fig. 3-3: Modular Neural Network

Fig. 3-4: Recurrent (Elman) Neural Network

Fig. 3-5: Radial Basis Function (RBF) Neural Network

Fig. 3-6: Probabilistic Neural Network.

Fig. 3-7: Jordan-Elman Neural Network, 1 Hidden Layer

Fig. 3-8: Time-Lag Recurrent Neural Network

Fig. 3-9: Prototype CANFIS Neural Network with Two-Inputs and One-Output

Fig. 3-10: CANFIS Neural Network

Fig. 3-11: Support Vector Machine (Regression)

Fig. 3-12: MLP Featuring Two Hidden Layers

Fig. 3-13: MLP Featuring One Hidden Layer

## Chapter 4

Fig. 4-1: EUR/USD Closing Price Time Series

Fig. 4-2: EUR/USD Closing Price Time Series Data (Log>Returns)

Fig. 4-3: P-P Plot

## Chapter 5

Fig. 5-1: Nikkei 225 Index Closing Price Time Series

Fig. 5-2: Nikkei 225 Index Closing Price Time Series (Absolute Log>Returns)

## **Chapter 6**

Fig. 6-1: U.S. Civilian Unemployment, (1972 – 2017), % Format

Fig. 6-2: U.S. Civilian Unemployment, (1972 – 2017), %Δ Format

## **Chapter 7**

Fig. 7-1: U.S. GDP, (1960 – 2016), \$Bn

Fig. 7-2: U.S. GDP, (1960 – 2016), %Δ Format

# Acknowledgements

I want to thank the University of Glasgow for giving me the opportunity to undertake this research. I would especially like to express my gratitude to my research supervisors, Dr Georgios Sermpinis and Dr Vasilios Sogiakas, for their patient guidance, support and useful critiques of this research.

I would also like to thank my in-laws, Larry Peel and Dr Deborah Peel for their support during my PhD years, being able to spend more time with my wife and children while performing this research was a privilege. I am also grateful to my other family members who have supported me along the way.

Last but not least, I would also like to thank my wife Kathryn for her moral support and tolerance.

Thank you all

Ross Gordon, March 2019

# Declaration

I declare that, except where explicit reference is made to the contribution of others, that this thesis is the product of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature:

Printed Name: Ross Gordon

# List of Abbreviations

<b>ADF</b>	Augmented Dickey-Fuller Test
<b>AI</b>	Artificial Intelligence
<b>ANN</b>	Artificial Neural Network
<b>AMA</b>	Adaptive Moving Average
<b>AMH</b>	Adaptive Market Hypothesis
<b>AR</b>	Annualised Returns
<b>BB</b>	Bollinger Bands
<b>B&amp;H</b>	Buy & Hold Strategy
<b>CANFIS</b>	Co-Active Neuro-Fuzzy Inference Systems
<b>CI</b>	Computational Intelligence
<b>DJIA</b>	Dow Jones Industrial Average
<b>DJRA</b>	Dow Jones Rail Average
<b>DM</b>	Diebold-Mariano Test
<b>MDM</b>	Modified Diebold-Mariano
<b>DT</b>	Dow Theory
<b>EMA</b>	Exponential Moving Average
<b>EMH</b>	Efficient Market Hypothesis
<b>FA</b>	Fundamental Analysis
<b>GA</b>	Genetic Algorithm
<b>GARCH</b>	Generalized Autoregressive Conditional Heteroskedasticity
<b>GDP</b>	Gross Domestic Product
<b>GFF</b>	Generalised Feed-Forward
<b>HONN</b>	Higher Order Neural Network
<b>IR</b>	Information Ratio
<b>IVS</b>	Input Variable Selection
<b>JB</b>	Jarque-Bera Test
<b>JE</b>	Jordan-Elman
<b>LM</b>	Levenberg Marquart
<b>MA</b>	Moving Average
<b>MACD</b>	Moving Average Convergence Divergence
<b>MAE</b>	Mean Absolute Error



<b>MAPE</b>	Mean Absolute Percent Error
<b>MDD</b>	Maximum Drawdown
<b>ME</b>	Mean Error
<b>MLP</b>	Multilayer Perceptron
<b>MSE</b>	Mean Squared Error
<b>PE</b>	Processing Element
<b>PNN</b>	Probabilistic Neural Network
<b>PT</b>	Pesaran-Timmermann
<b>RNN</b>	Recurrent Neural Network
<b>RMSE</b>	Root Mean Squared Error
<b>RSI</b>	Relative Strength Index
<b>RVF</b>	Radial Basis Function
<b>RW</b>	Random Walk
<b>RWT</b>	Random Walk Theory
<b>SMA</b>	Simple Moving Average
<b>SPF</b>	Survey of Professional Forecasters
<b>SR</b>	Sharpe Ratio
<b>SVM</b>	Support Vector Machine
<b>SVM-R</b>	Support Vector Machine-Regression
<b>TA</b>	Technical Analysis
<b>TDNN</b>	Time-Dependant Neural Network
<b>TI</b>	Technical Indicator
<b>TLRN</b>	Time-Lag Recurrent Network
<b>TRIX</b>	Triple Exponential Moving Average
<b>TTR</b>	Technical Trading Rule
<b>US</b>	United States
<b>WMA</b>	Weighted Moving Average

# Chapter 1

## Introduction

### 1.1 Background & Motivation

Forecasting financial markets from an empirical or quantitative perspective is perfectly described by McNelis (2005) who states: *“Questions of finance and market success or failure are first and foremost quantitative. Applied researchers and practitioners are interested not only in predicting the direction of change but also how much prices, rates of return, spreads, or likelihood of defaults will change in response to changes in economic conditions, policy uncertainty or waves of bullish and bearish behaviour in domestic or foreign markets. For this reason, the premium is on both the precision of the estimates of expected rates of returns, spreads and default rates as well as the computational ease and speed at which these estimates may be obtained. Financial market research is both empirical and computational”*. It is unquestionable that modern financial market forecasting is geared towards accurate forecasts that can be swiftly generated from quantitative computational models. Furthermore, the speed of computation is of growing importance in today’s fast-paced financial markets in which market participants ranging from international financial institutions to at-home day traders do not have extended periods to contemplate their buy or sell decision. Lastly, such analysis and forecasting are performed in a: high-frequency, chaotic, nonlinear, multivariate, and nonstationary environment generating noisy data. This is particularly true of interconnected financial markets such as the Foreign Exchange (FX) market where the collection of all macroeconomic variables affecting the respective economies of two nations merge into a single rate. Such markets are often volatile and prone to the butterfly effect where small changes in a single variable, or set of variables, of one nation, economic or otherwise, can have a significant effect on the value of the security. For example: the Japanese tsunami of 2011, an unpredictable noneconomic event, caused Japanese financial markets to crash, wiping 8% off the value of Nikkei 225 Index within days and causing currency pairs containing the Japanese Yen (JPY) to decline sharply. The definitive question of how one can distil useful information from such financial markets and their respective variables to

better aid their investment objectives is the essence of financial market forecasting and the problem domain of researchers. A common response would advocate quantitative models such as the linear regression model which has stood the test of time. However, such models are linear and not able to identify complex nonlinearities. These limitations question their compatibility and ultimately their utility in financial market analysis and forecasting. Furthermore, the global financial crisis of 2008 reinforced the belief stemming from the great depression of 1930 that traditional forecasting models are not robust tools of portfolio management especially in times of economic turmoil where financial markets become increasingly volatile and unpredictable. Lastly, adding to the already challenging task of financial market forecasting is globalisation and its by-product of increasingly interconnected and interdependent financial markets. It is evident that there exists the need to utilise forecasting models capable of capturing the complex, multivariate, nonlinearities of today's interconnected financial markets. The definitive question now becomes: how can one identify a set of relevant and informative variables from a universe of variables and accurately model the chaotic, multivariate, nonlinear environment in order to identify and exploit useful information to better aid their investment objectives? In response, researchers and practitioners have been increasingly utilising models offered by Computational Intelligence (CI), a sub-branch of Artificial Intelligence (AI), which is primarily driven by the laws of nature and adaptive mechanisms to facilitate intelligent behaviours in complex dynamic environments. The father of soft-computing, James Bezdek, defines CI as; *'A system is computational intelligent when it: deals with only numerical (low-level) data, has pattern recognition components, does not use knowledge in the artificial intelligence sense; and additionally when it (begins to) exhibit i) computational adaptivity, ii) computational fault tolerance, iii) speed approaching human-like turnaround, and iv) error rates that approximate human performance'*, Bezdek (1994). Common CI and AI models are: ANNs, Support Vector Machines (SVM), Genetic Algorithms (GAs) and Fuzzy Logic. These models have been applied to successfully solve a variety of real-world problems and have relatively recently been applied to financial market forecasting, see: Ahn & Kim (2009), Huang *et al.* (2013), Dunis *et al.* (2014), and Dunis *et al.* (2016). Furthermore, the Boston Consulting Group estimate that by the year 2025 the field of wealth management will be dominated by the interaction of AI and big data analytics.

ANNs were introduced by McCulloch & Pitts (1943), and are mathematical models inspired by the structure and functioning of biological neurons. Upon introduction, ANNs were regarded as a method to model the human brain and expectations from their application were high. However, due to the limitations of primitive computing power, the interest in ANNs diminished but was periodically renewed with cycles of development throughout the decades. It was not until the 1980s with the rapid growth of computer science, vastly increased computational power, and the works of Hopfield (1982) that the interest in ANNs was revived and significant theoretical progress occurred. Today, ANNs are a common technology applied in numerous domains - including financial market forecasting. Among the numerous advantages of ANNs compared to traditional linear models is that they comprise of numerous architectures, can accept a large set of input variables, can ignore redundant input variables, are inherently nonlinear, self-adaptive, data-driven, and can approximate any continuous function to any desired level of accuracy, Hornik *et al.* (1989). However, like all models, ANNs do hold limitations. Over-fitting is one problem that can arise and occurs when a supervised learning algorithm is trained to perform well in an in-sample or *training* dataset but fails to perform well in the vital out-of-sample or *testing* period. One solution is to segregate the sample period into an in-sample period and out-of-sample period, with modern sample period segregation further segregating the in-sample period to include a period of cross-validation, Sermpinis *et al.* (2013). Other anti-over-fitting solutions are the early stopping procedure; Lin *et al.* (2009) and Prechelt (2012) and parameter pruning approaches; Castellano *et al.* (1997) and Wang *et al.* (2010). Another limitation is dimensionality, commonly referred to as the *curse of dimensionality*. This limitation is minimised by IVS, with techniques such as dimensionality reduction; Jolliffe (2007), filtering techniques; Mundra & Rajapakse (2010), and embedded techniques; Hsieh *et al.* (2011). Lastly, ANNs are data-driven models therefore highly dependent on the set of input variables. It is therefore essential to identify appropriate and informative input variables to train an ANN in-sample and achieve accurate out-of-sample results. Therefore, the data collection stage is vital but essentially decided by the determination of the modeller.

## 1.2 Outline & Contribution

Considering the motivation expressed in the previous section, this thesis contributes to the field of Computational Finance and the existing body of knowledge by expanding on prior research in three ways: 1) by focusing on the data collection stage by compiling extensive candidate input variable pools, 2) evaluating a larger set of ANNs, and 3) utilising a broader set of forecast evaluation metrics. Regarding point 1, from the literature review, it is evident that prior research exhibits a deficiency in the data collection stage by failing to collect, so far as possible, all relevant input variables. This deficiency is especially true for prior research merging Technical Analysis (TA) with ANNs by utilising technical indicators as input variables. For example: the literature review reveals that prior research frequently utilises as little as fifty technical indicator models prior to IVS which is deficient given that the universe of technical indicators comprises of hundreds of individual models and significantly more derived models. It is argued that neglecting to include so far as possible the universe of technical indicator models is not only unrepresentative of TA and its subsection of technical indicators but also depriving the ANN of potentially informative data therefore likely hindering its performance, Sattler & Schallehn (2001). In solution, this thesis follows the data preparation scheme devised by Yu *et al.* (2007) with emphasis on the data collection stage, collecting as many relevant input variables as possible – this is especially applied to chapters 4 and 5 which both merge TA with ANNs by featuring extensive use of technical indicators in their respective candidate input variable pools. Regarding point 2, this thesis evaluates many different ANN architectures through the use of the premier ANN software package NeuroSolutions Pro. Prior research tends to evaluate small sets of ANNs, usually a maximum of four. While four is sufficient, there exists more ANN architectures each with their respective advantages and disadvantages. For example; in time series analysis and forecasting it is advisable to use temporal (recurrent) ANNs as they feature a memory component which is beneficial to such analysis. Neuro-Fuzzy systems such as CANFIS models have also demonstrated positive results. In solution, this thesis evaluates seven ANN architectures varied over eight activation functions, as well as an ANN combination forecast. For point 3, prior research tends to employ limited sets of forecast evaluation metrics that, although valid, do not yield informative insight into model performance. In solution, this thesis utilises a comprehensive set of forecast evaluation metrics. In total, a standard set of nine metrics are used to evaluate forecast performance, namely: Mean Error (ME), Mean Square Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE),

Mean Percent Error (MPE), Mean Absolute Percent Error (MAPE), Theils-U1, Theils-U2, Pesaran-Timmerman (PT), and Diebold-Mariano (DM). In addition, chapter 4: *Artificial Intelligence in Trading: A EUR/USD Example* also evaluates ANN trading performance, therefore, offers the additional statistics of: Sharpe Ratio (SR), Information Ratio (IR), Maximum Drawdown (MDD) and Annualised Return (AR). It is intended that the comprehensive candidate input variable pools, additional ANNs, and broader set of forecast evaluation metrics contribute in a meaningful way to the field of Computational Finance and the existing body of knowledge.

To outline this thesis, Chapter 1 has introduced the research and the motivation underpinning it, highlighted deficiencies of prior research, proposed solutions and stated how this thesis contributes to the existing body of knowledge. Chapter 2 presents a review of the forecasting techniques that serve as input variables to the ANNs throughout the thesis. The review largely describes TA and its subsection of technical indicators as they are heavily featured in the exchange rate forecasting and volatility forecasting studies and supports the argument presented by Aronson (2006) that if TA is to progress as a scientific discipline, it should merge with CI and IA methods. Chapter 3 describes the ANNs and benchmark models featured in this thesis so to avoid unnecessary repetition in the empirical chapters. Chapters 4-7 present the empirical research regarding exchange rate forecasting, volatility forecasting, and macroeconomic forecasting respectively.

In Chapter 4: *Artificial Intelligence in Trading: A EUR/USD Example*, a collection of ANNs is applied to forecast the logarithmic returns of the EUR/USD time series over a short-term forecast horizon of one period. The data collection stage is especially emphasised in this chapter as prior research has frequently underperformed this stage when utilising technical indicators as input variables. In solution, this chapter focuses on data collection by compiling an extensive candidate input variable pool of standard and advanced technical indicator models measuring all *technical* aspects of the EUR/USD time series for a total of 300 individual technical indicators and 10,568 derived models. The candidate input variable pool is then subjected to a two-stage IVS process, producing an informative subset of technical indicators to serve as inputs to the ANNs. A collection of ANNs is then trained and tested on the EUR/USD time series data with their forecasting performance and trading performance evaluated over a 5-year sample period (2012 to 2016), reserving the last two years for out of sample testing. A MACD model serves as a

benchmark with the empirical results demonstrating the MACD is a superior forecasting model across most forecast evaluation metrics.

In Chapter 5: *Volatility Modelling & Forecasting: A Nikkei 225 Example*, ANNs are applied to forecast the volatility of the Nikkei 225 Index over a short-term forecast horizon of one period. In the same manner as Chapter 4, an emphasis is placed on the data collection stage. An extensive candidate input variable pool is compiled consisting of implied volatility models and historical volatility models for a total of 3,504 models. The candidate input variable pool is then subjected to a two-stage IVS process. A collection of ANNs is then trained and tested on the Nikkei 225 Index time series data with their forecasting performance evaluated over a 4-year sample period (2014 to 2017), reserving the last year for out-of-sample testing. A GARCH (1,1) model serves as a benchmark with the empirical results demonstrating the GARCH (1,1) model to be the superior volatility forecasting model.

In Chapter 6: *Macroeconomic Forecasting with Artificial Neural Networks: U.S. Unemployment Forecasting*, ANNs are applied to forecast the monthly percent-change in U.S. civilian unemployment over a short-term forecast horizon of one month. Emphasis is again placed on the data collection stage by collecting a total of 266 relevant macroeconomic variables sourced from the Federal Bank of St Louis. The candidate input variable pool is then subjected to a two-stage IVS process. A collection of ANNs is trained and tested on U.S. unemployment time series data (UNEMPLOY) with their forecasting performance evaluated over a 45-year sample period (1972 to 2017), reserving the last nine years for out of sample testing. The performance of the ANNs is benchmarked against an SVR model and a Naïve forecast, with the empirical results finding a Modular Neural Network to be the superior forecasting model.

In Chapter 7: *Macroeconomic Forecasting with Artificial Neural Networks: U.S. GDP Forecasting*, ANNs are applied to forecast the quarterly percent-change in U.S. GDP over a short to medium-term forecast horizon of one to four quarters. Like chapter 6, emphasis is again placed on the data collection stage by sourcing as many relevant macroeconomic variables from the Federal Bank of St Louis for a total of 999 variables. The candidate input variable pool is then subjected to a two-stage IVS process. A collection of ANNs is trained and tested on U.S. GDP time series data with their forecasting performance evaluated over a 56-year sample period (1960 to 2016), reserving the last 11 years for out

of sample testing. The performance of the ANNs is benchmarked against an SVR model and a Naïve forecast, with the empirical results finding a Modular Neural Network to be the superior forecasting model.

Chapters 4-7 largely follow the same format and utilise the same forecast evaluation metrics. While a brief interpretation of the forecast evaluation metric set is provided prior to each empirical discussion, the full disclosure and description of each forecast evaluation metric is available in Appendix 1. To conclude the thesis, Chapter 8 presents a general conclusion and suggestions for future ANN-based financial market and macroeconomic forecasting.



# Chapter 2

## Financial Market Forecasting & Trading Strategies: A Review

### 2.1. Preliminaries in Financial Trading

Financial market forecasting is a challenging endeavour that has been extensively researched throughout the past century, especially in recent decades from an empirical perspective driven by the advancements in technology and data mining software. During the 19th century, two distinct forecasting methods were established: Technical Analysis (TA) and Fundamental Analysis (FA). Both methods share the common objective of successful analysis and forecasting but greatly differ in methods and opinion as to how financial markets operate - a difference in opinion that sparked one of the most heated debates in academic Finance. This thesis mainly utilises TA in two of the four empirical chapters; therefore the focus will be on its theory and application. The following sections present a detailed insight of TA, presenting an: overview of the discipline, a brief history, its strategies, the application of such strategies and a literature review pertaining to each strategy. For a recent and extensive chronological literature review of TA, see Farias Nazário *et al.* (2017).

To analyse and forecast financial markets and their seemingly random nature, a method was devised to analyse the *movements* of financial markets, giving market participants a context in which to operate and ultimately use the markets historical time series data to forecast future market movements and associated prices. This method is known as TA and is defined as a method which examines the directional movements of financial markets rather than the goods in which the market deals (Kirkpatrick & Dahlquist 2016).

Alternative definitions by notable practitioners and authors within the field define TA as; *“Technical Analysis is the study of market action, primarily through the use of charts, for the purpose of forecasting future price trends”* Murphy (1999, p3). From the above definitions, it is concluded that TA centres round the concept of trends which are a manifestation of a variety of variables driven by the psychology of mass market participants. TA believes that historical time series data holds predictive value and applies

its analytic methods to the chart of the security under observation to forecast the security's future movements.

Unknown to many, TA dates to Babylonian times. For a detailed history see (Lo & Hasanhodzic, 2013). However, modern day TA was devised by Charles H. Dow (1851 – 1902) who developed the Dow Jones Industrial Average (DJIA). Dow intended to use the DJIA to measure the *health* of the U.S. economy, but his theory quickly evolved into a financial market forecasting tool known as Dow Theory. For this reason, the next section presents a brief overview of Dow Theory and how it relates to modern day TA.

### **2.1.1 Dow Theory**

Dow Theory (DT) is described as the *granddaddy* of all technical studies (Edwards, Magee & Bassetti 2010, p11) and is the foundation on which modern day TA is built. Therefore, an understanding of DT, or at least the awareness of it, is essential to the understanding and interpretation of TA. As stipulated by Dow, the principles of DT are intended to serve as guidelines as opposed to rules. The principles of DT are as follows:

Principle 1: *The Averages Discount everything (Except Acts of God)*. This principle advocates that because the averages (market indexes) reflect the combined market activities and sentiment of market participants on a global scale, including market participants with excellent foresight and superior knowledge, that the averages in their daily fluctuations discount everything: known, foreseeable, and all variables that can influence the supply and demand of securities. Therefore, TA reasons there is no need to research the fundamentals of the security under observation. Consequently, most Technical Analysts do not examine fundamental data, except through the prism of price which TA believes reflects such fundamentals before such fundamentals are fully observable. This concept is reinforced by Engel & West (2005) who claim that changes in asset prices often precede observed changes in fundamentals. Furthermore, natural disasters (acts of god) are quickly appraised and their adverse impact discounted. For example, the Japanese tsunami of 2011 caused the Nikkei 225 Index and exchange rates containing the Japanese Yen (JPY) to decline sharply during that day. However, the markets appraised the freak event and discounted its affect. Within twenty-four hours, the Nikkei 225 Index and relevant exchange rates resumed normal trading operations at pre-tsunami rates.

Principle 2: *The Market has Three Trends*. This principle is central to TA as the existence of trends implies directionality, and ultimately, predictability. DT asserts there are three trends; major, intermediate and minor. The major trend is regarded as the dominant and most important trend and is used to denote bull or bear markets. It has a duration of at least one year but usually lasts far longer and is said to cause an appreciation or depreciation of price by at least twenty percent. TA employs a 200-day MA to represent the major trend. The intermediate trend is a counter trend that interrupts the major trend, i.e. if the major trend is bullish the intermediate trend is bearish. The intermediate trend has a duration of three weeks to three months, and its retracements can erode one to two thirds of the current major trend. TA advocates the use of a technique called Fibonacci Retracements to estimate retracement levels. However, as comforting as these suggested retracement levels are, DT advises not to participate in intermediate trends because they are difficult to analyse, therefore can be misleading. The minor trend is the minutest of all trends and is frequent bullish and bearish fluctuations, with a duration of three weeks but typically less than six days, that assemble the major trend. DT regards the minor trend as meaningless and the only trend of the three trends that can be somewhat manipulated. TA uses moving averages and trend lines to identify the three market trends.

Principle 3: *Major Trend Phases*. This principle governs the development of major bull and bear markets. Regarding major bull markets, DT states that three phases compile the major trend. The first phase is the *accumulation* phase in which economic conditions are poor and have manifested themselves in corporate financial statements. However, farsighted investors feel the current declining market is ending and is due to advance over the long run, therefore are willing to purchase the shares of disgruntled investors at low prices. The second phase involves a steady advance and increases in trading activity as economic conditions and corporate earnings improve. It is during this phase that the market participant can generate the most profit of the three phases. This phase is akin to wave three of Elliott Wave Theory, a TA method that asserts markets are cyclical and compiled of a five-wave structure. The third phase is the *distribution* phase during which the market is a hive of activity due to spectacular economic conditions and pleasing corporate earnings, both of which are driven by the financial media. It is in this stage that market participants are eager to buy shares; however the market has now been advancing for one year, possibly longer, and is due for a reversal with mass conservative sentiment poised to sell. The opposite is true for major bear markets. The last phase of a major bull market, the distribution phase, becomes the first phase of the major bear market. During this phase,

farsighted investors believe the spectacular economic conditions and pleasing corporate earnings have reached saturation, and they begin selling their shares. The mass public is still participating but exhibiting frustration as the hope for profits gradually diminishes. The next stage is the *panic* stage in which buyers decrease significantly, and sellers dominate the market. The result is a sudden and accelerated drop in price accompanied by heavy selling volume. Due to the *fear* present in bear markets the panic phase usually declines further than expected and is consequently followed by a period of recovery manifesting as a lengthy lateral movement. This lateral movement is regarded by TA as a consolidation zone and has a specific strategy to manage it. The third stage is characterised by discouraged sellers who held their position through the panic stage or bought shares during it because they were cheap in comparison to prices in prior months. Economic and business conditions continue to deteriorate, and mass selling ensues once again but with diminished momentum and magnitude, eventually developing as an accumulation phase and resetting the bull – bear market process. While the descriptions sound cyclical, or mechanical even, DT stresses that no two bull or bear markets are the same and no time duration can be set for any phase, it is merely a framework for market participants to operate in and devise their strategies around. Concerning TA, numerous technical indicators measure accumulation and distribution, the most prominent of which is the Accumulation / Distribution (A/D) indicator designed by Mark Chaikin which relates price to volume and serves as a leading indicator of price.

Principle 4: *Confirmation*. The principle of confirmation advises waiting for valid signals of a trend reversal and is compiled of two sub principles. The first sub principle asserts that the averages must confirm, meaning that no valid signal of a trend reversal is signalled by one average alone; both averages must produce a valid signal. In Dow's time, the two averages were the Dow Jones Industrial Average (DJIA) and the Dow Jones Rail Average (DJRA) however, other averages have been established since and are in use today, such as the S&P500 and Russell 2000. The second sub principle asserts that volume must confirm the trend, therefore volume must increase throughout bull and bear markets.

Principle 5: *Lines may substitute for intermediates*. DT regards a *line* as a lateral movement in a market index that has a duration of 3 weeks to 3 months in which price fluctuates approximately five percent, or less, of the average value. The presence of a line, or consolidation zone as TA terms it, signifies that buying and selling pressure is in equilibrium. However, the consolidative situation cannot last forever, and at some point,

directionality needs to ensue. If mass buyers enter the market consequently pushing the price up through the lateral range, this induces a buy signal. If mass sellers enter the market consequently forcing the price to break below the lateral range, this induces a sell signal. TA interprets the above as a breakout strategy, advising that the magnitude of the move preceding the consolidation zone is a measure of the magnitude one can expect to ensue following the directional break of the consolidation zone.

Principle 6: *Closing prices only*. DT only recognises closing prices; therefore, no recognition is given to the extreme high or low that occur throughout the trading day nor the opening price. DT asserts that for a trend to continue, the next day's closing price must exceed the current day's closing price. There is debate as to how much price must exceed the previous day's closing price; Dow advised as little as 0.01 as a valid signal, however, modern advocates suggest as much as an entire point (1.00).

Principle 7: *The Trend Remains Intact Until Otherwise Stated*. This last principle advises that a trend remains active until such a time its reversal is definitively signalled. This principle relates to principle 4 (confirmation) where the averages must confirm and is more a measure of probability in that a market participant is more likely to be successful in waiting for definitive confirmation of trend reversal rather than second guess the market and most likely prematurely enter or exit the market.

It is evident that DT advocates: the existence of trends, phases of trends, the interplay of market indexes, confirmation signals, the use of closing prices only, volume as a complementary variable to price, and the need to hold positions even through adversity until the market itself signals a trend reversal via price. It has been established that TA draws some of its tactics from DT, such as consolidation zones and the associated breakout strategy, has developed a range of technical indicators to measure aspects of DT such as accumulation / distribution, uses long-term moving averages to represent the major trend, applies moving averages and trendlines to identify trends, employs Fibonacci retracements to gauge the retracements of the intermediate trend, and uses the EWT to template major trend phases. However, DT is not without criticism and to fully appreciate its strengths, it is necessary to acknowledge its weaknesses.

### 2.1.2 Criticisms of Dow Theory

The main criticism of DT is its lagging nature. Edwards, Magee & Bassetti (2010) exemplify a perfect account of this criticism, citing the DJIA major bull market of 1942 commencing at 99.92 and ending four years later in 1946 at 212.50 for a total gain of 119.58 points. The strict application of DT using its principle of confirmation dictates that market entry was only possible at 125.88 and market exit not possible until 191.10, resulting in a mere 65 points gain or approximately 50% of the entire bull market. However, while DT advocates accept the above statement is irrefutable, they respond by asking critics to identify a market participant who entered that market at the optimal price of 99.92 and exited at the optimal price of 212.50 - a market participant has yet to come forward. DT is subject to interpretation therefore subject to the hazard of human interpretive ability. DT is also said to cast doubt as to the start and termination of a major trend when the stages of accumulation or distribution are active because there is no definitive sign of trend direction.

With the overview of TA stated and its roots originating from DT, in order to fully accept the principles of TA, it is necessary to differentiate it from its counterpart, Fundamental Analysis (FA). A brief description of FA and the interplay of TA and FA is presented in the following section.

### 2.1.3 Technical Analysis & Fundamental Analysis

As previously stated, TA believes the market discounts all available information before such information is fully observable in the market, and prices move in trends which are forecastable via certain methods. However, FA dismisses most of TA and is instead rooted in the economics of supply and demand and its deterministic factors. The difference between TA and FA is stated by Murphy (2012, p5): *'the fundamentalist studies the cause of market movement, while the technician studies the effect'*. FA is a method of valuing a security which entails measuring its *fair value* by examining relevant economic and financial factors. Economic factors are deemed to be either micro or macroeconomic factors relevant to the security. Financial factors are deemed to be firm-specific, for example, the firm's business ratios, projected earnings, and tax rates. The purpose of performing FA is to establish a fair value that an investor can compare with the security's

current market price to determine whether to buy or sell the security. A rational investor applying FA would buy a security if its current price is below its fair value and sell a security if its current price is above its fair value. A famous and successful proponent of FA is billionaire investor Warren Buffett, who is well known for his use of FA in selecting securities.

Historically, throughout academia, published literature, and industry existed the near constant division between TA and FA. TA was met with disdain by academics largely due to its conflict with the Efficient Market Hypothesis (EMH), although it has now been generally accepted. In industry, traders and investors typically describe themselves as either technical or fundamental with some describing themselves as both. The reality is both forms of analysis complement one another, and the categorisation should not be fundamental *or* technical but fundamental *and* technical because either form of analysis used in isolation is depriving the market participant of potentially valuable information. Furthermore, regarding Graham & Dodd's 1934 book *Security Analysis* which laid the foundations of FA, a close reading of their text reveals that Graham nor Dodd did not believe that FA alone determined stock prices: *"The influence of what we call analytical factors over the market price is both partial and indirect – partial, because it frequently competes with purely speculative factors which influence the price in the opposite direction; and indirect, because it acts through the intermediary of people's sentiments and decisions."* (Graham & Dodd 1934, p28). Pertaining to this view is Goodhart (1988) who argues that exchange rates are determined by a balance of technical and fundamentalist prediction: *"The interplay between speculation based on fundamentals, and on a random walk, or Chartist, approach, influences the outcome"*. Evidently, no form of analysis used in isolation is wholly capable of supplying the market participant with a complete information set. FA does not define trend, nor the belief in them, therefore does not subscribe to technical characteristics such as; levels of support and resistance, consolidation zones, trend channels, price targets or breakouts, despite such characteristics being abundantly evident in the chart of any security. Likewise, from a TA perspective little is said or implemented regarding FA. Although, Technical Analysts believe that fundamentals are inherent, or built-in, to a security's chart, therefore by applying TA they are indirectly and simultaneously employing FA.

Since FA is a product of individual estimates of economic and firm-specific factors, FA is not free from subjective-emotion and error on the part of the analyst. Therefore, proponents of FA could not only suffer losses from an incorrect estimation of fair value but also suffer losses from a correct estimation superseded by a counter market movement induced by mass market participants. In contrast, TA acts passively and only responds to its instruments which conform to the expectations of mass market participants. Proponents of FA also criticise TA and its subsection of Chartism because of its subjective nature and the fact it is based on a framework of patterns and trends renders it impossible to standardise the observations. Murphy (1999) affirms the subjectivity of Chartism, admitting that sometimes disagreement even regarding simple patterns exists. However, regarding technical indicators, such arguments are anchorless because definitive signals create an objective trading approach. As a result, technical indicator-based trading strategies can be implemented more efficiently than FA. While it is true that FA's evaluation of relevant internal and external company data can produce illuminating information as to the future development of a firm, it does not completely explain reactions to the company's share price on stock exchanges. Conversely, TA integrates the psychology of market participants, therefore at least partly accounts for such reactions. Ultimately, the critical question is whether forecasting, be it technical or fundamental, is profitable at all - or if an investor should spare themselves the analytic cost and apply the Buy & Hold strategy as advocated by the EMH.

#### **2.1.4 Technical Analysis - Application & Criticism**

Numerous researchers have documented the extent to which TA is applied to the FX market. Arnold & Moizer (1984) surveyed South African portfolio managers finding that 93% of respondents used FA with 56% of respondents also utilising TA. Allen & Taylor (1990) and Taylor & Allen (1992) conducted a survey on behalf of the Bank of England (BoE) among 400 Chief FX traders in London. The results showed extensive use of TA on short-term forecast horizons and diminished use on long-term forecast horizons. For short-term forecast horizons (intraday to one week), approximately 90% of respondents reported using TA in devising their forecast, with 60% judging TA to be equally important as FA. For longer forecast horizons (one month to one year) the weight given to FA increased. On forecast horizons exceeding one year, the skew toward FA was most pronounced, with approximately 33% of the respondents relying purely on FA and approximately 85%



judging FA to be more important than TA. Menkhoff (1997) expanded on Allen & Taylor (1990) finding comparable results, concluding that a weight of approximately 40% is placed on TA when making short to medium term forecasts and that the use of TA diminishes when longer forecast horizons are considered. Lui & Mole (1998) surveyed 153 FX traders in Hong Kong; the results showed that TA was superior to FA when identifying trend reversals but only marginally better than FA in forecasting trends. Also, the use of TA decreased when longer forecast horizons are considered which reinforces the findings of the surveys mentioned above. Cheung & Chinn (2001) find that 30% of US FX traders are best characterised as technical analysts and that an increasing percentage use TA. Cheung, Chinn & Marsh (2004) confirm previous findings that traders pay more attention to non-fundamental factors at short forecast horizons. More recent surveys have considered the educational background, experience and psychological biases of FX traders with Menkhoff (1997) refuting the notion that Technical Analysts lack the experience and or education of Fundamental Analysts. The surveyed German technicians did not differ from non-technicians regarding; age, education, position, seniority, their firm's trading turnover or assets under management. Menkhoff & Schmidt (2005) investigate the use of the B&H strategy, momentum and contrarian opinion strategies practiced by fund managers. The momentum traders were the least risk-averse, and the contrarian traders exhibited signs of overconfidence. Oberlechner & Osler (2008) use survey evidence from 400 North American FX traders to establish that respondents underestimate uncertainty and overestimate their abilities. They propose that their findings help to explain the high volatility of exchange rates, the profitability of trend following strategies, and the apparent irrationality of exchange rate forecasts. Clements (2010) interviews 13 prominent Technical Analysts regarding market efficiency, trading strategies, attitude towards risk, and intermarket analysis. Lastly, the 2010 CitiFX Pro Forex Trader Survey conducted online with an international pooling of 3,000 traders found that 53% of the respondents utilise a combination of TA and FA, with 36% declaring they only use TA and 8% stating they adhere strictly to FA.

The first major academic study that considered TA as a research subject was *Can Stock Market Forecasters Forecast?* (Cowles, 1933). Regarding the performance of TA on the FX market, it was Goodman (1979), Group of Thirty (1985), Frankel & Froot (1986, 1990a, 1990b), and Goodhart (1988) who first brought it to the attention of researchers and was initially met with disdain. It was not until later when Brock *et al.* (1992) with their pioneering paper presenting evidence on the profitability of several technical trading rules

(TTS) using the bootstrap methodology did researchers even begin to take notice. However, successful studies quickly came under scrutiny regarding risk-adjusted returns. Sweeney (1986) was the first to integrate the risk-adjusted performance of TTS compared to the B&H strategy on the FX market, finding no evidence nor explanation for excessive risk-adjusted returns generated by the TTS. Levich & Thomas (1993) applied a similar methodology finding similar results. Allen & Taylor (1990) record the accuracy of individual Technical Analyst's forecasts over a 1-4 week forecast horizon on three major exchange rates (USD/GBP, DEM/USD and YEN/USD) over a 10-month sample period benchmarked against alternative forecasting methods. They found that some Technical Analysts were able to outperform the alternative forecasting methods which included a: RW, VAR, and ARIMA model. However, they did report a large degree of heterogeneity among the TA-based forecasts. Qi & Wu (2006) tested 2,127 technical trading rules applied to seven exchange rates over the sample period 1973-1998. The rules included; filter rules, MAs, support & resistance rules and channel breakout rules. Their results indicated significant profitability of MA and channel breakout rules for the seven exchange rates. Qi & Wu then applied White's (2000) reality check bootstrap methodology to evaluate the rules and the effects of potential data-snooping bias. They find significant profitability at the 1% level for all seven exchange rates, even after data-snooping bias and transactions costs. Park & Irwin (2007), find similar results for EUR and JPY futures, concluding that TA is profitable in the FX and commodity futures markets but not in stock markets, a conclusion also reached by Silber (1994).

Considering the above, it is logical to wonder why TA remains under constant criticism. TA is extensively practised in the Finance industry yet heavily criticised by academics. The most vocal industry critic is perhaps Warren Buffett who jokingly said to an audience at Vanderbilt University in 2005: *'I realised that technical analysis didn't work when I turned the chart upside down and didn't get a different answer'*. The most vocal academic critic is perhaps Malkiel, exemplified by his following statement: *'Technical analysis is anathema to the academic world. We love to pick on it. Our bullying tactics are prompted by two considerations: (1) the method is patently false; and (2) it's easy to pick on. And while it may seem a bit unfair to pick on such a sorry target, just remember: it is your money we are trying to save'*, Malkiel (2007, p127-128). TA specifically attracts the attention of economists and academics as successful applications question the EMH and RWT, both cornerstones of Economics. Recall that the EMH states that current market prices fully reflect all relevant information. While this is generally accepted, the concept of

current market prices reflecting all available information is also a principle of DT (Principle 1: *The Averages Discount Everything*) and ingrained in TA, therefore highlights slight contradiction on the critic's part.

### 2.1.5 Random Walk Theory

Random Walk Theory (RWT) was initially devised in 1863 by the French broker Jules Regnault and has since undergone numerous revisions. It was subsequently adapted in 1900 by the French mathematician Louis Bachelier in his PhD dissertation titled; *The Theory of Speculation*, then later modified by MIT Sloan School of Management professor Paul Cootner in his 1964 book: *The Random Character of Stock Market Prices*. However, RWT is most associated with Professor Burton Malkiel of Princeton University and his 1973 book: *A Random Walk Down Wall Street*. RWT stipulates that financial markets are efficient and security prices evolve in a random and unpredictable path, therefore cannot be predicted. Price is defined in terms of fair value and an independently distributed random variable that reflects all available information regarding the security. RWT is expressed as:

$$X_t = \mu + X_{t-1} + \varepsilon_t \quad (1)$$

Where:

$X_t$  = Logarithm of the price of an asset at the time  $t$

$\mu$  = Drift of the random walk

$\varepsilon_t$  = Random increment variable

RWT holds implications for financial market forecasting by rendering forecasting methods such as TA ineffective as TA advocates that historical prices contain predictive value. Additionally, RWT holds implications for forecasting models, including those from AI and CI.

### 2.1.6 Efficient Market Hypothesis

Fama (1970) introduced the theory of financial market efficiency in his paper: *Efficient Capital Markets: A Review of theory and empirical work*. Fama's highly influential theory is ubiquitous to modern Finance, with the majority of financial market forecasting research referencing his work to some degree with empirical findings either proving or disproving his theory. According to Fama, a market is efficient if, at any given time, market prices fully reflect all available information and rapidly adjust to new information. Models of this type can be simplified to expected return theories where the following equation describes an efficient market:

$$P_t = P_t^* = E \left[ \left( \sum_{t+1}^{\infty} \frac{D_t}{(1+r)^t} \right) | \Omega_t \right] \quad (2)$$

Where:

$P_t$  = Market price of an asset at the time  $t$

$P_t^*$  = Real or 'intrinsic' price of an asset at the time  $t$

$D$  = Cash flows from the asset from  $t$  until infinity

$\Omega_t$  = Set of all publicly available information about an asset

$r$  = Expected constant rate of return for a risk-free asset

Equation 2 defines a relationship where the expected price of a stock at time  $t+1$  given a set of information  $\Omega_t$ , at time  $t$  is equal to 1 plus the return of the stock at time  $t+1$  multiplied by the price of the stock at time  $t$ . Given the price and return variables are random this implies that the future value of any stock given the set of available information is not predictable. The EMH implies that if the current price fully reflects all available information then the current price of a security is a good measure of its fair value and consequently no strategy can be devised to outperform the market, therefore, questioning forecasting models and the need for active portfolio management. The EMH and its implications for TA are well presented by Higgins (1992) who states: '*Market efficiency is a description of how prices in competitive markets respond to new information. The arrival of new information to a competitive market can be linked to the arrival of a lamb chop to a school of flesh-eating piranha, where investors are – plausibly enough – the piranha. The instant the lamb chop hits the water, there is a turmoil as the fish devour the meat. Very soon the meat is gone, leaving only the worthless bone behind, and the water returns to*

*normal. Similarly, when new information reaches a competitive market, there is much turmoil as investors buy and sell securities in response to the news, causing prices to change. Once prices adjust, all that is left of the information is the worthless bone. No amount of gnawing on the bone will yield any more meat, and no further study of old information will yield any more valuable intelligence".* By Higgins, it is clear the worthless bone is a metaphor for TA and its premise that historical data holds predictive value, with the proposition that the use of, or gnawing on, is an entirely fruitless task.

The basis of the EMH is the Economic theory of competitive markets, which holds that arbitrage competition among market participants and their profit motive will create efficient financial markets. As new information enters the marketplace, all participants will immediately appraise it, acting *rationally* to adjust the price to a revised fair value. Should the price deviate from the fair value, noise arbitrageurs will compete to bring the price back to its fair value. Fama (1970) promotes three forms of market efficiency:

*Weak Form Efficiency:* This form implies that all past information is fully reflected in the current market price. Therefore, if the current market price fully reflects all past information, the next day's price fluctuations would be the result of new information only. As new information arrives at random, the price fluctuations must also be random. Any trading based on this type of information is not expected to yield abnormal returns. Since TA uses historical data to forecast future price movements, if the weak form of efficiency holds, then TA is worthless. The weak form also applies to the semi-strong and strong and forms.

*Semi-Strong Efficiency.* The semi-strong form requires all publicly available information to be *instantly* reflected in the current market price. This form is based on competition among analysts who attempt to exploit new information generated from market actions. If this competition is perfect and fair, no analyst can outperform the market.

*Strong Efficiency.* Strong form efficiency states that the current market price reflects all available information, including private or *insider* information. This form of market efficiency is the most demanding because it dictates that profits cannot be achieved by insider information. If strong efficiency holds, then any attempt to outperform the market is futile. Given the restrictions and control over insider trading it is unlikely that such a

condition would hold true and in Fama et al. (1969) Fama does concede that there is evidence (though not a substantial amount) contradicting this form.

#### 2.1.6.1 Criticisms of the EMH

Although a cornerstone of Finance, the EMH is widely contested mainly due to its claims of the rational investor and the chart of any security exhibiting characteristics the EMH dismisses. For example: according to the EMH, if trends do not exist then how is it possible to distinguish between bull and bear markets? Leonard (2009) examines the S&P500 Index dating back to 1929, finding 23 separate bull and bear markets. The average bear market lasts 11 months, and the average bull market lasts 32 months. Ball & Brown (1968) and Chan *et al.* (1996) demonstrate that security prices not only react instantly to new fundamental information but continue to fluctuate thereafter. For example: regarding US markets, the American Non-Farm Payroll report is so significant that when published, volatile fluctuations occur instantly and set the *tone* of the market for days thereafter. The case for support of the EMH dismiss all efforts to forecast security prices, and even the EHM in its weak form stands in direct contradiction to TA.

There is general agreement that developed financial markets meet the conditions of semi-strong efficiency with the condition of strong efficiency not readily accepted especially considering robust anti-insider trading laws. While the EMH ultimately boils down to that of opinion, it remains one of the most robust and debated concepts in academic Finance and is still intensely researched to this day. The related literature is enormous, with notable papers being; LeRoy & Porter (1981), Malkiel (2003), Timmerman & Granger (2004), Yen & Lee (2008), and Guidi & Gupta (2013). The empirical results regarding the EHM are as equally controversial as the theory itself, especially during the 1980s and 1990s. However, recent studies do find empirical results reinforcing market efficiency. Despite the enormous literature and empirical testing, the existence of market efficiency is still heavily questioned. Observation alone reveals that trends exist and are *somewhat* forecastable, therefore strictly speaking the EMH is invalidated, Abu-Mostafa & Atiya, (1996).

The principle of *new information* is also highly questionable. For example, a well-documented characteristic of financial markets is the presence of asymmetric information in which one party has a competitive edge over the other party to the transaction. The

Enron scandal of 2000 is a prime example, where executives were concealing grossly manipulated performance figures from the public. Even when the news began to disseminate, Wall Street analysts continued to recommend buying the stock based on old projection figures; therefore the news did not hit the market instantly but was gradually disseminated. It is apparent that the process of disseminating new information is not instant and several problems exist in this respect. The first problem is *transmission*, i.e. the new information may be inaccurate. Second, the source may be dishonest. Third, the information may not be instantly disseminated even if it is time sensitive. Lastly, there is an inherent natural lag between when the news is announced and when the last recipient receives it. There is also the highly debatable concept of the *rational investor*. The science of Behavioural Finance researches investor behaviour and how investors interpret information. Results show irrational behaviour that would be undesirable in the marketplace such as comfort in crowds (Huberman & Regev 2001), overconfidence based on little information (Barber & Odean 2001), overreaction (DeBondt & Thaler 1985). As well as; psychological accounting, miscalibration of probabilities, hyperbolic discounting, and regret. An increasing amount of these studies are demonstrating that investors are not rational but act irrationally.

The EMH is often perceived to discouraging active participation in financial markets due to the martingale property of returns, whereby acting on new information does not ensure the additional benefit of abnormal returns. Therefore, a market participant should implement the B&H strategy, which in turn would cause liquidity to decrease and the markets would consequently come to a definite halt, Grossman (1976). However, this criticism is largely a misconstrued argument based on incorrect interpretation of the EMH. The EMH dictates convergence to equilibrium as a direct result of the competition between market participants. Therefore, it can be claimed that this line of criticism; “confuses a statement about an equilibrium ‘after the dust settles’ and the actions required to obtain that equilibrium”, Ball (2009, p10).

Additionally, it is argued the EMH should prevent, or at least predict market bubbles. This criticism is further fuelled by the ambiguity of the term *market bubble* defined as a prolonged and significant departure of prices from its fair value. However, identification of such circumstances is only possible post-event; therefore proponents of the EMH claim that until the market crashes, the bubble is a reflection of rationality, until the moment of sudden irrationality, which is followed by a rational correction. (Ball 2009).

Often cited is the issue of excess trading volume, volatility, and consistent violations of martingales and the RWH, Thaler (2010). The argument of excessive trading volume suggests there is too much trading and price changes to be justified solely based on variations of rational expectation in dividend, bond yields, or interest, Shiller (1981). The responses of EMH advocates are largely inconsistent and unconvincing, typically relying on micro and macroeconomic contradiction or exclaiming the impossibility to exploit these inefficiencies in a way contradictory to the hypothesis, Malkiel (2003).

