USING INTERNATIONAL DIVERSIFICATION TO ENHANCE PREDICTED EQUITY INDEX PERFORMANCE: A SOUTH AFRICAN PERSPECTIVE

Submitted in partial fulfillment of the requirements for the degree of

Master of Commerce in Financial Management

By

Jarryd Phillip

PHLJAR001

Supervisor: Associate Professor Francois Toerien

Faculty of Commerce

Plagiarism Declaration

I, Jarryd Asten Phillip, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 06/08/2019

Acknowledgments

To my research supervisor, Associate Professor Francois Toerien, for his valuable insights, honest and critical feedback, and ongoing support.

To my company of employment and line managers at Allan Gray (Pty) Ltd for the support in financial aid.

To my family for always motivating me.

And finally to my partner, Lipika, for her understanding, encouragement and support over the past few years.

Abstract

In the weak form, the Efficient Market Hypothesis (EMH) states that it is not possible to forecast the future price of an asset based on the information contained in the historical prices of that same asset. Under this assumption, the market behaves as a random walk and as a result, price forecasting is impossible. Furthermore, financial forecasting is a difficult task due to the intrinsic complexities of any financial system. The purpose of this study is to examine the potential of developing an international investment strategy using future index price predictions and offsetting predicted price declines by investing in negatively correlated international markets. Therefore, the first objective of this study was to examine the feasibility and accuracy of using a machine learning technique to model and predict the future price of stock market indices of South Africa (All Share Index) and a variety of other developed and developing international markets, which included South Africa, Brazil, Russia, India and China of the BRIC countries and Italy, France, Netherlands, Switzerland, Germany, Nigeria, Australia, Hong Kong, Saudi Arabia, Japan, the U.S., Turkey and the U.K., which were identified as South Africa's major trading partners. Secondly, an analysis of market correlation between each country's equity index and South Africa's ALSI was conducted to determine which of these international indices were positively and negatively correlated to the South African ALSI. This allowed an extrapolation of potential international diversification opportunities.

By using machine learning to predict future price trends of the South African All Share Index (ALSI) within a specified time period, the market correlation aspect of this study was able to suggest possible negatively correlated safe haven markets to invest in to offset predicted losses in an expected declining local market. The study's major limitations include a single method for regression analysis (GARCH(1, 1)) and a limited number of variables in the feature space when predicting future prices. Additional parameters could prove a more robust modelling technique. The data used was a series of past closing prices of each country's major index. The data was split into five periods, where each period was assigned an overarching theme based on the prevailing market conditions at the time. The ALSI data set was subjected to a unit root test and found to be non-stationary.

The analysis thereafter followed a two-step test, with the first being the determination of market correlation of the South African equity market with other markets, using a generalised autoregressive conditional heteroskedasticity (GARCH (1: 1)) approach given the non-stationary nature of the ALSI historic data. The results showed strong positive market correlations between South Africa and China, India, Nigeria, Russia and Saudi Arabia, and strong negative correlation between South Africa and Australia, Germany, the Netherlands, and the United Kingdom. Secondly, the specific area of machine learning employed in this study was support vector machines, as implemented using Python programming. The results compare the actual index price with those predicted by the model and showed that this technique has the ability to predict the future price of the Index within an acceptable accuracy. The accuracy measure used was the mean relative error which in most cases was calculated to be between 95 and 98 which is considered relatively high. However, the results of the investment approach described above was considered to be too inconsistent to consider this diversification strategy viable. From a South African perspective, this approach has not been documented previously.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction and Background

The field of stock return and stock price prediction is a continuously flourishing area of research that has attracted researchers for many years. It is centred on the idea that there is a predictive relationship between future prices and past publicly available information. Past public information can consist of a number of economic variables such as interest rates and exchange rates, industry specific data such as growth rates of industrial production, consumer prices and industry revenue, as well as company specific information such as financial statements and dividend yields. The general perception of efficient markets is a well-known theory which opposes the idea of stock price forecasting.