### **2.1.7 Behavioural Finance**

Behavioural Finance (BF) is the main contender and natural successor of the EMH. BF takes direct aim at the bounded rationality of market participants, makes allowances for market inefficiencies and attempts to understand market behaviour from a psychological perspective. The first occurrence of this theory pre-dates the EMH by several decades when Selden (1912) wrote *The Psychology of the Stock Market*, in which he argues that changes in market prices are a result of mental attitudes and biases of market participants and not a reflection of the underlying value of some asset. Tversky & Kahneman (1974) identified several cognitive biases which guide irrational behaviour. The most significant biases, both theoretically and empirically are:

*Representativeness.* A tendency to fit new and unrelated events to already observed events, with the increased likelihood of identifying common features, despite the overall divergence of the two events.

*Availability.* A bias whereby the importance of data is derived based on its availability.

*Overconfidence.* A tendency to overstate one's skills, aptitudes and knowledge about the markets, resulting in irrational behaviour, DeBondt & Thaler (1994).

*Herding.* A tendency of individuals to follow the behaviour of others rather than basing their actions on their analysis.

*Adjustment and Anchoring.* A bias of initiating the decision-making process within an informational realm, or anchor, which essentially presupposes the desired conclusion. Similarly, adjustment refers to the amendment of the initial 'starting point' before decision making, to reach the desired conclusion, Sewell (2007).



*Overreaction & underreaction.* Empirical evidence shows that reaction of market participants to new information is not adequate and therefore does not correctly reflect in the prices on an asset, DeBondt & Thaler (1985).

*Gambler's fallacy.* The false belief that previous occurrence of an event reduces its probability of it occurring in the future. Consequently, a market participant is likely to assume that stock prices will decline because previously they were rising for a prolonged period, Kudryavtsev *et al.* (2013).

In addition to the above biases, Kahneman & Tversky (1979) developed a theoretical framework called Prospect Theory which according to its authors violates the underlying assumption of Expected Utility Theory contained within the EMH. The basic premise of Prospect Theory states that a person replaces true probability distributions with adjusted probabilities, where they will treat extremely improbable events as impossible and extremely probable as certain. Therefore, the subsequent decision is likely to be based on subjectively perceived gains. Kahneman & Tversky then derived behavioural fallacies from such framework, which govern market participants decision making:

*Loss Aversion.* Through this fallacy, market participants allocate a lower tolerance toward risk of losses than to an equal risk of reward, as there is an inner stigma associated with loss.

*Regret Aversion.* This fallacy arises as a result of the negative emotional response associated with the realisation of a bad decision, which is not limited to the material matters (financial losses) but includes the responsibility for a bad decision. Regret Aversion means that people are reluctant to conclude a losing market action – a trade – and keep the asset for irrational periods, Zeelenberg (1999).

*Mental Accounting.* The last fallacy refers to the notion of separating our activities into separate mental compartments by the nature of the activities and treating each compartment separately and independently of each other. For instance, Gultekin & Gultekin (1983) showed that investors tend to treat January as a distinct period independent from the previous months, which leads to the so-called 'January effect'. Several studies have appeared in the financial literature discussing this field; however, for brevity, only a few of the more influential papers will be discussed. In Kahneman *et al.* (1979) the authors present a critique of utility theory (a theory essential for informationally efficient markets) and offer an alternative description of how investors make decisions. They argue that Prospect Theory offers a more accurate account of investor behaviour,

where investors value gains and losses relative to a reference point and that they are risk-averse with the possibility of a certain gain and risk-seeking when faced with a possible loss. Prospect theory can explain a well-documented behavioural bias called the disposition effect, which is an individual tendency to sell when in a positive return investment and hold on to an investment that is in a negative position. Thus, the utility functions for the same investor depend on the state of the investment itself. Weber *et al.* (1998) confirmed that the disposition effect, sometimes referred to as the reference point effect, does exist. The authors also discovered that the investment behaviour of their subjects was inconsistent with portfolio theory, as proposed in Markowitz (1952). The results showed that the subjects tended to have a more diversified portfolio and traded more often than what would be optimal. The authors also found that the subjects anticipated mean reversion in the market where they expected losers to become winners and vice-versa. This behavioural trait of being loss-averse is also observed in Coval *et al.* (2005) where day traders on the Chicago Board of Trade regularly assumed above-average afternoon risk to recover morning losses, which affected short-term prices. Another well-known market phenomenon that has received attention from BF is apparent over and under reaction of investors to new information which is contrary to how efficient markets would operate. A model of investor sentiment is offered in Barberis (1998) as an explanation for this behaviour. In underreaction, the evidence shows that over horizons of less than a year stock prices underreact to news, which would mean that current good news would have predictive power of positive stock returns in the future. For overreaction, the evidence shows that in time-horizons of 3-5 years, stock prices will overact (become overvalued/undervalued) to a series of good/bad news and subsequently the returns will be lower / higher over the coming time horizon. The authors draw on two psychological traits known as representativeness and conservatism to explain this behaviour. Representativeness is the tendency for people to see a situation as typical or representative of some group of behaviour and ignore the laws of probability in the process. In the market, this would lead to overreaction for a company that has had considerable growth over the last few years. It would become overvalued as investors are too optimistic about its future growth even though it is more likely to slow down rather than keep expanding forever

The second psychological trait is *conservatism*, which is the tendency of people to be slow in updating their models in the face of new evidence. This trait leads to the underreaction to news and the slow integration of it into stock prices. Behavioural economics, however, suffer from a lack of a unified approach to market behaviour concerning asset pricing, only

explaining various individual aspects, such as various market inefficiencies, which on its own cannot fully replace an overarching theory such as the hypothesis. Conversely, behavioural economics could be used to reconcile or bridge the hypothesis with the existing empirical challenges; however if a generalised framework for the interaction and inter-operation of those approaches is feasible, it is yet to be seen. In his review of the empirical evidence and literature, Fama (1998) dismissed most of the evidence as non-systematic and short-term only, hence rejecting the idea of behavioural economics providing a better understanding of the above-described market behaviours. Fama further stresses that the anomalies such as overreaction and underreaction have been shown to occur inconsistently, or by chance, and behavioural economics thus do not provide any new insight. Regardless of the criticism, behavioural economics still provide a superior approach to issues of bounded rationality and were the basis for the Adaptive Market Hypothesis (AMH) to be developed.

### **2.1.8 Adaptive Market Hypothesis**

In disagreement with the EMH, Campbell *et al.* (1997) contend it is likely that inefficiency occurs in financial markets based on different market conditions, and according to Lim & Brook (2011) market efficiency varies over time. Furthermore, perfectly functioning financial markets are impossible to achieve, with Grossman (1976) and Grossman & Stiglitz (1980) debating that markets cannot be perfectly efficient because if prices reflect all available information, market participants would have no incentive to obtain costly information. Additionally, despite its popularity in academia the EMH is not widely held in the Finance industry as it seems illogical that the time and resources allocated by financial institutions in modelling financial markets would be an exercise in futility. Many academics believe that markets need to be somewhat inefficient to be liquid and exist at all, Grossman (2012). Given that perfect efficiency appears to be impossible to achieve, especially since Campbell, Lo & MacKinlay (1996) proposed the concept of relative efficiency, research shifted from testing absolute market efficiency to measuring the degree of market efficiency.

With the questionable sociological framework of the EMH, Farmer & Lo (1999) and Farmer (2002), established a new theory incorporating biological evolutionary principles to account for complex market dynamics such as; cycles, trends, bubbles, crashes, and

manias, resulting in Lo (2004) devising the Adaptive Market Hypothesis (AMH). The most significant insights from this work is that market efficiency is not static but somewhat variable and cyclical, which allows for both the EMH and Behavioural Finance to coexist. Therefore, the AMH does not replace the EMH but instead regards it as a *frictionless ideal* and essentially merges the EMH with principles of Behavioural Finance. The AMH framework is based on established principles of evolutionary biology; competition, mutation, reproduction, and natural selection. The principles of the AMH are:

1. Individuals act on their self-interest.
2. Individuals make mistakes.
3. Individuals learn and adapt.
4. Competition drives adaptation and innovation.
5. Natural selection shapes market ecology.
6. Evolution determines market dynamics.

The EMH and AMH both agree with principle 1, but the two theories diverge in principles 2 and 3. Under the EMH, investors do not make mistakes, nor is there any learning and adaptation because the market environment is stationary and in a constant state of equilibrium. Conversely, under the AMH mistakes occur frequently, market participants are capable of learning from their mistakes and adapt their behaviour accordingly. Principle 4 states that adaptation does not occur independently of market forces but is driven by competition. The interactions among various market participants are governed by natural selection or *survival of the richest*, and principle 5 implies that the current market environment is a product of this selection process. Principle 6 states that the sum of the components of; selfish individuals, competition, adaptation, natural selection, and environmental conditions, is what is observed as market dynamics.

Lo (2004) argues that individuals learn from previous experiences in a heuristic approach based on prior negative or positive reinforcements, and that behaviours evolved through this approach may be optimal or suboptimal if the environment changes, in which behavioural biases occur because a previously relatively optimal behaviour is now suboptimal in the new environment, or maladaptive. Furthermore, market efficiency is defined by the size of the ecology (the market) and the species in the ecology (market participants). Although the AMH has an abstract and qualitative nature, it offers practical implications for portfolio management. The equity risk premium is connected to the risk-

reward relationship. The size of the market defines this relationship, and the preferences of market participants vary over time. Under the AMH, the aggregated preferences of market participants are constituted by the individual preferences of market participants. These preferences are shaped by natural selection. Due to the evolutionary aspect of the AMH, these preferences therefore change, and so does the risk-reward relationship and thus the risk premium profile. Thus, history influences current risk profiles based on the evolution of market participants. Given that some participants exit the market due to losses, new participants enter with different risk profiles and therefore stock prices can go through cycles of predictable patterns and unpredictable patterns, Lo (2004).

The same evolutionary framework to the exit and emergence of new species (market participants) in financial markets explains how arbitrage opportunities arise from time to time. Market conditions combined with the exit and entering of new species results in complex market dynamics where arbitrage opportunities emerge and become exploited before disappearing. In contrast to the EMH, the AMH emphasises these dynamics as evidence for periods of stock return predictability. This view is shared by Kent & Timan (1999) who highlight the possible co-existence of Behavioural Finance and the EMH by introducing the term *adaptive efficiency*. In adaptively efficient markets, profit opportunities do arise in historical data, but if investors learn from the past price history, these profit opportunities will gradually erode through time (Lo, 2005).

According to the AMH, investments undergo cycles of inferior and superior performance in response to changing economic conditions, the adaptability of investors, the number of competitors in the industry and the size of profit opportunities available. The evolution of markets and financial technologies and *survival of the richest* is eventually the only aspect that matters, Lo (2005, 2012). Therefore, market efficiency is expected to adapt over a long-term period in cyclical patterns due to changes in macro institutions, market regulations and information technologies (Kim, Shamsuddin and Lim, 2011). Therefore, the aspect of cyclical nature in regard to market efficiency contrasts with the EMH, which assumes market efficiency based on convergence in information symmetry reflected in asset prices, Butler & Kazakow (2012). The implications for empirically testing the AMH is two-fold; 1) market efficiency fluctuates over time, and 2) market efficiency is governed by market conditions, Kim *et al.* (2011).

The backbone of the AMH is the cyclical nature of market efficiency and the variable nature of market efficiency between mature and emerging markets, Butler & Kazakov (2012). The cyclical inefficiency is demonstrated by Lim (2008) who used the Portmanteau bicomrelation test (H-statistic) which tests for non-linear dependence in a time series. The implication being that the existence of non-linear dependence in a financial time series invalidates the EMH. Lim applied the H-statistic iteratively using a sliding window, with the results revealing that episodic non-linear dependence existed in all the financial markets considered. The results also indicate that emerging markets were less efficient than developed markets, but both exhibited a cyclical nature to market inefficiency. Ito *et al.* (2009) measure time-varying market efficiency based on the estimated autocorrelation of US stock returns in which the authors fit an AR model to returns data using the Kalman filter. The results found that time-varying market inefficiency exists and that the most inefficient period was in the 1980s. Zunino *et al.* (2009) examine market inefficiency based on forbidden patterns and permutation entropy. Their results showed that inefficiencies exist in markets around the world and that market efficiency is also time varying. Furthermore, Zunino *et al.* (2010) expand on their 2009 research but instead use the complexity-entropy causality plane. Following their previous studies, the results demonstrate that time-varying market inefficiencies can characterise the markets under study. Kim *et al.* (2011) evaluate the predictability of returns and their cyclical nature on the DJIA index over the sample period 1900 to 2009. They assess the adaptiveness of the DJIA using three methods; the variance ratio (VR), the portmantua test (H-statistic) to evaluate linear independence over time, and a generalised spectral analysis to test for nonlinear independence over time. They find that the predictability of returns varies over time and is dependent on market conditions. Furthermore, Kim *et al.* (2011) evaluate how predictability depends on market conditions by using dummy variables for economic bubbles, stock market crashes and economic or political crises. They find that the predictability of returns is governed by changing market conditions. No return predictability is found during market crashes while a high degree of predictability is found during economic and political crises. Charles, Darne & Kim (2012) examine the AMH on major foreign exchange rates using the VR test. In parity with the AMH, they find that foreign exchange rate predictability goes through periods of predictability and non-predictability. Furthermore, they evaluate different market conditions and conclude that significant events such as coordinated central bank interventions and a financial crisis coincide when foreign exchange rates show patterns of predictability. Urquhart & Hudson (2013) evaluate the adaptiveness of market efficiency and time-varying predictability of

three market indexes; DJIA, FT30, and TOPIX over 5-year subsamples. The varying predictability and dynamic efficiency is based on tests for independence and efficiency and is evaluated based on the criteria that the markets demonstrate efficiency or move towards efficiency if there is a reduction in historical dependency. A market is defined as adaptive when it has moved through three stages of dependence and defined as inefficient when there is no independence in returns throughout the sample. The VR test and the runs test are used to check for linear independence and McLeod Li test, Engle LM test, and BDS test check for nonlinear independence. The findings demonstrate that all three stock indices show market adaptiveness per the AMH. Zhou & Lee (2013) examine the US REIT market and find that certain market conditions have an impact on return predictability. Opposed to the findings of Urquhart & McGoarty (2016) they find that periods of high volatility are closely related to periods of return predictability. Urquhart & McGoarty (2016) link certain market conditions to returns predictability. Evaluation is based on market conditions of bullish, bearish and normal market returns. They examine the AMH on four stock indices; S&P500, FTSE100, Nikkei 225, and EURO STOXX 50 between 1990 and 2014, using fixed length moving subsamples, finding that returns predictability changes over time for each market, in line with the AMH. Furthermore, they find that different markets exhibit different predictability patterns based on various market conditions. For example the S&P 500 exhibits patterns of return predictability during bear markets and a low level of return predictability during bull markets. The EURO STOXX 50 exhibits the opposite patterns with a notably high level of return predictability during bull markets. The main contribution by Urquhart & McGoarty (2016) is the relationship between returns predictability, and market conditions are unique to each market and should be evaluated separately.

## **2.2 Trading Strategies**

This section presents an overview of the trading strategies offered by TA, hereafter referred to as Technical Trading Strategies (TTS). A complete discussion of the universe of TTS is beyond the scope of this thesis. However, the main categories of TTS (Moving Averages and Oscillators) will be discussed. To start, although not a TTS, it is necessary to discuss the strategy advocated by the EHM, RWT and FA - the Buy & Hold (B&H) strategy, as it is this strategy TA and its TTS strive to outperform and is referred to as *beating the market*.

### **2.2.1 The Buy & Hold Strategy**

The B&H strategy is a passive and simplistic investment strategy requiring a market participant to simply purchase the security and hold it long-term, with disregard for short to medium term counter-price fluctuations. The B&H strategy has the advantage of being a simplistic investment strategy capable of execution by any market participant. Furthermore, empirical evidence demonstrates its effectiveness and, at times, difficult to outperform by alternative methods. However, there exist bases for criticism. For example: The Great Depression of the 1930s severely questioned the B&H philosophy and proposed the need for active portfolio management. Furthermore, the B&H strategy is not without defeat: Levis *et al.* (1999), Fernández-Rodríguez *et al.* (2000), O'Neil (2001), Barber *et al.* (2006), and Szafarz (2012) perform comparisons of TTS benchmarked against the B&H strategy. Their empirical results demonstrate that in most cases the TTS outperform the B&H strategy.

### **2.2.2 Technical Trading Strategies**

TA has two main categories of TTS, charting and technical indicator-based strategies. Charting is viewed as subjective - even by Technical Analysts. For example: Technical Analysts often disagree over a certain chart pattern as being a triangle or a wedge. Technical indicators, on the other hand, provide predefined buy and sell signals derived from mathematical formulas. Therefore, technical indicators hold the advantages of eliminating the subjectivity of charting, suppressing the user's emotions during the decision-making process and elevating objectivity. Common technical indicators, their strategies and respective literature reviews will now be discussed.

#### **2.2.2.1 Moving Averages**

A Moving Average (MA) is a basic technical indicator, ubiquitous to TA, and according to Murphy (1999) is one of the most versatile and widely used technical indicators. From a trading perspective, an MA is employed to identify and track trends. When the price is trading above the MA, the market is in an uptrend, and when the price is trading below the MA, the market is in a downtrend. Trading signals are generated when price crosses above



or below the MA, with the price expected to close accordingly for such signals to be valid. For example: if price crosses and closes above the MA from below, a buy signal is generated. If price crosses and closes below the MA from above, a sell signal is generated. Following the development of a trading signal, the role of an MA switches to tracking the trend until a subsequent trading signal occurs. Pring likens this secondary function of an MA to a fluid version of another technical method known as support and resistance, describing a MAs secondary function as: ‘*dynamic levels of support and resistance*’, Pring (2014, p209). Therefore, when a market is trading above the MA, the MA serves as dynamic support. Conversely, when a market is trading below the MA, the MA serves as dynamic resistance. The following diagram features a screenshot of the EUR/GBP daily time series and the application of a 20-day Simple Moving Average (SMA):

Figure 2-1: EUR/GBP Daily Chart, November 2016 to May 2017, SMA Application.



From figure 2-1, it is observed that the SMA tracks the general advances and declines of the EUR/GBP time series. The SMA *smooths* the time series data via averaging, therefore, filtering market noise and clearly displaying the advances and declines irrespective of the day-to-day fluctuations and volatility. Regarding trading signals, at point A, price crosses and closes above the SMA from below, with the SMA generating a buy signal indicating an uptrend. As signalled, the uptrend ensues, with a duration of twenty days. The EUR/GBP then declines with price crossing and closing below the SMA from above at point B, generating a sell signal. Subsequent buy and sell signals occur at points; C, D and E. Pring’s description of MAs serving as dynamic levels of support and resistance is observable at point F where the EUR/GBP declines towards the MA but uses it as support to rebound and continue its advance.

As elegant as the SMA initially appears, closer inspection reveals problems in its application. For example, the SMA generated a sell signal at point B. However, the EUR/GBP had already been declining for six days prior, with the SMA generating its valid sell signal on day seven. The delay between the time series changing direction and the SMA generating a signal is known as lag. Lag is regarded as problematic and is inherent in the calculation of most technical indicators, especially MAs. Lag is regarded as one of the biggest problems for a trader and is equal to  $(N-1)/2$  of the MA. For example, a 21-day SMA has a lag of 10 days. Fortunately, TA has attracted the attention of Engineers, with some motivated to minimising or eliminating lag. John F. Ehlers is the pioneer in the search to solve the problem of lag and used his degrees in Electrical Engineering and his PhD in Filtered, Waves and Information Theory to solve the problem of lag by designing *zero-lag* technical indicators. Others have since followed Ehlers, and zero-lag technical indicators are now featured on premium trading platforms or as third-party add-ons. For example, this thesis features zero-lag technical indicators by Bowfrot Technologies.

Regarding MAs, the Hull MA (HMA) and Gaussian MA (GMA) both feature extremely low values of lag. Furthermore, advanced zero lag technical indicators feature *filter poles* that allow the user to exert control over the lag, adjusting it to their preference. Further inspection of figure 2-1 reveals another problem. At points 1 and 2, price crosses and closes above the SMA from below therefore generating valid buy signals, only for the EUR/GBP to then decline for six and five days respectively. The same counter movement also occurs at points 3 and 4. The counter movement is known as *whipsawing* and is an irritating problem amongst market participants because valid signals are generated, yet the time series quickly moves in the opposite direction. The problems of lag and whipsawing are also interconnected. In solution to whipsawing a longer length MA can be applied; however, a longer length MA increases lag. In solution to lag, a shorter length MA can be employed; however, this increases whipsawing. The solution is to find an optimal balance between the two problems and identify an MA length that generates timely trading signals while minimising the problems of lag and whipsawing. Therefore, an MA although versatile, simple to calculate and implement, begs critical questions such as: what length of MA to use? What type of MA to employ? Is it better to use a single MA or a combination of MAs? Which type of MA produces the best results? When does an MA perform best? When does an MA perform poorly? The last two questions are arguably of greatest significance as they are performance related. An MA performs best in trending market conditions and performs poorly in non-trending market conditions, i.e. in a consolidation zone in which oscillators are better suited. Figure 2-1 is illustrative of a non-trending,

range bound, market and the associated problems of MA application under such circumstances. There are numerous types of MA: triangular, weighted, linear weighted, variable length, adaptive, double exponential, triple exponential, hull, and gaussian. The three common types of MA are; simple, exponential, and weighted. If the length is  $n$  days, the current period  $t$ , and closing price  $P_t$ , they are defined as follows;

$$\text{Simple MA: } SMA_{t+1} = (1/n) (P_t + P_{t-1} + \dots P_{t-n+1}) \quad (3)$$

$$\text{Exponential MA: } EMA_{t+1} = EMA_t + \alpha(P_t - EMA_t) \quad (4)$$

$$\text{Weighted MA: } WMA_{t+1} = [nP_t + (n-1)P_{t-1} + \dots + 2P_{t-n+2} + P_{t-n+1}] / [n(n-1)/2] \quad (5)$$

According to Pring, the SMA is an average of values recalculated each period, usually daily, and is the most commonly used MA (2014, p209). An EMA adapts to changes in market prices via the smoothing parameter  $\alpha$ , where  $\alpha$  is a number bounded between 0 and 1. The smoothing parameter expresses the speed at which the EMA reacts to price changes. If  $\alpha$  is set towards 1, the EMA increases in sensitivity, and there is a quick reaction to price changes and vice versa. According to Pring, EMAs are better at identifying trend reversals and like SMAs signals are generated when price crosses and closes above or below the EMA. The WMA assigns weights to the prices used as lags. The weights are greater in recent periods, therefore, place more importance on recent prices. Due to the sensitivity of WMAs, they generate trading signals when the WMA changes direction as opposed to price crossing the WMA. Figure 2-2 expands on figure 2-1 by featuring an EMA and WMA, both calculated over 20 periods:

Figure 2-2: EUR/GBP Daily Chart, November 2016 to May 2017, SMA, EMA, WMA Examples.



Observation shows that both the EMA (red) and WMA (grey) are more responsive than the SMA (blue) as they change direction before the SMA due to featuring less lag. Furthermore, the EMA and WMA better track the time series as it progresses. However, as previously stated, the cost of reduced lag is increased whipsawing.

#### **2.2.2.2 Moving Average Literature Review**

Due to their popularity as confirmed by Cesari & Cremonini (2003) combined with their ease of empirical testing, MAs are extensively researched in the literature. Notable studies are Brock *et al.* (1992) and Hudson *et al.* (1996) who studied the utility of MAs on the DJIA and Financial Times Industrial Ordinary Index respectively. They conclude that MAs have predictive ability if a sufficient sample period is employed. Concerning the Brock *et al.* study, it transpired the optimal length MA was a 50-day MA, which generated an annual average return of 9.4%. LeBaron (1999) concludes that a 150-day MA generates Sharpe ratios of 0.60–0.98 after transaction costs in DEM and JPY markets from 1979 to 1992. LeBaron & Blake (2000) research MAs applied to the DJIA over a 100-year sample period (1987-1999). They found that the differences between conditional means during buy and sell periods changed dramatically over the last ten years of the sample period, but differences in conditional variances did not change much throughout the sample period. Robustness checks indicate that similar results could be obtained with simple momentum-based strategies. Gunasekarage & Power (2001) investigate the utility of the dual MA crossover strategy using a variable length MA and fixed length MA applied to forecasting four emerging Asian stock markets. Their findings indicate that MAs have predictive ability in those markets and returns exceeded those generated by the B&H strategy and that the variable length MA performed especially well.

In the interest of debate, numerous empirical studies dismiss the utility of MAs as prominent forecasting tools. Fong & Yong (2005) evaluate the fluctuations of internet stocks which led to the .com crash, investigate whether investors could have exploited the momentum in Internet stocks using MAs. They employed a recursive trading strategy applied to over 800 MAs. Their empirical results found no significant trading profits and align the internet stocks with the EMH. They also suggested that high volatility inflicted poor performance. Chiarella *et al.* (2006) studied the effect of long-term MAs on market

dynamics. When examining the disparity between FA and TA as being unbalanced, they present evidence that increasing the length of the MAs destabilised market prices. Zhu & Zhou (2009) evaluate the utility of MAs from an asset selection perspective, demonstrating that when stock returns are predictable, MAs add value to commonly used allocation rules. They combine MAs with fixed rules in order to identify market timing strategies that shift money between cash and risky assets. Their approach outperforms the simple rules and explains why both risk aversion and degree of predictability affect the optimal use of the MA. Milionis & Papanagiotou (2011) apply an alternative testing procedure to the significance of the predictive power of the MAs applied to the New York Stock Exchange (NYSE), the Athens Stock Exchange (ASE) and the Vienna Stock Exchange (VSE), over the sample period (1993-2005). Their alternative methodology considers the variability in the performance of the MA trading rule by considering cumulative joint returns of numerous MAs lengths. Their findings demonstrate that MA performance is a function of their length and their application outperforms the B&H strategy and rejects weak form market efficiency. Bajgrowicz & Scaillet (2011) revisit the historical success of TA applied to daily prices of the DJIA from 1897 to 2011, using the false discovery rate (FDR) for data snooping. Their persistence tests show that even with the more powerful FDR technique, an investor would not have been able to select ex ante the future best-performing rules. Furthermore, even in-sample, the performance is completely offset by the inclusion of favourable transaction costs. They conclude by calling into question the economic worth of MA strategies reported for earlier periods. For MA's integrated with CI is Gençay (1998 & 1999) who investigated the non-linear predictability of FX and index returns by combining ANNs and MAs. The results indicate that MA signals have market timing ability and provide statistically significant forecast improvements over the RW model. Psaradellis *et al.* (2018) investigate the performance of MAs on the crude oil market (WTI) in response to the debatable study of Sullivan, Timmermann, & White (1999). They use a sample period of 2006 to 2017, using the k-familywise error rate (k-FWER) and false discovery rate (FDR) to account for data snooping. Although in-sample performance was outstanding, out-of-sample performance found no persistent nature in performance, although they did note that small profits can be achieved in some periods.

### 2.2.2.3 Oscillators

Oscillators are technical indicators that generally measure momentum, although they do measure other characteristics. By definition, an oscillator is a technical indicator that measures momentum to identify market conditions in which the market has reached, or is reaching, the extremity of its current trend and is, therefore, due for an imminent retracement or reversal. Typically, oscillators are represented by a line oscillating above and below a zero value within a range bounded between +1 and -1, +100 and -100 or 0 and 100. According to Pring, the standard interpretation of momentum is through overbought and oversold levels (2014, p253). There are too many oscillators to discuss individually for the scope of this thesis; however, a detailed discussion of oscillators can be found in Pring's; *Definitive Guide to Momentum Indicators*. A prominent oscillator worth mentioning which utilises overbought and oversold levels, as well as convergence and divergence, is the Relative Strength Index (RSI). Developed by Wells Wilder in 1978, the RSI measures the relative internal strength of the market against itself and is defined as;

$$RSI_t = \frac{U_t}{U_t + D_t} \quad (6)$$

Where  $U_t$  denotes the cumulative advance i.e. the close-to-close increase on a day where the security under observation has closed higher than the previous day's closing price over a certain period, and  $D_t$  denotes the cumulative decline i.e. the close-to-close decrease on a day where the security under observation has closed lower than the previous day's closing price over the sample period  $x$  which by default is 14 trading days as prescribed by Wilder.

$$U_t = \sum_{i=1}^m l(S_{t-1} - S_{t-1-i} > 0)(S_{t-1} - S_{t-1-i}) \quad (7)$$

$$D_t = \sum_{i=1}^m l(S_{t-1} - S_{t-1-i} < 0)|S_{t-1} - S_{t-1-i}| \quad (8)$$

Decreasing the value of  $x$  results in increased sensitivity whereas increasing the value of  $x$  decreases sensitivity. The RSI is then normalised and presented on a vertical scale bound between 0 and 100 with values above 70 signalling overbought conditions and values below 30 signalling oversold conditions. A mid-range value of 50 signals equilibrium and

acts as a boundary between bullish and bearish market conditions, the crossing of which generates a respective minor buy or sell signal. However, comforting as these set levels are, in practice, securities do not instantly nor persistently reverse their direction when the RSI reaches the 70/30 levels. For example, Peter W. Aan in his article; *How RSI Behaves*, argues that the average value of an RSI top or bottom is closer to the 72/32 levels. Bull and bear markets also change the interpretation of the RSI. For example, bull markets are said to reverse closer to the 80 level with their intermediate declines reversing at the 40 level. Whereas bear markets terminate closer to the 20 level with their intermediate advances terminating at the 60 level. For a detailed discussion including examples of these levels, please see Brown's; *Technical Analysis for the Trading Professional*, Chapter 1, Oscillators Do Not Travel Between 0 and 100, p2-18. Another function of the RSI lies in identifying convergence and divergence between the RSI and the market. Convergence is observed when the RSI makes a higher low in relation to its previous low with the security under observation making a lower low in relation to its previous low. Divergence is observed when the RSI makes a lower high in relation to its previous high and the security under observation makes a higher high in relation to its previous high. Convergence is interpreted as additional confirmation of an impending advance whereas divergence is interpreted as additional confirmation of an impending decline. Figure 2-3 depicts the application of the RSI on the EUR/GBP:

Figure 2-3: EUR/GBP, Daily Chart, Relative Strength Index (RSI), Convergence Highlighted.



From figure 2-3, the RSI appears below the chart with its 70/30 levels represented by the dashed lines. The market top identified at point A recorded an RSI reading of 74; meanwhile, the market bottoms at points B and C recorded RSI readings of 27 and 36 respectively. Convergence is highlighted between the downward sloping line on the chart

and the corresponding upward sloping line on the RSI. As expected, the convergence correctly forecasts a significant advance.

As chapter 5 *Volatility Modelling and Forecasting, A Nikkei 225 Index Example* features a modified version of the RSI (RSIV) tailored to measure volatility, a description of its specification will now be presented. Expanding on Wells Wilder's RSI is the RSIV developed by Kirk Northington CMT in his book *Volatility-Based Technical Analysis*. Northington's RSIV differs from the original RSI in that it is specifically designed to measure volatility and adapted to account for changes in the linear regression of price. Therefore, volatility is calculated as a change of the closing price's distance from a linear regression line instead of the change in closing price between two consecutive periods. The second major difference is that  $U$  and  $D$  are calculated differently. The RSI calculates an EMA of the increases and decreases for  $U$  and  $D$  respectively, whereas the RSIV calculates a simple product of the two using multiplication. The process of multiplication emphasises larger movements more distinctly, therefore magnifying volatility. Consequently, the overbought and oversold threshold values are altered from 70/30 to 75/25. A 40-period linear regression and a 30-period historical window are the advised parameters. Chapter 4 *Artificial Intelligence in Trading, A EUR/USD Example* also features an adaptation of the RSI, featuring its original 14,70/30 parameters but smoothed with a 7-period SMA.

#### **2.2.2.4 Oscillators Literature Review**

The renowned Brock *et al.* (1992) study presents evidence on the profitability of the MACD and other technical indicators such as MAs. For oscillators combined with CI, Kim & Han (2000) propose a hybrid ANN model featuring a genetic algorithm (GA) to predict the Korean stock market (KOSPI) over a 10-year sample period (1989-1998). Their experimental results show the GA approach to the feature discretisation model outperforms the benchmarks in terms of accuracy by approximately 11%. Shen & Loh (2004) propose a trading system with rough sets to forecast the S&P 500 Index. Their candidate variable pool included the MACD, RSI and Stochastic Oscillator, with their resultant model outperforming the B&H strategy. Tanaka-Yamawaki & Tokuoka (2007) report that oscillators such as the MACD and RSI are not effective in forecasting various intra-day US stock prices. Chong & Ng (2008) examine the profitability of the MACD and RSI applied to the FT30 Index over a 60-year sample period (1935-1994). They find that the MACD



and RSI outperformed the B&H for most of the sample period up until 1980, after which the TF30 advanced considerably and B&H strategy triumphed. Dunis *et al.* (2011) use ANNs to forecast the EUR/USD exchange rate over an 8-year sample period (October 1994 to July 2001). The performance of their ANNs is benchmarked against a MACD model with their results finding the ANNs to be superior forecasting models. Sermpinis *et al.* (2013) use ANNs to forecast the EUR/USD, EUR/GBP, and EUR/JPY exchange rates over a 13-year sample period (1999-2011), using the MACD as a benchmark. Their featured ANN, an Adaptive Radial Basis Function with Particle Swarm Optimisation (ARBF-PSO) outperforms all other models in terms of statistical accuracy and trading efficiency for all three exchange rates. Asness *et al.* (2013), evaluate momentum-based trading strategies that generate positive risk-adjusted performance and examine the underlying factors of the results. Chong, Ng & Liew (2014) reassess the profitability of the MACD and RSI on the stock markets of five OECD countries over a 26-year sample period of daily data (1976 to 2002). They found that the MACD (12,26,0) and RSI (21,50) rules consistently generate abnormal returns in the Milan Comit General and the S&P/TSX Composite Index. The authors suggest this is because the Italian stock market is less developed compared to the stock markets of other major OECD countries and is therefore relatively inefficient. They also concluded the RSI with its default parameters of (14,30/70) is profitable on the DJIA. Stanković *et al.* (2015) evaluate the profitability of the MACD and RSI in the emerging markets of Croatia, Serbia, Romania, and Bulgaria over the sample period (2009-2015). They authors note that the MACD and RSI outperformed the B&H strategy in all markets; however, once transaction costs were included the results turned negative. Furthermore, the RSI performed poorly in all markets – the authors note that RSI trading strategies do not perform well in non-trending markets as previously concluded by Wong *et al.* (2010). The authors conclude by using the most profitable MACD and RSI strategies as inputs to a Least-Squares SVM (LS-SVM). They find that despite more frequent trading, the LS-SVM outperformed the best performing technical model by 44.23%. Macedo *et al.* (2017) compare the performance of two established multi-objective evolutionary algorithms, NSGA II and SPEA 2, combined with technical indicators for mean semi-variance portfolio optimisation. Four different strategies are used, namely; B&H, Bollinger Bands, MACD and RSI, all of them incorporating trading costs, for four different markets over the sample period (2000-2015). The purpose was twofold: to assess which of the algorithms performed better and to know if the chosen TA indicators influenced the results. NSGA II systematically outperforms SPEA 2. Results also show how the use of technical indicators influences the frontiers of non-dominated portfolios. In

most markets, Bollinger Bands present the best out-of-sample results. The MACD performed better than B&H in most markets, and the RSI displayed better performances in less developed markets over the B&H strategy.

#### 2.2.2.5 Technical Analysis & Computational Intelligence

The opening sentence in Brown's *Technical Analysis for the Trading Professional* asks why conventional technical indicators are failing. Brown states that market participants, including institutional level traders, are puzzled by the phenomenon, proposing the scenarios; have the indicators failed or have financial markets changed rendering old methods obsolete? Brown attributes the cause to technology, claiming it has changed the pace of TA and has, through the introduction of online trading platforms, allowed the small market participant to be equipped with the same analytic tools once confined to institutions; *'All online platforms use the same default settings and the less experienced trader rarely changes these settings. The effect is a mass group of amateur traders all acting on the same signals, therefore, all operating as one large institutional wildcard. The professional who fails to move past this group is, therefore, operating within this new technically armed and dangerous mass'*, (Brown, p4). Brown's view is reinforced by the market declines of 1987 and more recently 2011 with many commentators blaming algorithmic and high-frequency trading by mass systems encouraging the volatile declines by all acting on approximately the same technical signals triggering mass selling. Further to the above is the natural decline in the performance of a trading strategy once it has been globally dispersed and implemented. Kirk Northington, CMT of Northington Trading describes the decline in performance as a form a technical indicator arbitrage. Lastly, is the prediction of David Aronson CMT, who foresees the demise of TA in a similar fashion to the demise of Astrology and the establishment of the science of Astronomy. Aronson categorises TA in two distinct forms, subjective TA, and as he terms it *evidence based technical analysis*. Aronson makes no apology for identifying subjective forms of TA such as Elliot Wave Theory (EWT), Charting, Gann, Fibonacci ratios, and emphasises the need for objective and testable TA methods. In his vision, Aronson asserts that the future of TA would best be served by the partnership between TA practitioners and computers: *'The proper role of the TA practitioner in the envisioned human-computer partnership is twofold: (1) proposing an information-rich set of candidate inputs suitable for presentation to the data mining software, (2) specifying the problem to be solved by the data mining*

*software...In other words, the TA expert's proper role is to supply what the computer lacks, namely, expertise in the domain of TA and inventive talent*", (Aronson, 2007, p464). Integrating TA with CI would provide immediate benefits for TA. For example it promotes the use of TA and quantitative models such as ANNs, therefore, producing objective and measurable results. It also keeps TA relevant and modern by keeping pace with modern methods. Aronson concludes, stating: *"The TA practitioner who understands data-mining methods and the important issues in indicator design will be well positioned to play this crucial role in the human-machine partnership of twenty-first century technical analysis...Given that human intelligence is essentially unchanging, but computer intelligence is increasing at an exponential rate, no other approach to TA makes sense"*, (Aronson, p.473). Given that traditional TA (charting) is subjective therefore untestable, conventional technical indicators are failing due to mass used default settings coupled with performance decline from arbitrage, and the fact that the modern-day trading arena has gravitated towards scientific and computational methods, the partnership between TA and CI is evitable. It is perhaps time for TA, like Astrology, to shed itself of its biased, subjective and untestable methods and fully embrace the scientific approach by merging its useful aspects (technical indicators) with AI and CI.

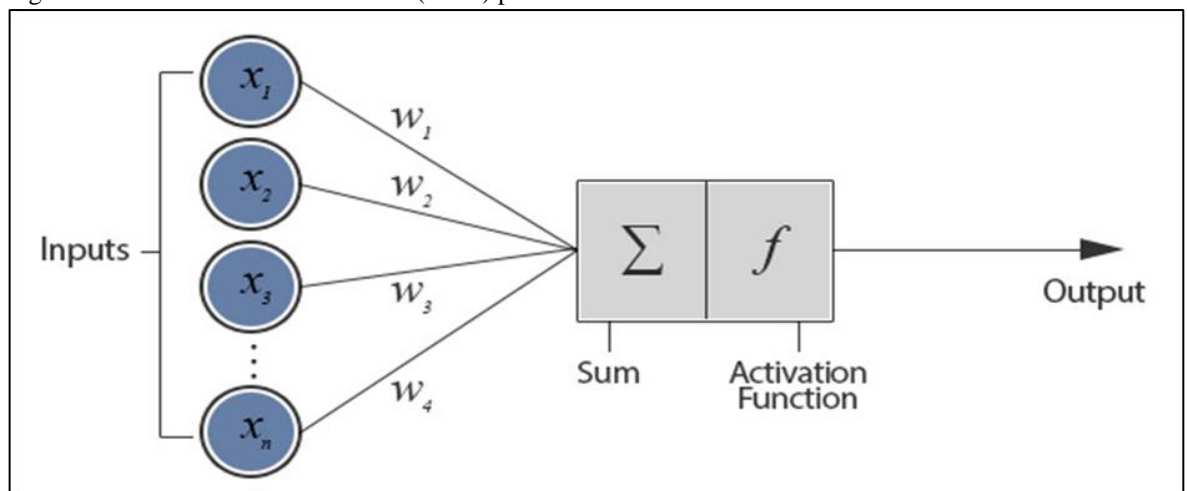
# Chapter 3

## Forecasting Techniques

### 3.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) are data analysis models that imitate the information processing capabilities of the human brain. An ANN relates a set of input variables  $\{X_i\}$ ,  $i = 1 \dots k$  to a set of one or more output variables  $\{Y_j\}$ ,  $j = 1 \dots k$ . The network architecture determines the relationship between  $X$  and  $Y$ , and it is the architecture that differentiates ANNs from other data analysis models. A standard ANN has at least three layers, see Figure 3-1:

Figure 3-1: Artificial Neural Network (ANN) process



Source: DataCamp

The first layer is called the input layer, in which the number of nodes corresponds to the number of input variables,  $X$ . The third layer is called the output layer, in which the number of nodes corresponds to the number of output variables. The second layer, better known as the hidden layer, is the key layer that separates the input layer from the output layer in which input variables are squashed or transformed by an activation function – usually a logistic or log sigmoid transformation. The number of nodes in the hidden layer also defines the complexity the ANN is capable of fitting. While the addition of this hidden

layer approach may seem complicated, it represents a very efficient and effective way to model a nonlinear statistical process. Regarding ANNs, there are many different architectures with unique processes and functions designed for certain tasks, such as forecasting and pattern recognition. These architectures will now be described.

### 3.1.1 Multilayer Perceptron (MLP)

The single hidden layer feed-forward network is the most basic of all ANNs and according to Kaastra & Boyd (1996), the most frequently used ANN in economic and financial applications. The following equation is used to describe a feed-forward ANN:

$$n_{k,t} = w_{k,o} + \sum_{i=1}^i w_{k,i} x_{i,t} \quad (9)$$

$$N_{k,t} = L(n_{k,t})$$

$$= \frac{1}{1 + e^{-n_{k,t}}}$$

$$y_t = \gamma_0 + \sum_{k=1}^k \gamma_k N_{k,t}$$

Where  $L(n_{k,t})$  represents a log-sigmoid activation function with the form:  $\frac{1}{1+e^{-n_{k,t}}}$ . In the system, there are  $i$  input variables  $\{x\}$  and  $k$  neurons. A linear combination of these input variables observed at time  $t$ ,  $\{x_{i,t}\}$ ,  $i = 1, \dots, i$ , with the coefficient vector or set of input weights  $w_{k,i}$ ,  $i = 1, \dots, i$ , as well as the constant term,  $w_{k,0}$ , from the variable  $n_{k,t}$ . This variable is squashed by the logistic function and becomes a neuron  $N_{k,t}$ , at time or observation  $t$ , the set of  $k$  neurons at time or observation index  $t$  are combined in a linear way with the coefficient vector  $\{\gamma_k\}$ ,  $k = 1, \dots, k$ , and taken with a constant term  $\gamma_0$ , to form the forecast  $y_t$  at time  $t$ . The feed-forward ANN coupled with the log sigmoid activation function is also known as the multilayer perceptron or MLP network. Figure 3-1 featured below generated by the ANN design software ‘Neuro Solutions Pro’ depicts a

feed-forward ANN with one hidden layer featuring the Levenberg Marquart (LM) learning algorithm and Tanh Axon activation function:

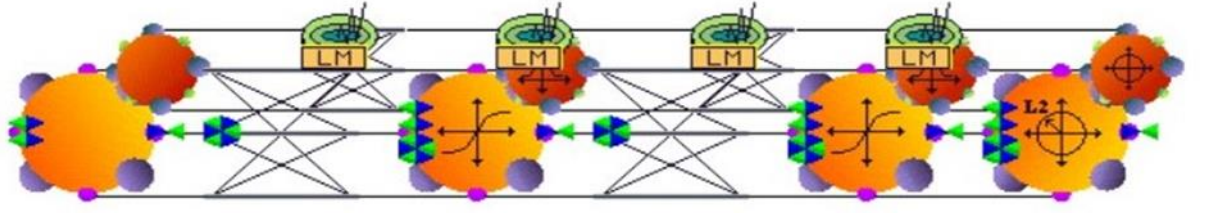


Figure 3-1: Feed Forward Artificial Neural Network (One Hidden Layer)

Thus far only ANN architectures with one hidden layer have been presented. It is, of course, possible to feature more than one hidden layer as found in multilayer feed-forward networks. For a problem displaying increased complexity, it can be approached by increasing the number of hidden layers to two or more. A feed-forward network comprising of two hidden layers can be represented by the following system with  $i^*$  inputs variables,  $k^*$  neurons in the first hidden layer and  $l^*$  neurons in the second hidden layer:

$$n_{k,t} = w_{k,0} + \sum_{i=1}^{i^*} w_{k,i} x_{i,t} \quad (10)$$

$$N_{k,t} = \frac{1}{1 + e^{-n_{k,t}}}$$

$$p_{l,t} = \rho_{l,0} + \sum_{k=1}^{k^*} \rho_{l,k} N_{k,t}$$

$$P_{l,t} = \frac{1}{1 + e^{p_{l,t}}}$$

$$y_t = \gamma_0 + \sum_{l=1}^{l^*} \gamma_l P_{l,t}$$

From the above, it is evident that an additional hidden layer increases the number of parameters to be estimated by the factor;  $(k^*+1)(l^*-1) + (i^*+1)$ , given a feed-forward network with one hidden layer with  $i^*$  inputs  $k^*$  neurons has;  $(i^*+1)k^* + (k^*+1)$  parameters whilst a similar network with two hidden layers with  $l^*$  neurons in the second hidden layer

has;  $(l^*+1)k^*+(k^*+1)l^*+(l^*+1)$  hidden layers. Feed-forward networks with multiple hidden layers increase model complexity and do so at the cost of additional parameters to estimate which in turn reduces degrees of freedom if the sample size is limited and increases training time. The additional parameters also introduce the likelihood that the parameter estimates may converge to a local rather than a global optimum. Figure 3-2 depicts an MLP with two hidden layers contrasted with figure 3-3 depicting an MLP with one hidden layer:

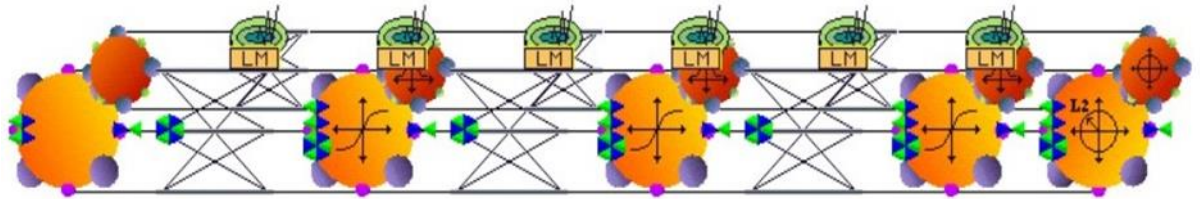


Figure 3-2: MLP Featuring Two Hidden Layers.

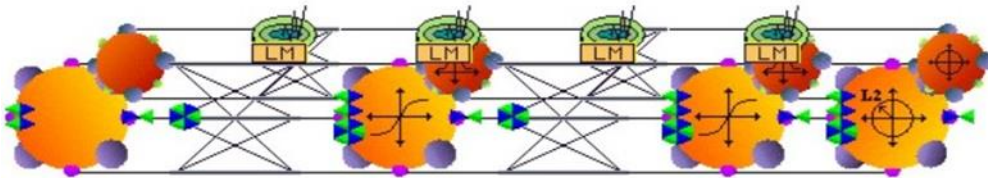


Figure 3-3: MLP Featuring One Hidden Layer.