In 1965, Eugene Fama developed the concept of the efficient market hypothesis (EMH) which, in its strong form, states that all public and private information that could affect the current stock price value is accounted for by the market before the public can trade on it. This means that it would be impossible to forecast future prices, as the current price is already reflective of everything that is known of the stock. Furthermore, Fama (1965) suggested that in an efficient market, prices will immediately adjust to good and bad news which comes into the market in a random fashion. This line of reasoning is based on the concept of the 'random walk', which implies that the best prediction of the next price would simply be the current price of the stock. In practice, the efficient market is a highly debated theory, as researchers have produced considerable evidence that markets are not fully efficient, and forecasting future stock and index prices is often possible with results that are more accurate that random guessing.

Research relating to evolving market conditions is largely centred on two broad areas, which is testing stock market efficiency and forecasting future stock prices and returns. The concept called efficient market hypothesis (EMH) clearly encapsulates the implications of an efficient market on stock forecasting accuracy. Throughout the years, a number of different modelling techniques have been developed and used to attempt to predict future stock and index prices. Most modelling techniques fall under two broad areas of prediction, namely fundamental analysis and technical analysis.

Market efficiency is not necessarily constant but rather dependent on a number of variables, including the prevailing market condition (Seetharam, 2016). This means that stock prediction accuracy could vary in different market conditions. In this study, an analysis of the accuracy of equity index forecasting using a machine learning technique is examined under varying market conditions. The motivation to predict stock prices is simple - investors seek financial gain and stable returns. Generally, individual stocks within a single stock market fluctuate in sync and are highly correlated (Leland, 1999). Therefore, when a market is undergoing a decline, most securities within that market follow a similar trajectory, and the same holds true for an upward moving market. A rational equity investor forecasting an incoming decline in a stock market will attempt to offset predicted losses. However, correlation between equity securities within a single market will suggest that all stocks are expected to decline. The use of international diversification could provide a potential safe haven. Thus, during a recession and a declining price forecast, investors could consider negatively correlated international markets to exploit in order to diversify and stabilise overall profit. On the other hand, during a bull market and increasing price forecasts, investors need not diversify as positively correlated markets should provide similar returns with similar risk.

A key assumption here is that international market correlation is never constant, but rather adapts to the current market condition, indicating that international diversification opportunities change as well (Chen, 2018). Correlation is a statistical measure that determines how assets move in relation to each other. Simple regression analysis allows users to determine how closely assets move against each other. Correlation is often used in portfolio management to measure the amount of diversification among the assets contained in a portfolio. Modern portfolio theory (MPT) uses the capital asset pricing model (CAPM) to determine the correlation of all assets in a portfolio to determine the most efficient frontier (Markowitz, 1991). This theory can be extended to an international level to determine the market correlation between country indices and determine how to weight investment in each country to determine the most efficient frontier. In this study, an analysis was conducted to determine how closely other country equity markets move in relation to South Africa’s All Share Index (ALSI). In order to represent the market, the largest equity market index of each country represented in this study was used.

Artificial intelligence is progressively being used within the finance industry as a relatively new method of forecasting price movement, and therefore as a means of testing the efficient market hypothesis. Machine learning, which is considered a subset of artificial intelligence, is the scientific study of algorithms and statistical models that computer systems use to effectively perform specific tasks by relying on past patterns and inference instead of explicit instructions, thereby learning from past historic data to predict future outcomes. This study makes use of a technical analysis approach, based on the fundamentals of machine learning, which seeks to determine a trend in historic price data movements and learn from previous price fluctuations (Pyo et al, 2017) to predict future prices.

The predicted values are examined in different market conditions and based on the predicted future forecasted trend, the correlation aspect of this study provides an indication of potential international markets to exploit in order to optimise profits or minimise losses, providing a useful technique for global portfolio creation. A fair assumption would be that a rational investor will attempt to minimise losses by considering negatively correlated markets as an indication of potential markets to exploit when future local prediction is expected to decline. When prices are predicted to go up, investors can then limit or reduce international diversification as positively correlated markets should provide similar returns and negatively correlated markets should decline. This strategy can be employed in any period or sub-period of the predicted time interval where prices are predicted to decline. This strategy of combining machine learning and international market correlation is summarised in Table 1.1 below.

Table 1.1. Investment Strategy using historic market correlation and future price prediction

Local Market Predicted Price Direction	Investment Approach	Expected Outcome
Predicted decline in prices	Identify negatively correlated international markets and consider investments	Offset losses by investing in negatively correlated markets
Predicted Increase in Prices	Limit or reduce international diversification	Take advantage of local returns as positively correlated markets will provide similar returns

The focus behind a fundamental analysis centres on fiscal and monetary policies as well as economic indicators such as GDP, imports, exports and other important factors that fall within a business cycle. The result of a fundamental analysis produces a very effective way to forecast general economic conditions but not necessarily exact market prices. The technique is usually used with a discounted cash flow (DCF) or dividend discount model (DDM) framework, which attempts to predict the intrinsic value of a share based on key assumptions on the company in question, and various market and industry related factors (Eriksson et al, 2011).

A technical analysis approach considers that historic market activity discloses important new information and understanding of the emotional factors influencing the share price in an effort to predict future prices and trends (Griffioen, 2003). A technical approach is based on the assumption that the share price is a reflection of mass psychology (referred to as the crowd) in action and attempts to predict the future price based on the concept that the crown psychology will always shift between fear, pessimism and panic on one side, and greed, undue optimism, and confidence on the other.

There are many studies that examine the effectiveness of these techniques, some of which will be briefly examined in the chapters to follow. Thereafter, this study focuses on the forecasting ability of support vector machines and examining the prediction power of this method within the South African market, using the All Share Index (ALSI) as a proxy for the market. Furthermore, an investigation is conducted of international correlation between South Africa and various developed and emerging markets, which provide potential safe haven markets to exploit under conditions of varying market forecasts. Negatively correlated markets provide an indication of which markets to turn to when local markets are expected to decline and vice versa. In this study, the underlying assumption is that predictions can be made based on historic stock price data alone, meaning that they do not follow a random walk in which consecutive changes have no correlation to each other.

1.1. Problem Statement

The ultimate goal for stock market price prediction is financial gain, and being able to provide consistent and stable returns. The ability to identify an algorithm or mathematical model that can consistently predict the direction and magnitude of future asset price moves would result in significant wealth creation. Therefore, investors and academic researchers are continuously developing new models in an attempt to produce greater returns and profits than their competitors. From an international portfolio creation perspective, efficient diversification between markets allows investors to maximise profit and produce stable and consistent returns.

The problem is that investors are continuously looking to identify opportunities for favourable investment, despite the suggestion by the EMH that such opportunities may not exist. Machine learning, in conjunction with an analysis of international regression, is a tool that can be used to potentially identify such opportunities

1.2. Research Gap

Many artificial intelligence and machine learning techniques have been applied to the research area of forecasting stock prices. However, most research in this field has focused on developed markets and has neglected emerging markets such as the South African Johannesburg Securities Exchange (JSE). Furthermore, combining the results of machine learning and regression in order to identify a potential investment strategy has not been explored in the South African context.

The first component of this study (equity index prediction using machine learning) will therefore provide an indication of the predicted direction the South African Market is expected to take, and the second aspect (correlation) will consider the most appropriate markets for a South African investor to diversify with based on historic correlation information.

1.3. Research Aims and Objectives

The aim of this study is to identify a potential investment strategy by conducting two separate analyses. Firstly, the study aims to identify the degree of correlation between the South African ALSI and the equity indices of some other developing and developed countries. Secondly, the study then applies a machine learning technique, called support vector machines, to forecast the price of a chosen equity index for these markets, with specific focus on South Africa and the All Share Index (ALSI). The results of each test is then assessed for accuracy.

The final aim is that potential investors can use negatively correlated international markets to offset losses when the JSE is predicted to decline. It should be noted, that this study should be considered as a decision-making support tool when investing, rather than a safe guideline.