### 3.1.2 Generalized Feedforward (GFF)

A generalised feed-forward (GFF) network is a variation of the basic feed-forward network in that it features jump connections. These jump connections allow for the inputs  $x$  to have direct linear links to the output  $y$  in addition to the output via the normal route of the hidden layer. Figure 3-4 depicts a GFF network featuring one hidden layer:

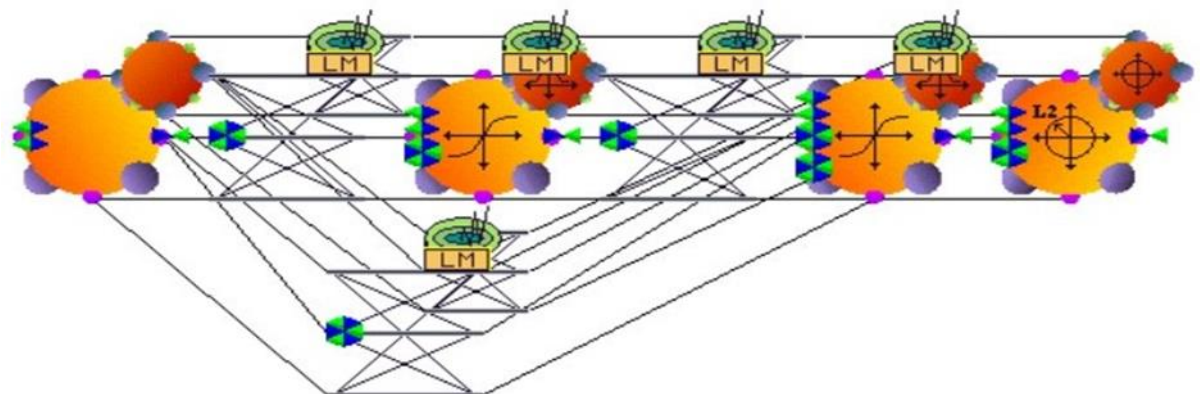


Figure 3-4: Generalized Feed-Forward Artificial Neural Network Featuring One Hidden Layer.

An advantage of the GFF network is that it embeds the pure linear model in addition to the feed-forward ANN, allowing for the possibility that a nonlinear function may also have a linear component. If the underlying relationship between the inputs and output is a pure linear relationship, then only the direct jump connectors will be significant within the architecture. However, if the underlying relationship between the inputs and output is a complex nonlinear relationship, then the jump connectors serve an insignificant role. The GFF network also allows for the relationship between input variables and the output to be decomposed into linear and nonlinear components. At a practical level, a good use for a GFF network is to test for neglected nonlinearities in the relationship between the input variables and the output. In relation to a basic feed-forward network, an MLP (in theory) can solve any problem that a GFF network can solve. In practice, however, GFF networks often solve the problem more efficiently. A classic example of this is the two-spiral problem in which a standard MLP requires hundreds of times more training epochs than a GFF network containing the same number of processing elements.

### 3.1.3 Modular Feed-Forward Networks

A Modular feed-forward (MFF) network is a special class of MLP. These networks process their inputs using several parallel MLPs then recombine the results. Figure 3-5 depicts an MFF network:

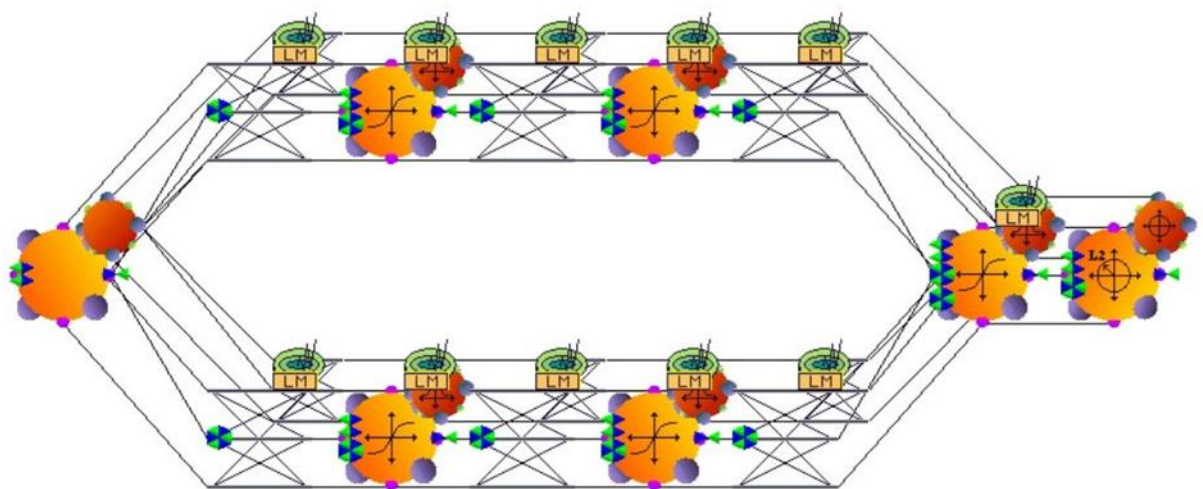


Figure 3-5: Modular Artificial Neural Network generated by Neuro Solutions Pro.

The parallel processing creates additional structure within the architecture which fosters specialisation of function in each sub-module. In contrast to a standard MLP, MFF



networks do not have full interconnectivity between their layers. Therefore, a smaller number of weights are required for the same size of network. This structure tends to decrease training time and reduce the number of required training exemplars.

### 3.1.4 Recurrent Neural Networks (RNNs)

A Recurrent Neural Network (RNN) is another commonly used architecture. An RNN allows the neurons to depend on the input variables  $x$  and their own lagged values, therefore creating a short-term memory. This function is similar to an MA process in which the dependent variable  $y$  is a function of observed inputs  $x$  as well as the current and lagged value of an unobserved disturbance term or random shock  $e$ . Similarly, RNNs utilise current and lagged unobserved unsquashed neurons in the hidden layer. In the estimation of an RNN, it is essential to use a multistep estimation procedure which starts by initialising the vector of lagged neurons with lagged neuron proxies from a simple feed-forward network. The coefficients are then calculated and used to recalculate the vector of lagged neurons. Parameter values are recalculated recursively, and the process continues until convergence occurs. The following represents an RNN:

$$n_{k,t} = w_{k,0} + w_{k,0} + \sum_{i=1}^i w_{k,i} x_{i,t} + \sum_{k=1}^k \phi_k n_{k,t-1} \quad (11)$$

$$N_{k,t} = \frac{1}{1 + e^{-n_{i,t}}}$$

$$y_t = \gamma_o + \sum_{k=1}^k \gamma_k N_{k,t}$$

RNNs are suited to data that has a time dimension such as financial time series data where RNNs are used to capture the memory in time series data, especially high-frequency time series data such as daily and intraday data. The advantages of using RNNs over feed-forward networks for modelling non-linear time series are well documented. However, as mentioned by Tenti (1996, p569): “*the main disadvantage of RNNs is that they require substantially more connections and more memory in simulation than standard backpropagation networks*”, thus resulting in a substantial increase in computational time.

Although, RNNs can yield better results in comparison to MLPs due to the additional memory inputs. Figure 3-6 depicts an RNN featuring one hidden layer:

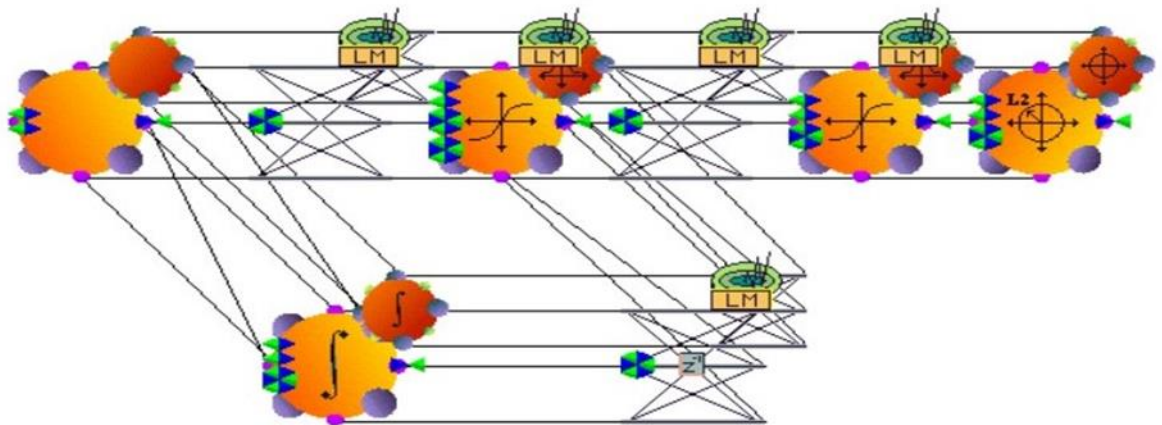


Figure 3-6: Recurrent Neural Network Featuring One Hidden Layer.

An extension of the RNN is the Time-Lag Recurrent Network (TLRN) which are MLPs extended with short-term memory structures. Most real-world data contain information in its time structure, yet most ANNs are purely static classifiers. TLRNs are state of the art in nonlinear time series prediction, system identification and temporal pattern classification.

### 3.1.5 Radial Basis Function

Initially proposed by Broomhead & Lowe (1988), a Radial Basis Function (RBF) is another commonly used ANN. The motivation behind the introduction of the RBF was to improve the accuracy of MLPs while decreasing training time and network complexity. An RBF network utilises a radial basis or Gaussian density function as the activation function. Also, the structure of the RBF is different from that of the MLP. McNelis (2005) describes the set up as follows: input neurons can be a linear combination of regressors just like other networks. However, there is only one input signal; therefore, only one set of coefficients of the input variables. The signal from the input layer is the same overall neurons which are Gaussian transformations, around  $k$  different means, of the input signals. Therefore, the input signals have different centres for the radial bases or normal distributions. The differing Gaussian transformations are combined in a linear function for forecasting the output. McNelis (2005) states that the following system describes an RBF network:

$$\text{Min}_{\langle w, \mu, \gamma \rangle} \sum_{t=0}^T (y_t - \hat{y}_t)^2 \quad (12)$$

$$n_t = w_0 + \sum_{i=1}^i w_i x_{i,t}$$

$$R_{k,t} = \phi(n_t; \mu_k)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_{n-\mu_k}} \exp\left(-\frac{[n_t - \mu_k]^2}{\sigma_n - \mu_k}\right)$$

$$\hat{y}_t = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k R_{k,t}$$

Where  $x$  represents the set of input variables, and  $n$  represents the linear transformation of the input variables based on weights  $w$ .  $k$  different centres for the RBF transformation,  $\mu_k, k = 1, \dots, k$ , calculate the  $k$  standard error implied by the different centres  $\mu_k$  and obtain the  $k$  different RBF,  $R_k$ . These functions in turn are combined linearly to forecast  $y$  with weights,  $\gamma$ . Optimization of the RBF network involves choosing the coefficient set  $w$  and  $\gamma$  and  $k$  centres of the RBF  $\mu$ . Figure 3-7 depicts an RBF network:

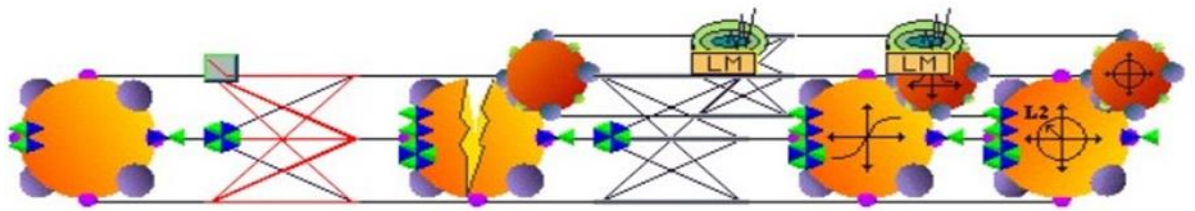


Figure 3-7: Radial Basis Function (RBF) Neural Network.

In application, Haykin (1994) identifies a number of significant differences between an RBF network and the common MLP. Firstly, an RBF network typically has one hidden layer at most whereas an MLP can have many hidden layers. Secondly, the activation function of an RBF network computes the Euclidean norm or distance between the signal from the input vector and the centre of that unit whereas the MLP computes the inner products of the inputs and the weights for that unit. RBF networks can improve accuracy and reduce training time due to their simpler and straightforward training phase which

features the location of an optimal number of hidden nodes and the optimisation of the parameters of the RBF. Mandic & Chambers (2001) highlight that both the MLP and RBF networks have good approximation properties but note that an MLP can always simulate a Gaussian RBF network. Also, RBFs are easily designed, have good generalisation and show strong tolerance to input noise, Yu *et al.* (2011). A disadvantage of the RBF network is finding the optimal parameter values for its RBF functions and their optimal number. Solutions exist in the form of clustering methods such as Self Organizing Maps (SoM) offered by Kohonen (1982) and the K-means offered by (Lloyd 1982). More recently, metaheuristic algorithms have been successfully applied to the task (Sermpinis *et al.* 2013).

### 3.1.6 Probabilistic Neural Network (PNN)

In a Probabilistic Neural Network (PNN), all the weights of the network are calculated analytically. The number of cluster centres is equal to the number of exemplars, and they are all set to the same variance (which may be optimised if cross validation is specified). A PNN usually performs best when the number of exemplars is small (<1000) or so dispersed that clustering is ill-defined. Figure 3-8 depicts a PNN:

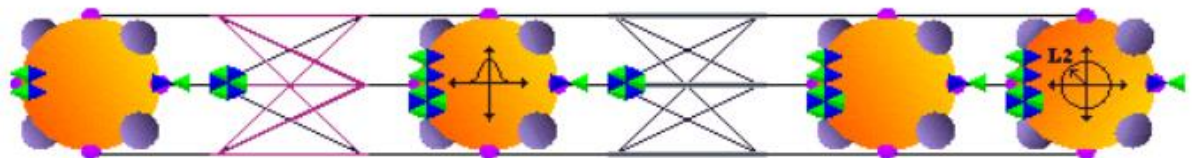


Figure 3-8: Probabilistic Neural Network

### 3.1.7 Co-Active Neuro-Fuzzy Inference System (CANFIS)

The Co-Active Neuro-Fuzzy Inference Systems (CANFIS) was proposed by J.-S. R. Jang, C.-T. Sun, and E. Mizutani, and integrates fuzzy inputs with an ANN to solve poorly defined problems. Fuzzy inference systems are valuable as they combine the explanatory nature of rules (membership functions) with the power of *black box* neural networks. The fundamental component of a CANFIS is a fuzzy axon which applies membership functions to the inputs. Two membership functions commonly used are generalized Bell and Gaussian. The output of a fuzzy axon is computed using the following formula:

$$f_j(x, w) = \text{Min} \forall_i (MF(x_i, w_{ij})) \quad (13)$$

Where  $I$  = input index,  $j$  = output index,  $x_i$  = input  $i$ ,  $w_{ij}$  = weights (MF parameters) corresponding to the  $j^{\text{th}}$  MF of input  $i$  and MF is the membership function of the subclass of the fuzzy axon. For model initiation a common rule set with  $n$  inputs and  $m$  if-then rules as follows (Jang 1993):

$$\begin{aligned} \text{Rule 1: If } z_i \text{ is } A_{11} \text{ and } z_2 \text{ is } A_{12} \dots \text{ and } z_n \text{ is } A_{1n} \\ \text{then } f_1 = p_{11}z_1 + p_{12}z_2 + \dots p_{1n}z_n q_1 \end{aligned} \quad (14)$$

$$\begin{aligned} \text{Rule 2: If } z_1 \text{ is } A_{21} \text{ and } z_2 \text{ is } A_{22} \dots \text{ and } z_n \text{ is } A_{2n} \\ \text{then } f_2 = p_{21}z_1 + p_{22}z_2 + \dots + p_{2n}z_n q_2 \end{aligned} \quad (15)$$

$$\begin{aligned} \text{Rule } m: \text{ If } z_1 \text{ is } A_{m1} \text{ and } z_2 \text{ is } A_{m2} \dots \text{ and } z_n \text{ is } A_{mn} \\ \text{then } f_m = p_{m1}z_1 + p_{m2}z_2 + \dots + p_{mn}z_n + q_n \end{aligned} \quad (16)$$

All layers in a CANFIS architecture are either adaptive or fixed. The function of each layer is described as follows:

#### *Layer 1 - Premise Parameters*

Every node in this layer is a complex-valued MF ( $\mu_{ij}$ ) with a node function:

$$O_{1,ij} = |\mu_{A_{ij}}(z_i)| \sqcup \mu_{A_{ij}}(z_i) \text{ for } (1 \leq i \leq n, 1 \leq j \leq m) \quad (17)$$

Each node in layer 1 is the membership grade of a fuzzy set ( $A_{ij}$ ) and specifies the degree to which the given input belongs to one of the fuzzy sets.

#### *Layer 2 - Firing Strength*

Each node in this layer is a product of all the incoming signals. This layer receives input in the form of the product of all the output pairs from the first layer:

$$O_{2,j} = w_j + \mu_{A_{i1}}(z_1)\mu_{A_{i2}}(z_2), \dots, \mu_{A_{in}}(z_n) \text{ for } (1 \leq i \leq m) \quad (18)$$

#### *Layer 3 - Normalized Firing Strength*

Every node in this layer calculates rational firing strength:

$$O_{3,j} = \bar{w}_j = \frac{w_j}{\sum_{j=1}^m w_j} \text{ for } (1 \leq j \leq m) \quad (19)$$

#### Layer 4 - Consequent Parameters

Each node in this layer is the multiplication of the normalised firing strength from the third layer and output of the neural network:

$$O_{4,j} = \bar{w}_j f_j = \bar{w}_j (p_{j1} z_1 + p_{j2} z_2 + \dots + p_{jn} z_n + q_j) \text{ for } (1 \leq j \leq m) \quad (20)$$

#### Layer 5 - Overall output

The node here computes the overall output of a CANFIS network:

$$O_{5,1} = \sum \bar{w}_j f_j \quad (21)$$

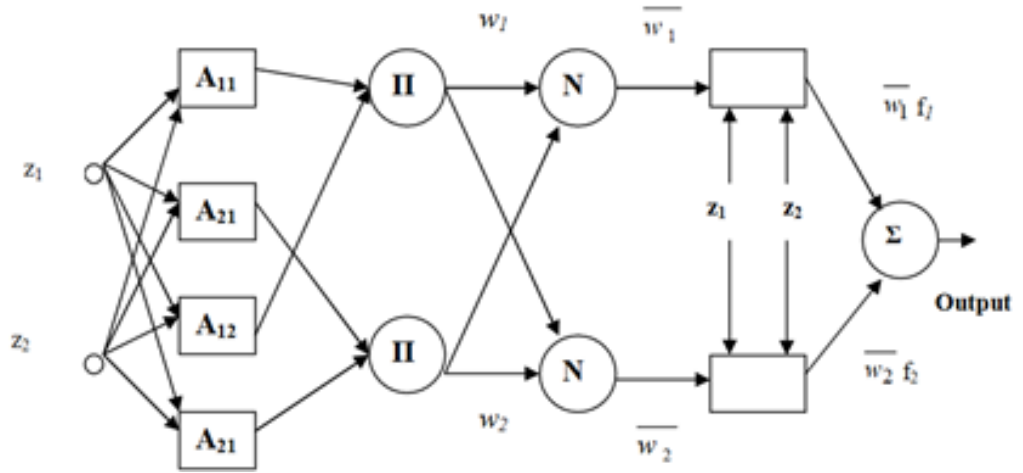


Figure 3-9: A Prototype CANFIS Network with Two-Inputs and One-Output.

The CANFIS powerful capability stems from pattern-dependent weights between the consequent layer and the fuzzy association layer. Membership values correspond to those dynamically changeable weights that depend on input patterns. The network also contains a normalisation axon to expand the output into a range of 0-1. Although the inside of each neural consequent turns out to be a black box, the whole CANFIS model retains the concept of fuzzy reasoning. In terms of the behaviour of the system, it enjoys transparency with respect to MF interpretability. One possible way of obtaining more precision in a

given mapping is to construct the neural consequents without increasing fuzzy rules. When neural consequents are combined to form neural rules, it helps to reduce the number of modifiable parameters in which a CANFIS with neural rules becomes equivalent to a modular network.

A disadvantage of a CANFIS is that it should not be used to predict values outside the extreme contained in the learning dataset. Another disadvantage is that sufficient dataset volume is required to build the model. As such it is not capable of direct prediction for sites which have a lack of archived observations. The CANFIS model is also more computationally intensive than most other models.

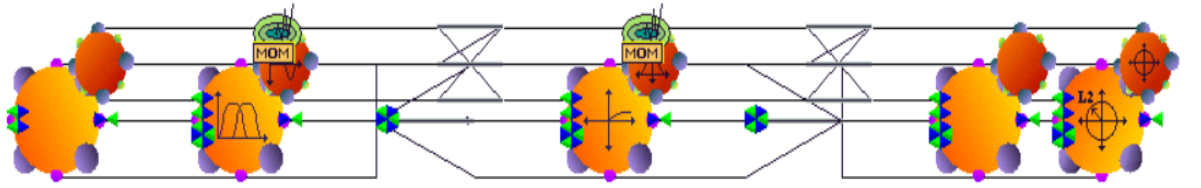


Figure 3-10: CANFIS (Tsukamoto) Neural Network.

### 3.1.8 Combination Forecast - Simple Average

The simple average forecast combination (Fc) is a basic but versatile forecasting technique that can be applied to most forecasting models to typically enhance forecasting robustness and accuracy as well as serving as an unbiased benchmark. For example: given a set of four ANN forecasts;  $f_{ANN1}^t, f_{ANN2}^t, f_{ANN3}^t, f_{ANN4}^t$  the combination forecast at time  $t$  is calculated as:

$$f_{c_{ANN}}^t = (f_{ANN1}^t + f_{ANN2}^t + f_{ANN3}^t + f_{ANN4}^t)/4 \quad (22)$$

## 3.2 Benchmark Models

### 3.2.1 Moving Average Convergence Divergence

In Chapter 4, *Artificial Intelligence in Trading - Forecasting the EUR/USD Exchange Rate with Artificial Neural Networks*, the performance of the ANNs are benchmarked against a Moving Average Convergence Divergence (MACD) model. The MACD is a practical benchmark because if the ANNs equipped with informative technical indicators cannot outperform a standalone technical indicator, then ANN modelling is costly and ineffective. The MACD is a well-established technical indicator invented in the late 1970s by Gerald Appel and showcased in his book; *Technical Analysis, Power Tools for Active Investors*. Technically, the MACD is an oscillator that measures trend and momentum and is comparable to the dual MA crossover strategy. Although observable as the interplay of two Exponential Moving Averages (EMAs), the MACD is calculated using three EMAs;

$$MACD\ Line = 12\ period\ EMA - 26\ period\ EMA \quad (23)$$

$$Signal\ Line = 9\ period\ EMA\ of\ MACD\ Line \quad (24)$$

The MACD line is a 12 period EMA minus a 26 period EMA. The signal line is a 9 period EMA of the MACD line and is plotted alongside the MACD line. Appel advises the use of the 12, 26, 9 period parameters and the use of EMAs as they are more responsive compared to SMAs. It is the interplay between the MACD line and the signal line in the form of crossovers and divergences that generates the trading signals. A buy signal is generated when the signal line rises below the MACD line. Conversely, a sell signal is generated when the signal line falls below the MACD line.

### 3.2.2 Naïve Forecast

A Naïve forecast is considered to be the simplest of all forecasting models. A Naïve forecast accepts as a forecast for time  $t+1$ , the value of time  $t$ , assuming that the best forecast is the most recent period change:

$$\hat{Y}_{t+1} = Y_t \quad (25)$$



### 3.2.3 Support Vector Regression (SVR)

Support Vector Machines (SVMs) were originally developed for solving classification problems in pattern recognition frameworks. However, the introduction of Vapnik's (1995)  $\epsilon$ -sensitive loss function extended their use to non-linear regression estimation problems. Support Vector Regression (SVR), is a robust technique for constructing data-driven, non-linear regression models and is commonly used in financial market and macroeconomic applications, see Lu *et al.* (2009) and Xu *et al.* (2013). SVR hold advantages such as; global solutions, insusceptible to local minima and provides an equilibrium between model accuracy and model complexity, Suykens (2005) and Kwon & Moon (2007). An SVR is specified as:

$$f(x) = \omega^T \varphi(x) + b \quad (26)$$

Where  $\omega$  and  $b$  are the regression parameter vectors of the function and  $\varphi(x)$  is the non-linear function that maps the input data vector  $x$  into a feature space where the training data exhibit linearity. Figure 3-11 is an example of the SVR generated in Neuro Solutions Pro, notice that in comparison to the ANNs mentioned above it has a simpler structure.

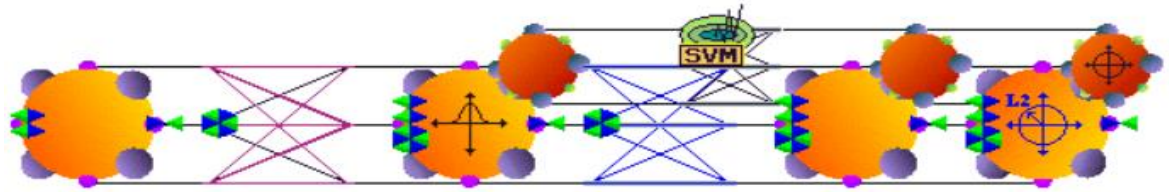


Figure 3-11: Support Vector Machine (Regression)

### 3.2.4 GARCH (1,1)

The Generalised Auto-Regressive Conditional Heteroscedasticity (GARCH) model of Engle (1982) and Bolerslev (1986) requires the joint estimation of the conditional mean model and the variance process. On the assumption that the conditional mean stochastic error,  $\epsilon_t$ , is normally distributed with zero mean and time-varying conditional variance,  $ht^2$ , the GARCH (1,1) model is given by:

$$h^2_{t+1} = \omega + \alpha \epsilon_t^2 + \beta h_t^2 \quad (27)$$

Where all the parameters are positive, and the sum of  $\alpha + \beta$  quantifies the persistence of shocks to volatility. The GARCH (1,1) model generates one-step-ahead forecasts of volatility as a weighted average of the constant long-run or average variance,  $\omega$ , the previous forecast variance,  $ht^2$ , and previous volatility reflecting squared news about the return,  $\epsilon t^2$ . Furthermore, as volatility forecasts are increased following a large return of either sign, the GARCH specification captures the volatility clustering effect.

# Chapter 4

## Artificial Intelligence in Trading - Forecasting the EUR/USD Exchange Rate with Artificial Neural Networks

### 4.1 Introduction

Financial market forecasting is known to be a challenging endeavour due to the time series exemplifying; chaotic, multivariate, nonlinear, noisy and nonstationary characteristics ultimately derived from the irrational investor that not only differ from market to market but are constantly evolving. To date, quantitative research has extracted useful information from the financial markets and devised a wealth of theories from it. For example, such information has yielded the; Capital Asset Pricing Model (CAPM), Arbitrage Pricing Theory, and the Black-Scholes / Merton option pricing model, all of which are predicated on the Efficient Market Hypothesis (EMH). Quantitative research has also produced an array of useful statistical models, for example; ARIMA and GARCH. The aforementioned statistical models are of unquestionable utility in financial market forecasting; however, their empirical performance and suitability are often called into question. Statistical performance and trading performance is often mixed and varies drastically across financial markets. More fundamentally, the nature of financial markets themselves often questions the suitability of such models. For example, can a linear model really interpret and accurately predict chaotic nonlinear data? How is the performance of such models impacted under changing market environments? Which statistical models are better than others? Those are questions only quantitative research can answer; however, it would initially be wise to rationally consider model selection based on the models specifications in relation to the data and may be more appropriate to employ a nonlinear model when analysing nonlinear data. In relation to financial market forecasting, such a proposal naturally introduces models derived from Computational Intelligence (CI), a sub-branch of Artificial Intelligence (AI), which are primarily driven by the laws of nature and adaptive mechanisms in order to facilitate intelligent behaviours in complex dynamic environments. Artificial Neural Networks (ANNs) are common CI models that have been successfully

applied to solve a variety of real-world problems and have relatively recently been applied to financial market forecasting, Huang *et al.* (2013). The advantage of such CI models over traditional models lies in their nonlinear adaptive nature, the variety of architectures and ability to accept a large number of input variables, therefore, making them well suited to financial market analysis and forecasting. Furthermore, promising empirical evidence permits them to remain at the centre of financial market research and debate.

## 4.2 Motivation

The motivation of this research is twofold. The principle motivation is to evaluate the performance of ANNs in their ability to forecast the logarithmic returns of the EUR/USD exchange rate one period in advance and compare their performance with a MACD model to conclude whether the ANNs offer superior forecasting ability. The secondary motivation is to address deficiencies of prior research in the data collection phase when utilising TA and its sub-section of technical indicators as input variables. Frequently, prior research has consistently failed to measure all *technical* aspects of a time series. This deficiency is not only failing to represent TA and its sub-section of technical indicators in their entirety but has the implication of depriving the ANN of potentially informative data. This implication is compounded when considering that ANNs are data driven models, highly dependent on the input data. In solution, an extensive candidate pool of 10,568 technical indicator models is generated, ensuring that all technical aspects of the EUR/USD time series were measured, i.e. trend, trend strength, momentum, volatility, breadth, support, resistance, relative strength, and cycles. The performance of the ANNs will not only be measured in comparison to one another but also benchmarked against the notable technical indicator, the Moving Average Convergence Divergence (MACD). After all, it would be unproductive to perform the task of ANN modelling and forecasting if ANNs cannot outperform a standalone technical indicator.

## 4.3 EUR/USD Time Series Data

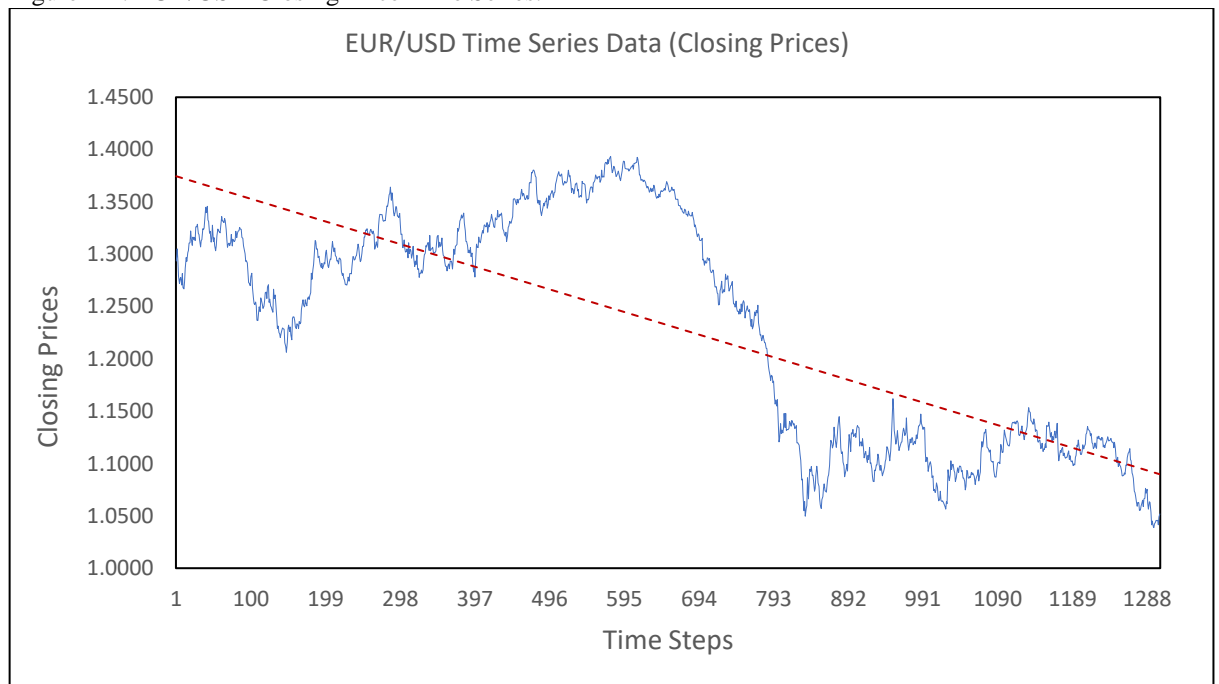
This section presents a statistical analysis of the EUR/USD time series data. Initially, descriptive statistics of the time series data in closing price format are presented, followed by descriptive statistics of the time series data transformed into log-returns format. The

statistical analysis includes statistical tests covering: normality (Jarque-Bera), stationarity (Augmented Dickey-Fuller), monotonic trend presence (Mann-Kendall) and the minimum number of explanatory variables required to model the time series data (Fractal Dimension), supplemented with summary statistics.

#### 4.3.1 EUR/USD Time Series Data (Closing Price)

Typically, users of financial time series data are accustomed to working with exchange rates in the form of; high, low, open, close - or just closing prices as prescribed by Dow Theory on the bases of representing the best measure of fair value for the given period. This investigation follows Dow Theory and utilises the closing price to represent the EUR/USD time series. Figure 4-1 depicts the daily closing prices of the EUR/USD time series data over the sample period of 2012 to 2016:

Figure 4-1: EUR/USD Closing Price Time Series.



Visually, an automated trend line shows the EUR/USD time series features an overall downtrend throughout the sample period, as the sample commences at 1.2935 on 2nd January 2012 and terminates at 1.0520 on 30th December 2016. The following table details key descriptive statistics regarding the EUR/USD time series, for their full disclosure and description, please see Appendix 3.

Table 4-1: EUR/USD Descriptive Statistics (Closing Prices)

<b>Descriptive Statistics</b>	<b>Confidence Level</b>	<b>P-Value</b>	<b>Result</b>
Mann-Kendall	99%	<0.0001	Monotonic Trend Presence
Jarque-Bera	99%	<0.0001	Not Normally Distributed
Dickey-Fuller	99%	0.7480	Non-Stationary
Fractal Dimension	N/A	N/A	2 (1.8)

From Table 4-1, the uptrend is empirically evidenced by the Mann-Kendal trend test, confirming the presence of a monotonic trend in the time series data. The Jarque-Bera test for normality finds the time series to be not normally distributed, and the augmented Dickey-Fuller test for stationarity categorises the time series as non-stationary. The test of fractal dimension generated the largest value of 1.8, therefore the time series requires a minimum of 2 explanatory variables for modelling purposes.

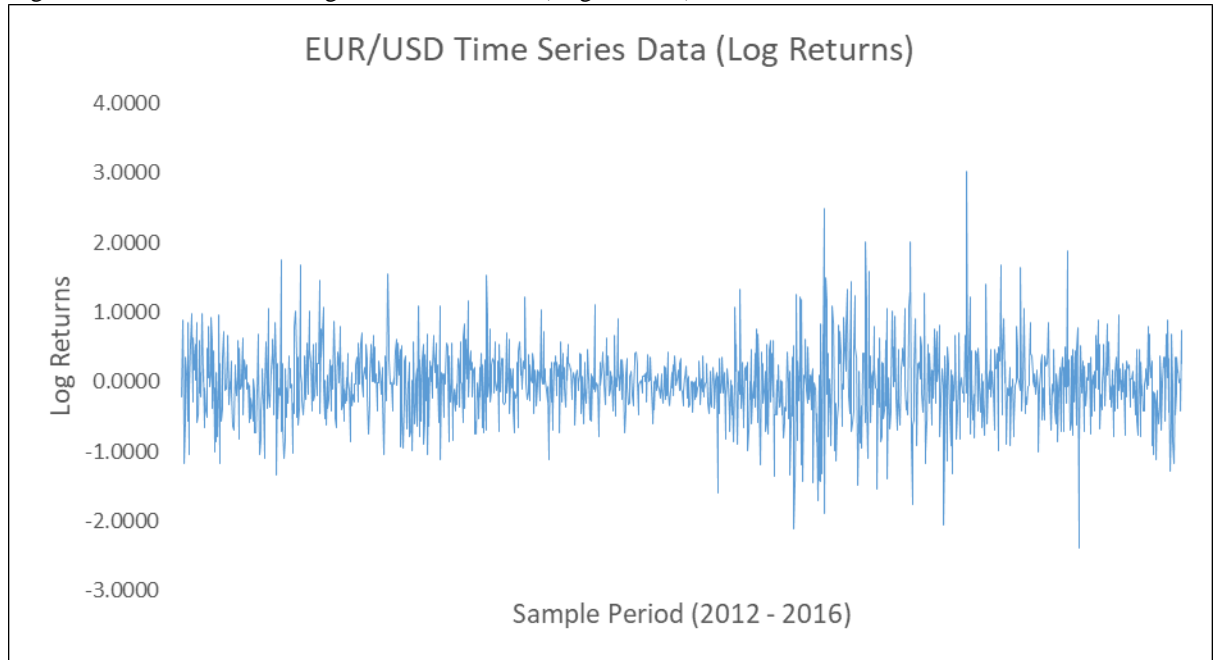
#### 4.3.2 EUR/USD Time Series (Log Returns)

The descriptive statistics mentioned thus far have covered the EUR/USD time series in raw-data format. This section will apply the same statistical tests to the EUR/USD time series data transformed into log-returns format:

$$R_i = \ln \frac{C_i}{C_{i-1}} \times 100 \quad (28)$$

The log-returns data transformation is implemented not only to reduce the magnitude of the data but also to solve the problem of non-stationarity. Figure 4-2 below depicts the EUR/USD transformed into log-returns format:

Figure 4-2: EUR/USD Closing Price Time Series (Log>Returns).



The following table details key descriptive statistics regarding the EUR/USD log-returns time series, for their full disclosure and description, please see Appendix 4.

Table 4-2: EUR/USD Descriptive Statistics (Log>Returns)

<b>Descriptive Statistics</b>	<b>Confidence Level</b>	<b>P-Value</b>	<b>Result</b>
Mann-Kendall	99%	<0.0001	Monotonic Trend Presence
Jarque-Bera	99%	<0.0001	Not Normally Distributed
Dickey-Fuller	99%	<0.0001	Stationary
Fractal Dimension	N/A	N/A	2 (1.9)

From Table 4-2, the Mann-Kendal trend test confirms the presence of a monotonic trend. This results is surprising as one of the properties of the log-returns data transform is the removal of the trend from the time series. The Jarque-Bera test for normality found the time series to be not normally distributed and the augmented Dickey-Fuller test for stationarity categorises the time series as stationary. The test of fractal dimension generated the largest value of 1.9, this value approximates the EUR/USD in closing-price format; therefore, both time series require a minimum number of 2 independent variables to model the time series in their respective formats.

## 4.4 Input Variable Selection

### 4.4.1 Filter Methods

A two-stage IVS method is implemented to identify the most informative variables from the candidate input variable pool. First, filter methods are applied as an initial screening process to identify and remove irrelevant variables by screening for constant, quasi-constant and duplicate variables. Of the original 10,568 variables, there were no constant, quasi-constant or duplicate variables. Pearson's correlation coefficient is then applied to test the correlation between the target variable (EUR/USD) and the predictor variables (technical indicators). A threshold of  $\pm 0.80$  (strong correlation) is set, and predictor variables that do not satisfy that criteria are removed, reducing the dataset by 9,970 variables from 10,568 variables to 624 variables. With the candidate variable pool reduced to 624 strongly correlated variables, the remaining variables progress to the second stage of IVS, Stepwise Multivariate Linear Regression, described in the next section.

### 4.4.2 Multivariate Linear Regression (Stepwise)

Stepwise Multivariate Linear Regression is employed as the second stage of the IVS process. The stepwise method of multivariate linear regression provides that the order of entry of predictor variables is based purely on statistical criteria. Predictor variables (technical indicators) that correlate most strongly with the dependent variable (log returns) are afforded priority of entry with no reference to theoretical considerations. The advantage of stepwise multiple linear regression is that it is applied primarily in exploratory work in which the researcher is unsure about the relative predictive power of the investigation's predictor variables.

The stepwise multivariate linear regression is performed in accordance with the Handbook of Univariate and Multivariate Data Analysis with IBM SPSS, authored by Ho. To perform robust IVS consisting of a maximum of 5 predictor variables, the following strict criteria are adhered to: the final model must have  $R$ ,  $R^2$ , and adjusted- $R^2$  values  $\geq 0.90$  (90%), a low standard error of the estimate, no multicollinearity between predictor variables, therefore, a Condition Index (CI) of  $\leq 15$ , Value Inflation Factor (VIF) values of  $\leq 5$  and all



predictor variables must be statistically significant at the 0.01 (99%) level. Following implementation, stepwise multivariate linear regression generated a total of 12 models. All models featured:  $R$ ,  $R^2$  and adjusted- $R^2$  values of 1.00, therefore achieving perfect linear correlation. However, all 12 models featured high levels of multicollinearity. An elimination process is employed to remove the multicollinearity, starting with the variable featuring the largest VIF value, then repeating the elimination process until the remaining variables feature VIF values of  $\leq 5$  and a CI of  $\leq 15$ . Following implementation, a 3 variable model is produced fully satisfying the above criteria. The 3 variable model also has the benefit of satisfying the fractal dimension requirement of a minimum of 2 variables required for modelling purposes as discussed in section 2.1.2. Table 4-3 presents the model summary, detailing the  $R$ ,  $R^2$ , adjusted- $R^2$ , standard error of the estimate and Durbin-Watson values.

Table 4-3: Model Summary

<i>Model Summary</i>				
<b>R</b>	<b>R<sup>2</sup></b>	<b>Adjusted R<sup>2</sup></b>	<b>Std. Error of the Estimate</b>	<b>Durbin-Watson</b>
1.000	1.000	1.000	0.0024718	1.93

From Table 4-3, the model features  $R$ ,  $R^2$  and adjusted  $R^2$  values of 1.000, thus achieving perfect linear correlation. The standardised error of the estimate is respectable at 0.0025470, and the Durbin-Watson test registers as 1.93 which is close to the optimal value of 2.0, therefore, indicating negligible autocorrelation in the residuals.

Table 4-4 details the One-Way Analysis of Variance (ANOVA) which is an extension of the independent t-test and is used when a research is interested in whether the means of two or more independent groups differ.

Table 4-4: Model ANOVA

<b>ANOVA</b>					
	<b>Sum of Squares</b>	<b>df</b>	<b>Mean Square</b>	<b>F</b>	<b>Sig.</b>
Regression	158.563	3	52.854	8650770.205	0.000
Residual	0.005	779	0.000		
Total	158.568	782			

From Table 4-4, it is evidenced that the Models ANOVA states it is a statistically significant model at the 0.01 (99%) level; therefore, a good predictor of the dependent

variable. The following table details the summarised model coefficients, displaying each predictor variables: standardised Beta, significance and VIF values.

Table 4-5: Model Coefficients

<b>Model</b>			
<b>Predictor Variable</b>	<b>Standardised Coefficients Beta</b>	<b>Sig.</b>	<b>VIF</b>
2PBFS1	0.997	0.000	3.297
RSI(SMA)7	0.002	0.000	1.893
LWMA(-L1(15))8	0.003	0.000	4.793

From Table 4-5, all predictor variables made a statistically significant unique contribution to the model at the 0.000 level, therefore satisfying the 0.01 (99%) significance criteria. All predictor variables have a VIF value of  $\leq 5$ ; thus, all predictor variables are free from multicollinearity. The predictor variable featuring the highest standardised Beta (model contribution) is predictor variable 2PBFS1 with a standardised Beta of 0.997. This satisfies the *a priori* assumption of model development where one, or more, variables contribute the most to the model. For a detailed description of the above selected inputs, and all other respective inputs of each empirical chapter, see appendix 5.

## 4.5 Methodology

NeuroSolutions Pro is utilised to build and test the ANNs, with network parameters held constant across all networks, i.e. the number of hidden layers, number of processing elements (PEs) in each hidden layer, number of epochs, weighting update method, activation function and learning rule. To prevent overfitting various procedures were adhered to, namely parameter pruning; Castellano *et al.* (1997) and Wang *et al.* (2010), cross-validation; Zhang *et al.* (1999), Amjady & Keynia (2009), and Sermpinis *et al.* (2012b), and early stopping; Lin *et al.* (2009) and Prechelt (2012). The rationale underpinning the methodology will now be discussed.

#### **4.5.1 Artificial Neural networks**

The evaluated ANNs are the: Multilayer Perceptron (MLP), Generalized Feedforward Network (GFF), Modular Network (M), Radial Basis Function (RBF), Co-Active Neuro-Fuzzy Inference System (CANFIS), Probabilistic Neural Network (PNN), and forecast combination (ANN Fc) of the featured ANNs.

#### **4.5.2 Number of Hidden Layers**

For each type of ANN, custom models are built featuring 1 to 4 hidden layers. Although any number of hidden layers is possible, it is widely accepted and supported by the general function approximation theorem that two hidden layers can approximate any complex nonlinear function and provides powerful computational properties, Dayhoff & DeLeo (2001). For that reason, this evaluation follows the 2 hidden layer suggestion and extends beyond it.

#### **4.5.3 Activation Function**

All ANN architectures are reproduced featuring the following activation functions: Tanh Axon, Sigmoid Axon, Linear Tanh Axon, Linear Sigmoid Axon, SoftMax Axon, Bias Axon, Linear Axon, and Axon. For a description of each activation function, see Appendix 6.

#### **4.5.4 Learning Rule**

All models feature the Levenberg-Marquardt (LM) learning algorithm as it is one of the most appropriate higher-order adaptive algorithms known for minimising the MSE, training significantly faster than the commonly employed momentum learning algorithm, and typically arrives at a solution with a significantly lower error.

#### **4.5.5 Processing Elements**

To establish the optimum number of processing elements (PEs) in each hidden layer, the *vary a parameter* function is employed which allows automated incremental testing to identify the optimal number of PEs defined in terms of lowest Mean Absolute Error (MAE). The optimum number of PEs to be identified is confined to a generous range of 1 to 50 with testing increments of 1. Therefore, a total of 50 PEs are tested for each hidden layer. During this phase, the network tests all PEs over three cycles, therefore a total of 150 simulations (50 PEs x 3 cycles) are performed on each hidden layer. This equates to 150 simulations (150 simulations x 1 hidden layer) for ANNs featuring one hidden layer, and 300, 450, and 600 simulations for ANNs featuring 2, 3, and 4 hidden layers respectively.

#### **4.5.6 Epochs**

The number of epochs is set to 1000 and programmed to terminate after 150 epochs without improvement, therefore implementing the early stopping procedure to guard against over-fitting as presented in Lin *et al.* (2009) and Prechelt (2012).

#### **4.5.7 Weights**

Each model features batch weighting which updates the weights after the presentation of the entire training set.

#### **4.5.8 Data Segregation**

This evaluation features a 5-year sample period, (2nd January 2012 - 30th December 2016), segregated into a training, cross-validation, and testing ratio of 45:15:40 summarised in table 4-6:

Table 4-6: EUR/USD, Data Segregation of Sample Period

Period	Start Date	End Date	% of Data	Trading Days
Sample Period	2nd January 2012	30th December 2016	100	1305
Training	2nd January 2012	1st April 2014	45	587
Cross Validation	2nd April 2014	31st December 2014	15	196
Testing	1st January 2015	30th December 2016	40	522

The use of cross validation is central to machine learning and is used to assess the stability of a model and how well it will generalise to new data in addition to preventing overfitting and underfitting.

#### 4.5.9 Benchmark: Moving Average Convergence Divergence (MACD)

The Benchmark model is a MACD model as described in section 3.3.1.

### 4.6 Forecast Evaluation

As it is standard in the literature, the forecast evaluation metrics of ME, MSE, RMSE, MAE, MPE, and MAPE are calculated. Regarding the interpretation of the above statistics, the lower the value, the better the forecasting accuracy of the model. Additionally, the advanced forecasting evaluation metrics of Theils-U1, Theils-U2, and Paseran-Timmerman (PT) are computed. A lower value of Theils-U1 is desirable, with values bound between 0 and 1. For Theils-U2, values  $<1$  indicate superior forecasting ability than a Naïve forecast, values of 1 equal a Naïve forecast, and values  $>1$  indicate the model has underperformed a Naïve forecast. For the PT test, larger values indicate superior performance in forecasting directional change. For the full description, mathematical notation, and interpretation of the statistics see Appendix 1.

#### 4.6.1 In-Sample Forecast Evaluation

Table 4-7 presents the in-sample forecasting evaluation of the ANNs and benchmark MACD model:

Table 4-7: In-Sample Forecast Evaluation

<i>In-Sample Forecasting Results</i>								
<b>Model</b>	<b>MLP</b>	<b>GFF</b>	<b>MNN</b>	<b>RBF</b>	<b>PNN</b>	<b>CANFIS</b>	<b>ANN Fc</b>	<b>MACD</b>
ME	0.0782	0.0511	0.0271	-0.0291	-0.0219	-0.0232	0.0137	<b>-0.0088</b>
MSE	0.2077	0.2043	0.2043	0.1983	0.2031	0.2022	0.1999	<b>0.1740</b>
RMSE	0.4558	0.4520	0.4520	0.4453	0.4507	0.4496	0.4470	<b>0.4171</b>
MAE	0.3469	0.3422	0.3436	0.3410	0.3385	0.3408	0.3382	<b>0.3144</b>
MPE	18.0560	10.1402	10.5493	-7.2021	-1.3668	<b>0.1081</b>	5.0474	3.5447
MAPE	18.5007	10.3808	15.3225	22.0456	<b>3.3376</b>	11.0622	7.3440	8.2448
Theils-U1	0.8398	0.8808	0.8701	<b>0.8045</b>	0.9721	0.8983	0.9112	0.8214
Theils-U2	1.0297	1.0277	0.9651	1.0735	0.9977	0.9870	0.9926	<b>0.9068</b>
PT	-1.0142	0.9886	-0.7015	0.8877	-0.9886	0.6264	1.4516	<b>10.8789</b>

\*\*1% Significance, \*5%Significance

From Table 4-7, the in-sample results demonstrate the MACD benchmark model is the superior forecasting model, acquiring 6 of the 9 forecast evaluation metrics: ME (0.01), MSE (0.17), RMSE (0.42), MAE (0.31), Theils-U2 (0.91), and PT (10.88). Amongst the ANNs, forecast evaluation results are mixed. The RBF achieved 3 of the 9 forecast evaluation metrics: MSE (0.20), RMSE (0.45), and Theils-U1 (0.80), indicating that of the ANNs it generated the most accurate and consistent forecasts. The PNN also achieved 3 of the 9 forecast evaluation metrics: ME (-0.02), MAE (0.34), and MAPE (3.34). The CANFIS achieved the best MPE value of 0.11. Meanwhile, the MNN achieved the best Theils-U2 value of 0.97, with the MLP, GFF, RBF, and PNN models all recording values >1 indicating they underperformed a Naïve forecast. The GFF secured the PT metric with a value of 0.99 which is notably better than other ANNs. The ANN Fc improved the metrics of ME and MAE with values of 0.01 and 0.34 respectively. The ANN Fc also secured the best PT value of 1.45 demonstrating a notable increase in directional accuracy over the individual ANN models; however, did not outperform the MACD model. To conclude the in-sample forecast results, the MACD is the superior model across 6 of the 9 forecast evaluation metrics, outperforming all ANN models and the ANN Fc model. Regarding the ANNs, the results are mixed with the RBF emerging as the best ANN model via RMSE, Theils-U1 and the second-best PT value. The ANN Fc offered a minor increase in performance in ME and MAE and did notably increase performance in forecasting directional change.

#### 4.6.2 Out-of-Sample Forecast Evaluation

Table 4-8 presents the out-of-sample forecast evaluation of the various ANNs and MACD benchmark model:

Table 4-8: Out-of-Sample Statistical Evaluation

<i>Out-of-Sample Forecast Results</i>								
<b>METRIC</b>	<b>MLP</b>	<b>GFF</b>	<b>MNN</b>	<b>RBF</b>	<b>PNN</b>	<b>CANFIS</b>	<b>ANN Fc</b>	<b>MACD</b>
ME	0.0641	0.0401	0.0128	-0.0367	-0.0406	-0.0383	<b>0.0003</b>	-0.0278
MSE	0.4144	0.4137	0.4149	0.4098	0.4173	0.4170	0.4098	<b>0.3571</b>
RMSE	0.6438	0.6432	0.6441	0.6402	0.6460	0.6458	0.6402	<b>0.5976</b>
MAE	0.4807	0.4766	0.4819	0.4802	0.4793	0.4826	0.4763	<b>0.4479</b>
MPE	8.0743	6.2367	2.3109	<b>0.1934</b>	-0.3055	0.3635	2.8122	0.8021
MAPE	8.2792	6.3470	5.0863	7.5730	<b>2.2770</b>	5.0682	3.6725	4.7003
Theils-U1	0.8618	0.8915	0.8760	0.8309	0.9809	0.8983	0.9161	<b>0.8208</b>
Theils-U2	0.9654	0.9872	0.9445	1.0793	1.0048	1.0144	0.9963	<b>0.8808</b>
PT	-0.3249	-0.9811	0.7919	*1.9655	-1.0232	0.7231	*1.9154	<b>8.0919</b>
DM	9.8585	8.2101	8.8543	5.2682	10.3366	8.5865	8.2765	N/A

\*\*1% Significance, \*5% Significance

From table 4-8, the out-of-sample performance is largely consistent with in-sample performance. The MACD benchmark model also dominates the out-of-sample period, securing 6 of the 10 forecast evaluation metrics: MSE (0.36), RMSE (0.6), MAE (0.45), Theils-U1 (0.82), Theils-U2 (0.88), PT (8.09) and DM. Although, the MACD did lose its in-sample ME metric to the out-of-sample ANN Fc. Regarding the ANNs, results are mixed, however, the RBF model emerges as the best performing ANN, acquiring 5 of the 10 forecast evaluation metrics of: MSE (0.41), RMSE (0.64), MPE (0.19), Theils-U1 (0.83), and PT (1.97). The MNN secured the best ME (0.01) and retained its Theils-U2 metric with a value of 0.94 which is an improvement on its in-sample value of 0.97. The RBF, PNN, and CANFIS models all recorded Theils-U2 values >1 indicating they underperformed a Naïve forecast; meanwhile the MLP and GFF models improved on their in-sample Theils-U2 values by recording values <1. The GFF model achieved the best MAE (0.48), and the PNN achieved the best MAPE (2.28). The ANN Fc model improved on its in-sample performance, acquiring 4 forecast evaluation metrics that offer improved performance over individual ANN models: ME (0.0003), MSE (0.41), RMSE (0.64), and MAE (0.48).

The Diebold-Mariano (DM) metric is applied to pairs of forecasts (ANN Vs MACD). A negative DM value indicates that the ANN model is more accurate than the benchmark MACD model. The lower the negative value, the more accurate the ANN forecast. As no ANN models recorded a negative DM value, the forecasting superiority of the MACD benchmark mode is reinforced.

To conclude the out-of-sample results, the MACD is the superior forecasting model, acquiring 6 of the 10 forecasting evaluation metrics. Regarding the ANNs, the RBF model reaffirmed its superiority amongst the ANNs by achieving the best out-of-sample results. The ANN Fc model improved upon its in-sample performance by securing the additional metrics of MSE and RMSE but did lose its in-sample PT metric to the RBF model.

#### 4.6.3 Point Forecast

Table 4-9 displays the point forecast, forecasts error, and rank. The actual value one day in advance is -0.4096.

Table 4-9: Point Forecasts

<i>Point-Forecast</i>			
<b>Model</b>	<b>Point Forecast (<i>Pf</i>)</b>	<b>Error (<i>Af - Pf</i>)</b>	<b>Rank</b>
GFF	-0.0718	-0.3378	1
MLP	-0.0619	-0.3477	2
RBF	-0.0560	-0.3536	3
ANN Fc	-0.0224	-0.3872	4
CANFIS	-0.0078	-0.4018	5
PNN	0.0136	-0.4232	6
MACD	0.0431	-0.4527	7
MNN	0.0496	-0.4592	8

Actual Value = -0.4096

From table 4-9, the results show the GFF model generated the most accurate point forecast with one day ahead prediction of -0.0718 compared with the actual value of -0.4096. Both the GFF and MLP models recorded trivial in-sample and out-of-sample performance but score highly in the point forecast, achieving rank 1 and 2 respectively. Although the



MACD recorded the best in-sample and out-of-sample forecasting performance, it ranks poorly in the point-forecast, achieving a rank of 7 of 8. Furthermore, the MACD also did not generate the correct directional sign. Regarding the ANN Fc model, although it recorded an exceptionally low out-of-sample ME of 0.0003, it achieved average point-forecast performance, ranking 4 of 8.

#### 4.6.4 Forecast Evaluation Under Increased Volatility

Table 4-10 displays the results of the models during a period of notable decline and increased volatility. The period of notable decline and increased volatility is 7<sup>th</sup> May 2014 to 13<sup>th</sup> March 2015, as depicted in figure 4-1 of section 4.3.1.

Table 4-10: Forecast Evaluation Results (07/05/2014 to 13/03/2015)

METRIC	<i>Volatile Period</i>							
	MLP	GFF	MNN	RBF	PNN	CANFIS	ANN Fc	MACD
ME	0.0388	0.0127	<b>-0.0098</b>	-0.0746	-0.0587	-0.0628	-0.0257	-0.0407
MSE	0.2035	0.2022	0.2035	0.2024	0.2081	0.2064	0.2008	<b>0.177</b>
RMSE	0.4511	0.4497	0.4512	0.4499	0.4562	0.4543	0.4481	<b>0.4207</b>
MAE	0.3291	0.3248	0.3274	0.3278	0.3242	0.3264	0.3221	<b>0.3014</b>
MPE	21.1688	10.5771	13.3303	-20.9867	<b>-2.0398</b>	-3.3037	3.1243	5.0119
MAPE	21.6702	10.828	16.0338	32.266	<b>3.9973</b>	11.9073	7.2824	8.2576
Theils-U1	0.828	0.8701	0.864	<b>0.8057</b>	0.9744	0.8945	0.9082	0.8229
Theils-U2	0.9391	0.9618	0.9512	1.1679	1.0141	1.0431	0.9996	<b>0.8749</b>
PT	1.0104	1.0104	-0.317	1.0986	-1.0104	1.434	1.5476	<b>8.0997</b>

\*\*1% Significance, \*5% Significance

From table 4-10, the MACD benchmark model reinforces its in-sample and out-of-sample performance by demonstrating superior performance during the volatile period. The MACD achieved 5 of the 9 forecast evaluation metrics: MSE (0.18), RMSE (0.42), MAE (0.30), Theils-U2 (0.87), and PT (8.10). Amongst the ANNs, results are mixed with no model emerging with distinct performance. The MLP records the best Theils-U2 value of 0.94, with the RBF, PNN, and CANFIS models all recording Theils-U2 values >1, therefore, underperforming a Naïve forecast. The GFF model secures the metrics of MSE and RMSE with values of 0.20 and 0.45 respectively. The PNN model secures the metrics of MAE (0.32), MPE (-2.04), and MAPE (4.00). The MNN records the best ME value of -0.01, the RBF records the best Theils-U1 value of 0.80, and the CANFIS model records the best PT value of 1.43. The ANN Fc model offers improvements over individual ANNs regarding MSE (0.20), RMSE (0.45), MAE (0.32), and PT (1.55), therefore it can be

claimed that the ANN Fc model provides greater stability in volatile periods by recording lower errors and greater directional change – although the ANN Fc model, like all models, failed to outperform the MACD benchmark model.

## 4.7 Trading Performance

Industry standard performance metrics are employed to evaluate the trading performance of the ANNs, namely the; Sharpe ratio (SR), Information ratio (IR), Maximum Drawdown (MDD) and Annualized Return (AR). Regarding interpretation, for the SR, IR, and AR, positive values are desired. For MDD, the smallest negative value is best. For the full description of the performance metrics, see Appendix 2.

### 4.7.1 In-Sample Trading Performance

Table 4-11 presents the in-sample trading results for the various ANNs and benchmark MACD model.

Table 4-11: In-Sample Trading Evaluation

<i>In-Sample Trading Performance</i>				
<b>Model</b>	<b>SR</b>	<b>IR</b>	<b>MDD</b>	<b>AR</b>
MLP	-3.13	-1.58	-2.64	-2.38
GFF	-3.12	-0.88	-3.92	-4.46
MNN	-1.38	-0.63	-39.74	-12.19
RBF	-0.36	0.18	-6.63	-3.74
PNN	-1.48	0.23	-0.07	-4.85
CANFIS	-0.75	0.20	-8.69	-5.25
Fc	-1.70	-0.39	-2.64	-3.24
MACD	-0.89	N/A	-2.43	-0.04

B&H Strategy = -152.64%

From table 4-11, all SRs are negative, indicating the risk-free rate (3M T-bill) is a better investment model. The worst SR values are recorded by the MLP and GFF models with SR value of -3.13 and -3.12 respectively. The model featuring the best SR is the RBF model with a value of -0.36 which is likely attributable to its in-sample and out-of-sample Theils-U1 metric. Regarding the IR, the RBF, PNN, and CANFIS models offer the best IR values of 0.18, 0.23, and 0.20 respectively, indicating they achieved consistent returns over the benchmark MACD model. Regarding MDD, the model featuring the lowest in-sample

MDD is the PNN model with a value of  $-0.07\%$ . For annualised returns, like the SRs, all models recorded negative ARs. The best AR is attributed to the benchmark MACD model with a value of  $-0.04\%$ . All models outperform the B&H strategy which recorded an AR of  $-152.64\%$ .

#### 4.7.2 Out-of-Sample Trading Performance

Table 4-12 presents the out-of-sample trading performance:

Table 4-12: Out-of-Sample Trading Evaluation

<i>Out-of-Sample Trading Performance</i>				
<b>Model</b>	<b>SR</b>	<b>IR</b>	<b>MDD</b>	<b>AR</b>
MLP	-1.77	-1.18	-3.17	1.59
GFF	-1.65	-0.71	-11.05	2.35
MNN	-1.34	-0.49	-48.73	-15.78
RBF	-0.91	0.06	-9.15	-6.93
PNN	-1.13	0.16	-0.08	-4.88
CANFIS	-1.06	0.11	-4.48	-5.67
ANN Fc	-1.36	-0.34	-3.16	-2.02
MACD	-1.06	N/A	-1.91	0.12

B&H Strategy =  $-409700\%$

From table 4-12, out-of-sample SRs all remain negative but do improve upon their in-sample values. The best SR remains attributed to the RBF model. However, the negative SRs indicate that the risk-free rate remains a better investment model. For the IR, there is a slight improvement from the in-sample. However, 4 of the 7 models recorded negative IR values, indicating the benchmark MACD model is a superior model which is reinforced by its dominance in the out-of-sample forecast results. Regarding MDD, the model featuring the lowest out-of-sample MDD is again the PNN model with a value of  $-0.08\%$ . However, there is a slight worsening in out-of-sample MDD across all models, especially the Modular MDD was  $-48.73\%$ . All out-of-sample MDDs fall within the range of  $-0.08\%$  to  $-48.73\%$ . The out-of-sample AR values are all negative but improve upon their in-sample values. All models did outperform the B&H strategy which recorded an AR of  $-409700\%$ , although it should be noted that a significant decline occurred in the EUR/USD during the

out-of-sample-period. Overall, the out-of-sample performance of all models is unsatisfactory, with the risk-free rate (3M T-Bill) proving to be a better investment model.

## 4.8 Conclusion

This research has evaluated the use of ANNs applied to forecasting the logarithmic returns of the EUR/USD exchange rate over a one-period forecasting horizon. The IVS procedure reduced a candidate input variable pool consisting of 10,568 models to an informative subset of 3 models to serve as inputs to the ANNs. A collection of ANN was then trained and tested on the EUR/USD time series data over a 5-year sample period (2012-2016), reserving the last 2-years for out-of-sample testing. The ANN models were evaluative in terms of forecasting accuracy, point-forecast, performance during increased volatility, and trading performance. Regarding forecast evaluation, the MACD benchmark model dominated both the in-sample and out-of-sample periods, clearly demonstrating its superior forecasting performance over the ANNs and ANN Fc model. The MACD reinforced its superior forecasting performance by also demonstrating the best results during the period of increased volatility. Although, the MACD did rank poorly on the point-forecast. The ANN Fc model provided notable out-of-sample forecasting performance by recording the best ME, MSE, RMSE, and MAE values amongst the ANNs, but is outperformed by the MACD model.

Trading performance is unsatisfactory both in-sample and out-of-sample for all models including the MACD, with the risk-free rate proving to be a superior investment. The research results question the forecasting and trading ability of ANNs on the EUR/USD as the evaluated ANNs failed to outperform the benchmark MACD model across the majority of forecast evaluation metrics both in-sample and out-of-sample. A potential extension of this research could investigate the use of different learning rules, extend the number of hidden layers beyond 4, implement on-line weighting opposed to batch weighting, and evaluate additional architectures. Furthermore, at the time of writing, aggregate volume data was made available regarding exchange rates. Previously, aggregate volume data was not available as the FX market is decentralised. The availability of aggregate volume data would allow for the inclusion of volume data and technical indicators that incorporate volume data in their calculation to be included in candidate variable pools. This may produce interesting IVS results and may improve ANN forecasting as volume data is

central to Dow Theory, market analysis in general, and is typically highly correlated with log-returns data.

# Chapter 5

## Volatility Forecasting with Artificial Neural Networks: A Nikkei 225 Index Example

### 5.1 Introduction

Stock market volatility is a central subject in Finance and related literature for which an enormous body of research now exists. The motivation for the hive of interest surrounding volatility largely began following the October 1987 stock market crash, referred to as Black Monday, where stock markets such as the Dow Jones Industrial Average (DJIA) declined by 508 points (22.6%) in a single day. By the end of October 1987, stock markets had declined in Hong Kong by 45.5%, Australia 41.8%, Spain 31%, the United Kingdom 26.45%, and Canada 22.5%. New Zealand's stock market suffered the worst, declining by approximately 60% from its 1987 peak. Furthermore, during the financial crisis of 2007-2009, the DJIA fell over 1,874 points (18%), in its worst weekly decline in history on a points and percentage basis. Diverging from volatility induced by a financial crisis, natural disasters have also proved capable of inducing volatile movements. For example the Japanese tsunami of 2011 caused Japan's Nikkei 225 Index to decline by 10.4% over 3 days. More recently, in 2018 the DJIA declined by 1,500 points in a single day. Volatile movements are not a new phenomenon as throughout history and especially since the post-war period, stock markets and other financial markets have recorded volatile movements, and there is no reason to believe they will not continue to do so. However, stock market volatility has been increasing over time. Consequently, the demand for volatility modelling and forecasting has also been increasing. In response, in 1982 Robert Engle devised the ARCH / GARCH model. Years later in 1989, Sir Dennis Weatherstone of JP Morgan devised the RiskMetrics approach initially comprised of the Exponentially Weighted Moving Average (EWMA), launching the service to the marketplace in 1992. During the same year, consultant Robert Whaley with the Chicago Board Options Exchange (CBOE) would develop the VIX Index with the CBOE publishing real-time VIX data in 1993. All three methods were largely well received by industry and academia, with Engle receiving the 2003 Nobel Memorial Prize in Economic Sciences for his contribution. The above mentioned methods are still heavily relied upon today, and more models have been devised

since. However, any model contains inherent limitations and with today's increasing stock market complexity coupled with new models derived from Artificial Intelligence (AI), although volatility researchers and practitioners have yet to reach a consensus, they continue to strive to improve the accuracy and robustness of volatility modelling and forecasting.

For those who deal with financial markets, accurately modelling and forecasting volatility is of substantial importance. According to Figlewski (2004), the term *volatility* implies risk, with increased volatility regarded as a symptom of market instability. Furthermore, the influence of volatility extends beyond financial markets, also impacting the functioning of the economy. According to Beckett & Sellon (1991), governments and banks recognise that the ability to accurately predict volatility is a decisive factor in making sound economic decisions, especially in periods of economic turmoil. Policymakers impacted by increased volatility may pursue regulatory reforms to directly reduce volatility or assist financial institutions such as central banks in adapting to increased volatility. In periods of instability, volatility forecasting becomes even more important as governments, banks, institutional and individual investors attempt to manage increased risk and increased volatility with diminished resources. Knowledge, understanding and the ability to estimate and accurately forecast volatility is a determining factor for endurance not only during turbulent times but also during periods of economic expansion, giving an advantage, to whoever can successfully manage future volatility. Evidently, the importance of accurately modelling and forecasting volatility is fundamental to the functioning of an economy and those involved in Finance.

The term *stock market volatility* refers to the characteristic of a stock market to sharply advance or decline in price within a short-term period, usually on a daily or weekly basis. Campbell, Lo & Mackinlay (1997) observed that one pattern that mainly distinguished Financial Economics from microeconomics was the role that uncertainty played in both theory and empirics. In the absence of uncertainty, problems of Financial Economics reduce to simple microeconomic exercises. Furthermore; Andersen, Bollerslev, Christoffersen, & Diebold (2005) observed that the main characteristic of financial data, as compared to microeconomic data, was the inherently unobserved character of volatility that evolves stochastically throughout time. A complete definition of volatility is given by Andersen *et al.* (2005, p1); "*Volatility within economics is used slightly more formally to describe without a specific implied metric, the variability of the random variable (unforeseen) component of a time series. More precisely, in financial econometrics,*

volatility is often defined as the (instantaneous) standard deviation (or  $\sigma$  “sigma”) of the random Wiener-driven component in a continuous-time diffusion model. Expressions such as “implied volatility” from option prices rely on this terminology”. By any definition, the degree of uncertainty coupled with the unobservable character of volatility transformed the return variance estimation and forecasting problem into a filtering problem in which the volatility cannot be precisely determined but instead extracted with some degree of error. In an early review essay by Cochrane (1991, p463) several concepts are considered; “*what, ultimately, is behind day-to-day movements in prices? Can we trace the source of movements back in a logical manner to fundamental shocks affecting the economy...? Are price movements due to changes in opinion or psychology, that is, changes in confidence, speculative enthusiasm...?*” Furthermore, economists and academics were concerned about the Efficient Market Hypothesis (EMH) and volatility; questioning if volatility tests reject efficiency. Cochrane (1991, p464) state that volatility tests do not prove that markets are inefficient, stating; “*Volatility tests are in fact only tests of specific discount-rate models, and they are equivalent to conventional return-forecasting tests...Thus, the bottom line of volatility tests is not ‘markets are inefficient’ since ‘prices are too volatile’, but simply ‘current discount-rate models leave a residual’ since (discounted) returns are forecastable.*” According to Shiller (1988), the determinants of financial market volatility are challenging to define because economists and researchers do not have a theorem of financial fluctuations and existent theories are often unconvincing. One explanation of financial market volatility offered by Shiller (1988) is market psychology, in which investors appear to react to each other instead of the unfolding fundamental events, with the interaction triggering sizable market swings. Shiller proved via a survey that market psychology was a key factor behind the stock market crash of October 1987, suggesting that on the day of the crash investors were not responding to any specific fundamental news items but to news of the crash itself. Mishkin (1988) agreed with Shiller that stock market volatility is difficult to explain, and although he did not fully agree with Shiller’s survey evidence, he too believed that factors other than underlying economic fundamentals could have contributed the stock market crash of October 1987.

Speculation has also been cited concerning the adverse impact of volatility. There is debate concerning speculators and their impact on volatility, suggesting that increased volatility is undesirable and reductions in volatility are desirable. However, Antoniou & Holmes (1995) state that this is misleading as it fails to recognise the link between information and volatility. Within the EMH literature, there is a direct positive relationship, with a rapid reaction between the arrival of new information and price fluctuations. Consequently, if the



flow of information increases in an efficient market, price movements will be more volatile, Antoniou *et al.* (1997). Furthermore, volatility could increase because of innovation, by reflecting the actual variability of information regarding fundamental values. Ross (1989, p17) using a simple model under the condition for no arbitrage, proved that the variance of price change is equal to the rate (or variance) of information flow, stating; *“In an arbitrage-free economy, the volatility of prices is directly related to the rate of flow of information to the market. In a simple model the two were found to be identical. This result links volatility tests to the efficient market hypothesis which specifies the information set the market uses for pricing”*. It is, therefore, reasonable to conclude that the volatility of an asset price and consequently the volatility of the market, will increase as the rate of information increases. For example: during the financial crisis of 2007-2009, much blame was placed on algorithmic trading. Although such criticisms have been downplayed, today's financial institutions employ supercomputers operating 24/7 in real-time, continually analysing price, volume and volatility levels, trading pairs, correlations and foreign exchange rates right down to the tick data on every available security. Vendors such as Thomson Reuters provide international market data feeds that propagate data into high-speed databases. For example Reuter's FastTick analytics suite can capture, sort and archive 100,000 tick transactions per second in real-time, with the data then instantaneously used by quantitative trading algorithms. The same vendor connects the institutional client directly to the trading exchange via T3 communication lines for automated order execution, a hyper-efficient process that measures its end-to-end performance in milliseconds. Furthermore, Thomson Reuters also offers software called NewScope, which reformats an electronic news document with meta tags and can process the document in less than 300 milliseconds. The meta tags mark and specifically classify the pertinent facts and figures in the news document, pass them to an automated quantitative programme to quantify the news items impact on the specific asset class. Targeted securities are then bought or sold in real-time based on the logic of the algorithm, with no human interaction, all performed in under 3 seconds. Such innovation is unquestionably impressive. However, the impact of its ill-use has already been suffered on multiple occasions. For example: the flash crash of May 6<sup>th</sup> 2010, referred to as the crash of 2:45 to reflect its incredibly short-term duration, wiped \$1 Trillion from US markets within minutes using the very technology described above by a 36-year-old London based trader operating from his parent's house, skilled enough to alter the software's code. Given the above, it is difficult to argue that innovation, increased information and the rate of information does not play a role in the volatility process.

Finance literature has focused on the ability to forecast volatility of asset returns, with Walsh & Yu-Gen Tsou (1998) stating there are many reasons why forecasting volatility is important. For example, option pricing has traditionally suffered without accurate volatility forecasts. Or Controlling for estimation error in portfolios constructed to minimise ex-ante risk, with accurate forecasts providing the ability to take advantage of the correlation structure between assets. Also, when constructing and understanding asset pricing models, the nature of volatility and its ability to be predicted must be considered, since risk preferences will be based on a market assessment of volatility. Evidently, volatility is important and has a far reach, extending beyond financial markets and impacting the economy.

## **5.2 Outline & Contribution**

This chapter contributes to the body of knowledge by empirically evaluating the use of ANNs applied to volatility modelling and forecasting. In doing so, a collection of notable ANNs of the static and temporal type are applied to forecast the volatility of the Nikkei 225 Index over a short-term forecasting horizon of one period. Initially, a pool of candidate volatility models is generated consisting of historical volatility models and implied volatility models. The pool of candidate volatility models is then subjected to a two-stage IVS process to identify the most informative models to serve as inputs to the ANNs. The ANNs are then applied to the Nikkei 225 Index over a four-year sample period with year four reserved for out-of-sample testing. The performance of the ANNs is benchmarked against a GARCH (1,1) model. Forecasting performance is evaluated using the following set of forecast evaluation metrics: ME, MSE, RMSE, MAE, MPE, MAPE, Theils-U1, Theils-U2, PT, and DM. The out-of-sample empirical results demonstrate that the GARCH (1,1) model is the superior volatility forecasting model.

## 5.3 Volatility Modelling & Forecasting

### 5.3.1 Volatility Modelling

Volatility modelling is the main challenge academics, and practitioners face. Over the decades, several models have been devised; however, the search for more robust and accurate volatility models continues. This section presents some of the primitive volatility models before progressing to sophisticated models and their application.

#### 5.3.1.1 Primitive Volatility Models

To model volatility, it is first necessary to define volatility. Figlewski (2004), lays the foundation for understanding the concept of volatility using existing Finance models. Starting with the Random Walk (RW) model, in which asset price movements are described by the following equation:

$$r_t = \frac{S_t - S_{t-1}}{S_t} = \mu_t + \varepsilon_t \quad (29)$$

Where;  $E[\varepsilon_t] = 0$ , and  $Var[\varepsilon_t] = \sigma_t^2$

Figlewski states: “the return at time  $t$ ,  $r_t$ , is the percentage change in the asset price  $S$ , over the period from  $t-1$  to  $t$ . This is equal to  $\mu_t$ , a non-random mean return for period  $t$ , plus a zero mean random disturbance  $\varepsilon_t$ , that is independent of all past and future  $\varepsilon_t$ ’s. It is the lack of serial correlation in the random  $\varepsilon_t$ ’s that is the defining characteristic of efficient market pricing: past price movements give no information about the sign of the random component of return in period  $t$ ” Figlewski (2004, p3).

Modern option pricing theory began with Black & Scholes (1973), where volatility is a key component in determining the fair value of an options contract. The input parameters required to produce a Black–Scholes option price are the current stock price  $S$ , the option strike price  $K$ , the risk-free interest rate  $r$ , the option’s remaining time to maturity  $t$ , and the future volatility of the underlying asset  $\sigma$ . All parameters except the latter one can be obtained easily by the market, Figlewski (2004).

$$\frac{dS}{S} = \mu dt + \sigma dz \quad (30)$$

Options and futures benefit from price fluctuations of the underlying asset as well as of securing a portfolio against loss or hedging a planned purchase against a possible price increase Maris *et al.* (2004). In deriving the option pricing formula, Black & Scholes needed to model stock price movements over short time intervals to adjust their trading strategy after continually rebalancing a portfolio consisting of an option and its underlying stock. The formula they adopted (above) is a logical extension of the RW model over time. This is a limiting RW process as the time interval decays to zero, keeping the mean and the variance of returns per year constant. The result is a lognormal diffusion model where the  $dS$  is the asset price change over infinitesimal time interval  $dt$ ,  $\mu$  is the mean return at an annual rate,  $dz$  is a time-independent random disturbance term with mean of zero and variance of one at  $dt$ , and  $\sigma$  is the volatility calculated as the standard deviation of the annual return, Figlewski (2004).

Volatility modelling and forecasting has been the subject of a hive of theoretical and empirical research over the decades as volatility has become one of the most important concepts in Finance and the surrounding literature. Frequently, volatility is measured by the standard deviation of log-returns as a simple risk measure. Other models such as Value at Risk (VaR) modelling for measuring market risk, and the previously mentioned Black-Scholes model for pricing options, require the estimation of volatility. To consider the returns process given by:

$$r_t = m_t + \varepsilon_t \quad (31)$$

where  $m_t$  is the conditional mean process (which could include autoregressive (AR) and moving average (MA) terms), where the error term can be decomposed as  $\varepsilon_t = \sigma_t z_t$  with  $z_t$  an idiosyncratic zero-mean and constant variance noise term, and  $\sigma_t$  is the volatility process to be estimated and forecast, with forecast values denoted  $h_t^2$ . The sample data is split between an in-sample period,  $t=1, \dots, T$ , and an out-of-sample period  $t=T, \dots, \tau$ . In order to generate a historical ‘actual volatility’ series on the basis of which volatility forecasts may be generated using the statistical models described below, the methodology by Pagan & Schwert (1990) is followed in representing historical volatility by the squared residuals,  $h_{t+1}^2$ , from a conditional mean model, for returns estimated over the in-sample period.

### 5.3.1.2 Historical Volatility

Historical volatility (HV) models make few assumptions regarding the source or changes in volatility levels. Such models are limited to assuming that past volatility is the best estimate of future volatility. Additionally, HV models imply that volatility is constant, at least as far as the forecasting horizon is concerned. Despite their limited set of assumptions, HV models still require a number of parameters to be set, for example:

- The sample size over which to estimate HV.
- Frequency of the sample data, i.e. daily, weekly, monthly.
- The time series to be employed (H, L, O, C) and which volatility calculation method to employ.

Points one and two are somewhat related given that sufficient sample size is required to achieve meaningful estimates and minimise the impact of noise. However, the question of sample size is common, and a definitive answer is not agreed in the literature. The selection of sample size appears to be a function of the specific assumptions made by researchers of what constitutes a sufficient data set. The objective is to select a sample size that reflects the current or future volatility level and not to include observations so old as to no longer be relevant to the security. Regarding sample frequency, the selection of frequency appears to depend on the personal preferences of researchers in addition to the objectives of the research, with a clear preference for any one frequency not emerge from the literature. This is perhaps due to the conflicting evidence between theoretical research and practitioners' experiences. For example Poon & Granger (2003) state that: *"In general, volatility forecast accuracy improves as data sampling frequency increases relative to forecast horizon (Andersen, Bollerslev & Lange, 1999)"*. It is noteworthy, as discussed by the same authors that Figlewski (1997) found that long-term forecasts benefit from aggregation (lower sampling frequencies). This is contrary to the theoretical discussion by Drost & Nijman (1993), who show that aggregation should preserve the features of volatility. However, Poon & Granger (2003) state; *"it is well known that this is not the case in practice...and that this further complicates any attempt to generalize volatility patterns and forecasting results."* In light of these results and the conclusions reached by these widely-cited authors, the question arises if any results from previous research are applicable to a specific problem. The estimation and evaluation methodologies rather than the resulting findings are the likely contribution of volatility modelling research. The above discussions regarding aggregation, sampling frequency and sample size are of some importance to HV models and stochastic volatility (SV) models.

They both typically assume a single observation, i.e. a single price or return, on which to base the volatility estimate. However, this univariate approach is only one of the choices and is commonly applied to daily, weekly or monthly returns or closing prices. Typically, volatility is calculated as the standard deviation of log-returns over the sample period. Alternatively, the trading range can be used, a technique frequently implemented before the availability of high-frequency data. Parkinson (1980) devised a volatility estimator utilising high and low prices instead of the closing price. Conversely, Garman & Klass (1980) devised a combination of the trading range and the close where the range serves as the current observation and the change in closing price as a reference point. For related literature, see Haug (2007). The HV models presented below are traditional models, most of which are still implemented today. The list of HV models is extensive. Therefore, only notable models will be described.

#### **5.3.1.3 Random Walk (RW)**

If volatility randomly fluctuates, the optimal forecast of next period's volatility is simply the current period's actual volatility:

$$h_{t+1}^2 = \sigma_t^2 \quad (32)$$

The RW model suggests that the optimal forecast of volatility is for no change since the last observation.

#### **5.3.1.4 Historical Average**

The historical average is perhaps the most basic means of forecasting future volatility. If the distribution of volatility has a constant mean, all variation in estimated volatility can be attributed to measurement error, and the historical average gives an optimal forecast for all future periods;

$$h_{t+1}^2 = \frac{1}{t-T} \sum_{t=1}^T \sigma_t^2 \quad (33)$$

#### 5.3.1.5 Simple Moving Averages (SMA)

Under the SMA method, volatility is forecast by an unweighted average of previously observed volatilities over a specified moving window of arbitrary but fixed length:

$$h_{t+1}^2 = \frac{1}{p} \sum_{j=1}^p \sigma_j^2 \quad (34)$$

Where:  $p$  is the moving window.

#### 5.3.1.6 Exponential Smoothing

Under exponential smoothing, a one-step-ahead volatility forecast is a weighted function of the previous volatility forecast and actual volatility:

$$h_{t+1}^2 = \varphi h_t^2 + (1 - \varphi) \sigma_t^2 \quad (35)$$

Where  $\varphi$  is a smoothing parameter bound between 0 and 1, such that for  $\varphi=0$ , the exponential smoothing model reduces to an RW model. While for  $\varphi=1$ , weight is given only to the prior period's forecast. The value of  $\varphi$  is determined empirically by that value which minimises the in-sample sum of squared prediction errors.

#### 5.3.1.7 Exponentially Weighted Moving Average (EWMA)

Perhaps the most notable HV model is the EWMA model devised by JP Morgan in 1989 as part of their RiskMetrics approach to volatility which received wide acceptance in Finance and academia. The EWMA model is similar to the exponential smoothing model; however, HV is replaced with an EMA forecast:

$$h_{t+1}^2 = \varphi h_t^2 + (1 - \varphi) \frac{1}{p} \sum_{j=1}^p \sigma_j^2 \quad (36)$$

The EWMA is an extension of the HV measure permitting recent observations to have a larger influence on volatility forecasting than past observations. When applying the EWMA, the current observation carries the largest weight, with weights of prior observations declining exponentially over time, with a decay factor ideally between 0.94 and 0.99, with a suggestion of 0.96. The EWMA model is best applied when the objective is to capture the short-term movement of volatility. Conversely, due to its short-term bias, it does hold limitations when accounting for the volatility characteristic of long-term memory.

#### 5.3.1.8 Smooth Transition Exponential Smoothing method

These models allow parameters to adapt to changes in the characteristics of the time series. Taylor (2004) proposes the use of a logistic function of a user specified variable adaptive smoothing parameter:

$$h_{t+1}^2 = \alpha_t + \varepsilon_t^2 + (1 - \alpha_t)h_t^2 \quad (37)$$

Where:  $\alpha = 1/(1 + \exp(\beta + \gamma V_t))$

The smoothing parameter varies between zero and one, adapting to changes in the transition variable  $V_t$  where  $\varepsilon_t$  and  $[\varepsilon_t]$  serve as transition variables in a similar way the sign and size of past shocks have been used as transition variables in non-linear GARCH models.

#### 5.3.1.9 Standard Deviation

It was previously mentioned that a measure of HV is the standard deviation of log-returns. In a review paper by Poon & Granger (2003) they make a distinction between standard deviation, volatility and risk. Standard deviation ( $\sigma$ ), or variance ( $\sigma^2$ ), is computed from a set of observations as:



$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{t=1}^N (R_t - \bar{R})^2 \quad (38)$$

where  $\bar{R}$  is the average return.

They argue that standard deviation is only the correct dispersion measure for the normal dispersion measure for the normal distribution. The link between volatility and risk is a questionable one. Risk is usually associated with negative returns, whereas most measures of dispersion make no distinction. The examples mentioned by Poon & Granger (2003) are: 1) the Sharpe Ratio, defined as excess returns over the risk free rate divided by the standard deviation which is frequently used as an investment performance measure occasionally penalizes infrequent large returns and 2) the *semi-variance*, a concept developed by Markowitz (1991), where only the squared returns below the mean are utilised, but this method is not easy to apply, and it is not widely used.

More recently, advancements in the field of Technical Analysis have been made concerning its measurement of volatility. Kirk Northington CMT, believing that: *“Current technical analysis leaves the individual and private trader ill-prepared to participate in today’s markets because it does not adequately utilize volatility measurement”*, developed a new set of technical indicators, incorporating statistical techniques to estimate volatility better. For example the RSIV, best described as an adaptation of the classic Relative Strength Index (RSI) but adapted to utilise regression in its calculation, for more details see; *Volatility-Based Technical Analysis* by Kirk Northington CMT.

#### 5.3.1.10 Stochastic Volatility (SV)

Despite the various choices and parameters to be set in the process of developing an HV model, they are comparatively basic and consistent. However, over time, researchers realised they could not reflect various characteristics that are frequently present in observed volatility, i.e. heteroscedasticity, volatility clustering, and the long memory of volatility shocks. In response, and combined with the general view by Finance professionals that volatility represents an asset in its own right, led to the development of more complex models that address forecasting power and the statistical properties of the model and its estimation procedures. The two notable classes are stochastic volatility (SV)

models and autoregressive conditional heteroscedasticity (ARCH) models (Engle, 1982), in particular. While neither is particularly new, they are still actively researched. The purpose is to introduce them only to the degree necessary for a reference model implementation.

### 5.3.2 ARCH/GARCH Modelling

This section describes a selection of the most popular models in the Finance literature models belonging to the ARCH family. The ARCH model was developed by Engle (1982), in his attempt to test for the ARCH effects on the variance of the United Kingdom inflation, and later reviewed by Engle and Bollerslev (1986). The model accounts for the difference between the unconditional and conditional variance of a stochastic process. While traditional econometric models operate under the assumption of a constant variance, the ARCH process allows the conditional variance to vary over time, leaving the unconditional variance constant. To model for ARCH effects in the conditional variance of a random error,  $\varepsilon_t$  we have:

$$h_t^2 = \text{Var}(\varepsilon_t | \Omega_{t-1}) \quad (39)$$

Where;  $h_t^2$  is the conditional volatility and  $\Omega_{t-1}$  is the information set.

The ARCH (q) specification is given by:

$$h_t^2 = \omega + \sum_{i=1}^q a_i \varepsilon_{t-i}^2 \quad (40)$$

#### 5.3.2.1 GARCH (Generalised ARCH)

The GARCH model of Engle (1982) and Bollerslev (1986) requires the joint estimation of the conditional mean model and the variance process. On the assumption that the conditional mean stochastic error,  $\varepsilon_t$ , is normally distributed with zero mean and time-varying conditional variance,  $h_t^2$ , the GARCH (1,1) model is given by:

$$h_{t+1}^2 = \omega + \alpha \varepsilon_t^2 + \beta h_t^2 \quad (41)$$

Where all the parameters must be positive, while the sum of  $\alpha + \beta$  quantifies the persistence of shocks to volatility. The GARCH (1,1) model generates one-step-ahead forecasts of volatility as a weighted average of the constant long-run or average variance,  $\omega$ , the previous forecast variance,  $h_t^2$ , and previous volatility reflecting squared *news* about the return,  $\varepsilon_t^2$ . In particular, as volatility forecasts are increased following a large return of either sign, the GARCH specification captures the well-known volatility clustering effect. The GARCH models are also capable of capturing leptokurtosis and skewness which are the features most often observed in empirical analysis.

The comprehensive review by Poon & Granger (2003), also discussed by Poon & Granger (2007), revealed that implied volatility (IV) is usually superior to models from the ARCH-family, but the difference is not as clear in the case of HV compared with IV and even less so when comparing HV and ARCH-family models concluding: “*as a rule of thumb, historical volatility methods work equally well compared with more sophisticated ARCH class and SV models*” (Poon & Granger, 2003). The authors stress that it is critical to understand the objective of the volatility model and set selection criteria accordingly. They also emphasised that model performance is dependent on the security being examined. Lastly, the review revealed that the GARCH (1,1) model is among the most commonly implemented in the ARCH family, a point also made across the wider literature.

### 5.3.3 Implied Volatility (IV)

As previously stated, the Black-Scholes option pricing formula dictates that the option price is a function of; the price of the underlying asset, the strike price, the risk-free interest rate, the time to option maturity and the volatility of the underlying asset. Given that the above parameters are observable, once the market has produced a price for the option, volatility could be derived using backward induction, and then use the volatility value that the market used as input. This measure of volatility is called option implied volatility (OIV). OIV is often interpreted as a market’s expectation of volatility over the option’s maturity. Because each asset can have only one volatility measure, difficulties arise when options with similar maturities but different strikes produce different IV estimates for the same asset, Granger & Poon (2005). Examples of earlier studies where the basic Black-

Scholes option pricing model was used are Latane & Rendleman (1976), Chiras & Manaster (1978) and Beckers (1981). Studies also examined IV as a source of information, for examples studies by Day & Lewis (1990) conclude that time-series models of conditional volatility outperform IV, and Lamoureux & Lastrapes (1993) find that information contained in HV is superior to that contained in IV. Furthermore, Canina & Figlewski (1993) find that IV does not correlate with future volatility. Conversely, many studies are found in favour of IV; Jorion (1995), Fleming (1998), Christensen & Prabhala (1998), and Christensen & Hansen (2002). More recently, Christensen, Hansen & Prabhala (2001), Blair, Poon & Taylor (2001), Ederinton & Guan (2002), Pong *et al.* (2004) and Jiang & Tian (2005) employ larger sample sizes, high frequency data and account for structural changes, concluding that IV is a more efficient forecast for future volatility than HV. Lastly, in a review paper by Granger & Poon (2005), it is concluded that the predictive ability of IV cannot be ignored nor underestimated compared to other volatility models; *“implied volatility appears to have superior forecasting capability, outperforming many historical price volatility models and matching the performance of forecasts generated from time series models that use a large amount of high frequency data”* (p. 489-490).

#### **5.4 Stylised Facts: Returns & Volatility**

Regarding financial time series data, a stylised fact is regarded as a characteristic that is consistent enough to be generally accepted as fact. Market participants tend to hold an event-based approach to understanding financial markets in which movements are rationalised by connecting them to fundamental data. As different assets classes are impacted by class-specific news items (with a small degree of overlap) under this model, it would be expected to see little or no commonality in their respective time series. However, researchers have demonstrated remarkable amounts of similarity in time series data across asset classes. Although the qualifier is important - they are generalities. Gaining generality permits the recognition of broad similarities, but comes at the cost of diminished capacity to make precise statements. Consequently, many stylised facts are of qualitative nature. Furthermore, it is surprisingly difficult to incorporate such characteristics into a single model. Fortunately, model developers do not seek to develop models that capture all these characteristics. Instead, they strive to develop models that capture one or a small number of these characteristics, which perform well in accurately estimating volatility. Following the

works of Mandelbrot (1963) and Fama (1965), stylised facts include the following characteristics:

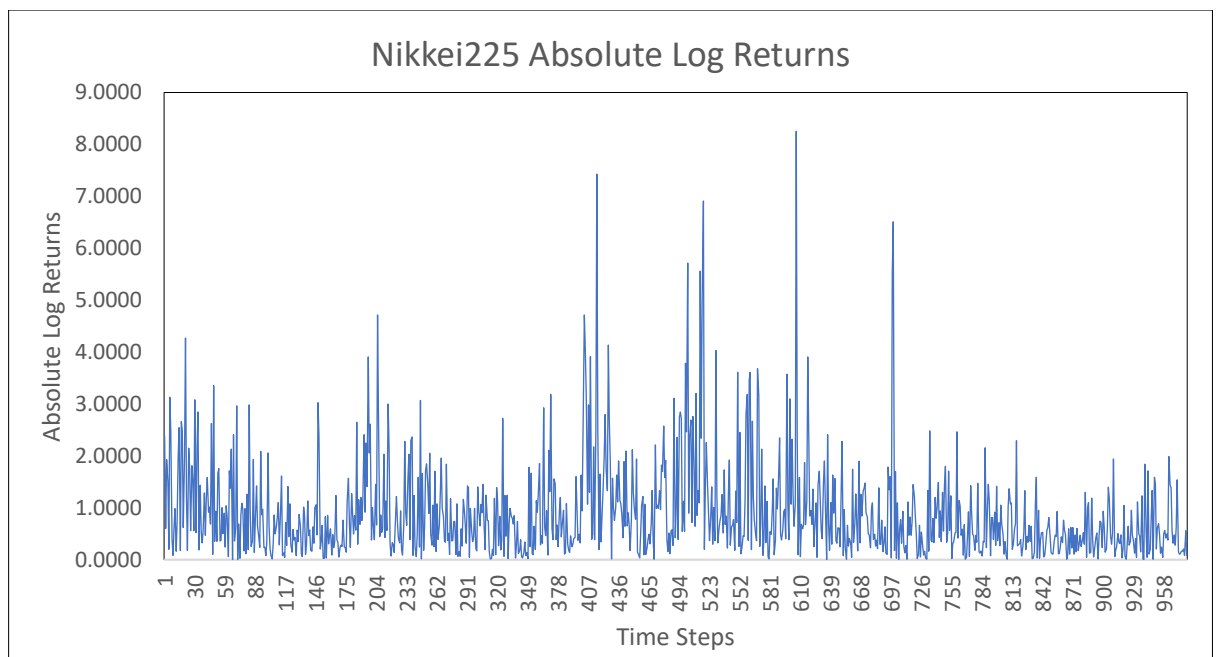
- Volatility is not constant, mean-reverts, clusters and possesses long memory.
- Large returns occur frequently followed by subsequent aftershocks.
- Returns and volatility are negatively correlated.
- Volatility and volume share a strong positive relationship.
- The distribution of volatility is close to log-normal.