For this research, the following objectives have been identified, and the application of suitable research methodologies will attempt to provide insights into each one:

1. To determine which countries' equity markets were negatively or positively correlated to South Africa in each period being investigated.
2. To determine whether Support Vector Machines can be used to forecast the future stock market index prices within an acceptable accuracy.
3. To determine potential markets to exploit when increasing or declining South African market prices are expected, by investing in markets suitably correlated to South Africa with strong prediction power.

1.4. Structure of the Report

This report is divided and structured into chapters as indicated below.

Chapter 2 assesses the existing literature around each area of research. A number of assumptions and model choices were used in this study that are based on prior research. These papers and research articles are critically examined.

In Chapter 3 the data used in this paper is defined. The sample time period used spans a number of years from 2000 to 2018. The historic data of each country was split into five separate periods for analysis based on the overarching market conditions at the time. Furthermore, the countries used in the study are also discussed.

The techniques used in this paper (GARCH) assumes non-stationarity in the historic data of the All Share Index (ALSI). Chapter 3 then continues by assessing the ALSI data set (noted as the dependent variable) for unit roots and confirm whether the data is considered stationary or non-stationary.

Chapter 4 follows from the results obtained in Chapter 3 and makes use of the generalised autoregressive conditional heteroscedasticity (GARCH) model to examine the correlation between South Africa and each of the countries tested. The results obtained in this section is then referenced in chapter 5.

In Chapter 5, support vector machines are employed to assess the accuracy of future index price prediction and a future price trend is determined. Furthermore, following from the results obtained in Chapter 4, the potential markets that can be exploited from an international diversification perspective is analysed for each time period. It is important to note that this study is conducted from a South African market perspective and focuses on the Johannesburg Securities Exchange (JSE) as the basis for the interpretation of the results. Primarily, the price prediction is conducted on the All Share Index (ALSI) in order to determine a future trend. Based on this predicted trend, international correlation is considered to enhance return.

Finally, the results and implications of this study is discussed and concluded in Chapter 6. Limitations of this study is examined and future research opportunities to expand the subject is suggested.

Chapter 2: Literature Review

There are two bodies of literature that is reviewed in this section. Firstly, research around the efficient market hypothesis and its implication to stock price prediction is covered in Sections 2.1 and 2.2 and secondly, international market correlation research and methods of measurement are covered in Section 2.3. The review of the relevant literature starts by investigating the evidence around the efficient market hypothesis and its relationship to the South African market. As briefly discussed in Chapter 1, stock market efficiency is strongly related to stock price forecasting, as a result of which existing literature and methods of stock prediction is explored in the next section. Some of the current major methods of stock price prediction, including fundamental analysis and technical analysis, is examined, before looking at literature that explored the most efficient methods of machine learning in stock price prediction. Thereafter, the empirical evidence on the effectiveness of support vector machine prediction is addressed within the context of this study. Finally, the existing investigation of market correlation, how it can be tested using generalised autoregressive conditional heteroskedasticity, and what it means from a diversification perspective, is considered.

2.1. The Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH) has been extensively tested by researchers globally. The concept was first introduced by Fama (1965) in a paper where it was concluded that typically, competition will result in the effects of new investor information to be immediately reflected in the actual prices of listed stocks. When new investor information arises, the word is spread very quickly and is instantaneously reflected in the price of the stock with no delay. In this type of market, the implication is that technical and fundamental analysis of securities would not produce excess returns any more than a simple buy-and-hold strategy. The concept of EMH is described as a “random walk”, which suggests that future stock prices fluctuate in a random fashion and past prices have no effect on future prices. The comparison to a “random walk” is fitting as it asserts that, since new information is already reflected in the stock price, any future prices will be determined by new information that comes in randomly. In the absence of any prediction power in efficient markets, the most appropriate strategy for investors would be a buy-and-hold strategy. Fama (1965) identified three types of EMH:

Weak-Form Efficiency – This type of efficiency states that information in past prices is fully reflected in the current price, resulting in no prediction power using past price data. This renders the concept of technical analysis futile and suggests that predicting returns based on technical analysis is impossible.