Given the above-stated characteristics, a discussion will now follow relating the characteristics to the Nikkei 225 Index used in this investigation as well as citing relevant prior research.

#### 5.4.1 Volatility is not Constant

Volatility is not constant, a characteristic confirmed in the early but notable studies of Akgiray (1989) and Turner & Weigel (1992). The characteristic is pertinent to any estimation of volatility and is best demonstrated visually. Figure 5-1 displays the volatility of the Nikkei 225 Index over this investigations sample period using absolute log-returns as a proxy for volatility:

Figure 5-1: Nikkei 225 Index Absolute Log-Returns (2014-2017)



From figure 5-1, it is observable that not only does volatility change, it also changes in specific ways. Observation reveals that volatility clusters which was initially observed by Mandelbrot (1963) who stated: *‘Large changes tend to be followed by large changes...and small changes tend to be followed by small changes’*. Furthermore, volatility clustering is not unique to any specific security, holding true across all financial markets (Taylor 1986). It is also true that absolute log-returns are more highly correlated than squared returns. The slow decay of autocorrelation is also referred to as long memory. Ding *et al.* (1993, p83) state that: *“... not only is there substantially more correlation between absolute returns than returns themselves, but the power transformation of the absolute return also has quite high autocorrelation for long lags”*. It can, therefore, be argued that the long memory characteristic appears to be present for which models capturing volatility need to account for. Clustering dictates that an accurate estimate of future volatility is current volatility. Volatility clustering also tends to display the following:

- Clustering tends to be greater in developed markets than in emerging markets.
- Clustering is more pronounced in bear markets than bull markets, with faster decay of autocorrelation in bear markets.
- Autocorrelation decays fastest during a financial crisis.

#### **5.4.2 Mean Reversion**

A characteristic that aids volatility forecasting is mean-reversion. As positive autocorrelation decays, short-term volatility reverts to its long-term mean. A robust test for mean-reversion is the variance ratio by Campbell, Lo & MacKinlay (1997). When the volatility measured from daily returns is greater than that which is measured from longer-term returns, mean reversion tends to occur. However, while volatility mean reverts, it is often not clear what the current mean is, as short-term fluctuations oscillate around a slowly changing mean value. The changing mean value of volatility can be significant as depicted in figure 5-2:

Figure 5-2: Changing Mean Values, Nikkei 225 Index

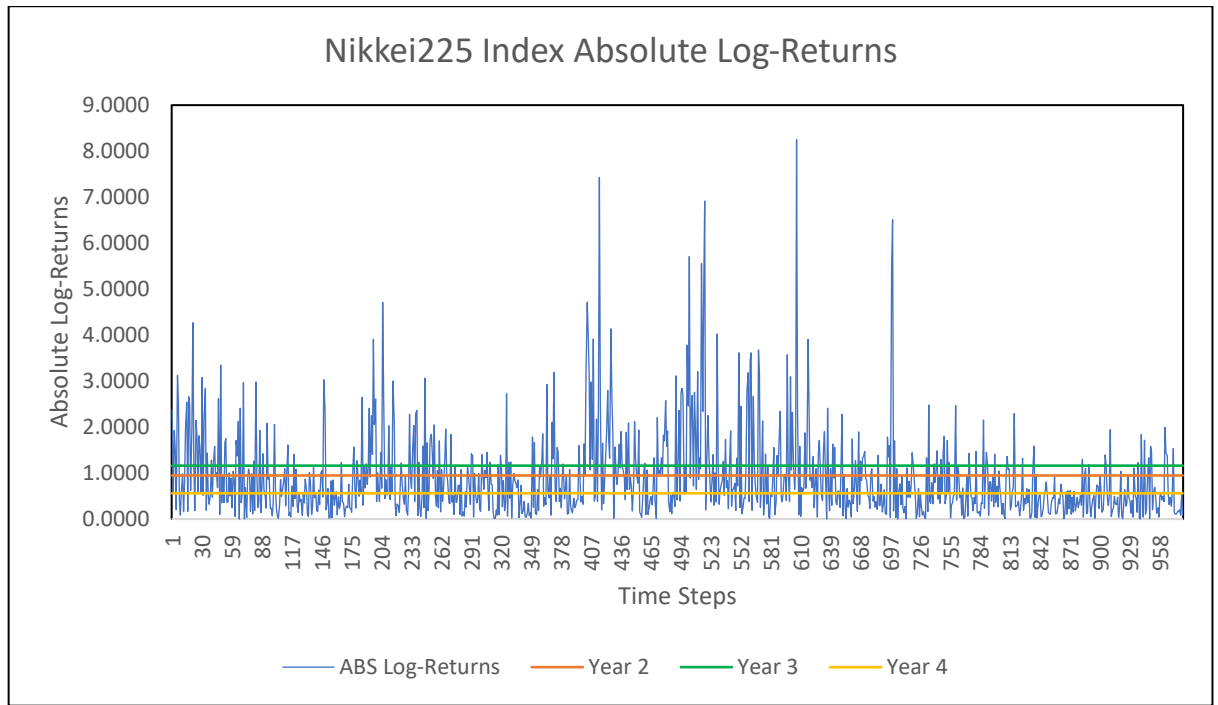


Figure 5-2 displays the volatility of the Nikkei 225 Index taken during years 2, 3 and 4 of this investigations sample period. The mean volatility of years 1 and 2 are similar with values of 0.93 and 0.94 respectively. However, the difference in mean volatility between years 2, 3 and 4 is large (0.94, 1.16 and 0.55 respectively). A practical solution is to take the mean volatility of the sample period, in this case, 0.90. Campbell, Lo, & MacKinlay (1997, p.481) state that: *“it is both logically inconsistent and statistically inefficient to use volatility measures that are based on the assumption of constant volatility over some period when the resulting series moves through time”*. This statement relates to the characteristic of volatility clustering in which large changes are followed by large changes and small changes followed by small changes as observed by Mandelbrot (1963). Therefore, given that volatility is not constant, such a characteristic naturally brings into question the utility of the historical average model. As has been demonstrated, the historical average is sample-specific and can dramatically change when recalculated on a periodic basis. However, it can be argued that the utility of the historical average model resides in the characteristic of mean reversion, aiding mean-reverting volatility models. The interplay of volatility clustering and mean reversion dominates the dynamics of volatility research.

### 5.4.3 Returns Distribution

Stock market returns are not normally distributed; they are typically negatively skewed and feature excessive kurtosis. This characteristic was initially documented by Mills (1927) but has since been documented by: Mandelbrot (1963), Fama (1963, 1965), Clark (1973), and Blattberg & Gonedes (1974). The presence of excess kurtosis or *fat tails* means that large returns occur more frequently than one would expect if returns were normally distributed. For example: using S&P500 data Sinclair (2013, p45) states: *'Between 1950 and 2011, the daily logarithmic returns of the S&P500 have had an excess kurtosis of 21.3. There have been 24 days with returns less than negative 5 percent and 17 days with returns greater than 5 percent. The worst day (19<sup>th</sup>, October 1987) had a return of -20.47 percent. If the returns were normally distributed with a volatility of 20 percent, the probability of such a decline is  $10^{-88}$ .*' Another characteristic of kurtosis is that most of the kurtosis in stocks establishes during overnight returns. This is because many of the fundamental events that significantly impact stock prices are released when the market is closed, i.e. earnings announcements. Cornell (1978), Patel & Wolfson (1979, 1981) show that stock market volatility increases around earnings announcements, with Harvey & Huang (1991, 1992) find that fixed income and FX volatility increases during periods of macroeconomic announcements.

### 5.4.4 The Leverage Effect

Another characteristic of volatility is that it tends to increase when stock prices decline - this is referred to as the leverage effect. The leverage effect dictates that a fall in stock price will cause a company's financial leverage to increase which subsequently increases its risk which consequently leads to increased volatility. While this sounds plausible, it does not explain the effect in practice, Figlewski & Wang (2000). The leverage effect is not a new observation, it is mentioned in Black (1976), and Christie (1982) and numerous studies have been published since; Edderington & Guan (2010) and Abura & Wagner (2010). The effect is prevalent for stock indices but also true for individual stocks, bonds and commodities. It seems to be a characteristic of any security in which participants invest and expect a positive return. For example, it does not apply to the Foreign Exchange (FX) market which is predominantly speculative. The asymmetry is also robust with respect to how it is measured. Most studies have looked at correlations between returns and



variances (or volatilities), but other methods show the same broad effect. For example: in a similar approach to the first exit time volatility, Jensen, Johansen & Simonsen (2003) analyse the first exit times for an asset to breach a corridor on the upside and the downside. They show that the first exit time is smaller for downside exits than it is for upside exits. Another way of quantifying asymmetry is to calculate the average size of positive and negative returns. For example: over this investigations sample period (2014 to 2017), the average positive daily return for the Nikkei 225 Index is 0.87, and the average negative return is -0.93. Therefore the average negative returns are approximately 7% larger.

#### **5.4.5 Volatility & Volume**

Trading volume is highly correlated with all measures of volatility. Establishing the casualty of the relationship is difficult. However, valid arguments can be made for trading volume moving the underlying price and inducing volatility, but also volatility attracting market participants thus increasing volume. Whatever the causation is, the relationship itself is strong and holds over all timeframes, Tauchen & Pitts (1982) and Lee & Rui (2002). For example, the correlation between daily absolute log-returns and daily volume over this investigations sample period is 0.49 indicating a moderate to strong positive relationship. Furthermore, trading volume is associated with information flow with many studies demonstrating that the performance of volatility models can be significantly improved with the inclusion of volume data in their model specification, Taylor (2008).

#### **5.4.6 Volatility Distribution**

Research has discovered that the distribution of volatility is log-normal; Anderson, Bollerslev, Diebold & Ebbens (2001) and Cizeau *et al.* (1997). The particular distribution is less important; the important aspect is that the distribution is skewed to the right with more periods of increased volatility than would be expected if the distribution was normal. An alternative way of expressing this is that volatility spends more time at low levels than high levels. Consequently, the volatility distribution is significantly different during bull and bear markets. For example Sinclair (2013, p45) states: *‘From 1990 to 2011, if the S&P500 was below its 200-day MA the median 30-day volatility was 21.6% If it was above*

*its 200 day MA, it was 12.1%''*. This characteristic is robust concerning how volatility is modelled, and the length of the MA employed to define the market as bullish or bearish.

#### **5.4.7 Co-Movements in Volatilities**

Black (1976), mentions that in general when volatilities change, they all tend to change in the same direction. Several studies support this argument of the existence of common factors explaining volatility movements in stock and exchange rates. Engle *et al.* (1990) demonstrate that US bond volatility changes are closely linked across maturities. The commonality of volatility changes holds not only across assets within a market but also across different markets. For example Schwert (1989) found that US stock and bond volatilities move together. Furthermore, Engle & Susmel (1993) and Hamao *et al.* (1990) discovered links between volatility changes across international stock markets. As Bollerslev *et al.* (1994) state, the fact that volatilities move together should be encouraging to model developers, as it indicates that a few common factors may explain much of the temporal variation in the conditional variances and covariances of asset returns, which is the basis of the ARCH modelling.

#### **5.4.8 Macroeconomic Variables & Volatility**

Stock prices are considered to be related and closely linked to the health of the economy, therefore, it is logical to expect that measures of economic uncertainty such as conditional variances of industrial production and interest rates, should help explain changes in stock market volatility Bollerslev *et al.* (1994). Schwert (1989), finds that although stock market volatility increases during recessions and decreases during expansions, the correlation between macroeconomic uncertainty and stock volatility is surprisingly weak. Conversely, Glosten *et al.* (1993), uncover a strong positive relationship between stock market volatility and interest rates. Engle & Rangel (2005) is a prominent example of exploring links between macroeconomic fundamentals and stock market volatility. They implement a spline-GARCH model to isolate low-frequency volatility, using the model to explore the links between macroeconomic fundamentals and low-frequency volatility across fifty nations of varying sample periods using daily data. They find the low-frequency component of volatility is greater when the macroeconomic factors

of GDP, inflation, and short-term interest rates are more volatile or when inflation is high, and GDP is low. Volatility is higher not only for emerging markets and markets with small numbers of listed companies and market capitalisation relative to GDP but also for large economies. Engle, Ghysels & Sohn (2006) revisit the relation between stock market volatility and macroeconomic activity using a new class of component models that distinguish short-run from long-run movements by blending the spline-GARCH approach with a mixed data sampling (MIDAS) approach. They find that including macroeconomic data in volatility models benefits both short-term and long-term forecasting. Furthermore, Diebold (2007) examines a broad international cross-section of stock markets, finding a clear link between macroeconomic fundamentals and stock market volatilities, with volatile fundamentals translating into volatile stock markets. Lastly, Chan *et al.* (2014) demonstrate that: interest rates, inflation, exchange rate, government debt and industrial production have significantly influenced the Nikkei 225 Index over the long-term.

## 5.5 Literature Review

Although a popular research topic, the literature has not reached a consensus regarding the best model in forecasting volatility; this is because results are sample-specific and vary between bull and bear markets. The lack of a universal set of standard performance metrics perhaps also contributes to the lack of consensus. However, a variety of volatility models have been proposed over the years from the standard deviation of log-returns to simple models such as the: RW, Historical Average, SMAs, Exponential Smoothing, EWMA, Regression, ARCH-family models and even hybrid models. In 1982 the introduction of the ARCH model by Engle and its subsequent generalisation, the GARCH model, by Bollerslev (1986) was an important milestone in the volatility modelling and forecasting literature. The GARCH model triggered the development of ARCH/GARCH generalisations, with the most influential generalisation being the EGARCH by Nelson (1991). Relatively recently a further model belonging to the ARCH-family was developed, the CEV-ARCH by Fornary & Mele (2005). The list of ARCH-family models continues to grow, with more generalisations being proposed, including the YAARCH model – an acronym for Yet Another ARCH model. More recently, ANNs have been applied to volatility forecasting with the expectation that their nonlinear and data adapting capabilities coupled with the memory components of certain ANN architectures would enhance the domain of volatility modelling and forecasting. In respect of the ARCH-family

and their contribution to volatility modelling and forecasting, this literature review will commence with a chronological review of the relevant literature by notable authors. The next section will present a review of ANNs applied to volatility forecasting.

### 5.5.1 ARCH-Family Volatility Forecasting

Financial market volatility has been an important research topic throughout past decades, with much research examining volatility modelling and volatility forecasting across all financial markets. Developments in financial econometrics advocated the use of nonlinear time series models to model investor sentiment towards risk and expected returns. For example, Bera & Higgins (1993, p315) state that: *“A major contribution of the ARCH literature is the finding that apparent changes in the volatility of economic time series may be predictable and result from a specific type of nonlinear dependence rather than exogenous structural changes in variables.”* This led to the further development of new models of non-linear nature. Frances & Van Dijk (1996) studied the performance of ARCH-family models, focusing on two non-linear generalisations of the GARCH model to forecast stock market volatility on a weekly basis. The models were the GJR-ARCH by Glosten, Jagannathan & Runkle (1992), also known as the Threshold GARCH (TGARCH) and the Quadratic GARCH (QGARCH) initially introduced by Engle & Ng (1993). These models best describe the frequently observed negative skewness found in stock market indexes. The dataset consisted of national stock indexes; DAX (Germany), EOE (Netherlands), MAD (Spain), MIL (Italy) and VEC (Sweden), with each index spanning a nine year sample period (1986 to 1994). Based on MSE, the authors find the QGARCH model to be superior to the GJR model when the sample does not contain extreme events such as those induced by a financial crisis. Walsh & Yu-Gen Tsou (1998) compared four methods of volatility forecasting including; the naïve approach which uses HV to forecast future volatility, the Improved Extreme-Value (IEV) model by Kunitomo (1992) which accommodates extreme observations and a drift term in the stochastic process, ARCH/GARCH models and the EWMA model. The authors conclude that the EWMA model is the superior forecasting model, closely followed by the GARCH models. The authors note that both the IEV and HV models are inferior by comparison and the diversification benefit that arises from indexes with larger numbers of stocks appears to increase the accuracy of volatility forecasting. However, as the sampling interval is decreased, the non-trading effects evident in the larger stock indexes begins to counteract

this benefit. Chong *et al.* (1999) investigate the performance of the GARCH model and its generalisations, using daily returns of the Kuala Lumpur Stock Exchange (KLSE) and related indexes including the; Composite Index, Tins Index, Plantations Index, Properties Index, and Finance Index. The models are the standard GARCH, unconstrained GARCH, non-negative GARCH, GARCH-M, EGARCH and IGARCH, with model parameters and variance processes estimated jointly using the maximum likelihood method. The performance of the in-sample estimation is diagnosed using several goodness-of-fit statistics. The results observe that, among the models, even though the EGARCH is not the best model in the goodness-of-fit statistics, it presents superior performance in describing the skewness in stock market indexes and in out-of-sample forecasting at a one-step-ahead horizon. The IGARCH model proved to be the worst model in both respects. Tseng *et al.* (2009) proposed a new hybrid model for stock market volatility forecasting using a Grey-GARCH model and concluded that proposed model enhanced the one-period-ahead volatility forecasts of the GARCH model evaluated using MAE, RMSE, and MAPE metrics. However, the model failed to outperform the GARCH (1,1) model in certain cases. Chia *et al.* (2012) employ long-memory models applied to sixteen agricultural commodity futures. The class of fractional GARCH models, namely the FIGARCH, FIEGARCH and FIAPARCH are estimated and compared with GARCH, EGARCH and APARCH models. The estimated  $d$  parameters, indicating long-term dependence, suggest that fractional integration is found in most of the agricultural commodity futures returns series. Additionally, the FIGARCH (1, $d$ ,1) and FIEGARCH (1, $d$ ,1) models are found to outperform their GARCH (1,1) and EGARCH (1,1) counterparts.

Conversely, in the ongoing debate to find the best volatility forecasting model, early empirical studies examining the forecasting ability of ARCH-family and simple models often concluded in favour of the simple models, see Cumby *et al.* (1993) and Jorion (1995 & 1996). Figlewski represents the academics in favour of simple models, arguing that volatility models based on SMAs of HV are superior to GARCH models, Figlewski (1997). Engle (2002) reviewed the utility of GARCH models since their introduction and found mixed results. For example; when trading options, ARCH-family models acted as indicators of options mispricing leading to trading opportunities - this was evident initially, but more recent data has failed to support the view that ARCH-family models lead to significant trading opportunities. As Engle (2002) states, this is not surprising since ARCH-family models have a limited information set and have long since been made available to mass market participants; therefore, eroding their competitive advantage. Since

2002, Engle examined the Finance literature, finding high-frequency volatility models dominant throughout. The availability of high frequency and intra-daily data introduced the concept of HV. Volatility models derived from high-frequency data should be more accurate, allowing for forecast efficiency gains, however, Engle & Gallo (2006) state that the dependence of the model upon the frequency of data makes it difficult to conclude. A major identified problem is serial correlation in the returns. According to the same authors, although the literature regarding HV does present promising results, the question of how to improve the accuracy of volatility forecasting remains. According to Hansen *et al.* (2010), high-frequency data and the inception of HV initiated several HV models including the realised variance, bipower variation and realised kernel model to name a few. These models are proven to be more informative concerning the current level of volatility rendering HV useful in volatility modelling and forecasting. As a result, the estimation of GARCH models that include an HV measure in the GARCH specification was put forward by Hansen *et al.* (2010). Poon & Granger (2003, 2005) present a review examining the wide variety of models employed in volatility modelling and forecasting. They review ninety-three papers on the topic and after classifying the different models into the four categories of; HV models, GARCH-family models, IV and SV models they conclude the following. HV models outperform GARCH-family models in twenty-two studies with the GARCH models found to be superior in seventeen studies. HV models were found to be superior to IV models in eight studies, and twenty-six IV models were found to be superior. Lastly, GARCH-family models outperformed IV models in one study, with the opposite result found in seventeen papers. Evidently, the results are mixed. Overall, the IV models appear to provide superior forecasts ability than HV models and GARCH-family models. An important aspect of the review is the authors concluding that financial market volatility is forecastable. However, the problem resides in identifying the appropriate models and the relevant parameters that would assist in generating more accurate volatility forecasts.

Eventually, the literature shifted from introducing new models to improving parameter estimation of the existing models. Comparisons within the GARCH-family in a study by Hansen & Lunde (2005) founds the GARCH (1,1) to be superior to the other family models. Model specification comparisons have also been conducted within a risk management framework, examining VaR, Kuester *et al.* (2006) and Dimitrakopoulos *et al.* (2010) in which the authors specifically examine emerging economies. Furthermore, Brownlees & Gallo (2010) use a selection of volatility models such as; GARCH,

unconditional variance, HV, and RiskMetrics models, finding that these models are outperformed by the VaR model when Ultra High-Frequency Data (UHFD) is used. According to McAleer & Caporin (2011), when drawing comparisons, the problem of unfair comparison arises; how can comparisons and rankings take place when models are characterised by different architectures? In their review they provide an empirical comparison of a set of models over eighty-nine US stocks, using a range of direct and indirect model comparison metrics also taking into account cross-sectional influences. They concluded that more research is required on the topic focusing on the methodological approach to model comparison and rankings.

### **5.5.2 Artificial Neural Network Based Volatility Forecasting**

Ormoneit & Neuneier (1996) used an MLP and a density estimating ANN to forecasting the volatility of the DAX Index. Their results showed that the density estimating ANN leads to better performance when compared to that of MLP. Donaldson & Kamstra (1997) proposed the use of ANNs combined with the GARCH model for capturing the impact of volatility on stock returns, applying their models to the daily log-returns of the S&P500, TSEC, Nikkei 225 and the FTSE100 Index spanning a sample period from January 1<sup>st</sup>, 1969 to December 31<sup>st</sup>, 1990. In sample and out-of-sample evaluation show that the ANN models capture volatility effects that are not captured by GARCH, EGARCH or GJR models. Additionally, they also document important differences between volatility in international markets, such as the substantial persistence of volatility effects in Japan relative to North American and European markets. Hamid & Iqbal (2004) predicted the volatility of commodities using an ANN. Then compare their results to the IV obtained by using the Barone-Adesi and Whaley options pricing model (1987) to contrast both predictions with the actual volatility. The results indicate that the predictions made by the ANN substantially improved the predictions obtained through the IV. Aragonés *et al.* (2008) analyse the use of ANNs to improve ‘traditional’ volatility forecasts models, as well as IV obtained from options on futures on the Spanish stock market index (IBEX-35). They explore the predictive ability of ANNs that incorporate both IV and historical time-series data. Their results show that the general regression ANN forecasts improve the information content of IV and enhance the predictive ability of the models. Their analysis is also consistent with the results from prior research showing that IV is an unbiased forecast of future volatility and that traditional time-series models have lower explanatory

power than IV. Conversely, according to Mantri, Gahan, & Nayak (2010), ANNs are no different in their ability to forecast stock market volatility compared to GARCH, EGARCH, IGARCH, and GJR-GARCH models. Using the H,L,O,C values of two Indian stock indexes (BSE and NIFTY) spanning 1995 to 2008, the ANN forecasts are less volatile but using analysis of variance (ANOVA) to compare the annual observations; no difference could be found. Mohsen *et al.* (2010) reported that GRNN outperformed GARCH (1,1) regarding Root Mean Squared Error (RMSE) in forecasting volatility of crude oil markets (Brent and WTI). Monfared & Enke (2014) use ANNs for volatility forecasting to enhance the ability of an asset allocation strategy based on the target volatility. The target volatility level is achieved by dynamically allocating between a risky asset and a risk-free cash position. Given the limited availability of data during times of high volatility, such as during financial crises, a stability-oriented approach to compare data for the current period to historical data for a period of low volatility, providing a more abundant source of data for comparison was employed. To comparatively evaluate the proposed model, the VIX Index, HV, EWMA, and a GARCH model serve as benchmarks. Trading metrics are used to evaluate the performance of the models in forecasting volatility. The sample period spans 2006 to 2016 using data from the Korea Composite Stock Price Index 200 (KOSPI 200). They concluded that ANN-based asset allocation strategies based on target volatility perform better than the fixed-allocation strategy since it generates less maximum drawdown and daily loss. Monfareda & Enkeb (2015) devise a sample portfolio consisting of five stocks from the S&P 500 Index over a sample period spanning April 2007 to December 2010, implementing noise-reducing adaptive filters to improve volatility forecasting performance. Adaptive filters hold the advantage of adapting to a changing environment, a desirable feature when incorporated with a model for error reduction purposes. The authors utilised Nonlinear Autoregressive (NAR) ANNs, and although NAR-ANNs have simple structures, they are efficient error reducing models when applied to a non-stationary or random walk noise time series. An adaptive threshold filter is designed to respond to changes in the time series when the GARCH(1,1) model generates forecasting errors. The results demonstrated that the adaptive filter can forecast the noise (errors) in the GARCH(1,1) forecasts, reducing the MSE by 42.9%. Hronec (2015) used ANNs to perform a multiple-step-ahead forecast of the Nasdaq Composite Index returns and daily range-based volatility analysed in four different time horizons; 1 day, 1 week, 2 weeks and 1 month. The sample period spans January 1990 to March 2015 for a total of 6339 observations. The sample period is partitioned by a ratio of 60:20:20 to take advantage of 2-fold non-exhaustive cross-validation and uses; MSE, RMSE, MAE



and Diebold-Mariano evaluation metrics. Based on MSE the performances of all ANNs over all time horizons are superior to a simple AR(1) model. The author suggests that the ANNs superiority is a result of their non-linearity and number of lags featured in the input variables. Lahmiri (2016) models and forecasts the volatility of currency pairs. The purpose of the study is to present a simple and effective approach for predicting HV of exchange rates. The approach utilises a limited set of technical indicators as inputs to the ANNs. The data consisted of the USD/CAD and EUR/USD exchange rates. The forecasting results show that the simple approach outperformed the traditional GARCH and EGARCH models, and also the hybrid GARCH and EGARCH ANN models regarding MAE, MSE, and Theil's inequality coefficient. The simplicity and effectiveness of the approach demonstrates the utility of ANN in forecasting FX volatility. Pradeepkumar & Ravia (2017) present a Particle Swarm Optimization (PSO)-trained Quantile Regression Neural Network (PSOQRNN), to forecast volatility of the following time series; USD/JPY, GBP/USD, EUR/USD, Gold, Crude Oil, and the S&P500 Index. They benchmarked the PSOQRNN with that of traditional models; GARCH and three ANNs including an MLP, General Regression Neural Network (GRNN) and Quantile Regression Neural Network (QRNN). Their results indicate that the proposed PSOQRNN outperformed the competing models in terms of Mean Squared Error (MSE). Its superior performance was further verified by the Diebold–Mariano test of statistical significance. It also performed well in terms of other important measures such as directional change and Theil's Inequality Coefficient. The superior performance of the PSOQRNN can be attributed to the role played by the PSO algorithm in obtaining better solutions. Therefore, they conclude that the PSOQRNN can be used as a viable alternative in forecasting volatility. Tana, Wang & Wanga (2017) introduce a new Adaptive-Network-Based Fuzzy Inference System (ANFIS) which adaptively adjusts fuzzy inference rules by using the Fruit Fly Optimization Algorithm (FOA). The research data used in this study are stock data from the Shanghai Stock Exchange from January 2006 to July 2016. Input variables are macroeconomic variables, microeconomic variables and technical indicators. Their results demonstrate their model can accurately and successfully forecast stock market volatility.

## 5.6 Nikkei 225 Index Time Series Data

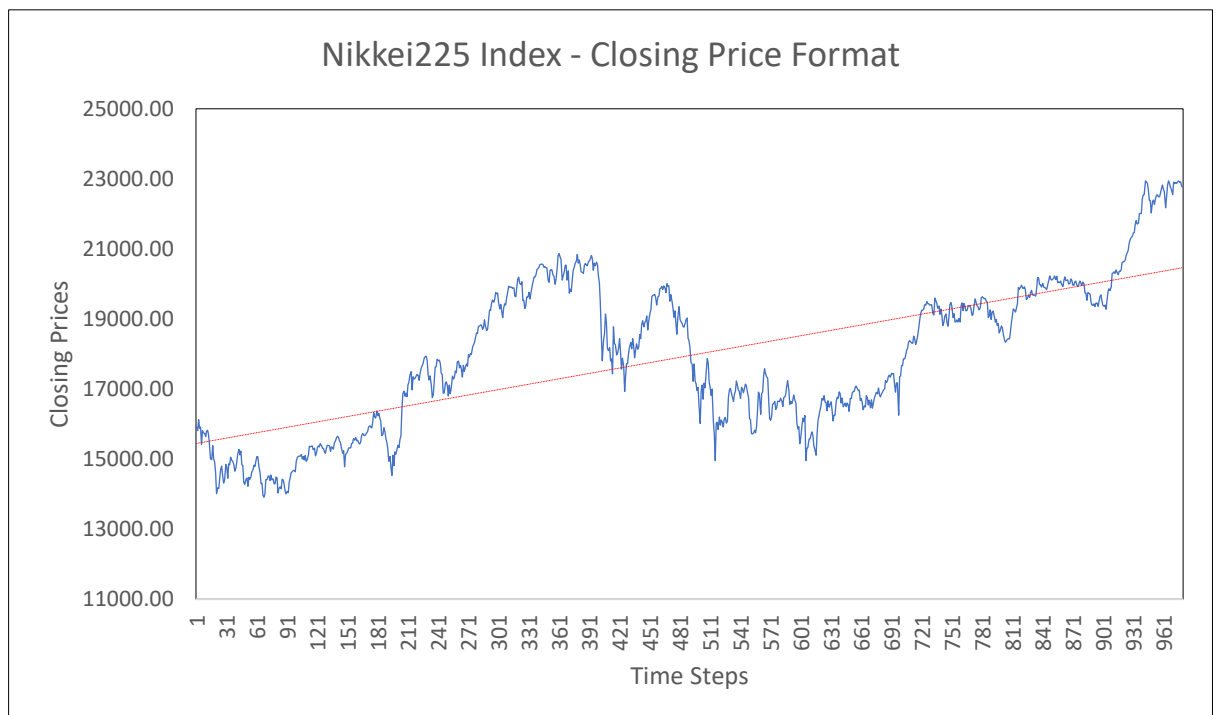
This section presents a statistical analysis of the Nikkei 225 Index time series data.

Initially, descriptive statistics of the time series data in closing price format are presented, followed by descriptive statistics of the time series data transformed into absolute log-returns format. The statistical analysis includes statistical tests covering: normality (Jarque-Bera), stationarity (Augmented Dickey-Fuller), monotonic trend presence (Mann-Kendall) and the minimum number of explanatory variables required to model the time series data (Fractal Dimension). Summary statistics pertaining to each data format can be found in Appendix 7 and 8 respectively.

### 5.6.1 Nikkei 225 Index Time Series Data (Closing Price)

Typically, users of financial time series data are accustomed to working with stock market data in the form of; high, low, open, close - or just closing prices as prescribed by Dow Theory on the bases of representing the best measure of fair value for the given period. This investigation follows Dow Theory and utilises the closing price to represent the Nikkei 225 Index time series. Figure 5-3 depicts the daily closing prices of the Nikkei 225 Index time series data over the sample period of 2014 to 2017:

Figure 5-3: Nikkei 225 Index Closing Price Time Series (2014-2017)



Visually, an automated trend line shows the Nikkei 225 Index time series features an overall uptrend throughout the sample period as the sample commences at 15908.88 on January 6<sup>th</sup>, 2014 and terminates at 22764.94 on December 29<sup>th</sup>, 2017. The following table details key descriptive statistics regarding the Nikkei 225 Index, for their full disclosure and description, please see Appendix 7:

Table 5-1: Nikkei 225 Index Descriptive Statistics (Closing Price Format)

<b>Descriptive Statistics</b>	<b>Confidence Level</b>	<b>P-Value</b>	<b>Result</b>
Mann-Kendall	99%	<0.0001	Monotonic Trend Presence
Jarque-Bera	99%	<0.0001	Not Normally Distributed
Dickey-Fuller	99%	0.7480	Non-Stationary
Fractal Dimension	N/A	N/A	2 (1.8)

From table 5-1, the Mann-Kendal trend test confirming the presence of a monotonic trend in the data. The Jarque-Bera test for normality finds the time series to be not normally distributed with a positive skewness of 0.14 and leptokurtic kurtosis of -0.78. The augmented Dickey-Fuller test for stationarity categorises the time series as non-stationary. The test of fractal dimension generated the largest value of 1.8, therefore, the time series requires a minimum of 2 independent variables for modelling purposes.

### 5.6.2 Nikkei 225 Index Time Series Data (Absolute Log Returns)

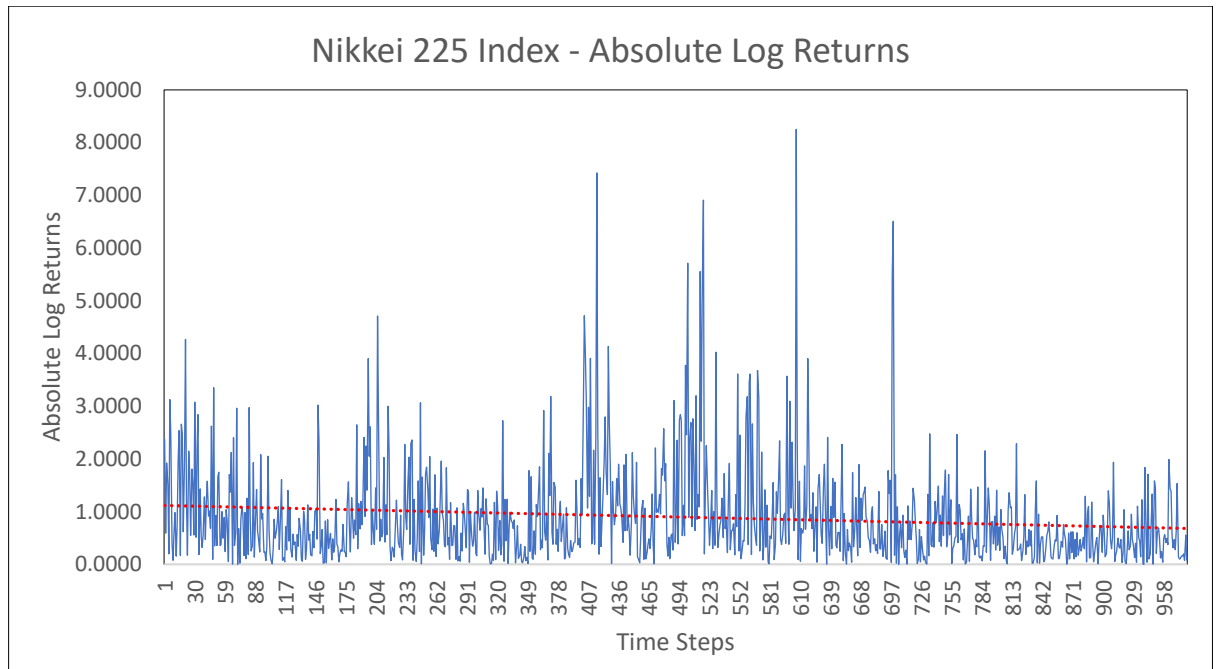
The descriptive statistics mentioned thus far have covered the Nikkei 225 Index time series in closing price format. This section will apply the same statistical tests to the Nikkei 225 Index time series data transformed into percent absolute log-returns format via the following formula:

$$R_i = \ln \frac{C_i}{C_{i-1}} \times 100 \quad (42)$$

The absolute log-returns data transformation is implemented not only to model volatility but to reduce the magnitude of the data and solve the problem of non-stationarity.

Figure 5-2 below depicts the Nikkei 225 Index in absolute log-returns format:

Figure 5-2: Nikkei 225 Index Closing Price Time Series (Absolute Log>Returns format)



The following table details key descriptive statistics regarding the Nikkei 225 Index transformed into absolute log-returns, for their full disclosure and description, please see Appendix 8.

Table 5-4: Nikkei 225 Index, Descriptive Statistics, (% Log>Returns)

Descriptive Statistics	Confidence Level	P-Value	Result
Mann-Kendall	99%	<0.0001	Monotonic Trend Presence
Jarque-Bera	99%	<0.0001	Not Normally Distributed
Dickey-Fuller	99%	<0.0001	Stationary
Fractal Dimension	N/A	N/A	2 (1.9)

From table 5-2, the Mann-Kendal trend test confirms the presence of a monotonic trend. The Jarque-Bera test for normality found the time series to be not normally distributed with a positive skewness of 2.60 and leptokurtic kurtosis of 10.67. The augmented Dickey-Fuller test for stationarity categorises the time series as stationary. The test of fractal dimension generated the largest value of 1.9, this value approximates the Nikkei 225 Index in closing price format, therefore, both time series require a minimum number of 2 independent variables to model the time series in their respective formats.

## 5.7 Input Variable Selection

### 5.7.1 Filter Methods

A two-stage IVS process is implemented to identify the most informative variables from the candidate input variable pool. First, filter methods are applied as an initial screening process to identify and remove irrelevant variables by screening for constant, quasi-constant and duplicate variables. Of the original 3,504 variables, there were no constant, quasi-constant or duplicate variables. Pearson's correlation coefficient is then applied to test the correlation between the Nikkei 225 Index and the predictor variables volatility models. A threshold of  $\pm 0.80$  (strong correlation) is set, and predictor variables that do not satisfy those criteria are removed, reducing the dataset by 3,146 variables to 358 variables. With the candidate variable pool reduced to 358 strongly correlated variables, the remaining variables progress to the second stage of IVS, Stepwise Multivariate Linear Regression, described in the next section.

### 5.7.2 Multivariate Linear Regression (Stepwise)

The stepwise multivariate linear regression is performed in accordance with the Handbook of Univariate and Multivariate Data Analysis with IBM SPSS, authored by Ho. To perform robust IVS the following strict criteria are adhered to; the final model must have  $R$ ,  $R^2$  and adjusted- $R^2$  values  $\geq 0.90$  (90%), a low standard error of the estimate, no multicollinearity between predictor variables therefore, a Condition Index (CI) of  $\leq 15$ , Value Inflation Factor (VIF) values of  $\leq 5$  and all predictor variables must be statistically significant at the 0.01 (99%) level. Following implementation, stepwise multivariate linear regression generated a total of 8 models. All models featured:  $R$ ,  $R^2$  and adjusted- $R^2$  values of 1.00, therefore achieving perfect linear correlation. However, all 8 models featured high levels of multicollinearity. An elimination process is employed to remove the multicollinearity, starting with the variable featuring the largest VIF value, repeating the elimination process until the remaining variables feature VIF values of  $\leq 5$  and a CI of  $\leq 15$ . Following implementation, a 5 variable model is produced fully satisfying the above criteria. The 5 variable model also has the benefit of satisfying the fractal dimension requirement of a minimum of 2 variables required for modelling purposes as discussed in section 5.3.2.

Table 5-3 presents the model summary, detailing the R,  $R^2$ , adjusted- $R^2$ , standard error of the estimate and Durbin-Watson values.

Table 5-3: Model Summary

Model Summary				
<b>R</b>	<b><math>R^2</math></b>	<b>Adjusted <math>R^2</math></b>	<b>Std. Error of the Estimate</b>	<b>Durbin-Watson</b>
1.0000	1.0000	1.0000	0.0246059	2.50

From table 5-3, the model features R,  $R^2$  and adjusted  $R^2$  values of 1.000, therefore achieving perfect linear correlation. The standardised error of the estimate is respectable at 0.0246059, and the Durbin-Watson test registers as 2.50 which is close to the preferred value of 2, therefore indicating negligible autocorrelation in the residuals. The following table details the models ANOVA:

Table 5-4: Model 5 ANOVA

ANOVA					
	<b>Sum of Squares</b>	<b>df</b>	<b>Mean Square</b>	<b>F</b>	<b>Sig.</b>
Regression	792.167	5	158.443	261677.766	0.000
Residual	0.440	727	0.001		
Total	792.607	732			

From table 5-4, it is evidenced that the models ANOVA states it is a statistically significant model at the 0.01 (99%) level; therefore, a good predictor of the dependent variable. The following table details the summarised coefficients, displaying each predictor variables; standardised Beta, significance and VIF values:

Table 5-5: Model 5-5 Coefficients

Model 5			
<b>Predictor Variables</b>	<b>Standardized Coefficients Beta</b>	<b>Sig.</b>	<b>VIF</b>
EWMA(0.95)	0.991	0.000	4.79
N225VI(C)	0.057	0.000	5.31
RSIV	0.001	0.000	1.79
Garman-Klass	0.001	0.000	1.33
CV(L9)ROC12	0.001	0.000	1.01

From table 5-5, all predictor variables made a statistically significant unique contribution to the model at the 0.000 level; therefore, satisfying the 0.01 (99%) significance criteria. Most predictor variables have a VIF value less than the strict value of 5, therefore, most predictor variables are free from multicollinearity. The only exception is predictor variable N225VI(C) with a VIF value of 5.31, which although above the strict value of 5, is well below the threshold of 15. Furthermore, most predictor variables have a VIF value closer to 1 which is the optimal value. The predictor variable featuring the highest standardised Beta (model contribution) is predictor variable; EWMA(0.95) with a standardised Beta of 0.991. This satisfies the *a priori* assumption of model development where one, or more, variables contribute the most to the model. For a detailed description of the above selected input variables see Appendix 5.

## **5.8 Methodology**

NeuroSolutions Pro is utilised to build and test the ANNs, with network parameters held constant across all networks, i.e. the number of hidden layers, number of processing elements in each hidden layer, number of epochs, weighting update method, activation function and learning rule. To avoid the issue of overfitting, various procedures were adhered to, namely parameter pruning; Castellano *et al.* (1997) and Wang *et al.* (2010), cross-validation; Zhang *et al.* (1999), Amjady & Keynia (2009) and Sermpinis *et al.* (2012b) and early stopping; Lin *et al.* (2009) and Prechelt (2012). The rationale underpinning each parameter will now be discussed.

### **5.8.1 Artificial Neural networks**

This investigation will assess the performance of a collection of ANNs, namely the Multilayer Perceptron (MLP), Generalized Feedforward Network (GFF), Modular Network, Radial Bias Function (RBF), Jordan-Elman (JE), Neuro-Fuzzy (CANFIS), Probabilistic, Recurrent, Support Vector Machine (SVM) and Time-Lag Recurrent network (TLRN).

### 5.8.2 Number of Hidden Layers

For each type of ANN, custom models are built featuring one and two hidden layers. Although any number of hidden layers is possible, it is widely accepted and supported by the general function approximation theorem that two hidden layers can approximate any complex nonlinear function and provides powerful computational properties; Dayhoff & DeLeo (2001). For that reason, this investigation limits the number of hidden layers at two hidden layers per model.

### 5.8.3 Processing Elements

To establish the optimal number of processing elements (PEs) in each hidden layer, the *vary a parameter* function is employed which allows automated incremental testing to identify the optimal number of PEs defined in terms of lowest MAE. The optimum number of PEs to be identified is confined to a generous range of 1 to 50 with testing increments of 1. Therefore, a total of 50 PEs are tested for each hidden layer. During this phase, the network tests all PEs over three cycles, therefore a total of 150 simulations (50 PEs x 3 cycles) are performed on each hidden layer. This equates to 150 simulations (150 simulations x 1 hidden layer) for ANNs featuring one hidden layer and 300 simulations (150 simulations x 2 hidden layers) for ANNs featuring two hidden layers.

### 5.8.4 Epochs

The number of epochs is set to 1000 and programmed to terminate after 150 epochs without improvement, therefore, implementing early stopping to guard against over-fitting as presented in Lin *et al.* (2009) and Prechelt (2012).

### 5.8.5 Weights

Each model was reproduced featuring two different weight update methods, batch weighting and online weighting. Batch weighting updates the weights after the presentation of the entire training set. On-Line weighting updates the weights after the presentation of



each exemplar. The only exemptions to this are the RBF and SVM models in which online weighting is not available.

### 5.8.6 Activation Function

All ANN models feature the Tanh Axon activation function, see appendix 6 for details.

### 5.8.7 Learning Rule

All ANN models feature the Levenberg-Marquardt (LM) learning rule as it is one of the most appropriate higher-order adaptive algorithms known for minimising MSE, training significantly faster than the commonly employed momentum learning rule, and typically arrives at a solution with a significantly lower error.

### 5.8.8 Data Segregation

This investigation features a 4-year sample period, (6th January 2014 – 29th December 2017), segregated into training, cross-validation, and testing periods for a ratio of 65:15:20 summarised as follows:

Table 5-6: Nikkei 225 Index, Data Segregation of Sample Period.

<i>Data Segregation of Sample Period</i>				
<b>Period</b>	<b>Start Date</b>	<b>End Date</b>	<b>% of Data</b>	<b>Periods</b>
Sample Period	6th January 2014	29th December 2017	100	980
Training	6th January 2014	1st April 2016	65	637
Cross Validation	2nd April 2016	31st December 2016	15	147
Testing	2nd January 2017	29th December 2017	20	247

## 5.9 Forecast Evaluation

As it is standard in the literature, the forecast evaluation metrics of ME, MSE, RMSE, MAE, MPE, and MAPE are calculated. Regarding the interpretation of the above statistics, the lower the value, the better the forecasting accuracy of the model. Additionally, the advanced forecasting evaluation metrics of Theils-U1, Theils-U2, and PT are computed. A lower value of Theils-U1 is desirable, with values bound between 0 and 1. For Theils-U2, values <1 indicate superior forecasting ability than a Naïve forecast, values of 1 equal a Naïve forecast, and values >1 indicate the model has underperformed a Naïve forecast. For the PT test, larger values indicate superior performance in forecasting directional accuracy. For the full description, mathematical notation, and interpretation of the statistics, please see Appendix 1.

### 5.9.1 In-Sample Forecast Evaluation

Table 5-7, below, details the in-sample forecast evaluation of the various ANNs benchmarked against a GARCH (1,1) model:

Table 5-7: In-Sample Forecast Evaluation

<i>In-Sample Forecast Evaluation</i>								
<b>Model</b>	<b>ME</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b>MPE</b>	<b>MAPE</b>	<b>U1</b>	<b>U2</b>
CANFIS	<b>*0.01</b>	0.71	0.84	0.67	-74.27	101.38	0.41	0.73
GFF	-0.03	0.73	0.86	0.69	-87.30	113.82	0.41	0.72
JE	-0.06	0.69	0.83	0.69	-89.08	114.56	0.39	0.69
MLP	0.16	0.65	0.81	0.57	-42.06	76.21	0.42	0.73
PNN	-0.14	0.70	0.84	0.73	-106.99	130.87	0.38	0.65
RNN	-0.03	<b>0.06</b>	<b>0.24</b>	<b>0.22</b>	<b>-3.95</b>	<b>60.41</b>	<b>0.23</b>	0.68
RBF	-0.45	0.23	0.48	0.45	-127.52	127.52	0.35	1.45
SVR	0.62	3.44	1.86	1.45	217.96	415.36	0.81	4.31
TLRN	0.06	0.26	0.51	0.40	-7.76	95.97	0.50	1.37
GARCH(1,1)	-0.57	1.04	1.02	1.01	-197.96	215.39	0.39	<b>0.53</b>

\*5% Significance, \*\*1% Significance

From Table 5-7, it is evident the Recurrent Model dominates the in-sample period, acquiring 6 of the 9 forecast evaluation metrics – including the key forecast evaluation metrics of RMSE and Theils-U1. In-sample ANN forecast evaluation results are mixed. The CANFIS model achieved the best ME value of 0.01, and the benchmark GARCH(1,1) model achieved the best Theils-U2 value of 0.53. The: GFF, JE, PNN, R, RBF, and

GARCH (1,1) model display negative ME values, therefore tend to understate their forecasts. Whereas the CANFIS, MLP, SVR, and TLRN models have positive ME, values, therefore, tend to overstate their forecasts. Regarding the Theils-U2 metric, models: RBF, SVR, and TLRN display values >1 indicating that a Naïve forecast would have performed better.

### 5.9.2 Out-of-Sample Forecast Evaluation

Table 5-8 details the out-of-sample forecast evaluation of the various ANNs against the GARCH (1,1) benchmark:

Table 5-8: Out-of-Sample Statistical Evaluation

<i>Out-of-Sample Forecast Evaluation</i>									
<b>Model</b>	<b>ME</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b>MPE</b>	<b>MAPE</b>	<b>U1</b>	<b>U2</b>	<b>DM</b>
CANFIS	0.48	1.42	1.19	0.69	-13.93	60.51	0.57	0.84	-18.19
GFF	0.39	1.31	1.14	0.72	-20.44	69.21	0.51	0.79	0.00
JE	0.48	1.45	1.21	0.71	-15.64	64.03	0.58	0.85	-21.22
MLP	0.47	1.45	1.21	0.72	-13.32	57.78	0.57	0.85	-3.66
PNN	0.48	1.44	1.20	0.73	-14.10	61.65	0.58	0.85	<b>-220.87</b>
RNN	0.53	1.31	1.14	<b>0.62</b>	9.82	<b>38.81</b>	0.56	0.81	-2.71
RBF	0.4	1.30	1.14	0.67	-22.11	60.98	0.53	0.80	-2.56
SVR	0.34	1.45	1.23	0.73	-29.64	72.79	0.53	0.83	0.00
TLRN	0.55	1.47	1.21	0.74	<b>0.01</b>	59.82	0.62	0.84	0.00
<b>GARCH(1,1)</b>	<b>0.31</b>	<b>1.31</b>	<b>1.14</b>	0.73	-48.72	86.19	<b>0.51</b>	<b>0.79</b>	N/A

\*5% Significance, \*\*1% Significance

From Table 5-8, the out-of-sample forecast performance changed dramatically, with the benchmark GARCH (1,1) model dominating the forecast evaluation results. The GARCH (1,1) model achieved 5 of the 8 forecast evaluation metrics including RMSE, Theils-U1 and Theils-U2. ANN results were mixed with no notable performance across all models. However, the PNN model did achieve the best DM value. All models display positive ME values, therefore, have the tendency to overstate their forecasts, a mild change from the in-sample results.

## 5.10 Conclusion

This research has evaluated ANNs applied to forecast the volatility of the Nikkei 225 Index one period in advance. The empirical results of the ANNs are disappointing. Regarding in-sample forecast evaluation, the Recurrent class of ANNs dominated the various forecast evaluation metrics categories, proving to be superior in 6 of the 8 metrics. However, out-of-sample forecast evaluation changed dramatically with the GARCH (1,1) benchmark model proving to be the superior forecasting model over 5 of the 8 evaluation metrics including RMSE, Theils-U1, and Theils-U2. To conclude, this research has demonstrated that ANNs do not outperform a GARCH (1,1) model when forecasting the volatility of the Nikkei 225 Index. A potential extension of this research could investigate alternative ANN architectures, extend the hidden layers past 2, and incorporate alternative transfer functions and learning algorithms.

# Chapter 6

## Macroeconomic Forecasting with Artificial Neural Networks: U.S. Unemployment Forecasting

### 6.1 Introduction

Macroeconomic forecasting methods have improved considerably over the decades, with current methods capable of generating relatively accurate forecasts. However, current methods mainly utilise traditional forecasting models that contain undesirable properties and limitations such as the constraints of linearity. Despite the considerable advancements, there remains room to improve macroeconomic forecasting. Recently, econometricians have been exploiting advancements in machine learning through the application of Artificial Neural Networks (ANNs). The major attraction of ANNs resides in their flexible nonlinear modelling capabilities. For example, there is no need to stipulate a functional form. Instead, the model is adaptively formed based on the features presented in the data, making them appropriate for situations where *a priori* theoretical expectations do not hold. Furthermore, ANNs offer their benefits in periods of extreme structural instabilities such as recessions, market bubbles, and periods of increased volatility - circumstances in which traditional models tend to underperform. The popularity of ANNs for forecasting time series data that contains potential nonlinearities continues to receive considerable attention in the literature especially over the last two decades, see; Zhang & Qi (2005), De Gooijer & Hyndman (2006), Crone, Hibon, & Nikolopoulos (2011), Huang *et al.* (2013), Olmedo (2014), and Giusto & Piger (2017). However, despite the popularity and advantages of ANNs, they have been sparingly applied to macroeconomic forecasting, with most of the literature concentrated on utilising ANNs for stock market and exchange rate forecasting. Therefore, this research explores the utility of ANNs applied to macroeconomic forecasting, focusing on U.S. civilian unemployment which is a significant macroeconomic indicator and vital for domains such as policymaking.

Forecasting unemployment rates is well documented in the literature. For example; Swanson & White (1997) forecast several macroeconomic time series, including U.S. unemployment, with linear models and ANNs. In their approach, ANNs present promising

empirical evidence against the linear VAR models. Johnes (1999) reports the results of a forecasting competition between linear auto-regressive, GARCH, threshold autoregressive and ANNs, applied to the UK monthly unemployment rate series. In his application, ANNs are superior over an 18 and 24 month forecasting horizon but fail to outperform the other models over shorter forecasting horizons. Moshiri & Brown (2004) apply ANNs to forecast post-war unemployment rates in the; USA, Canada, UK, France and Japan, with their out-of-sample results confirming the forecasting superiority of ANNs over traditional linear and non-linear models. Liang (2005) applies ANNs to forecast unemployment in West Germany; the empirical results indicate that ANNs present significantly better forecasts than traditional autoregressive models. Wang (2010) combines several rival individual U.S. unemployment forecasts with directed acyclical graphs. The results indicate that models that are not directly causally linked can be combined to project a more accurate composite forecast. Chua *et al.* (2012) present a latent variable approach to the same forecasting task. Their model exploits the time series properties of U.S. unemployment while satisfying the economic relationships specified by Okun's law and the Phillips curve. The specification is advantageous because it provides an unemployment forecast consistent with both theories, and is less computationally intensive than equivalent atheoretical models such as VAR and BVAR. Sermpinis *et al.* (2013) investigate the efficiency of the Psi Sigma Network (PSN), in forecasting U.S. unemployment and compare the utility of the Kalman Filter and Support Vector Regression (SVR) model in combining ANN forecasts. An ARMA, MLP and RNN are used as benchmarks. The sample period of (1972-2012) is employed, reserving the last seven years for out-of-sample testing. The empirical results found the PSN outperformed all models. Their forecast combinations did improve accuracy, and the SVR outperformed the Kalman Filter. Olmedo (2014) performs a competition between non-linear models, including ANNs to forecast European unemployment rates with the best results achieved by a vector autoregressive model. Stasinakis *et al.* (2016) investigate the efficiency of the RBF network in forecasting U.S. unemployment and explore the utility of the Kalman filter and SVR model as forecast combination techniques. An ARMA model, STAR model and three ANNs, namely an MLP, RNN and PSN models are used as benchmarks. The sample period of (1972–2012) is employed, reserving the last 7 years for out-of-sample testing. The results show that the RBF network outperforms all benchmark models. The forecast combination techniques were successful as the Kalman filter, and SVR models improved the forecast accuracy. However, the SVR model outperformed all models. Cook & Hall (2017) evaluate ANNs in their ability to forecast U.S. civilian unemployment. They

benchmark the ANNs against a VAR model and consensus forecast, finding that ANNs outperform both benchmarks over a four-quarter forecasting horizon. This research determines if ANNs provide superior forecasting performance when compared to an SVR model and a Naïve forecast. As U.S. civilian unemployment data is released on a monthly basis, this research features a 45-year sample period (1972 to 2017). In-sample forecast evaluation is performed using the following metrics: ME, MSE, RMSE, MAE, MPE, MAPE, Theils-U1, Theils-U2, and PT. Out-of-sample forecasting evaluation features the same performance metrics but includes the additional metric of the Diebold-Mariano (DM) test to assess the forecasting accuracy the ANNs Vs the SVR benchmark.

## **6.2 Motivation**

The research presented in this chapter is motivated by the desire to address the challenges faced when performing macroeconomic forecasting, in particular, challenges related to macroeconomic modelling. The motivation is twofold. The primary motivation is to evaluate the performance of ANNs in their ability to forecast U.S. civilian unemployment and evaluate if they offer superior forecasting ability over an SVR model and a Naïve forecast. The secondary motivation is to contribute to the relatively smaller body of knowledge concerning macroeconomic forecasting using ANNs. For example; at the time of writing, a Google Scholar search for ‘artificial neural networks, exchange rate forecasting’ and ‘artificial neural networks, stock market forecasting’ generated 68,200 and 47,800 results respectively. Whereas a search for ‘artificial neural networks, macroeconomic forecasting’ generated only 24,800 results. Evidently, there is a strong bias towards employing ANNs for financial market forecasting over employing ANNs for macroeconomic forecasting.

## **6.3 U.S. Unemployment Time Series Data**

This research examines the macroeconomic indicator of U.S. civilian unemployment (UNEMPLOY). This measure is published on a monthly basis by the US Bureau of Labour and Statistics and measures the number of the U.S. workforce that is currently unemployed. Unemployment forecasting, as opposed to another macroeconomic variable, was selected for two reasons. First, unemployment is a fundamentally significant indicator and is integral to domains such as policymaking. Second, unemployment data typically

undergoes limited revision following its initial release. This is an important characteristic as it allows circumventing collecting and assembling appropriate *vintages* of data. For example; according to Cook & Hall (2017) the largest discrepancy in U.S. civilian unemployment data between the original vintage of data and its final release is approximately 23 basis points with an average discrepancy of 9 basis points. It is assumed the impact of these discrepancies on the predictive accuracy of a model is negligible. The UNEMPLOY data is sourced from the Economic Research section of the Federal Bank of St Louis (FRED), available at; [www.stlouisfed.org](http://www.stlouisfed.org). The observations are of monthly frequency, measured in the number of persons and percent-change (%Δ) format, and seasonally adjusted to avert ANN overfitting.

The following sections provide a statistical analysis of the UNEMPLOY time series data. The statistical analysis includes statistical tests covering; normality (Jarque-Bera), stationarity (Augmented Dickey-Fuller), monotonic trend presence (Mann-Kendall) and the minimum number of explanatory variables required to model the time series data (Fractal Dimension). The full description and disclosure of each statistic are available in appendix 1.

### **6.3.1 Descriptive Statistics, UNEMPLOY (Persons)**

This section presents a statistical analysis of UNEMPLOY in its original format to gain an understanding of the data and its characteristics. The following statistical tests are performed on the data; the Mann-Kendall test initially assesses the presence of a monotonic trend. The Jarque-Bera test of normality coupled with tests of skewness and kurtosis assess the distribution. The augmented Dickey-Fuller tests for stationarity and fractal dimension tests for the minimum number of explanatory variables required to model the time series. Figure 6-1 depicts the UNEMPLOY time series over the sample period (1972 – 2017) followed by table 6-1 detailing the results of the abovementioned statistics;



Figure 6-1: U.S. Civilian Unemployment (UNEMPLOY), (1972 – 2017) Persons Format.

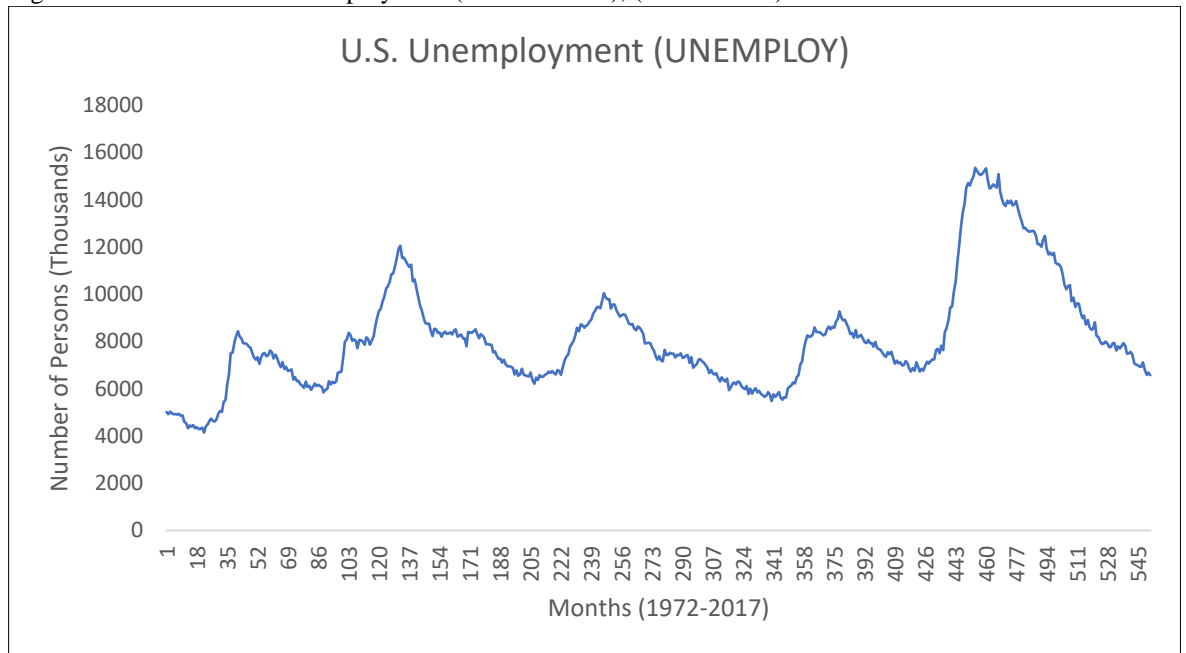


Table 6-1: US Civilian Unemployment (UNRATE), Descriptive Statistics, (Percent)

Descriptive Statistics	Confidence Level	P-Value	Result
Mann-Kendall	99%	<0.0001	Monotonic Trend Presence
Jarque-Bera	99%	<0.0001	Not Normally Distributed
Dickey-Fuller	99%	0.0726	Non-Stationary
Fractal Dimension	N/A	N/A	2 (2.3359)

From table 6-1, a trend is empirically evidenced by the Mann-Kendal trend test confirming the presence of a monotonic trend. The Jarque-Bera test for normality finds the time series to be not normally distributed with positive skewness of 0.73 and platykurtic kurtosis of -0.16. The augmented Dickey-Fuller test for stationarity categorises the time series as non-stationary. The test of fractal dimension generated the largest value of 2.3, therefore, the time series requires a minimum of 2 explanatory variables for modelling purposes.

### 6.3.2 Descriptive Statistics, UNEMPLOY (%Δ Format)

The descriptive statistics mentioned thus far have covered the UNEMPLOY time series data in number of persons format. This section will apply the same statistical tests to the UNEMPLOY time series data transformed into per cent-change (%Δ) format. Figure 6-2

displays the seasonally adjusted UNEMPLOY data transformed into %Δ format using the following formula in MS Excel:

$$\% \Delta = \left( \left( \frac{x_t}{x_{t-1}} \right) - 1 \right) \times 100 \quad (43)$$

The %Δ data transformation is implemented to reduce the magnitude of the data and solve the problem of non-stationarity.

Figure 6-2: U.S. Civilian Unemployment (UNEMPLOY) %Δ Format

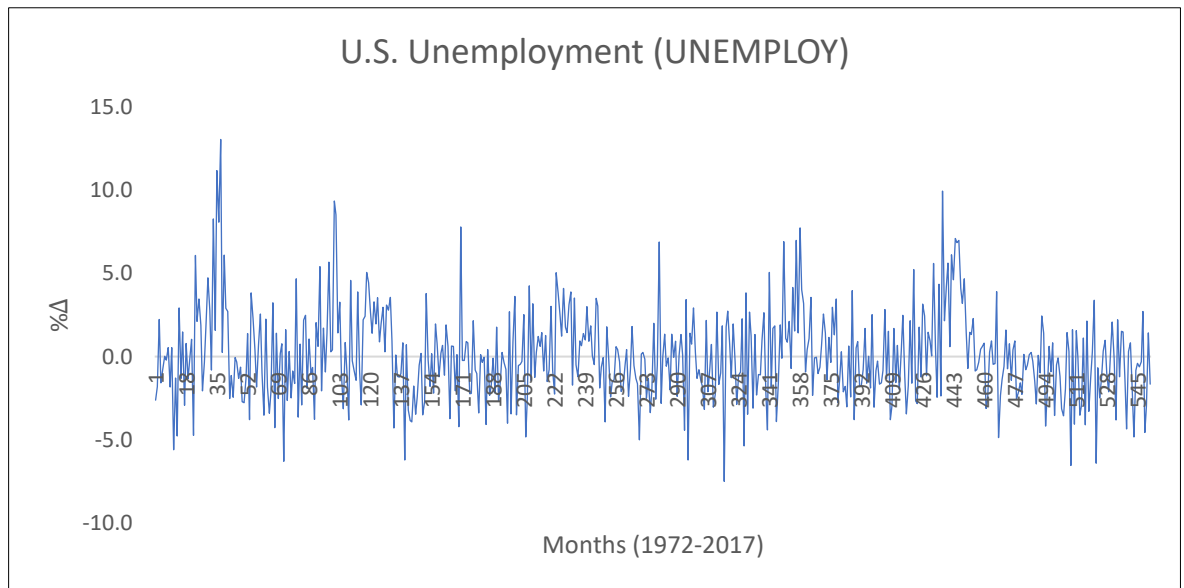


Table 6-2: Descriptive Statistics, U.S. Civilian Unemployment (%Δ)

Descriptive Statistics	Confidence Level	P-Value	Result
Mann-Kendall	99%	<0.03107	No Monotonic Trend
Jarque-Bera	99%	<0.0001	Not Normally Distributed
Dickey-Fuller	99%	<0.0001	Stationary
Fractal Dimension	N/A	N/A	2 (2.2834)

From table 2, the Mann-Kendal trend test confirms there is no monotonic trend contained in the data. The Jarque-Bera test for normality found the time series to be not normally distributed with positive skewness of 0.64 and leptokurtic kurtosis of 1.29. The augmented Dickey-Fuller test for stationarity categorises the time series as stationary. The test of fractal dimension generated the largest value of 2.28 which approximates UNEMPLOY in

percent format; therefore, both time series require a minimum number of 2 explanatory variables for modelling purposes.

## **6.4 Input Variable Selection**

### **6.4.1 Filter Methods**

A two-stage IVS method is implemented to identify the most informative variables from the candidate input variable pool. First, filter methods were applied as an initial screening process to identify and remove irrelevant variables. Initially, constant, quasi-constant and duplicated variables were screened for. Of the original 168 variables, 32 were identified as duplicate variables and removed, reducing the dataset to 136 variables. Pearsons correlation coefficient was then applied to test the correlation between the target variable (UNEMPLOY) and the predictor variables. A threshold of 0.80 (strong relationship) is set and predictor variables that did not satisfy that criteria were removed, further reducing the dataset by 96 variables to 40 variables. With duplicate variables removed and the candidate variable pool reduced to 40 highly correlated variables, the remaining variables progress to the second stage of IVS, LASSO regression, described in the next section.

### **6.4.2 Least Absolute Shrinkage and Selection Operator (LASSO)**

The Least Absolute Shrinkage and Selection Operator, LASSO, (Tibshiranni, 2011) is a suitable and well-researched approach to the problem of variable selection. Compared with traditional estimation methods, LASSO's major advantage is its simultaneous execution of both parameter estimation and variable selection. Furthermore, LASSO has excellent computational properties, for example; in an ordinary least squares (OLS) setting, Efron *et al.* (2004) demonstrated that the solution path of LASSO is piecewise linear and moves in a predictable manner. Path-finding algorithms are computationally light and have been proved valuable in both theory and practice.

LASSO Regression is a class of Shrinkage or Regularization Regressions, which applies when multicollinearity exists among variables (Sundberg, 2006). LASSO regression

performs L1 regularization, which adds a penalty equal to the absolute value of the magnitude of coefficients. This variant of regularization can result in sparse models with few coefficients, with some coefficients reduced to zero and eliminated from the model. Larger penalties result in coefficient values closer to zero, which is ideal for producing simpler models. LASSO solutions are quadratic programming problems where the objective of the algorithm is to minimize:

$$\sum_{i=1}^n (y_i - \sum_j x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (44)$$

Which is the same as minimizing the sum of squares with constraint  $\sum |\beta_j| \leq$ . Some of the  $\beta$ s are shrunk to exactly zero, resulting in a regression model that's easier to interpret. A tuning parameter,  $\lambda$ , controls the strength of the L1 penalty or the amount of shrinkage. When  $\lambda = 0$ , no parameters are eliminated, and the estimate is equal to linear regression. As  $\lambda$  increases, an increasing number of coefficients are set to zero and eliminated (when  $\lambda = \infty$ , all coefficients are eliminated). As  $\lambda$  increases, bias increases. As  $\lambda$  decreases, variance increases. LASSO best applies in samples of few variables with medium to large effect (Hastie *et al.*, 2017). For more details on the mathematical specifications of LASSO see Wang *et al.* (2007). Following the implementation of LASSO, seven variables were selected. For the list and description of each variable, see Appendix 5.

## 6.5 Methodology

In order to perform a fair and extensive investigation, the following methodology was devised and implemented. NeuroSolutions Pro was utilized to build and test the ANNs, with network parameters initially held constant across all networks, i.e. number of hidden layers, number of processing elements in each hidden layer, number of epochs, weighting update method, transfer function and learning rule. To avoid the issue of overfitting, various procedures were adhered to, namely parameter pruning; Castellano *et al.* (1997) and Wang *et al.* (2010), cross-validation; Zhang *et al.* (1999), Amjady & Keynia (2009) and Sermpinis *et al.* (2012b) and early stopping; Lin *et al.* (2009) and Prechelt (2012). The rationale underpinning each parameter will now be discussed.

### 6.5.1 Artificial Neural networks

This investigation will assess the performance of a collection of ANNs, namely the Multilayer Perceptron (MLP), Generalised Feed-Forward (GFF), Modular Neural Network (MNN), Radial Basis Function (RBF), Probabilistic Neural Network (PNN), Co-Active Neuro-Fuzzy Inference System (CANFIS), and an ANN forecast combination (Fc).

### 6.5.2 Number of Hidden Layers

For each type of ANN, custom models are built featuring 1 to 4 hidden layers. Although any number of hidden layers is possible, it is widely accepted and supported by the general function approximation theorem that two hidden layers can approximate any complex nonlinear function and provides powerful computational properties; Dayhoff & DeLeo (2001). For that reason, this investigation incorporates the two hidden layer principle and extends beyond it.

### 6.5.3 Processing Elements

To establish the optimum number of processing elements (PEs) in each hidden layer, the *vary a parameter* function is employed which allows automated incremental testing to identify the optimal number of PEs defined in terms of lowest MAE. The optimum number of PEs to be identified is confined to a generous range of 1 to 50 with testing increments of 1. Therefore, a total of 50 PEs are tested for each hidden layer. During this phase, the network tests all PEs over 3 cycles, therefore a total of 150 simulations (50 PEs x 3 cycles) are performed on each hidden layer. This equates to 150 (150 simulations x 1 hidden layer) simulations for ANNs featuring one hidden layer and 600 (150 simulations x 4 hidden layers) simulations for ANNs featuring four hidden layers.

#### **6.5.4 Epochs**

The number of epochs was set to 1000 and programmed to terminate after 150 epochs without improvement, therefore, implementing early stopping to guard against over-fitting as presented in Lin *et al.* (2009) and Prechelt (2012).

#### **6.5.5 Weights**

Each model features batch weighting which updates the weights after the presentation of the entire training set.

#### **6.5.6 Activation Function**

All ANN models are replicated to feature the following activation functions; Tanh Axon, Sigmoid Axon, Linear Tanh Axon, Linear Sigmoid Axon, SoftMax Axon, Bias Axon, Linear Axon, Axon. For a description of each activation function, see appendix 6.

#### **6.5.7 Memory**

For Recurrent ANNs that feature a memory component, focused and unfocused; Gamma Axon, Laguerre Axon and TDNN Axons are used.

#### **6.5.8 Learning Rule**

All ANN models will feature the Levenberg-Marquardt (LM) learning rule as its one of the most appropriate higher-order adaptive algorithms known for minimizing the MSE of a neural network, training significantly faster than the commonly employed momentum learning rule and typically arrives at a solution with a significantly lower error.

### 6.5.9 ANN Competition and Selection

Following training and testing, the best ANN from each class defined in terms of RMSE is selected to compete against an SVR model and Naïve forecast. Therefore, the best performing MLP, GFF, MNN, RBF, PNN, and CANFIS will compete against the abovementioned benchmarks.

### 6.5.10 Data Segregation

This investigation features a 45-year sample period, (1972-2017). The sample period is segregated into training, cross-validation , and testing periods for a ratio of 65:15:20 as detailed by table 6-3:

Table 6-3: U.S. Civilian Unemployment, Data Segregation of Sample Period.

<i>Data Segregation</i>				
<b>Period</b>	<b>Start Date</b>	<b>End Date</b>	<b>% of Data</b>	<b>Months</b>
Sample Period	1st January 1972	1st December 2017	100	552
Training	1st January 1972	1st November 2001	65	359
Cross Validation	1st December 2001	1st October 2008	15	83
Testing	1st November 2008	1st December 2017	20	109

### 6.5.11. Benchmark Models

The benchmark model is the SVR model as described in section 3.2.3.

## 6.6 Forecast Evaluation

To statistically evaluate the performance of the various ANNs, the commonly employed and robust statistics of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) will be used to identify the best model. The additional statistics of; Theils-U1, Theils-U2, and Mean Absolute Percent Error (MAPE) will be presented to yield more insight regarding model performance. Regarding interpretation, the lower the statistic, the better the model.

### 6.6.1 In-Sample Forecast Evaluation

The following tables present the in-sample and out-of-sample statistical evaluation of the various model. Table 4 details the in-sample statistical evaluation of the various ANNs and competing models:

Table 6-4: In-Sample Forecast Evaluation

<i>In-Sample Forecast Evaluation</i>								
<b>Model</b>	<b>MLP</b>	<b>GFF</b>	<b>MNN</b>	<b>RBF</b>	<b>PNN</b>	<b>CANFIS</b>	<b>Fc</b>	<b>SVR</b>
ME	0.1294	0.0267	<b>-0.0077</b>	-0.0478	0.1013	-0.9063	-0.1174	0.0100
MSE	<b>6.8022</b>	7.3088	7.2907	7.6007	7.7132	8.5159	7.2152	8.0195
RMSE	<b>2.6081</b>	2.7035	2.7001	2.7569	2.7773	2.9182	2.6861	2.8319
MAE	<b>1.9840</b>	2.0709	2.0969	2.1287	2.1270	2.3148	2.0751	2.2490
MPE	-12.2917	-3.0715	<b>0.7163</b>	2.4221	-0.4725	-18.2616	-5.1598	-10.6865
MAPE	14.4067	5.2233	<b>1.3600</b>	2.6134	2.4973	21.4909	7.3101	14.4209
Theils-U1	<b>0.6884</b>	0.7626	0.7750	0.7892	0.9551	0.7366	0.7982	0.6086
Theils-U2	0.5551	1.0055	0.9783	0.9596	0.9185	0.1036	0.7196	<b>0.0271</b>
PT	<b>4.8241</b>	3.1078	*0.3952	-0.4699	1.0045	1.0045	*2.6428	*3.5976

\*1% Significance, \*\*5% Significance

From table 6-4, the in-sample forecasting results demonstrate the MLP and MNN networks are the best performing models, with the MLP emerging as the superior model. The MLP acquires 5 of the 9 forecasting evaluation metrics of MSE (6.80), RMSE (2.61), MAE (1.98), Theils-U1 (0.69), and PT (4.42). The MNN emerged as the next best performing model, securing 3 of the 9 forecast evaluation metrics of ME (-0.008), MPE (0.72), and MAPE (1.36). The remaining ANNs and Fc model failed to record any notable performance. The SVM benchmark model achieved the best Theils-U2 value of 0.03. Most ANN models recorded Theils-U2 values of <1, indicating they outperformed a Naïve forecast. However, the GFF network recorded a Theils-U2 value of 1, indicating it underperformed a Naïve forecast.

### 6.6.2 Out-of-Sample Forecast Evaluation

In addition to the standard set of forecasting evaluation metrics, to further verify the forecasting ability of the ANN, the Modified Diebold-Mariano (MDM) test is applied as proposed by Harvey *et al.* (1997). The null hypothesis of the test is the equivalence in forecasting accuracy between competing forecasting models. The use of the MDM test is common practice in forecast evaluation as it is found to be robust in assessing the



significance of observed differences between competing forecasts (Barhoumi *et al.*, 2010), and overcomes the problem of over-sized DM values in moderate samples (Dreger and Kholodilin, 2013). The MDM test is an extension of the Diebold-Mariano (1995) test and is presented below:

$$MDM = T^{-1/2}[T + 1 - 2k + T^{-1}k(k - 1)]^{1/2}DM \quad (45)$$

Where  $T$  is the number of out-of-sample observations and  $k$  is the number of the step-ahead forecasts. In this case the MDM test is applied to competing forecasts (ANN vs. SVR). A negative MDM value indicates that the ANN model is more accurate than SVR benchmark model. The lower the negative value, the more accurate are the ANN forecasts.

Table 6-5: Out-of-Sample Statistical Performance

<i>Out-of-Sample Forecast Evaluation</i>								
<b>Model</b>	<b>MLP</b>	<b>GFF</b>	<b>MNN</b>	<b>RBF</b>	<b>PNN</b>	<b>CANFIS</b>	<b>ANN Fc</b>	<b>SVR</b>
ME	-0.4988	-0.5051	-0.5473	-0.5882	-0.4305	-1.3847	-0.6591	<b>-0.2746</b>
MSE	5.6912	5.4316	<b>5.3469</b>	5.4680	6.4133	7.8348	5.7647	9.0686
RMSE	2.3856	2.3306	<b>2.3123</b>	2.3384	2.5325	2.7991	2.4010	3.0114
MAE	1.8160	<b>1.7742</b>	1.7747	1.7970	1.9201	2.1757	1.8320	2.3692
MPE	<b>0.8263</b>	1.1061	0.9693	1.0014	1.0188	1.2643	1.0310	1.3980
MAPE	1.3894	1.1354	<b>0.9693</b>	1.0498	1.0188	1.9961	1.0598	2.2467
Theils-U1	0.6810	0.7112	0.7226	0.7202	0.9603	0.7673	0.7685	<b>0.6673</b>
Theils-U2	1.0386	0.9919	1.0167	1.0746	1.0260	1.3181	1.0688	<b>0.9338</b>
PT	<b>1.5565</b>	0.9580	-0.2125	0.8914	-1.1351	-1.1351	1.3879	0.2320
DM	-3.3015	<b>-3.5659</b>	-3.2173	-3.4027	-1.4405	-0.6883	-2.9050	N/A

\*1% Significance, \*\*5% Significance

From Table 6-5, the out-of-sample forecasting results are mixed. Of the ANNs, the MNN emerges as the best performing model based on an RMSE value of 2.31. The MNN also achieves the forecasting evaluation metrics of MSE (5.35), and MAPE (0.97). The MLP achieves the forecasting evaluation metrics of MPE (0.83), and PT (1.56). The GFF achieves the forecast evaluation metrics of MAE (1.77), and MDM (-3.57). The Theils-U2 metric is notably worse for most models with values >1, indicating they underperformed a Naïve forecast. The only models to record a Theils-U2 value <1 are the GFF networks and the SVR benchmark model. The benchmark SVR model improved its in-sample performance, retaining the Theils-U2 metric of 0.93 and acquiring the metrics of Theils-U1 (0.67), and ME (-0.27).

### 6.6.3 Point Forecast

Table 6-6: Point Forecast Results

<i>Point-Forecast</i>			
<b>Model</b>	<b>Point Forecast (<math>Pf</math>)</b>	<b>Error (<math>Af - Pf</math>)</b>	<b>Rank</b>
CANFIS	0.9805	0.0195	1
MLP	1.4591	-0.4591	2
SVR	1.6604	-0.6604	3
Fc	0.3356	0.6644	4
PNN	0.0711	0.9289	5
GFF	0.0501	0.9499	6
MNN	0.0101	0.9899	7
RBF	-0.5572	1.5572	8

Actual Value = 1.00

From table 6-6, the CANFIS model records the most accurate point forecast with a value of 0.98 which is in contrast to its poor in-sample and out-of-sample performance. The MLP ranked 2<sup>nd</sup>, reinforcing its generally acceptable forecasting performance throughout the sample period. The SVR benchmark model ranks 3<sup>rd</sup>, likely attributable to its out-of-sample Theils-U1 metric. Although the MNN is the overall best ANN and the GFF network outperforms the SVR model as per the DM metric, both models rank low in the point forecast.

## 6.7 Conclusion

The motivation of this research was to investigate the utility of ANNs in forecasting U.S. civilian unemployment. A collection of ANNs, namely an; MLP, GFF, MNN, RBF, PNN, CANFIS, and Fc was benchmarked against an SVR model. The forecasting performance of the models is evaluated throughout the sample period of 1972–2017, reserving the last 9 years for out-of-sample testing. The in-sample forecasting evaluation results found the MNN network to be the best performing ANN, closely followed by the MLP, with the SVR benchmark demonstrating no notable performance apart from outperforming a Naïve forecast. The out-of-sample results confirm the in-sample results, finding the MNN to be the best performing network based on RMSE as is standard in the literature. However, the GFF network achieved the best DM value, indicating it is more accurate than the SVR benchmark model. The forecasting performance of the SVR model did improve out-of-

sample, however, is outperformed by the ANNs over most evaluation metrics. It is concluded that ANNs offer superior forecast performance over the SVR benchmark model.

# Chapter 7

## Macroeconomic Forecasting with Artificial Neural Networks: U.S. GDP Forecasting

### 7.1 Introduction

Macroeconomic forecasting methods have improved considerably over the decades, with current methods capable of generating relatively accurate forecasts. However, current methods mainly utilise traditional forecasting models that contain undesirable properties and limitations such as; sensitivity to model specification and the constraints of linearity. Despite the considerable advancements, there remains room to advance the state of macroeconomic forecasting. This could be achieved by leveraging recent advancements in machine learning through the application of Artificial Neural Networks (ANNs). The popularity of ANNs for forecasting time series data that contains nonlinearities continues to receive considerable attention in the literature, especially over the last two decades, see; Terasvirta, Van Dijk, & Medeiros (2005), De Gooijer & Hyndman (2006), Crone, Hibon, & Nikolopoulos (2011), Huang *et al.* (2013), Giusto & Piger (2017). The main attraction of ANNs resides in their flexible nonlinear modelling capabilities. For example, there is no need to stipulate a functional form. Instead, the model is adaptively formed based on the features presented in the data, making them appropriate for situations where *a priori* theoretical expectations do not hold. However, despite the popularity and advantages of ANNs, they have been sparingly applied to macroeconomic forecasting, with most of the literature concentrated on utilising ANNs for stock market and exchange rate forecasting. For example; at the time of writing, a Google Scholar search for ‘artificial neural networks, exchange rate forecasting’ and ‘artificial neural networks, stock market forecasting’ generated 68,200 and 47,800 results respectively. Whereas a search for ‘artificial neural networks, macroeconomic forecasting’ generated only 24,800 results. Evidently, there is a strong bias towards researching ANNs applied to financial market forecasting over researching ANNs for macroeconomic forecasting.

This research explores the utility of ANNs applied to macroeconomic forecasting, focusing specifically on GDP as GDP is the principle measure economic of performance therefore descriptive of other macroeconomic and microeconomic indicators. Being able to forecast GDP accurately allows a top-down approach to economic forecasting which; businesses, policy makers, and governments depend on. Tkacz (2001) employs ANNs to forecast Canadian GDP, finding that ANNs do not provide satisfactory performance at the one-quarter forecasting horizon and are unable to outperform a Naïve forecast. However, at the four-quarter forecasting horizon, Tkacz finds that ANNs are superior predictors of GDP. Ao & Tang (2007) perform a simulation forecast of annual GDP based on chaotic attractors. They were motivated by the limitation that chaotic time series cannot sufficiently fit the actual fluctuation of small sample discrete data, especially for long-term economic forecast errors. In solution, they use ANNs to predict the fitting errors and correct the results based on the forecast. Their ANN models correct the errors of the GDP growth rate during 2004 and 2006, improving the three-year average relative error rate to 0.553%, which is more accurate compared with Wang (2007) of 3.38%, Liu (2004) of 1.21%, and Su (2003) of 1.42%. Curak, Klime & Curak (2009) investigate the performance of ANNs benchmarked against linear regression models applied to forecast GDP for EU member nations. Their results show that at the 1-year forecasting horizon, according to three out of four forecasting evaluation metrics, ANNs improve forecasting accuracy. Sinclair, Stekler & Kitzinger (2010) perform separate analyses of the Fed's forecasts of GDP and inflation. They present a method for jointly evaluating the directional change of forecasts, utilising the metric devised by Pesaran & Timmermann (2004). They conclude that some of the inflation forecasts, examined separately, were not valuable. However, the joint pattern of GDP and inflation projections were generally in accord with the economy's movements. Diverging from national GDP, the accuracy of macroeconomic forecasts can also be assessed at a regional level. Lehmann *et al.* (2013) overcome the problem of a data-poor environment at the regional level by complementing various regional economic indicators with more than 200 national and international economic indicators. They calculate single-indicator, multi-indicator, pooled and factor forecasts in a pseudo-real-time setting. Their results show that forecast accuracy was significantly increased when compared with an autoregressive benchmark model, for both short-term and long-term predictions.

Combination forecasts are also a promising technique, see; Bates & Granger (1969), Newbold & Granger (1974), and Deutsch *et al.* (1994). More recently, Swanson & Zeng

(2001) perform forecast combinations, concluding that a SIC-based approach can be a useful alternative to combination methods such as simple averaging. Conversely, Stock & Watson (2003) forecast U.S. inflation and advise that the best forecast performance is achieved through simple averaging. Teräsvirta *et al.* (2005) examine the forecast accuracy of linear autoregressive, smooth transition autoregressive and ANN models for 47 macroeconomic variables, including GDP, of G7 economies. Their results show that the forecasting accuracy of ANNs is improved when combined with autoregressive models. Kapetanios *et al.* (2008) report that combinations of forecasts from several models (RW, STARs, ARs, VARs etc.) generate good forecasts of inflation and GDP growth. They also note that forecast combinations could serve as an unbiased benchmark. Inoue & Kilian (2008) combine forecasts of U.S. CPI using the bootstrap forecast aggregation method. Their results confirm the superiority of the method compared to Bayesian averaging or Bayesian shrinkage estimators used by other researchers such as Groen *et al.* (2010) and Stock & Watson (2012). Vasnev *et al.* (2013) combine forecasts of models of macroeconomic time series to predict the monetary operations of the Reserve Bank of Australia with their findings confirming the benefits of forecast combination models. Lastly, Sermpinis *et al.* (2015) investigate an SVR model, ANNs and combination forecasts applied to unemployment forecasting. They conclude the combination forecasts are successful, with SVR and Kalman filter models improving the forecast accuracy.

## 7.2 Motivation

The research presented in this chapter is motivated by the desire to address the challenges faced when performing macroeconomic forecasting, in particular, challenges related to macroeconomic modelling. The motivation is twofold. The primary motivation is to evaluate the performance of ANNs in their ability to forecast U.S. GDP over a one to four quarters forecasting horizon and evaluate if they offer superior forecasting ability over an SVR model and Naïve forecast. The secondary motivation is to explore the use of ANN forecast combinations and evaluate if the forecast combination provides enhanced accuracy. A specific commentary is given on the ANN combination point-forecast over the 1-4 quarters forecasting horizon and during the financial crisis of 2008. Forecast evaluation is performed over a 57-year sample period (1960-2016), reserving the last 20% of data for out-of-sample testing. In-sample and out-of-sample forecast evaluation is performed using the following forecast evaluation metrics; Mean Error (ME), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) Mean Absolute Error (MAE), Mean Percent Error

(MPE), Mean Absolute Percent Error (MAPE), Theils-U1, Theils-U2 and the modified Diebold-Mariano test.

This research should extend the literature concerning ANNs applied to macroeconomic forecasting. Additionally, the performance of the ANN combination forecasts adds validity to existing research that reports the benefits of such forecasting techniques. The remainder of this chapter is presented as follows; section two presents descriptive statistics of the U.S. GDP time series data used in this investigation, initially in its original (\$Bn) format then percent-change (% $\Delta$ ) format. Section three details the IVS methods and subset selection. Section four presents the empirical results and section five finishes with the conclusion.

### **7.3 U.S GDP Time Series Data**

The decision to forecast GDP over alternative macroeconomic indicators is twofold. First, GDP is regarded as the principle measure of economic performance, therefore, is arguably the most meaningful macroeconomic indicator. For example, it is GDP that is used by institutions such as the IMF, World Bank and the UN to rank nations. Second, GDP typically undergoes limited revision after its initial release. This is an important characteristic as it allows for the sidestepping of the problem of collecting and assembling appropriate vintages of data. This study uses the last release of U.S. GDP for all training and testing. For example, the largest discrepancy between the original vintage of the data and final release of the data is about 23 basis points with the average discrepancy being 9 basis points. It is assumed the impact of these discrepancies on the predictive accuracy of the forecasts is negligible. The following sections present descriptive statistics of the U.S. GDP time series data. The descriptive statistics include statistical tests covering; normality (Jarque-Bera), stationarity (Augmented Dickey-Fuller), monotonic trend presence (Mann-Kendall) and the minimum number of explanatory variables required to model the time series data (Fractal Dimension). The full description and disclosure of each statistic are available in appendix 10.

### 7.3.1 Descriptive statistics, U.S. GDP (\$Bn)

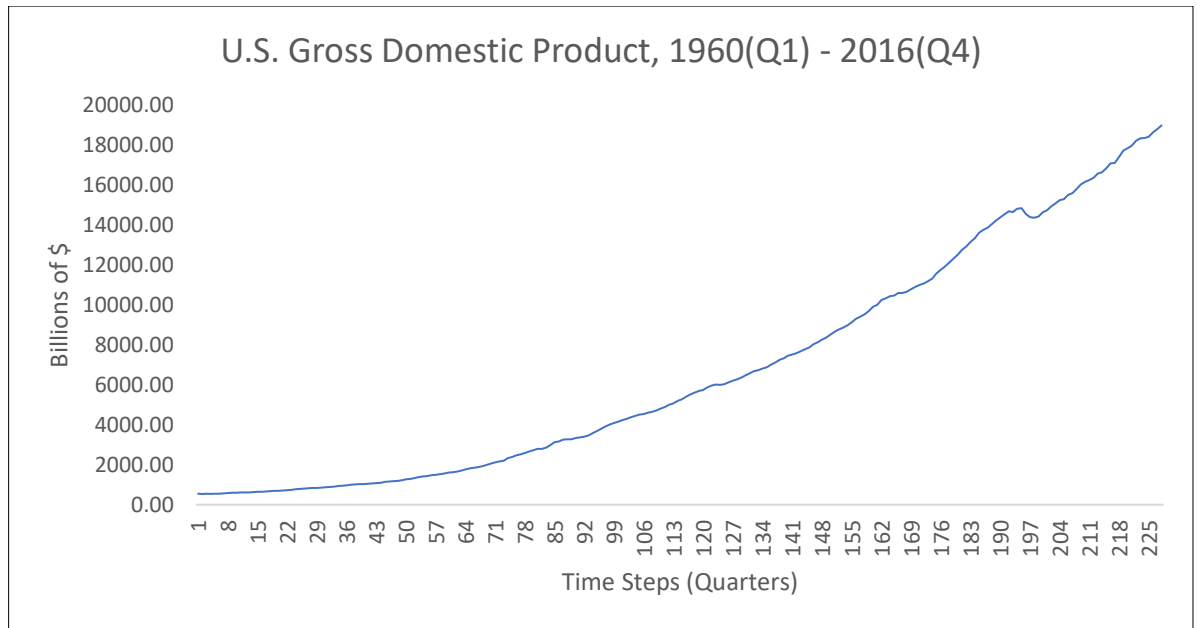
There are three methods of determining a nation's GDP all of which should, in principle, yield the same result. The favoured method is the expenditure method where GDP ( $Y$ ) is calculated according to the following equation, with figure 7-1 depicting U.S. GDP calculated over the sample period (1960-2016):

$$Y = C + I + G + (C - M) \quad (46)$$

Where:

$C$  = Consumption,  $I$  = Investment,  $G$  = Government Spending,  $X$  = Imports, and  $M$  = Exports

Figure 7-1: U.S. GDP Time Series Data (1960 – 2016), \$Bn Format.



Visually, an automated trend line illustrates that U.S. GDP features an overall uptrend, commencing at \$542Bn on 1st January 1960 and terminating at \$19.98Tr on the 1st October 2016. In the last quartile, there is a noticeable disruption to the curve located between 1/1/2008 and 1/4/2009 where U.S. GDP turned negative for the first time since 1990. This notable disruption is the recession of 2008 formally known as the banking crisis. In addition to the overall analysis, this period will be examined in isolation to evaluate the performance of ANNs under such circumstances. The GDP data was sourced from the economic data section of the Federal Bank of St Louis (FRED), available at;



www.stlouisfed.org. The observations are of quarterly frequency, measured in billions of USD, annual rate and seasonally adjusted to deter ANN overfitting.

Table 7-1: Descriptive Statistics, U.S. GDP (\$Bn).

<b>Descriptive Statistics</b>	<b>Confidence Level</b>	<b>P-Value</b>	<b>Result</b>
Mann-Kendall	99%	<0.0001	Monotonic Trend Presence
Jarque-Bera	99%	<0.0001	Not Normally Distributed
Dickey-Fuller	99%	0.9286	Non-Stationary
Fractal Dimension	N/A	N/A	2 (2.1)

From table 7-1, the uptrend is empirically evidenced by the Mann-Kendal trend test, confirming the presence of a monotonic trend – this is visually evident from Figure 7-1, displaying a smooth incline over the sample period and reinforced by the automated trendline (red). The Jarque-Bera test for normality finds the time series to be not normally distributed with positive skewness of 0.61 and platykurtic kurtosis of -0.95. The augmented Dickey-Fuller test for stationarity categorises the time series as non-stationary and the test of fractal dimension generated the largest value of 2.1, therefore, the time series requires a minimum of 2 explanatory variables for modelling purposes.

### 7.3.2 Descriptive Statistics, U.S. GDP (%Δ)

The descriptive statistics mentioned thus far have covered the U.S. GDP time series in \$Bn format. This section will apply the same descriptive statistics to the U.S. GDP time series transformed into %Δ format. Figure 7-2 displays the U.S. GDP data transformed into %Δ format using the following transformation:

$$\% \Delta = \left( \left( \frac{x_t}{x_{t-1}} \right) - 1 \right) \times 100 \quad (47)$$

The percent-change data transformation is implemented to reduce the magnitude of the data and solve the problem of non-stationarity.

Figure 7-2: U.S. GDP Data - % $\Delta$  Format (1960 – 2016).

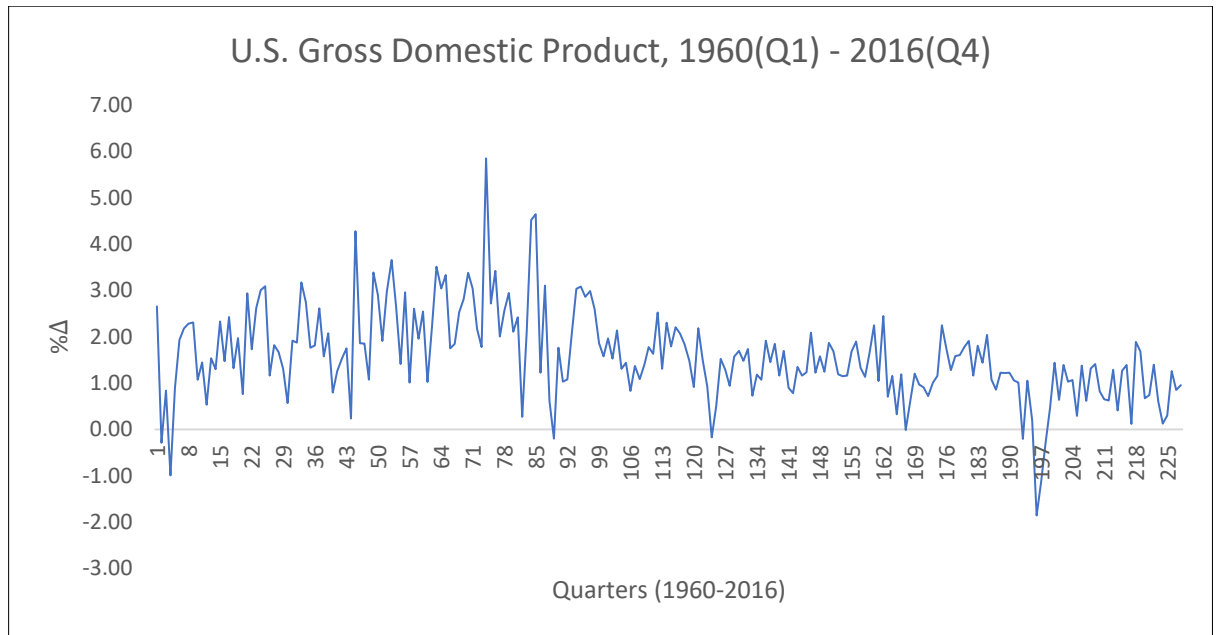


Table 7-2: Descriptive Statistics, U.S. GDP (% $\Delta$ )

Descriptive Statistics	Confidence Level	P-Value	Result
Mann-Kendall	99%	<0.0001	Monotonic Trend Presence
Jarque-Bera	99%	<0.0001	Not Normally Distributed
Dickey-Fuller	99%	0.0008	Stationary
Fractal Dimension	N/A	N/A	2 (2.2)

From table 7-2, the Mann-Kendal trend test confirms the presence of a monotonic trend. This is surprising as one of the properties of the % $\Delta$  data transform is the removal of the trend from the time series. The Jarque-Bera test for normality found the time series to be not normally distributed with positive skewness of 0.46 and leptokurtic kurtosis of 2.24. The augmented Dickey-Fuller test for stationarity categorises the time series as stationary, and the test of fractal dimension generated the largest value of 2.2 which approximates U.S. GDP in \$Bn format, therefore, both formats require a minimum number of 2 explanatory variables for modelling purposes.

## 7.4 Input Variable Selection

### 7.4.1 Filter Methods

A two-stage IVS method is implemented to identify the most informative variables from the candidate input variable pool. First, filter methods are applied as an initial screening process to identify and remove constant, quasi-constant and duplicated variables. Of the original 780 variables, 1 variable was identified as a duplicate variable and removed, reducing the dataset to 779 variables. Pearsons correlation coefficient is then applied to test the correlation between the target variable (GDP) and the predictor variables. An arbitrary threshold of  $\pm 0.80$  (strong relationship) is set and predictor variables that do not satisfy that criteria are removed, further reducing the dataset by 752 variables to 27 variables. With duplicate variables removed and the candidate variable pool reduced to 27 highly correlated variables, the remaining variables progress to the second stage of IVS, LASSO regression, described in the next section.

### 7.4.2 Least Absolute Shrinkage and Selection Operator (LASSO)

For a description of LASSO, refer to chapter 6, subsection 6.4.2. Following the implementation of LASSO, two variables were selected; GDP (% $\Delta$  from the preceding period) and GDP Per Capita (\$Bn) as described in Appendix 5.

## 7.5 Methodology

In order to perform a fair and extensive investigation, the following methodology was devised and implemented. Neuro Solutions Pro was utilized to build and test the ANNs, with network parameters initially held constant across all networks, i.e. number of hidden layers, number of processing elements in each hidden layer, number of epochs, weighting update method, activation function and learning rule. To avoid the issue of overfitting, various procedures were adhered to, namely parameter pruning; Castellano *et al.* (1997) and Wang *et al.* (2010), cross-validation; Zhang *et al.* (1999), Amjady & Keynia (2009) and Sermpinis *et al.* (2012b) and early stopping; Lin *et al.* (2009) and Prechelt (2012). The rationale underpinning each parameter will now be discussed.

### 7.5.1 Artificial Neural networks

This investigation will assess the performance of a collection of ANNs, namely the Multilayer Perceptron (MLP), Radial Basis Function (RBF), Recurrent Neural Network (RNN), Generalised Feed-Forward (GFF), Probabilistic Neural Network (PNN), Time-Lag Recurrent Network (TLRN) and CANFIS Network.

### 7.5.2 Number of Hidden Layers

For each type of ANN, custom models are built featuring 1 to 4 hidden layers. Although any number of hidden layers is possible, it is widely accepted and supported by the general function approximation theorem that two hidden layers can approximate any complex nonlinear function and provides powerful computational properties; Dayhoff & DeLeo (2001). For that reason, this investigation incorporates the two hidden layer principle and extends beyond it.

### 7.5.3 Processing Elements

To identify the optimum number of processing elements (PEs) in each hidden layer, the *vary a parameter* function is employed which allows automated incremental testing to identify the optimal number of PEs defined in terms of lowest MAE. The optimum number of PEs to be identified is confined to a generous range of 1 to 50 with testing increments of 1. Therefore, a total of 50 PEs are tested for each hidden layer. During this phase, the network tests all PEs over 3 cycles, therefore a total of 150 simulations (50 PEs x 3 cycles) are performed on each hidden layer. This equates to 150 (150 simulations x 1 hidden layer) simulations for ANNs featuring one hidden layer and 600 (150 simulations x 4 hidden layers) simulations for ANNs featuring four hidden layers.

### **7.5.3 Epochs**

The number of epochs was set to 1000 and programmed to terminate after 200 epochs without improvement, therefore, implementing early stopping to guard against over-fitting as presented in Lin *et al.* (2009) and Prechelt (2012).

### **7.5.5 Weights**

Batch weighting updates the weights after the presentation of the entire training set.

### **7.5.6 Activation Function**

All ANN models are replicated to feature the following activation functions; Tanh Axon, Sigmoid Axon, Linear Tanh Axon, Linear Sigmoid Axon, SoftMax Axon, Bias Axon, Linear Axon, Axon. For a description of each activation function, see appendix 6.

### **7.5.7 Memory**

For Recurrent ANNs that feature a memory component, focused and unfocused; Gamma Axon, Laguerre Axon and TDNN Axons are used.

### **7.5.8 Learning Rule**

All ANN models will feature the Levenberg-Marquardt (LM) learning rule as it is one of the most appropriate higher-order adaptive algorithms known for minimizing the MSE of an ANN. It is also known for training significantly faster than the commonly employed momentum learning rule and typically arrives at a solution with a significantly lower error.

### 7.5.9 ANN Competition and Selection

Following training and testing, the best ANN from each class defined in terms of lowest RMSE is selected to compete against an SVR model and Naïve forecast. Therefore, the best performing MLP, RNN, GFF, PNN, RBF, TLRN, and CANFIS model will compete against the abovementioned benchmarks.

### 7.5.10 Data Segregation

This investigation features a 57-year sample period, (2nd January 2012 – 30th December 2016). For ANN data analysis, the sample period is segregated as follows; three years are allocated for training (2nd January 2012 – 31st December 2014), one year for cross-validation (1st January 2015 – 31 December 2015) and one year for testing (1st January 2016 – 30th December 2016). This equates to a ratio of 60:20:20. Table 7-3 displays this information:

Table 7-3: U.S. GDP, Data Segregation of Sample Period (1960-2016)

<i>Data Segregation</i>				
<b>Period</b>	<b>Start Date</b>	<b>End Date</b>	<b>% of Data</b>	<b>Quarters</b>
Sample Period	1st January 1960	1st October 2016	100	228
Training	1st January 1960	1s October 1997	65	148
Cross Validation	1st January 1998	1st April 2006	15	34
Testing	1st July 2006	1st October 2016	20	45

## 7.6 Empirical Results

As it is standard in the literature, to statistically evaluate the forecasts, the ME, MSE, MAE, MPE, and RMSE are calculated. Regarding the interpretation of the above statistics, the lower the value, the better the forecasting accuracy of the model. Additionally, the advanced forecasting evaluation metrics of Theils-U1, Theils-U2, and MDM are computed. A lower value of Theils-U1 is desirable, with values bound between 0 and 1. For Theils-U2, values <1 indicate superior forecasting ability than the Naïve forecast, values of 1 equal the Naïve forecast, and values >1 indicate the model has underperformed the Naïve forecast. The mathematical formulas of these statistics are given in Appendix 1.

### 7.6.1 In-Sample Forecast Evaluation

Table 7-4 displays the in-sample results of the ANNs and benchmark models.

Table 7-4: In-Sample Results, U.S. GDP Forecasting, ANNs and Benchmark Models.

<i>In-Sample Forecast Evaluation</i>								
<b>Model</b>	<b>MLP</b>	<b>GFF</b>	<b>MNN</b>	<b>RBF</b>	<b>PNN</b>	<b>CANFIS</b>	<b>Fc</b>	<b>SVR</b>
ME	0.2440	0.2087	0.8144	0.4330	-0.0785	<b>0.0446</b>	0.2777	-0.2310
MSE	0.9515	0.9530	1.4342	0.9351	0.7002	0.7556	0.8096	<b>0.3868</b>
RMSE	0.9755	0.9762	1.1976	0.9670	0.8368	0.8692	0.8998	<b>0.6220</b>
MAE	0.7082	0.7063	0.9164	0.7234	0.6361	0.6423	0.6488	<b>0.4877</b>
MPE	0.4277	0.4669	0.7136	<b>0.4031</b>	0.4410	0.4859	0.4897	0.4544
MAPE	0.9422	1.0308	0.8651	<b>0.7523</b>	1.2406	1.1736	0.9500	1.1136
Theils-U1	0.2663	0.2667	0.4001	0.2798	0.2146	0.2303	0.2531	<b>0.1498</b>
Theils-U2	0.9073	1.0264	0.6014	1.0630	1.6355	1.5178	1.1107	<b>0.6607</b>

From Table 4-7, the in-sample results demonstrate the SVR benchmark model it is the superior forecasting model, acquiring 5 of the 8 forecast evaluation metrics: MSE (0.39), RMSE (0.62), MAE (0.49), Theils-U1 (0.15), and Theils-U2 (0.66). The remaining metrics of ME, MPE, and MAPE are acquired by the CANFIS model with an ME of 0.04, and the RBF with an MPE and MAPE of 0.40 and 0.75 respectively. Amongst the ANNs, forecast evaluation results are mixed. The PNN achieved 3 of the 8 forecast evaluation metrics: MSE (0.70), RMSE (0.84), and MAE (0.64), indicating that of the ANNs it generated the most accurate and consistent forecasts. The RBF achieved 2 of the 8 forecast evaluation metrics: MPE (0.40), and MAPE (0.75). The CANFIS achieved the best ME value of 0.04 which is notably better than all other models. The MNN achieved the best Theils-U2 value of 0.60 meanwhile the GFF, RBF, PNN, CANFIS, and Fc models all recorded Theils-U2 >1 indicating they underperformed a Naïve Forecast. The MLP, GFF, and Fc models did not achieve any in-sample forecast evaluation metrics.

### 7.6.2 Out-of-Sample Forecast Evaluation

In addition to the standard set of forecasting evaluation metrics, to further verify the forecasting ability of the ANN, the Modified Diebold-Mariano (MDM) test is applied as proposed by Harvey *et al.* (1997). The null hypothesis of the test is the equivalence in forecasting accuracy between competing forecasting models. The use of the MDM test is common practice in forecast evaluation as it is found to be robust in assessing the

significance of observed differences between competing forecasts (Barhoumi *et al.*, 2010), and overcomes the problem of over-sized DM values in moderate samples (Dreger and Kholodilin, 2013). The MDM test is an extension of the Diebold-Mariano (1995) test and is presented below:

$$MDM = T^{-1/2}[T + 1 - 2k + T^{-1}k(k - 1)]^{1/2}DM \quad (49)$$

Where  $T$  is the number of out-of-sample observations and  $k$  is the number of the step-ahead forecasts. In this case the MDM test is applied to competing forecasts (ANN vs. SVR). A negative MDM value indicates that the ANN model is more accurate than SVR benchmark model. The lower the negative value, the more accurate are the ANN forecasts. Table 7-5 displays the out-of-sample forecasting evaluation results:

Table 7-5: Out-of-Sample Forecasting Results

<i>Out-of-Sample Forecast Evaluation</i>								
<b>Model</b>	<b>MLP</b>	<b>GFF</b>	<b>MNN</b>	<b>RBF</b>	<b>PNN</b>	<b>CANFIS</b>	<b>ANN Fc</b>	<b>SVR</b>
ME	-0.2360	-0.1638	<b>-0.1511</b>	-0.1904	-0.9400	-0.7403	-0.4036	-1.5288
MSE	0.6919	0.5619	<b>0.4784</b>	0.5648	1.3418	1.0490	0.6476	3.1049
RMSE	0.8318	0.7496	<b>0.6916</b>	0.7516	1.1584	1.0242	0.8047	1.7621
MAE	0.5983	0.5094	<b>0.4712</b>	0.5205	0.9492	0.7971	0.5409	1.5290
MPE	-0.3196	<b>-0.1447</b>	-0.1919	-0.1463	-1.1974	-0.9530	-0.4921	-1.9179
MAPE	1.2644	1.0277	<b>0.9979</b>	1.0856	2.0982	1.8359	1.2964	3.1786
U1	0.3814	0.3676	<b>0.3439</b>	0.3621	0.4135	0.3926	0.3545	0.5124
U2	0.9332	0.9332	<b>0.8445</b>	0.8955	1.1074	1.0404	0.9169	1.6690
MDM	-4.7448	-5.1180	-4.8268	-5.0320	<b>-5.4050</b>	-5.3287	-5.2852	N/A

\*1% Significance, \*\*5% Significance

From table 7-5 it is evident that out-of-sample forecasting performance changed dramatically. The SVR benchmark which claimed best in-sample performance failed to acquire any forecasting evaluation metrics and recorded the worst Theils-U2 value of 1.67 indicating it failed to outperform a Naïve forecast. Out-of-sample, the MNN emerges as the best model, achieving 7 of the 9 forecasting evaluation metrics of: ME (0.15), MSE (0.48), RMSE (0.69), MAE (0.47), MAPE (0.99), Theils-U1 (0.34), and Theils-U2 (0.84). The best ME was achieved by the GFF model with a value of 0.14. The Modified Diebold-Mariano (MDM) statistic as proposed by Harvey *et al.* (1997) is applied to couples of forecasts (ANN Vs Benchmark Model). A negative value of the MDM test statistic indicates the ANN model is more accurate than the benchmark model. The lower the negative value, the more accurate the ANN forecast. The out-of-sample statistical



superiority of the PNN is confirmed by the MDM test, with a value of -5.40, closely followed by the CANFIS model with a value of -5.29.

### 7.6.3 Point Forecasts

Table 7-6: Point Forecasts, U.S. GDP, Q1-Q4 (2017).

<i>Point Forecast</i>										
Date	Quarter	GDP	MLP	GFF	MNN	RBF	PNN	CANFIS	SVR	ANN Fc
01/01/2017	Q1	0.97	0.73	0.64	0.89	0.79	1.64	1.40	2.56	<b>1.01</b>
01/04/2017	Q2	1.03	1.31	0.79	<b>0.98</b>	1.28	1.73	1.65	2.92	1.29
01/07/2017	Q3	1.18	1.05	0.95	0.92	0.83	1.74	1.58	2.27	<b>1.18</b>
01/10/2017	Q4	1.24	1.11	0.92	0.93	1.02	1.74	1.61	3.06	<b>1.22</b>

From table 7-6, the results of the point forecast demonstrate the ANN Fc is the superior model as it achieves the most accurate point forecasts across in quarters 1, 3, and 4. Furthermore, the ANN Fc achieves 100% forecast accuracy in Q3. The MNN reinforces its out-of-sample performance by achieving the most accurate Q2 forecast of 0.98 Vs 1.03.

### 7.6.4 Crisis Period Forecast Evaluation

Table 7-7 presents the forecasting performances during the financial crisis of 2007-2009, which according to the National Bureau of Economic Research (the official arbiter of U.S. recessions) began in December 2007 and ended in June 2009. The sample period containing the crisis is, therefore, 1/10/2007 to 1/4/2009.

Table 7-7: Financial Crisis, (2007-2009), Forecasting Performance.

<i>Crisis Period</i>								
Model	MLP	GFF	MNN	RBF	PNN	CANFIS	Fc	SVR
ME	-1.2512	-1.3041	<b>-1.1091</b>	-1.3637	-1.8786	-1.7463	-1.4422	-2.6792
MSE	2.3294	2.5998	<b>2.1773</b>	2.6105	4.4584	3.9142	2.9301	8.6674
RMSE	1.5263	1.6124	<b>1.4756</b>	1.6157	2.1115	1.9784	1.7118	2.9441
MAE	1.2512	1.3041	<b>1.1552</b>	1.3637	1.8786	1.7463	1.4422	2.6792
MPE	1.4119	<b>1.2947</b>	1.3768	1.4995	1.6519	1.5919	1.4711	1.7662
MAPE	2.8677	2.8455	<b>2.4168</b>	3.0514	4.2357	3.9712	3.2240	6.5167
Theils-U1	<b>0.7215</b>	0.7544	0.7621	0.7337	0.7799	0.7666	0.7522	0.8275
Theils-U2	1.2451	1.2832	<b>1.1647</b>	1.2320	1.5388	1.4803	1.3202	2.0562
MDM	-3.0723	-3.3468	<b>-3.4237</b>	-3.1552	-3.2128	-3.0766	-3.2428	N/A

From table 7-7, the MNN reinforces its out-of-sample forecasting performance by demonstrating it is the superior forecasting model during the crisis period. The MNN achieved 7 of the 9 forecast evaluation metrics: ME (-1.11), MSE (2.17), RMSE (1.48), MAE (1.56), MAPE (2.42), Theils-U2 (1.16), and MDM (-3.42). The GFF model achieved the best MPE of 1.29, and the MLP achieved the best Theils-U1 value of 0.72. During the crisis period, all models recorded Theils-U2 values >1 indicating they underperformed a Naïve forecast.

## 7.7 Conclusion

The motivation of this research is to investigate the use of ANNs applied to forecast the  $\% \Delta$  in U.S. GDP over a 1-4 quarter forecasting horizon, evaluating if ANNs offer superior forecasting ability than an SVR model and Naïve forecast. The investigation also evaluates the performance of an ANN combination forecast model (Fc) and the forecasting ability of the featured models during times of economic crisis. The statistical evaluation is recorded over a 57-year sample period (1960-2016) reserving the last 11-years (20%) for out-of-sample-testing.

The results demonstrate that ANN, especially the Modular network, are superior to the benchmark models in the critical out-of-sample period, achieving the best forecasting performance metrics and point-forecast values. The performance of the ANN forecast combination model was especially demonstrated in the point-forecasts, achieving close prediction in 3 of the 4 quarters and achieving 100% forecast accuracy in 2017(Q3). The SVR benchmark model dominated the in-sample results; however, its performance diminished significantly in the out-of-sample period and underperformed the Naïve forecast, recording the worst Theils-U2 value of all models. The Modular network dominated the out-of-sample period, securing 6 of the 9 forecast evaluation metrics – including; RMSE, Theils-U1 and Theils-U2. The Modular network also exhibited the best ANN performance during the financial crisis, also outperforming the Fc model which is said to offer greater forecast stability during crisis periods. However, the Naïve forecast outperformed all models during the crisis period, therefore questioning the claims that ANNs offer superior forecasting ability during times of economic crisis. A potential extension of this research could feature additional macroeconomic variables external to the

Federal bank of St Louis data source, such as stock market returns, relevant commodities such as oil, and regional macroeconomic indicators as opposed to only national-level macroeconomic indicators also available from the Federal Bank of St Louis.

# Chapter 8

## General Conclusion

Our current understanding of financial markets and financial market forecasting justifies the need for CI and AI models, such as ANNs. This is reinforced by our interconnected and increasingly complicated financial markets. Furthermore, periods of economic turmoil, such as the 2008 global recession, command more robust and accurate forecasting models that can adapt to changing market environments and exploit nonlinearities in the data.

Chapter 4 evaluated ANNs applied to exchange rate forecasting in which a collection of ANNs was applied to forecast the logarithmic returns of the EUR/USD over a short-term forecast horizon of one period. Despite considerable effort to equip the ANNs with informative technical indicators by compiling an extensive candidate input variable pool consisting of standard and advanced technical indicators measuring all technical aspects of the EUR/USD times series for a total of 10,568 models, and with the best models selected through a two-stage IVS process, the ANNs failed to outperform a MACD model. The ANNs forecasting performance was worsened by the MACDs domination of both the in-sample and out-of-sample periods, with some ANNs underperforming a Naïve forecast. This chapter also investigated the performance of ANNs during increased market volatility, finding that an ANN forecast combination did offer enhanced stability under increased volatility by offering lower errors and increased directional accuracy, however, was still outperformed by the MACD model.

Chapter 5 evaluated ANNs applied to volatility forecasting in which a collection of ANNs was applied to forecast the volatility of the Nikkei 225 Index over a short-term forecast horizon of one period. Like chapter 4, considerable effort was taken to compile an extensive candidate variable pool consisting of historical volatility and implied volatility models. A wider range of ANNs was also evaluated. Of the various ANNs, the Recurrent Neural Network demonstrated robust in-sample performance, however, was superseded in the out-of-sample period by the benchmark GARCH (1,1) model.

Chapters 6 and 7 evaluated ANNs applied to macroeconomic forecasting, with chapter 6 applying ANNs to forecast U.S. civilian unemployment and chapter 7 applying ANNs to forecast U.S. GDP. In both studies, the ANNs generally demonstrated superior performance over the SVR benchmark model, with the Modular Neural Network performing notably well in both studies, especially the GDP study in which it offered greater stability during the 2008 financial crisis period. The use of an ANN forecast combination was also evaluated in the GDP study, finding that extremely close prediction was achieved over the 4 quarters forecast horizon, with 100% accuracy achieved in Q3.

This thesis addressed a specific deficiency found in prior research concerning candidate input variable pools in that they are typically not compiled, so far as possible, of all available and relevant variables. This is especially true of research that combines ANNs with TA by utilising technical indicators as input variables. In solution, considerable effort was undertaken to compile extensive candidate input variable pools, largely applicable to chapters 4 and 5. Despite the extensive candidate input variable pools and best inputs selected through a two-stage IVS process, the ANNs failed to outperform the benchmark models in both studies, with their underperformance most observable in chapter 4 where a MACD model dominated both the in-sample and out-of-sample periods. This is compounded by the fact that a MACD model was included in the candidate input variable pool of chapter 4 but deemed to be a lesser model in comparison to other models and not selected by IVS. This result suggests it could be more efficient to equip ANNs with common technical indicators and their default parameters. However, the effort to compile extensive candidate input variable pools was redeemed in chapters 6 and 7 regarding macroeconomic forecasting. With the Federal Bank of St Louis offering their MS Excel add-in giving direct access to their data bank and the download of such data subject to user specified date ranges and data transforms, allows for the efficient gathering of data and the construction of extensive candidate variable pools.

Future research could extend on this thesis by including exchange rate volume data in candidate variable pools. Previously, exchange rate volume data has typically been omitted from prior research due to exchange rate volume data being decentralised therefore inaccurate. However, such data is now available in centralised, aggregate, format from financial data providers such as Quandl. The inclusion of exchange rate volume data may serve as an informative input variable as volume data is typically highly correlated with volatility and a central principal of DT. The use of autoregressive terms and lagged values

may also be beneficial, particularly for macroeconomic data. The learning algorithm was set to Levenberg-Marquart across all empirical chapters, alternative learning algorithms such as: Resilient Backpropagation (RProp), Step / Momentum, Delta Bar Delta, Quickprop, or Conjugate Gradient could also be investigated.

This thesis concludes by submitting additional evidence to the field of Computational Finance and the existing body of knowledge. Although universal approximations can never be embraced in scientific research, the empirical evidence of chapter 4-7 holds strong implications for decision making. Governments and banks should be of interest as to the capabilities of ANNs in relation to macroeconomic forecasting as it was found that ANNs are perhaps better suited to macroeconomic forecasting, were data is more readily available, revised, and of lower frequency. Nonetheless, there remains additional research to be performed regarding the calibration of AI and CI models for financial and economic forecasting applications.

## Appendices

### Appendix 1: Forecast Evaluation Metrics

#### Forecast Accuracy

For the following statistics, Denoting the series of interest as  $y_t$  and a forecast of it as  $\hat{y}_t$ , the resulting forecast error is given as  $e_t = (y_t - \hat{y}_t)$ , for  $t = 1 \dots T$ . Per this notation, the set of forecast evaluation statistics are presented below:

$$MAE = \frac{1}{T} \sum_{t=1}^T |e_t| \quad (50)$$

$$MSE = \frac{1}{T} \sum_{t=1}^T e_t^2 \quad (51)$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T e_t^2} \quad (52)$$

$$MAPE = \frac{1}{T} \sum_{t=1}^T 100 \times \left( \left| \frac{e_t}{y_t} \right| \right) \quad (53)$$

#### Theils-U1

The Theils-U1 statistic is bounded between 0 and 1, with values closer to 0 indicating greater forecasting accuracy. The Theils-U1 measure was applied by comparing the ANN forecasts to that of a Random Walk. Denoting the series of interest as  $y_t$  and a forecast as  $\hat{y}_t$ , the resulting forecast error is given as  $e_t = (y_t - \hat{y}_t)$ , for  $t = 1 \dots T$ . Using this notation, Theils-U1 is presented as:

$$Theils - U1 = \frac{\sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - f_t)^2}}{\sqrt{\frac{1}{N} \sum_{t=1}^N y_t^2} + \sqrt{\frac{1}{N} \sum_{t=1}^N f_t^2}} \quad (54)$$

## Theils-U2

A value less than 1 for Theils-U2 confirms the superiority of the compared forecast, while a value greater than 1 shows a higher accuracy for the benchmark forecast. Denoting the series of interest as  $y_t$  and the forecast as  $f_t$ , the resulting forecast error is given as  $e_t = (y_t - f_t)$ , for  $t = 1 \dots T$ . Using this notation, Theils-U1 is presented as:

$$Theils - U2 = \frac{\sqrt{\frac{1}{T} \sum_{t=1}^{T-1} \left( \frac{f_{t+1} - y_{t+1}}{y_t} \right)^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^{T-1} \left( \frac{y_{t+1} - y_t}{y_t} \right)^2}} \quad (55)$$

## Pesaran & Timmermann (1992)

The Pesaran & Timmermann (1992) test, stated as PT in the performance tables is a non-parametric test to examine the ability of a forecast to predict the direction of change in a series of interest. Denoting the series of interest as  $y_t$  and its forecast  $x_t$  the PT test is defined as:

$$Sn = \frac{\hat{P} - \widehat{Pt}}{[\widehat{V}(\hat{P}) - \widehat{V}(\widehat{P}_*)]^{0.5}} \quad (56)$$

Where;

$$\hat{P} = n^{-1} \sum_{t=1}^n I(y_t x_t)$$

$$\widehat{P}_t = \hat{P}_y \hat{P}_x + (1 - \hat{P}_y)(1 - \hat{P}_x)$$



$$\hat{V}(\hat{P}) = n^{-1} \hat{P}_*(1 - \hat{P}_*)$$

$$\hat{V}(\hat{P}_*) = n^{-1}(2\hat{P}_y - 1)^2 \hat{P}_x(1 - \hat{P}_x) + n^{-1}(2\hat{P}_x - 1)^2 \hat{P}_y(1 - \hat{P}_y) + 4n^{-2} \hat{P}_y \hat{P}_x(1 - \hat{P}_y)(1 - \hat{P}_x)$$

$$\hat{P}_y = n^{-1} \sum_{t=1}^n I(y_t)$$

$$\hat{P}_x = n^{-1} \sum_{t=1}^n I(x_t)$$

$$I(\cdot) = \begin{cases} 1 & \text{IF } \cdot > 0 \\ 0 & \text{otherwise} \end{cases}$$

**Diebold-Mariano**

$$DM = \frac{\bar{Z}}{\bar{c}} \sim (0,1), H_0: E(z_t) = 0 \quad (57)$$

Where;

$$Errors = \{\hat{\epsilon}_T\}, \{\hat{\eta}_T\}$$

$$Absolute\ Difference\ (Z_T) = |\hat{\eta}_T| - |\hat{\epsilon}_T|$$

$$Mean\ (\bar{z}) = \frac{\sum_{t=1}^{T^*} z_t}{T^*}$$

$$Covariogram(c) = [Cov(z_t, z_{t-p}), Cov(z_t, z_t), Cov(z_t, z_{t+p})]$$

$$Mean(\bar{c}) = \sum c/(p+1)$$

## Appendix 2: Trading Evaluation Metrics

### Annualized Return

Annualised return is the return earned by an investment each year over a given time period:

$$R^A = 252 \times \frac{1}{N} \sum_{t=1}^N R_t, \quad (58)$$

Where;

$R_t$  = Daily Returns

### Maximum Drawdown

Maximum Drawdown (MDD) is an indicator of maximum downside risk over a specified time period:

$$MDD = \text{Min}[R_t - \text{Max}(\sum_{t=1}^N R_t)] \quad (59)$$

Where;

$R_t$  = Daily Return

### Sharpe Ratio

One of the most commonly cited statistics in financial analysis is the Sharpe ratio (SR). The SR is the ratio of the excess expected return of an investment to its return volatility or standard deviation. The SR is defined as the ratio of the excess expected return to the standard deviation of return:

$$SR = \frac{r_p - r_f}{\sigma_p} \quad (60)$$

Where;  $r_p$  = portfolio return,  $r_f$  = risk free rate and  $\sigma_p$  = standard deviation of returns

The SR can also help explain whether a portfolio's excess returns are due to intelligent investment decisions or a product of excessive risk. A negative SR indicates that a risk-less asset such as the risk free rate would perform better than the security being analysed.

### Information Ratio

The information ratio (IR) is a ratio of portfolio returns in relation to a benchmark -- usually a related index. The IR also attempts to identify the consistency of the investor with the greater the IR value, the better:

$$IR = \frac{r_p - r_b}{s_{p-b}} \quad (61)$$

Where;  $r_p$  = return of the portfolio,  $r_b$  = return of a benchmark and  $s_{p-b}$  = tracking error (standard deviation of the difference between returns of the portfolio and the returns of the index).

## Appendix 3: EUR/USD Descriptive Statistics (Closing Price)

### Mann-Kendall Trend Test

The Mann-Kendall (MK) test (Mann 1945, Kendall 1975, Gilbert 1987) is a non-parametric, distribution-free, test that statistically assesses if there is a monotonic trend contained in the time series data which may or may not be linear:

Table 1: Mann-Kendall Test Results

Mann-Kendall Trend Test	
Kendall's tau	-0.411
S	-349884
Var(S)	247221596.7
p-value (Two-tailed)	< 0.0001
Alpha, $\alpha$	0.01

### Normality

The Jarque-Bera (JB) test was implemented to assess whether or not the EUR/USD time series data follows a normal distribution. The JB test is summarized as follows;  $H_0$ : the time series follows a normal distribution.  $H_a$ : the time series does not follow a normal distribution. The significance level was set to 0.01 (99%) and the resulting p-value registered as <0.0001 as evidenced from the results table below:

Table 2: Jarque-Bera Test Results

Jarque-Bera Test	
JB (Observed value)	139.714
JB (Critical value)	9.21
DF	2
p-value (Two-tailed)	< 0.0001
Alpha, $\alpha$	0.01

As the computed p-value of <0.0001 is lower than the significance level of 0.01, the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted. Therefore the EUR/USD time series data does not follow a normal distribution. For more details on the Jarque-Bera statistic see Jarque & Bera (1980).

## Stationarity

Regarding univariate time series analysis, non-stationarity is a problem for the time series (Moody, 1995) and also problematic for ANN data analysis. For example; the presence of a trend may have an undesired impact on prediction performance (Weigend & Gershenfeld, 1994). Similarly, Tseng *et al.* (2002) demonstrated that seasonal data has a significant impact on ANN prediction. For trending, seasonal and non-stationary data, difference or log-difference is a simple and effective widely used solution. The Augmented Dickey-Fuller (ADF) test was used to assess data stationarity. The test is summarized as follows;  $H_0$ : there is a unit root for the time series, therefore, the series is nonstationary,  $H_a$ : there is no unit root for the series, and the series is stationary. The following table summarises the ADF test results;

Table 3: ADF Test Results

Augmented Dickey-Fuller (ADF) Test	
Tau (Observed value)	-1.842
Tau (Critical value)	-0.26
p-value (one-tailed)	0.685
alpha	0.01

The significance level was set to 0.01 (99%) and the p-value registered as 0.685. Therefore as the computed p-value is greater than the alpha, the null hypothesis  $H_0$  is accepted, and the alternative hypothesis  $H_a$  is rejected. The EUR/USD time series data is therefore classed as non-stationary.

## Fractal Dimension

Fractal Dimension is a measure of the complexity of the phase portrait of a time series. The complexity number it generates is an indication of the minimum number of independent variables (technical indicators) required to model the time series data. However, it is not known what these variables are; such variable identification is assigned to IVS methods. Knowing the fractional dimension of a time series can help with ANN design because it helps answer the often subjective and troubling question of how many input variables to use. In this case, the fractional dimension of the EUR/USD time series data requires a minimum of 3 independent variables to model the time series data.

## Summary Statistics

For a summary of descriptive statistics regarding the EUR/USD time series data, see table 5 below. For statistics requiring a confidence or significance level, a 99% level was set:

Table 4: EUR/USD Summary Statistics

EUR/USD Summary Statistics	
Number of observations	1305
Number of missing values	0
Minimum	1.039
Maximum	1.394
Range	0.355
Median	1.267
Mean	1.232
Variance (n)	0.012
Variance (n-1)	0.012
Standard deviation (n)	0.108
Standard deviation (n-1)	0.108
Skewness (Pearson)	-0.204
Skewness (Fisher)	-0.204
Skewness (Bowley)	-0.429
Kurtosis (Pearson)	-1.550
Kurtosis (Fisher)	-1.551
Standard error of the mean	0.003
Geometric mean	1.227

## Appendix 4: EUR/USD Descriptive Statistics (Log>Returns)

### Mann-Kendall Trend Test

Table 5: MK Test Results:

Mann-Kendall Test	
Kendall's tau	-0.017
S	-14056.000
Var(S)	247199757.333
p-value (Two-tailed)	0.371
Alpha, $\alpha$	0.01

The MK test is summarised as follows;  $H_0$  there is no monotonic trend in the time series,  $H_a$  there is a monotonic trend in the time series. From the results, as the computed p-value is higher than the significance level ( $\alpha=0.01$ ), the null hypothesis  $H_0$  is accepted, and the alternative hypothesis  $H_a$  is rejected. Therefore, there is no monotonic trend in the EUR/USD log returns time series data.

### Normality

Table 6: Jarque-Bera Normality Test (Log>Returns);

Jarque-Bera test (Log Returns):	
JB (Observed value)	284.467
JB (Critical value)	9.21
DF	2
p-value (Two-tailed)	< 0.0001
Alpha	0.01

As the computed p-value of <0.0001 is lower than the significance level of 0.01 (99%), the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted. Therefore, the EUR/USD log returns time series data does not follow a normal distribution. It is therefore evident that the EUR/USD time series data in closing price format and log-returns format both do not follow a normal distribution.

## Stationarity

Table 7: ADF Test Results

Augmented Dickey-Fuller (ADF) Test	
Tau (Observed value)	-11.158
Tau (Critical value)	-0.26
p-value (one-tailed)	< 0.0001
Alpha	0.01

The significance level (alpha) was set to 0.01 (99%) and the p-value registered as <0.0001, therefore, the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted. Therefore, the EUR/USD log-returns time data series is classified as stationary.

## Fractal Dimension

The fractional dimension of the EUR/USD log returns time series data requires a minimum of 3 independent variables to model the time series data. Interestingly, this was also the same result for the EUR/USD time series data in closing price format.

## Summary Statistics

For a summary of descriptive statistics regarding the EUR/USD log returns time series data see table 8 below. For statistics requiring a confidence or significance level, a 99% level was set:

Table 8: EUR/USD Summary Statistics

EUR/USD Log Returns Summary Statistics	
Number of observations	1305
Number of missing values	0
Minimum	-0.0238
Maximum	0.0302
Range	0.0540
Median	-0.0001
Mean	-0.0002
Variance (n)	0.0000



Variance (n-1)	0.0000
Standard deviation (n)	0.0054
Standard deviation (n-1)	0.0054
Skewness (Pearson)	0.1492
Skewness (Fisher)	0.1494
Skewness (Bowley)	-0.0476
Kurtosis (Pearson)	2.2677
Kurtosis (Fisher)	2.2810
Standard error of the mean	0.0001

---

## Appendix 5: Input Variables

### Chapter 4: Input Variables

#### Relative Strength Index (Smoothed, Simple Moving Average)

This indicator attempts to smooth the Relative Strength Index indicator by calculating its moving average. Since it uses a simple moving average (which gives equal importance to all the periods within the moving average's time frame), this method of smoothing may cause more of a lag time than the exponential variant.

**Abbreviation:** RSI(SMA)7

**Category:** Momentum

#### Input Parameters:

Time Series = EUR/USD % Log Returns

RSI Period = 14

Average Period = 7

#### Calculation:

$$\text{Average}(X, \text{Average Periods}) \quad (62)$$

Where:

$X =$  RSI(Closing Price, RSI Periods)

$RSI =$  Relative Strength Index

$Average =$  Simple Moving Average

### Advanced Linear Weighted Moving Average

This indicator attempts to quantify movements in a time series by calculating the difference between a moving average and a different sized lagged moving average.

**Abbreviation:** LWMA(-L1(15))8

**Category:** Trend

#### Input Parameters:

Time Series = EUR/USD % Log Returns

LWMA1 Periods = 15

Lag = 1

LWMA2 Periods = 8

#### Calculation:

$$LWMA(X, n1) - Lag(LWMA(X, n2), L) \quad (63)$$

Where:

$X =$  Time Series

$n1 =$  LWMA1 Periods

$L =$  Lag Amount

$n2 =$  LWMA2 Periods

$LWMA =$  Linearly Weighted Moving Average

$Lag =$  Lag

## 2-Pole Butterworth Filter

The Butterworth filter was introduced to Finance by John F. Ehlers and is a low-pass filter borrowed from the field of Electrical Engineering. Butterworth filters operate by means of averaging and are thought of as a similar but superior version of an EMA. In Butterworth filters, a higher degree of filtering is accompanied by a higher degree of lag and vice versa. Given that lag is regarded as a serious defect in most technical indicators, it is preferable to minimize lag and accept the resultant smoothing. The low-frequency lag of the Butterworth filter is exemplified by the following equation where  $N$  is the number of poles in the filter and  $P$  is the longest cycle period to pass through the filter:

$$Lag = N \times P / \pi^2 \quad (64)$$

The application of the Butterworth filter is simply the multiple application of an EMA. The transfer response of an EMA is given by:

$$H(z) = \alpha \times g + (1 - \alpha) \times f[1] \quad (65)$$

Where:  $f$  = output, and  $g$  = input

Using a Z transform notation this becomes:

$$H(z) = \alpha \times g + (1 - \alpha) \times f \times Z^{-1} \quad (66)$$

Solving the output in terms of the input obtains:

$$H(z) \times (1 - (1 - \alpha)Z^{-1}) = \alpha \times g \quad (67)$$

$$H(z) = \left( \frac{\alpha}{1 - (1 - \alpha)Z^{-1}} \right) \times g \quad (68)$$

This equation is an expression of the signal output of filter in terms of the data input to the filter times the Transfer Response of the filter itself. Therefore, the transfer response of an EMA is denoted by:

$$H(z) = \frac{\alpha}{1 - (1 - \alpha)z^{-1}} \quad (69)$$

Applying an EMA ‘ $N$ ’ times gives a  $N$  Pole filter response as:

$$H(z) = \frac{\alpha^N}{1 - (1 - \alpha)z^{-1})^N} \quad (70)$$

At zero frequency  $Z^{-1}=1$ , so this low pass filter has unity gain. Also, the denominator assumes the value of  $\alpha^N$  at zero frequency. The corner frequency of the filter is defined as that point where the transfer response is down by 3 dB, or .707 in amplitude. When this occurs, we have the relationship:

$$(1 - (1 - \alpha)Z^{-1})^N = 1.414\alpha^N \quad (71)$$

Where  $Z^{-1} = e^{-j\omega}$  and  $\omega = 2\pi/P$

Following arithmetic, the solution for Alpha is:

$$\alpha = -\beta + SQR(\beta^2 + 2\beta) \quad (72)$$

Where:  $\beta = (1 - \cos(\omega))/(1.414^{\frac{2}{N}} - 1)$

## Chapter 5: Input Variables

### EWMA(0.95)

The RiskMetrics Volatility Model is a special case of the Exponential Weighted Moving Average Model (EWMA). The EWMA suggests that the variance of a financial asset can be calculated using the formula:

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1 - \lambda)r_{t-1}^2 \quad (73)$$

Where  $\sigma_{t-1}^2$  is the EWMA variance at time  $t=1$ ,  $r_{t-1}^2$  the squared returns at  $t-1$ , and  $\lambda$  a weight between 0 and 1. Following IVS, a RiskMetrics Volatility Model with a weight of 0.95 was selected therefore calculating daily volatility as:

$$RiskMetricsVol = \sqrt{0.95\sigma_{t-1}^2 + 0.05r_{t-1}^2} \quad (74)$$

### **Nikkei 225 Volatility Index**

The Nikkei 225 VI indicates the expected degree of fluctuation of the Nikkei 225 Index in the future. The greater the index values are, the larger fluctuation investors expect in the market. Nikkei 225 VI Futures are based on the Nikkei 225 VI, which is an index, calculated by Nikkei Inc., estimating the degree of expected fluctuation in the Nikkei 225. The Nikkei225 Volatility Index displays the following characteristics:

**High Volatility.** Nikkei 225 VI has a regular range of 22 ~ 28 points, however when the Nikkei 225 declines sharply, the Nikkei 225 VI may advance rapidly.

**Mean Regression.** The Nikkei 225 VI has a tendency to regress to its regular range of 22 ~ 28 points following a rapid increase without sticking in a high point range.

### **Calculation Method:**

The Nikkei 225 Volatility Index is calculated using prices of Nikkei 225 futures and Nikkei 225 options on the Osaka Exchange (OSE). In the calculation, taking near-term future price as the basis of ATM, the volatility of near-term options and next-term options are calculated with OTM option prices of each delivery month. Next, the index value is calculated using linear interpolation or linear extrapolation between the volatilities of each delivery month to set the time to expiration as 30 days.

### **Base Date:**

The commencement date of the calculation was November 19, 2010, which had been retroactively calculated in the past on the end-of-day basis, to June 12th, 1989. The index is currently calculated every 15 seconds during the day session of the Nikkei 225 options on the OSE.

### Relative Strength Index (Volatility)

$$RSI(V)_t = \frac{U_t}{U_t + D_t} \quad (75)$$

Where  $U_t$  denotes the cumulative advance i.e. the close-to-close increase on a day where the security under observation has closed higher than the previous day's closing price over a certain period, and  $D_t$  denotes the cumulative decline i.e. the close-to-close decrease on a day where the security under observation has closed lower than the previous day's closing price over the sample period  $x$  which by default is 14 trading days as prescribed by Wilder.

$$U_t = \sum_{i=1}^m l(S_{t-1} - S_{t-1-i} > 0)(S_{t-1} - S_{t-1-i}) \quad (76)$$

$$D_t = \sum_{i=1}^m l(S_{t-1} - S_{t-1-i} < 0)|S_{t-1} - S_{t-1-i}| \quad (77)$$

Where  $S$  = Standard Deviation of the Close

Decreasing the value of  $x$  results in increased sensitivity whereas increasing the value of  $x$  decreases sensitivity. The RSI is then normalised and presented on a vertical scale bound between 0 and 100 with values above 70 signalling overbought conditions and values below 30 signalling oversold conditions. A mid-range value of 50 signals equilibrium and acts as a boundary between bullish and bearish market conditions, the crossing of which generates a respective minor buy or sell signal.

### Garman Klass Volatility

The Garman Klass Volatility estimator incorporates intraday information stored at daily frequency and is calculated as follows. Initially a scaled vector is determined equal to  $N$ , the number of trading days in a year, and  $n$  the chosen sample size. If  $n = N$  no scaling is required. Thereafter, the scaling factor is multiplied with the variance of the sample. The results obtained is the annualized Garman Klass variance. Lastly, taking the square root of this result produces the Garman Klass volatility estimator:

$$\sigma_t = \sqrt{\frac{N}{n}} \cdot \sum_{i=1}^n \frac{1}{2} \cdot \left( \log\left(\frac{H_i}{L_i}\right) \right)^2 - (2 \cdot \log(2) - 1) \cdot \log\left(\frac{C_i}{O_i}\right) \quad (78)$$

Where:

$N$  = Number of trading days in 1 year

$n$  = Sample Period

$H$  = High

$L$  = Low

$O$  = Open

$C$  = Close

### **CV(L9)ROC12**

Chaikin Volatility (CV), designed by Marc Chaikin, is a technical indicator employed to measure volatility. CV compares the spread between an asset's high and low prices and quantifies volatility as a widening of the range between the high and the low price. CV is calculated by initially calculating an EMA of the difference between the periods' high and low prices:

$$H - L \text{ Average} = EMA(H - L) \quad (79)$$

Where:

$H$  = High

$L$  = Low

$EMA$  = Exponential Moving Average

Lastly, the percent that the EMA has changed over a specified period is calculated as:

$$\left( \frac{(H - L \text{ Average}) - (H - L \text{ Average } n - \text{periods ago})}{(H - L \text{ Average } n - \text{periods ago})} \right) \times 100 \quad (80)$$

## **Chapter 6: Input Variables**

For Chapter 6: Macroeconomic Forecasting with Artificial Neural Networks: U.S. Unemployment Forecasting, all variables of the candidate variable pool were sourced from the (FRED) Economic Data section of the Federal Bank of St Louis. Following 2-stage IVS, seven variables were selected:

### **Variable 1**

Title: Civilian Unemployment Rate

Series ID: UNRATE

Unit Measurement: Percent (%)

Frequency: Monthly

Adjusted: Seasonally Adjusted

### **Variable 2**

Title: Unemployment Rate: Ages 15-64, All Persons for the U.S.

Series ID: LRUN64TTUSM156S

Unit Measurement: Percent (%)

Frequency: Monthly

Adjusted: Seasonally Adjusted

### **Variable 3**

Title: Unemployment Level: Men

Series ID: LNS13000001

Unit Measurement: Thousands of Persons

Frequency: Monthly

Adjusted: Seasonally Adjusted

### **Variable 4**

Title: Unemployment Level: Looking for Full-Time Work

Series ID: LNS13100000

Unit Measurement: Thousands of Persons

Frequency: Monthly

Adjusted: Seasonally Adjusted



**Variable 5**

Title: Unemployment Level: White

Series ID: LNS13000003

Unit Measurement: Thousands of Persons

Frequency: Monthly

Adjusted: Seasonally Adjusted

**Variable 6**

Title: Unemployment Level: 20 Years +

Series ID: LNS13000024

Unit Measurement: Thousands of Persons

Frequency: Monthly

Adjusted: Seasonally Adjusted

**Variable 7**

Title: Unemployed Population: Ages 15-64, All Persons for the U.S.

Series ID: LFUN64TTUSM647S

Unit Measurement: Persons

Frequency: Monthly

Adjusted: Seasonally Adjusted

**Chapter 7: Input Variables**

For Chapter 7: Macroeconomic Forecasting with Artificial Neural Networks: U.S. GDP Forecasting, all variables of the candidate variable pool were sourced from the (FRED) Economic Data section of the Federal Bank of St Louis. Following 2-stage IVS, two variables were selected:

**Variable 1**

Title: GDP

Series ID: A191RP1Q027SBEA

Unit Measurement: % Change from Preceding Period

Frequency: Monthly

Adjusted: Seasonally Adjusted

## Variable 2

Title: GDP Per Capita

Series ID: A939RC0Q052SBEA

Unit Measurement: USD (\$)

Frequency: Quarterly

Adjusted: Seasonally Adjusted

## Appendix 6: Activation Functions

ANNs are constructed by interconnecting PEs which mimic the biological nerve cell, or neuron. NeuroSolutions divides the functionality of a neuron into two disjoint operations: a nonlinear instantaneous map, which mimics the neuron's threshold characteristics; and a linear map applied across an arbitrary discrete time delay, which mimics the neuron's synaptic interconnections. The Axon family implements common variations on the nonlinear instantaneous maps employed by neural models. Each axon represents a layer, or vector, of PEs. All axons will also be equipped with a summing junction at their input and a splitting node at their output. This allows multiple components to feed an axon, which then processes their accumulated activity. It is important to notice the difference between this sum of activity vectors, and the weighted sum of products depicted by the McCulloch-Pitts neuron model. The latter is implemented as a linear map by the other functional division of the neuron, the Synapse family.

For generality, an axon's map may be either linear or nonlinear. However, components in the Axon family typically apply a nonlinear instantaneous map, as given by the equation:

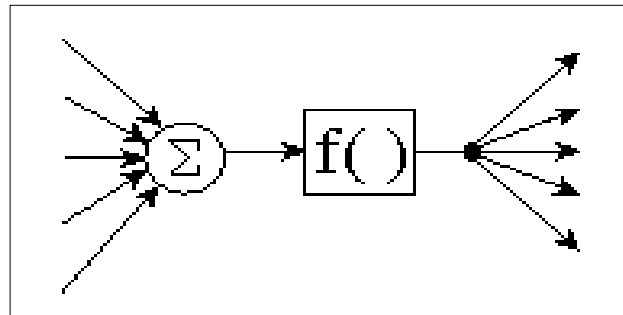
$$y_i(t) = f(x_i(t), w_i) \quad (81)$$

where  $y_i(t)$  is the axon's output,  $x_i(t)$  is an accumulation of input activity from other components,  $w_i$  is an internal weight or coefficient and  $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  represents an arbitrary functional map. We call  $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  the activation function.

All members of the Axon family accumulate input from, and provide output to, an arbitrary number of activation components. Therefore, each axon has a summing junction at its

input and a splitting node at its output. This functionality is illustrated by the following diagram:

Diagram A-1: Mapping for the PE of the Axon Class



Axons can receive input from and provide output to both axons and synapse within the network.

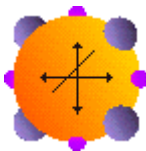
### Axon



The Axon performs an identity map between its input and output activity. The Axon is the first member of the Axon family, and all subsequent axon members subclass its functionality. Furthermore, each subclass will use the above icon with a graph of their activation function superimposed on top.

$$f(x_i, w_i) = x_i \quad (82)$$

### Bias Axon



The Bias Axon provides a bias term. Most nonlinear axons are subclasses of this component in order to inherit this bias characteristic. The Weights access point of the Bias Axon provides access to the Bias vector ( $\omega_i$ )

$$f(x_i, w_i) = x_i + w_i \quad (83)$$

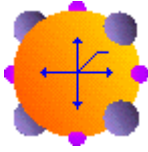
### Linear Axon



The Linear Axon implements a linear axon with slope and offset control. It is therefore more powerful than the Bias Axon because it implements an affine transform. The bias is inherited from the Bias Axon and the slope is controlled by an additional parameter  $\beta$ , which is not adaptive.

$$f(x_i, w_i) = \beta x_i + w_i \quad (84)$$

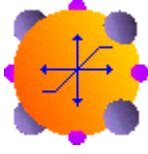
### Linear Sigmoid Axon



The Linear Sigmoid Axon substitutes the intermediate portion of the sigmoid by a line of slope  $\beta$ , making it a piecewise linear approximation of the sigmoid. This component is more computationally efficient than the Sigmoid Axon (it is much easier to compute the map).

$$f(x_i, w_i) = \begin{cases} 0 & x_i^{lin} < 0 \\ 1 & x_i^{lin} > 0 \\ x_i^{lin} & else \end{cases} \quad (85)$$

## Linear Tanh Axon

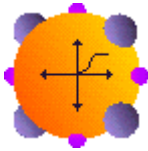


The Linear Tanh Axon substitutes the intermediate portion of the tanh by a line of slope  $\beta$ , making it a piecewise linear approximation of the tanh. This component is more computationally efficient than the Tanh Axon (it is much easier to compute the map). The Weights access point of the Linear Tanh Axon provides access to the Bias vector ( $\omega_i$ ).

$$f(x_i, w_i) = \begin{cases} -1 & x_i^{lin} < -1 \\ 1 & x_i^{lin} > 1 \\ x_i^{lin} & else \end{cases} \quad (86)$$

where  $x_i^{lin} = \beta x_i$  is the scaled and offset activity inherited from the Linear Axon.

## Sigmoid Axon



The Sigmoid Axon applies a scaled and biased sigmoid function to each neuron in the layer. The scaling factor and bias are inherited from the Linear Axon. The range of values for each neuron in the layer is between 0 and 1. Such nonlinear elements provide a network with the ability to make soft decisions. The Weights access point of the Sigmoid Axon provides access to the Bias vector ( $\omega_i$ ).

$$f(x_i, w_i) = \frac{1}{1 + \exp(-x_i^{lin})} \quad (87)$$

where  $x_i^{lin} = \beta x_i$  is the scaled and offset activity inherited from the Linear Axon.

## Soft Max Axon



The Soft Max Axon is a component used to interpret the output of the neural net as a probability. For a set of numbers to constitute a probability density function, their sum must equal one. Often the output of an ANN produces a similarity measure. In order to convert this similarity measure to a probability, the Soft Max Axon is used at the output of the network. The Weights access point of the Soft Max Axon provides access to the Bias vector ( $\omega_i$ ).

$$f(x_i, w_i) = \frac{\exp(x_i^{lin})}{\sum_j \exp(x_j^{lin})} \quad (88)$$

where  $x_i^{lin} = \beta x_i$  is the scaled and offset activity inherited from the Linear Axon.

## Tanh Axon



The Tanh Axon applies a bias and tanh function to each neuron in the layer. This will squash the range of each neuron in the layer to between -1 and 1. Such nonlinear elements provide a network with the ability to make soft decisions. The Weights access point of the Tanh Axon provides access to the Bias vector ( $\omega_i$ ).

$$f(x_i, w_i) = \tanh(x_i^{lin}) \quad (89)$$

where  $x_i^{lin} = \beta x_i$  is the scaled and offset activity inherited from the Linear Axon.

## Appendix 7: Nikkie 225 Index Descriptive Statistics (Closing Price)

### Mann-Kendall Trend Test

Table 9: Mann-Kendall Test

Mann-Kendall Trend Test	
Kendall's tau	0.493
S	235917.000
Var(S)	104416550.333
p-value (Two-tailed)	< 0.0001
Alpha, $\alpha$	0.01

The MK test is summarised as follows;  $H_0$  there is no trend in the time series,  $H_a$  there is a trend in the time series. From the results, as the computed p-value is lower than the significance level ( $\alpha=0.01$ , 99%), the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted therefore there is a trend in the Nikkei 225 Index time series.

### Normality

Table 10: Jarque-Bera Test Results

Jarque-Bera Test	
JB (Observed value)	27.601
JB (Critical value)	9.210
DF	2
p-value (Two-tailed)	< 0.0001
Alpha, $\alpha$	0.01

As the computed p-value of <0.0001 is lower than the significance level of 0.01, the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted. Therefore, the Nikkei225 Index time series data does not follow a normal distribution.

## Stationarity

Table 11: ADF Test Results

Augmented Dickey-Fuller (ADF) Test	
Tau (Observed value)	-1.706
Tau (Critical value)	-3.921
p-value (one-tailed)	0.748
alpha	0.01

The significance level was set to 0.01 (99%) and the p-value registered as 0.748.

Therefore, as the computed p-value is greater than the alpha, the null hypothesis  $H_0$  is accepted, and the alternative hypothesis  $H_a$  is rejected. The Nikkei225 Index time series data is therefore classed as non-stationary.

## Fractal Dimension

The fractal dimension of the Nikkei225 Index time series data requires a minimum of 2 independent variables to model the time series data.

## Summary Statistics

Table 12: Nikkei 225 Index Summary Statistics

Nikkei225 Index Summary Statistics	
Number of observations	980
Number of missing values	0
Minimum	13910.160
Maximum	22939.180
Range	9029.020
Median	17872.870
Mean	17954.297
Variance (n)	4636929.993
Variance (n-1)	4641666.387
Standard deviation (n)	2153.353
Standard deviation (n-1)	2154.453



Skewness (Pearson)	0.136
Skewness (Fisher)	0.137
Skewness (Bowley)	0.026
Kurtosis (Pearson)	-0.776
Kurtosis (Fisher)	-0.773
Standard error of the mean	68.821
Geometric mean	17825.031

## Appendix 8: Nikkei 225 Index Descriptive Statistics (Absolute Log>Returns)

### Mann-Kendall Trend Test

Table 13: MK Test Results:

Mann-Kendall Test	
Kendall's tau	-0.014
S	-49726.000
Var(S)	104736683.333
p-value (Two-tailed)	<0.0001
Alpha, $\alpha$	0.01

The MK test is summarised as follows;  $H_0$  there is no monotonic trend in the time series,  $H_a$  there is a monotonic trend in the time series. From the results, as the computed p-value is lower than the significance level ( $\alpha=0.01$ ), the null hypothesis  $H_a$  is accepted and the alternative hypothesis  $H_0$  is rejected. Therefore, there is a monotonic trend in the Nikkei 225 Index absolute log-returns time series data.

## Normality

Table 14: Jarque-Bera Normality Test (Log>Returns);

Jarque-Bera test (Log Returns):	
JB (Observed value)	5750.177
JB (Critical value)	9.210
DF	2
p-value (Two-tailed)	< 0.0001
Alpha	0.01

As the computed p-value of <0.0001 is lower than the significance level of 0.01 (99%), the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted. Therefore the Nikkei 225 Index absolute log returns time series data does not follow a normal distribution. It is therefore evident that the Nikkei 225 Index time series data in closing price format and absolute log-returns format both do not follow a normal distribution.

## Stationarity

Table 15: ADF Test Results

Augmented Dickey-Fuller (ADF) Test	
Tau (Observed value)	-6.409
Tau (Critical value)	-3.921
p-value (one-tailed)	< 0.0001
Alpha	0.01

The significance level (alpha) was set to 0.01 (99%) and the p-value registered as <0.0001, therefore, the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted. Therefore, the Nikkei225 Index absolute log-returns time data series is classified as stationary.

## Fractal Dimension

The fractal dimension of the Nikkei 225 Index absolute log returns time series data requires a minimum of 2 independent variables to model the time series data. Interestingly, this was also the same result for the Nikkei 225 Index time series data in closing price format.

## Summary Statistics

Table 16: Nikkei 225 Index (Absolute Log>Returns format) Summary Statistics

Nikkei225 Index Absolute Log Returns Summary Statistics	
Number of observations	980
Number of missing values	0
Minimum	0.0000
Maximum	8.2530
Range	8.2530
Median	0.610
Mean	0.900
Variance (n)	0.911
Variance (n-1)	0.912
Standard deviation (n)	0.954
Standard deviation (n-1)	0.955
Skewness (Pearson)	2.596
Skewness (Fisher)	2.600
Skewness (Bowley)	0.280
Kurtosis (Pearson)	10.670
Kurtosis (Fisher)	10.731
Standard error of the mean	0.031

## Appendix 9: U.S. Civilian Unemployment (UNEMPLOY) Persons

### Mann-Kendall Trend Test

Table 17: MK Test Results

Man-Kendall Test	
Kendall's tau	0.2374
S	35965.0000
Var(S)	18637545.0000
p-value (Two-tailed)	< 0.0001
alpha	0.0100

The MK test is summarised as follows;  $H_0$  there is no trend in the time series,  $H_a$  there is a trend in the time series. From the results, as the computed p-value is lower than the significance level ( $\alpha=0.01$ , 99%), the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted therefore there is a trend in the UNEMPLOY Index time series.

### Normality

Table 18: Jarque-Bera Test Results

Jarque-Bera Test	
JB (Observed value)	184.5862
JB (Critical value)	9.2103
DF	2.0000
p-value (Two-tailed)	< 0.0001
alpha	0.0100

As the computed p-value of <0.0001 is lower than the significance level of 0.01, the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted. Therefore, the UNEMPLOY time series data does not follow a normal distribution.

### Stationarity

Table 19: ADT Test Results

Dickey-Fuller Test	
Tau (Observed value)	-3.1662
Tau (Critical value)	-3.9550
p-value (one-tailed)	0.0910
alpha	0.0100

## Fractal Dimension

In this case, the fractal dimension of the UNEMPLOY time series data requires a minimum of 2 independent variables to model the time series data.

## Summary Statistics

Table 20: Summary Statistics

U.S. Unemployment	
Number of observations	551.00
Number of missing values	0.00
Minimum	4144.00
Maximum	15352.00
Frequency of minimum	1.00
Frequency of maximum	1.00
Range	11208.00
1st Quartile	6728.00
Median	7771.00
3rd Quartile	8748.00
Mean	8170.70
Variance (n)	5326328.22
Variance (n-1)	5336012.45
Standard deviation (n)	2307.88
Standard deviation (n-1)	2309.98
Skewness (Pearson)	1.22
Skewness (Fisher)	1.23
Skewness (Bowley)	-0.03
Kurtosis (Pearson)	1.43
Kurtosis (Fisher)	1.46
Standard error of the mean	98.41
Lower bound on mean (1%)	8169.46
Upper bound on mean (1%)	8171.93
Lower bound on variance (1%)	5338448.96
Upper bound on variance (1%)	5346528.71
Standard error(Skewness (Fisher))	0.10
Standard error(Kurtosis (Fisher))	0.21
Mean absolute deviation	1662.44
Median absolute deviation	1020.00

## Appendix 10: U.S. Unemployment (UNEMPLOY) % $\Delta$

### Mann-Kendall Trend Test

Table 21: Mann-Kendall Test Results

Mann-Kendal Test	
Kendall's tau	-0.0374
S	-5645.0000
Var(S)	18536375.0000
p-value (Two-tailed)	0.1898
alpha	0.0100

### Normality

Table 22: Jarque-Bera Test Results

Jarque-Bera Test	
JB (Observed value)	103.6849
JB (Critical value)	9.2103
DF	2.0000
p-value (Two-tailed)	< 0.0001
alpha	0.0100

### Stationarity

Table 23: ADF Test Results

Dickey-Fuller	
Tau (Observed value)	-5.2146
Tau (Critical value)	-3.9200
p-value (one-tailed)	< 0.0001
alpha	0.0100

### Fractal Dimension

The fractal dimension of the UNEMPLOY time series data in % $\Delta$  format requires a minimum of 2 independent variables to model the time series data.

## Summary Statistics

Table 24: U.S. UNEMPLOY (%Δ)

Summary Statistics	
Number of observations	550.00
Number of missing values	0.00
Minimum	-7.49
Maximum	13.03
Frequency of minimum	1.00
Frequency of maximum	1.00
Range	20.52
1st Quartile	-1.74
Median	-0.06
3rd Quartile	1.58
Mean	0.09
Variance (n)	7.54
Variance (n-1)	7.56
Standard deviation (n)	2.75
Standard deviation (n-1)	2.75
Skewness (Pearson)	0.69
Skewness (Fisher)	0.69
Skewness (Bowley)	-0.01
Kurtosis (Pearson)	1.62
Kurtosis (Fisher)	1.65
Standard error of the mean	0.12
Lower bound on mean (1%)	0.09
Upper bound on mean (1%)	0.09
Lower bound on variance (1%)	7.56
Upper bound on variance (1%)	7.57
Standard error(Skewness (Fisher))	0.10
Standard error(Kurtosis (Fisher))	0.21
Mean absolute deviation	2.09
Median absolute deviation	1.66

## Appendix 11: U.S. GDP Time Series Data (£Bn Format)

### Mann-Kendall Trend Test

Table 25: Mann-Kendall Test Results

Mann-Kendall Trend Test	
Kendall's tau	0.9970
S	25800.0000
Var(S)	1325528.6667
p-value (Two-tailed)	< 0.0001
alpha	0.0100

The MK test is summarised as follows;  $H_0$  there is no trend in the time series,  $H_a$  there is a trend in the time series. From the results, as the computed p-value is lower than the significance level ( $\alpha=0.01$ , 99%), the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted therefore there is a monotonic trend in the GDP time series.

### Normality

Table 26: Jarque-Bera Test Results

Jarque-Bera Test	
JB (Observed value)	22.8756
JB (Critical value)	9.2103
DF	2.0000
p-value (Two-tailed)	< 0.0001
alpha	0.0100

As the computed p-value of <0.0001 is lower than the significance level of 0.01, the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted. Therefore, the GDP time series data does not follow a normal distribution.



## Stationarity

Table 27: ADF Test Results

Augmented Dickey-Fuller Test (Stationarity)	
Tau (Observed value)	-1.0052
Tau (Critical value)	-3.9388
p-value (one-tailed)	0.9286
alpha	0.0100

The significance level was set to 0.01 (99%) and the p-value registered as 0.685.

Therefore, as the computed p-value is greater than the alpha, the null hypothesis  $H_0$  is accepted, and the alternative hypothesis  $H_a$  is rejected. The GDP time series data is therefore classed as non-stationary.

## Fractal Dimension

The fractional dimension of the U.S. GDP time series data requires a minimum of 2 independent variables to model the time series data.

## Summary Statistics

For a summary of descriptive statistics regarding the US GDP time series data see table 28 below. For statistics requiring a confidence or significance level, a 99% level is set:

Table 28: U.S. GDP (\$Bn) Summary Statistics (1960-2016)

U.S. GDP (\$Bn) Summary Statistics	
Number of observations	228.00
Number of missing values	0.00
Minimum	540.20
Maximum	18979.25
Freq. of minimum	1.00
Freq. of maximum	1.00
Range	18439.05
1st Quartile	1520.34
Median	5236.44
3rd Quartile	11011.90

Mean	6769.94
Variance (n)	31702878.28
Variance (n-1)	31842538.54
Standard deviation (n)	5630.53
Standard deviation (n-1)	5642.92
Skewness (Pearson)	0.61
Skewness (Fisher)	0.62
Skewness (Bowley)	0.22
Kurtosis (Pearson)	-0.95
Kurtosis (Fisher)	-0.94
Standard error of the mean	373.71
Lower bound on mean (1%)	6765.25
Upper bound on mean (1%)	6774.63
Lower bound on variance (1%)	31898693.78
Upper bound on variance (1%)	31973928.72
Standard error(Skewness (Fisher))	0.16
Standard error(Kurtosis (Fisher))	0.32
Mean absolute deviation	4860.82
Median absolute deviation	4159.72

---

## Appendix 12: U.S. GDP Time Series Data (%Δ Format)

### Mann-Kendall Trend Test

Table 29: MK Test Results

Mann-Kendall Test	
Kendall's tau	-0.3178
S	-8224.0000
Var(S)	1325528.6667
p-value (Two-tailed)	<0.0001
Alpha, $\alpha$	0.0100

The MK test is summarised as follows;  $H_0$  there is no monotonic trend in the time series,  $H_a$  there is a monotonic trend in the time series. From the results, as the computed p-value is lower than the significance level ( $\alpha=0.01$ ), the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted. Therefore, there is a monotonic trend in the US GDP  $\Delta\%$  time series data.

### Normality

Table 30: Jarque-Bera Normality Test (%Δ)

Jarque-Bera test (Log Returns):	
JB (Observed value)	55.6428
JB (Critical value)	9.2103
DF	2.0000
p-value (Two-tailed)	< 0.0001
Alpha	0.0100

As the computed p-value of <0.0001 is lower than the significance level of 0.01 (99%), the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted. Therefore, the US GDP  $\Delta\%$  time series data does not follow a normal distribution.

## Stationarity

Table 31: ADF Test

Augmented Dickey-Fuller (ADF) Test	
Tau (Observed value)	-4.6442
Tau (Critical value)	-3.9388
p-value (one-tailed)	0.0008
Alpha	0.0100

The significance level (alpha) was set to 0.01 (99%) and the p-value registered as 0.0008, therefore, the null hypothesis  $H_0$  is rejected, and the alternative hypothesis  $H_a$  is accepted. Therefore, the US GDP % $\Delta$  time data series is classified as stationary.

## Fractal Dimension

The fractional dimension of the US GDP % $\Delta$  time series data requires a minimum of 2 independent variables to model the time series data. Interestingly, this was also the same result for the US GDP time series data in \$Bn format.

## Summary Statistics

For a summary of descriptive statistics regarding the US GDP time series data in % $\Delta$  format see table 32 below. For statistics requiring a confidence or significance level, a 99% level was set:

Table 32: U.S. GDP (% $\Delta$ ) Summary Statistics, (1960 – 2016).

U.S. GDP (% $\Delta$ ) Summary Statistics	
Number of observations	228.00
Number of missing values	0.00
Minimum	-1.86
Maximum	5.85
Freq. of minimum	1.00
Freq. of maximum	1.00
Range	7.71
1st Quartile	1.05
Median	1.48

3rd Quartile	2.08
Mean	1.59
Variance (n)	0.94
Variance (n-1)	0.94
Standard deviation (n)	0.97
Standard deviation (n-1)	0.97
Skewness (Pearson)	0.46
Skewness (Fisher)	0.47
Skewness (Bowley)	0.17
Kurtosis (Pearson)	2.24
Kurtosis (Fisher)	2.31
Standard error of the mean	0.06
Lower bound on mean (1%)	1.59
Upper bound on mean (1%)	1.59
Lower bound on variance (1%)	0.94
Upper bound on variance (1%)	0.95
Standard error(Skewness (Fisher))	0.16
Standard error(Kurtosis (Fisher))	0.32
Mean absolute deviation	0.72
Median absolute deviation	0.48

---

## References

- Abu-Mostafa, Y. and Atiya, A. (1996). Introduction to financial forecasting. *Applied Intelligence*, 6(3), pp.205-213.
- Ahn, H. and Kim, K. (2009). Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach. *Applied Soft Computing*, 9(2), pp.599-607.
- Allen, H. and Taylor, M. (1990). Charts, Noise and Fundamentals in the London Foreign Exchange Market. *The Economic Journal*, 100(400), p.49.
- AMJADY, N. and KEYNIA, F. (2009). Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. *Energy*, 34(1), pp.46-57.
- Andersen, T., Bollerslev, T., Christoffersen, P. and Diebold, F. (2005). Volatility Forecasting. *PIER Working Paper*, pp.5-11.
- Aragonés, J., Blanco, C. and Estévez, P. (2007). Neural network volatility forecasts Intell. Syst. Account. Finance. Manage, (15), pp.107-121.
- Aronson, D. (2007). *Evidence-Based Technical Analysis: Applying the Scientific Method and Statistical Inference to Trading Signals*. 1st ed. New Jersey: John Wiley & Sons, pp.461-465.
- Asness, C., Frazzini, A. and Pedersen, L. (2013). Low-Risk Investing Without Industry Bets. *SSRN Electronic Journal*.
- Bajgrowicz, P. and Scaillet, O. (2011). Technical Trading Revisited: False Discoveries, Persistence Tests, and Transaction Costs. *SSRN Electronic Journal*.
- Barber, B. and Odean, T. (2001). Boys will be Boys: Gender, Overconfidence, and Common Stock Investment. *The Quarterly Journal of Economics*, 116(1), pp.261-292.
- Barber, B., Lehavy, R., McNichols, M. and Trueman, B. (2005). Buys, Holds, and Sells: The Distribution of Investment Banks' Stock Ratings and the Implications for the Profitability of Analysts' Recommendations. *SSRN Electronic Journal*.

Bates, J. and Granger, C. (1969). The Combination of Forecasts. *OR*, 20(4), p.451.

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.

Brown, C. (2011). *Technical Analysis for the Trading Professional, Second Edition*. Blacklick: McGraw-Hill Publishing.

Campbell, J., Lo, A. and MacKinlay, A. (1997). *Econometrics of Financial Markets*. New Jersey: Princeton University Press.

Castellano, G., Fanelli, A. and Pelillo, M. (1997). An iterative pruning algorithm for feedforward neural networks. *IEEE Transactions on Neural Networks*, 8(3), pp.519-531.

Cesari, R. and Cremonini, D. (2003). Benchmarking, portfolio insurance and technical analysis: a Monte Carlo comparison of dynamic strategies of asset allocation. *Journal of Economic Dynamics and Control*, 27(6), pp.987-1011.

Charles D. Kirkpatrick II and Julie R. Dahlquist. (2016). *Technical Analysis: The Complete Resource for Financial Market Technicians*. Pearson.

Charles, A., Darné, O. and Kim, J. (2012). Exchange-rate return predictability and the adaptive markets hypothesis: Evidence from major foreign exchange rates. *Journal of International Money and Finance*, 31(6), pp.1607-1626.

Cheung, Y. and Chinn, M. (2001). Currency traders and exchange rate dynamics: a survey of the US market. *Journal of International Money and Finance*, 20(4), pp.439-471.

Chiarella, C., He, X. and Hommes, C. (2006). A dynamic analysis of moving average rules. *Journal of Economic Dynamics and Control*, 30(9-10), pp.1729-1753.

Chong, T. and Ng, W. (2008). Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30. *Applied Economics Letters*, 15(14), pp.1111-1114.

- Chong, T., Ng, W. and Liew, V. (2014). Revisiting the Performance of MACD and RSI Oscillators. *Journal of Risk and Financial Management*, 7(1), pp.1-12.
- Chua, C., Lim, G. and Tsiaplias, S. (2011). A latent variable approach to forecasting the unemployment rate. *Journal of Forecasting*, 31(3), pp.229-244.
- Cook, T. and Smalter Hall, A. (2017). Macroeconomic Indicator Forecasting with Deep Neural Networks. *The Federal Reserve Bank of Kansas City Research Working Papers*.
- COVAL, J. and SHUMWAY, T. (2005). Do Behavioral Biases Affect Prices?. *The Journal of Finance*, 60(1), pp.1-34.
- Cowles, A. (1933). Can Stock Market Forecasters Forecast?. *Econometrica*, 1(3), p.309.
- Crone, S., Hibon, M. and Nikolopoulos, K. (2011). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting*, 27(3), pp.635-660.
- Crone, S., Hibon, M. and Nikolopoulos, K. (2011). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting*, 27(3), pp.635-660.
- Daniel, K. and Titman, S. (1999). Market Efficiency in an Irrational World. *Financial Analysts Journal*, 55(6), pp.28-40.
- Dayhoff, J. and DeLeo, J. (2001). Artificial neural networks. *Cancer*, 91(S8), pp.1615-1635.
- De Gooijer, J. and Hyndman, R. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22(3), pp.443-473.
- De Gooijer, J. and Hyndman, R. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22(3), pp.443-473.
- Deutsch, M., Granger, C. and Teräsvirta, T. (1994). The combination of forecasts using changing weights. *International Journal of Forecasting*, 10(1), pp.47-57.



- Donaldson, R. and Kamstra, M. (1997). An artificial neural network-GARCH model for international stock return volatility. *Journal of Empirical Finance*, 4(1), pp.17-46.
- Dunis, C. (2014). *Computational intelligence techniques for trading and investment*. London: Routledge.
- Dunis, C., Laws, J. and Sermpinis, G. (2011). Higher order and recurrent neural architectures for trading the EUR/USD exchange rate. *Quantitative Finance*, 11(4), pp.615-629.
- Dunis, C., Middleton, P., Karathanasopoulous, A. and Theofilatos, K. (2016). *Artificial Intelligence in Financial Markets*. London: Palgrave Macmillan UK.
- Edwards, R., Magee, J. and Bassetti, W. (n.d.). *Technical analysis of stock trends*.
- Ehlers, J. (2013). *Cybernetic analysis for stocks and futures*. Hoboken, N.J.: Wiley.
- Elman, J. (1990). Finding Structure in Time. *Cognitive Science*, 14(2), pp.179-211.
- Engel, C. and West, K. (2005). Exchange Rates and Fundamentals. *Journal of Political Economy*, 113(3), pp.485-517.
- Enke, D. and Monfared, S. (2014). Volatility Forecasting Using a Hybrid GJR-GARCH Neural Network Model. *Procedia Computer Science*, 36, pp.246-253.
- Enke, D. and Monfared, S. (2015). Noise Canceling in Volatility Forecasting Using an Adaptive Neural Network Filter. *Procedia Computer Science*, 61, pp.80-84.
- Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), p.383.
- Fama, E., Fisher, L., Jensen, M. and Roll, R. (2003). The Adjustment of Stock Prices to New Information. *SSRN Electronic Journal*.
- Farias Nazário, R., e Silva, J., Sobreiro, V. and Kimura, H. (2017). A literature review of technical analysis on stock markets. *The Quarterly Review of Economics and Finance*, 66, pp.115-126.

- Farmer, J. (2002). Market force, ecology and evolution. *Industrial and Corporate Change*, 11(5), pp.895-953.
- Farmer, J. and Lo, A. (1999). Frontiers of finance: Evolution and efficient markets. *Proceedings of the National Academy of Sciences*, 96(18), pp.9991-9992.
- Fernández-Rodríguez, F., González-Martel, C. and Sosvilla-Rivero, S. (2000). On the profitability of technical trading rules based on artificial neural networks:. *Economics Letters*, 69(1), pp.89-94.
- Figlewski, S. (2012). What is Risk Neutral Volatility?. *SSRN Electronic Journal*.
- Fong, W. and Yong, L. (2005). Chasing trends: recursive moving average trading rules and internet stocks. *Journal of Empirical Finance*, 12(1), pp.43-76.
- Gençay, R. (1998). Optimization of technical trading strategies and the profitability in security markets. *Economics Letters*, 59(2), pp.249-254.
- Gençay, R. (1999). Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules. *Journal of International Economics*, 47(1), pp.91-107.
- Giusto, A. and Piger, J. (2017). Identifying business cycle turning points in real time with vector quantization. *International Journal of Forecasting*, 33(1), pp.174-184.
- Giusto, A. and Piger, J. (2017). Identifying business cycle turning points in real time with vector quantization. *International Journal of Forecasting*, 33(1), pp.174-184.
- Goodhart, C. (1988). The Foreign Exchange Market: A Random Walk with a Dragging Anchor. *Economica*, 55(220), p.437.
- Graham, B. and Dodd, D. (2009). *Security analysis*. [New York]: McGraw-Hill.
- Grossman, S. (1976). On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information. *The Journal of Finance*, 31(2), p.573.
- Guidi, F. and Gupta, R. (2013). Market efficiency in the ASEAN region: evidence from multivariate and cointegration tests. *Applied Financial Economics*, 23(4), pp.265-274.

- Gunasekarage, A. and Power, D. (2001). The profitability of moving average trading rules in South Asian stock markets. *Emerging Markets Review*, 2(1), pp.17-33.
- Hamid, S. and Iqbal, Z. (2004). Using neural networks for forecasting volatility of S&P 500 Index futures prices. *Journal of Business Research*, 57(10), pp.1116-1125.
- Hastie, T., Friedman, J. and Tibshirani, R. (2017). *The elements of statistical learning*. New York: Springer.
- Haykin, S. and Hakin, S. (1994). *The Neural networks a comprehensive foundation*. Macmillan College Publishing Company.
- Higgins, R. (1992). *Analysis for financial management*.
- Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8), pp.2554-2558.
- Hornik, K., Stinchcombe, M. and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), pp.359-366.
- Hronec, M. (2015). Forecasting stock market returns and volatility in different time horizons using Neural Networks. *Charles University, Digital Depository*.
- Hsieh, T., Hsiao, H. and Yeh, W. (2011). Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm. *Applied Soft Computing*, 11(2), pp.2510-2525.
- <https://www.bcg.com>. (2019). *AI in the Factory of the Future: The Ghost in the Machine*. [online] Available at: <https://www.bcg.com/publications/2018/artificial-intelligence-factory-future.aspx> [Accessed 25 Feb. 2019].
- Huang, S., Wang, N., Li, T., Lee, Y., Chang, L. and Pan, T. (2013). Financial Forecasting by Modified Kalman Filters and Kernel Machines. *Journal of Statistics and Management Systems*, 16(2-03), pp.163-176.

- Huberman, G. and Regev, T. (2001). Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar. *The Journal of Finance*, 56(1), pp.387-396.
- Hudson, R., Dempsey, M. and Keasey, K. (1996). A note on the weak form efficiency of capital markets: The application of simple technical trading rules to UK stock prices - 1935 to 1994. *Journal of Banking & Finance*, 20(6), pp.1121-1132.
- Inoue, A. and Kilian, L. (2008). How Useful Is Bagging in Forecasting Economic Time Series? A Case Study of U.S. Consumer Price Inflation. *Journal of the American Statistical Association*, 103(482), pp.511-522.
- Inoue, A. and Kilian, L. (2008). How Useful Is Bagging in Forecasting Economic Time Series? A Case Study of U.S. Consumer Price Inflation. *Journal of the American Statistical Association*, 103(482), pp.511-522.
- Ito, M. and Sugiyama, S. (2009). Measuring the degree of time varying market inefficiency. *Economics Letters*, 103(1), pp.62-64.
- Johnes, G. (1999). Forecasting unemployment. *Applied Economics Letters*, 6(9), pp.605-607.
- Jolliffe, I. (2007). *Principal component analysis*. New York: Springer.
- Kaastra, I. and Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10(3), pp.215-236.
- Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), p.263.
- Kapetanios, G., Labhard, V. and Price, S. (2008). Forecast combination and the Bank of England's suite of statistical forecasting models. *Economic Modelling*, 25(4), pp.772-792.
- Kim, J., Shamsuddin, A. and Lim, K. (2011). Stock return predictability and the adaptive markets hypothesis: Evidence from century-long U.S. data. *Journal of Empirical Finance*, 18(5), pp.868-879.

- Kim, K. and Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Systems with Applications*, 19(2), pp.125-132.
- Lahmiri, S. (2017). Modeling and predicting historical volatility in exchange rate markets. *Physica A: Statistical Mechanics and its Applications*, 471, pp.387-395.
- LeBaron, B. (1999). Technical trading rule profitability and foreign exchange intervention. *Journal of International Economics*, 49(1), pp.125-143.
- Lehmann, R. and Wohlrabe, K. (2013). Forecasting GDP at the Regional Level with Many Predictors. *German Economic Review*, 16(2), pp.226-254.
- Lehmann, R. and Wohlrabe, K. (2013). Forecasting GDP at the Regional Level with Many Predictors. *German Economic Review*, 16(2), pp.226-254.
- LeRoy, S. and Porter, R. (1981). The Present-Value Relation: Tests Based on Implied Variance Bounds. *Econometrica*, 49(3), p.555.
- Levich, R. and Thomas, L. (1993). The significance of technical trading-rule profits in the foreign exchange market: a bootstrap approach. *Journal of International Money and Finance*, 12(5), pp.451-474.
- Levis, M. and Liodakis, M. (1999). The Profitability of Style Rotation Strategies in the United Kingdom. *The Journal of Portfolio Management*, 26(1), pp.73-86.
- Liang, F. (2005). Bayesian neural networks for nonlinear time series forecasting. *Statistics and Computing*, 15(1), pp.13-29.
- Lim, K. and Brooks, R. (2011). The Evolution of Stock Market Efficiency Over Time: A Survey of the Empirical Literature. *Journal of Economic Surveys*, 25(1), pp.69-108.
- Lim, K., Brooks, R. and Hinich, M. (2008). Nonlinear serial dependence and the weak-form efficiency of Asian emerging stock markets. *Journal of International Financial Markets, Institutions and Money*, 18(5), pp.527-544.

Lin, C., Wang, J., Chen, C., Chen, C. and Yen, C. (2009). Improving the generalization performance of RBF neural networks using a linear regression technique. *Expert Systems with Applications*, 36(10), pp.12049-12053.

Lo, A. and Hasanhodzic, J. (2013). *The evolution of technical analysis*. Hoboken, N.J.: Bloomberg Press.

Lui, Y. and Mole, D. (1998). The use of fundamental and technical analyses by foreign exchange dealers: Hong Kong evidence. *Journal of International Money and Finance*, 17(3), pp.535-545.

Malkiel, B. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, 17(1), pp.59-82.

Malkiel, B. (2007). *A random walk down Wall Street*.

Mantri, J., Gahan, P. and Nayak, B. (2010). Artificial neural networks an application to stock market volatility. *International Journal of Engineering Science and Technology*, (2), pp.1451-60.

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.

McNelis, N. (2005). *Neural networks in finance*. London: Elsevier Academic Press.

Menkhoff \*, L. and Schmidt, U. (2005). The use of trading strategies by fund managers: some first survey evidence. *Applied Economics*, 37(15), pp.1719-1730.

Menkhoff, L. (1997). Examining the use of technical currency analysis. *International Journal of Finance & Economics*, 2(4), pp.307-318.

Menkhoff, L. (1997). Examining the use of technical currency analysis. *International Journal of Finance & Economics*, 2(4), pp.307-318.

Milionis, A. and Papanagiotou, E. (2011). A test of significance of the predictive power of the moving average trading rule of technical analysis based on sensitivity analysis: application to the NYSE, the Athens Stock Exchange and the Vienna Stock Exchange.

Implications for weak-form market efficiency testing. *Applied Financial Economics*, 21(6), pp.421-436.

Moizer, P. and Arnold, J. (1984). Share Appraisal by Investment Analysts—Portfolio vs. Non-Portfolio Managers. *Accounting and Business Research*, 14(56), pp.341-348.

Moshiri, S. and Brown, L. (2004). Unemployment variation over the business cycles: a comparison of forecasting models. *Journal of Forecasting*, 23(7), pp.497-511.

Mundra, P. and Rajapakse, J. (2010). SVM-RFE With MRMR Filter for Gene Selection. *IEEE Transactions on NanoBioscience*, 9(1), pp.31-37.

Murphy, J. (2012). *Charting Made Easy*. Hoboken: Wiley.

Murphy, J. and Murphy, J. (n.d.). *Technical analysis of the financial markets*.

Newbold, P. and Granger, C. (1974). Experience with Forecasting Univariate Time Series and the Combination of Forecasts. *Journal of the Royal Statistical Society. Series A (General)*, 137(2), p.131.

Oberlechner, T. and Osler, C. (2008). Overconfidence in Currency Markets. *SSRN Electronic Journal*.

Olmedo, E. (2014). Forecasting Spanish Unemployment Using Near Neighbour and Neural Net Techniques. *Computational Economics*, 43(2), pp.183-197.

Park, C. and Irwin, S. (2007). WHAT DO WE KNOW ABOUT THE PROFITABILITY OF TECHNICAL ANALYSIS?. *Journal of Economic Surveys*, 21(4), pp.786-826.

Pelckmans, K., Suykens, J. and De Moor, B. (2005). Building sparse representations and structure determination on LS-SVM substrates. *Neurocomputing*, 64, pp.137-159.

Pesaran, M. and Timmermann, A. (1992). A Simple Nonparametric Test of Predictive Performance. *Journal of Business & Economic Statistics*, 10(4), pp.461-465.

Pesaran, M. and Timmermann, A. (1992). A Simple Nonparametric Test of Predictive Performance. *Journal of Business & Economic Statistics*, 10(4), pp.461-465.

- Pradeepkumar, D. and Ravi, V. (2017). Forecasting financial time series volatility using Particle Swarm Optimization trained Quantile Regression Neural Network. *Applied Soft Computing*, 58, pp.35-52.
- Prechelt, L. (2012). Early Stopping - but when?. *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science*, 7700, pp.35-67.
- Pring, M. (n.d.). *Technical analysis explained*.
- Psaradellis, I., Laws, J., Pantelous, A. and Sermpinis, G. (2018). Pairs Trading, Technical Analysis and Data Snooping: Mean Reversion vs Momentum. *SSRN Electronic Journal*.
- Qi, M. and Wu, Y. (2006). Technical Trading-Rule Profitability, Data Snooping, and Reality Check: Evidence from the Foreign Exchange Market. *Journal of Money, Credit, and Banking*, 38(8), pp.2135-2158.
- Sattler, K. and schallehn, E. (2001). A Data preperation framework based on multi-database language. *Proceeding of the international Symposium on Database Engineering & Application*, pp.219-228.
- Sermpinis, G., Stasinakis, C., Theofilatos, K. and Karathanasopoulos, A. (2014). Inflation and Unemployment Forecasting with Genetic Support Vector Regression. *Journal of Forecasting*, 33(6), pp.471-487.
- Sermpinis, G., Stasinakis, C., Theofilatos, K. and Karathanasopoulos, A. (2015). Modeling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms—Support vector regression forecast combinations. *European Journal of Operational Research*, 247(3), pp.831-846.
- 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.
- 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.

Shen, L. and Loh, H. (2004). Applying rough sets to market timing decisions. *Decision Support Systems*, 37(4), pp.583-597.

Sinclair, T., Stekler, H. and Kitzinger, L. (2010). Directional forecasts of GDP and inflation: a joint evaluation with an application to Federal Reserve predictions. *Applied Economics*, 42(18), pp.2289-2297.

Sinclair, T., Stekler, H. and Kitzinger, L. (2010). Directional forecasts of GDP and inflation: a joint evaluation with an application to Federal Reserve predictions. *Applied Economics*, 42(18), pp.2289-2297.

Stock, J. and Watson, M. (2003). Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature*, 41(3), pp.788-829.

Stock, J. and Watson, M. (2003). Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature*, 41(3), pp.788-829.

Stock, J. and Watson, M. (2012). Generalized Shrinkage Methods for Forecasting Using Many Predictors. *Journal of Business & Economic Statistics*, 30(4), pp.481-493.

Stock, J. and Watson, M. (2012). Generalized Shrinkage Methods for Forecasting Using Many Predictors. *Journal of Business & Economic Statistics*, 30(4), pp.481-493.

Swanson, N. and White, H. (1997). A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks. *Review of Economics and Statistics*, 79(4), pp.540-550.

Swanson, N. and Zeng, T. (2001). Choosing among competing econometric forecasts: Regression-based forecast combination using model selection. *Journal of Forecasting*, 20(6), pp.425-440.

Sweeney, R. (1986). Beating the Foreign Exchange Market. *The Journal of Finance*, 41(1), p.163.

Szafarz, A. (2012). Financial crises in efficient markets: How fundamentalists fuel volatility. *Journal of Banking & Finance*, 36(1), pp.105-111.

Tanaka-Yamawaki, M. and Tokuoka, S. (2007). Adaptive use of technical indicators for the prediction of intra-day stock prices. *Physica A: Statistical Mechanics and its Applications*, 383(1), pp.125-133.

Tenti, P. (1996). Forecasting foreign exchange rates using recurrent neural networks. *Applied Artificial Intelligence*, 10(6), pp.567-582.

Teräsvirta, T., van Dijk, D. and Medeiros, M. (2005). Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting*, 21(4), pp.755-774.

Teräsvirta, T., van Dijk, D. and Medeiros, M. (2005). Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting*, 21(4), pp.755-774.

Thaler, R. and DeBondt, W. (1982). A Balanced Perspective on Economic Psychology. *Contemporary Psychology*, 27(5), pp.394-395.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3), pp.273-282.

Tkacz, G. (2001). Neural network forecasting of Canadian GDP growth. *International Journal of Forecasting*, 17(1), pp.57-69.

Tkacz, G. (2001). Neural network forecasting of Canadian GDP growth. *International Journal of Forecasting*, 17(1), pp.57-69.

Urquhart, A. and Hudson, R. (2013). Efficient or adaptive markets? Evidence from major stock markets using very long run historic data. *International Review of Financial Analysis*, 28, pp.130-142.

Urquhart, A. and McGroarty, F. (2016). Are stock markets really efficient? Evidence of the adaptive market hypothesis. *International Review of Financial Analysis*, 47, pp.39-49.

- Vapnik, V. (2005). *The nature of statistical learning theory*. New York: Springer.
- Vasnev, A., Skirtun, M. and Pauwels, L. (2013). Forecasting Monetary Policy Decisions in Australia: A Forecast Combinations Approach. *Journal of Forecasting*, 32(2), pp.151-166.
- Walsh, D. and Tsou, G. (1998). Forecasting index volatility: sampling interval and non-trading effects. *Applied Financial Economics*, 8(5), pp.477-485.
- Wang, H., Li, G. and Jiang, G. (2007). Robust Regression Shrinkage and Consistent Variable Selection Through the LAD-Lasso. *Journal of Business & Economic Statistics*, 25(3), pp.347-355.
- Wang, T., Qin, Z., Jin, Z. and Zhang, S. (2010). Handling over-fitting in test cost-sensitive decision tree learning by feature selection, smoothing and pruning. *Journal of Systems and Software*, 83(7), pp.1137-1147.
- Wang, Z. (2010). Directed graphs, information structure and forecast combinations: an empirical examination of US unemployment rates. *Journal of Forecasting*, p.n/a-n/a.
- Weber, M. and Camerer, C. (1998). The disposition effect in securities trading: an experimental analysis. *Journal of Economic Behavior & Organization*, 33(2), pp.167-184.
- Yen, G. and Lee, C. (2008). Efficient Market Hypothesis (EMH): Past, Present and Future. *Review of Pacific Basin Financial Markets and Policies*, 11(02), pp.305-329.
- Yu, L., Wang, S. and Keung Lai, K. (2007). *Foreign-Exchnage-Rate Forecasting with Artificial Neural Networks*. 1st ed. New York: Springer Science, pp.44-56.
- Yung-Keun Kwon and Byung-Ro Moon (2007). A Hybrid Neurogenetic Approach for Stock Forecasting. *IEEE Transactions on Neural Networks*, 18(3), pp.851-864.
- Zhang, G. and Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2), pp.501-514.
- Zhang, G., Y. Hu, M., Eddy Patuwo, B. and C. Indro, D. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, 116(1), pp.16-32.

Zhou, J. and Lee, J. (2013). Adaptive market hypothesis: evidence from the REIT market. *Applied Financial Economics*, 23(21), pp.1649-1662.

Zhu, Y. and Zhou, G. (2009). Technical analysis: An asset allocation perspective on the use of moving averages☆. *Journal of Financial Economics*, 92(3), pp.519-544.

Zunino, L., Zanin, M., Tabak, B., Pérez, D. and Rosso, O. (2010). Complexity-entropy causality plane: A useful approach to quantify the stock market inefficiency. *Physica A: Statistical Mechanics and its Applications*, 389(9), pp.1891-1901.

Zunino, L., Zanin, M., Tabak, B., Pérez, D. and Rosso, O. (2009). Forbidden patterns, permutation entropy and stock market inefficiency. *Physica A: Statistical Mechanics and its Applications*, 388(14), pp.2854-2864.

Zurada, J., Marks, R. and Robinson, C. (1994). *Computational intelligence*. New York: Institute of Electrical and Electronics Engineers.