Semi-Strong Form Efficiency – This type of efficiency states that any public information is fully incorporated in the current price of a security. Public information is considered to include past prices, as well as data reported in a company's financial statements, earnings announcements, dividend announcements, expectation on macroeconomic factors and all public information pertaining to company competitors.

Strong Form Efficiency – This type of efficiency states that all public and private information is already incorporated in the current price of a security. This suggests that no investor can consistently produce profitable returns, even if they had access to information that is not publicly available.

In contrast, a number of studies have emerged that oppose the view of EMH. For example, Giot (2005) found that significantly large volatility levels are a possible indication to consider a long position in the market. In periods of financial ambiguity, very high implied volatility is observed, and investors are inclined to desperately sell-off their securities in order to manage their losses in a declining market. This is a possible reasoning that was described by Giot opposing the view of EMH. This behaviour of investors is described as "herding", where investors act irrationally in the short term, resulting in a steady decline or increase of stock prices. A long position allows investors to take advantage of temporary cheaper prices, and vice versa.

In an efficient market, stock price movement is considered to follow a "random walk". This theory assumes that consecutive stock price changes are independent and past prices cannot be used to predict future prices. In an inefficient market however, it is assumed that stock prices do not follow a random walk. For example, the phenomenon of mean reversion in stock prices (Grater and Struweg, 2015) is observable in an inefficient market. It is possible to predict future stock prices using past prices in a market considered to be in mean reversion, as one would expect a period of high returns to be followed by a period of low returns, as it settles towards the mean and vice versa. This suggests that the use of technical analysis in order to outperform a buy-and-hold strategy is only possible in an inefficient market. In this study, the equity market efficiency will be tested by predicting stock prices of developed and developing markets, using the South African ALSI as the base case.

Evidence on the efficiency of the Johannesburg Stock Exchange

The research around the efficiency of the Johannesburg Stock Exchange (JSE) is very mixed. Grater and Struweg (2015) produced a recent empirical study which tests the market efficiency of the JSE and in specific whether stock prices reflect a “random walk”. The study covers the period between 1994 and 2014, and it was found that the returns for the period are stationary. This suggest that when an extraordinary or shock event occurs, the return did not deviate from its average in the future events. The study concludes that the JSE is inefficient in the weak form and that future stock price prediction is possible based on input data of historical price movement. This is consistent with mean reversion or non-randomness of stock prices.

Kruger (2011) supported the findings of Grater and Struweg (2015). Kruger (2011) found that returns were non-random as there is serial dependence in the JSE’s return generating process and therefore the possibility of return prediction on the JSE. The study covered the period between February 2000 and December 2009 for the JSE All Share Index (ALSI). However, it is clearly noted that the findings of serial dependence is sporadic in nature, meaning that periods of price prediction is inconsistent and intersected by white noise. The findings confirmed that investors that consider market timing in their investment approach could possibly outperform the market.

The occurrence of the random walk hypothesis on the JSE All Share Index returns were tested by Seetharam (2016) for the period between 1997 and 2014. In addition to the ALSI, Seetharam included 50 individual South African stocks in his investigation, using different periods and frequencies. The idea of varying the period and frequency was vital, as it was found to have a significant impact on the results. It was found that when using lower frequency data such as quarterly or semi-annual prices, the returns do not seem to follow a random walk. On the other hand, when higher frequency data, such as daily prices, was used, the random walk hypothesis was supported by the evidence.

Seetharam's (2016) investigation confirmed that market efficiency or inefficiency was associated with the time period and frequency of data used. On the other hand, Phiri (2015), considered a different approach and associated the results with the type of market index used, as well as the linearity of the testing procedure. By using six comprehensive market indices, which included the All Share Index (ALSI), the JSE Top 40 index, the Industrials Index, the Mining Index, the Gold Index and the Financial Index, a test for weak form efficiency was conducted. The study comprised the period between January 2000 and December 2014 and concluded that when nonlinearities were accounted for in the unit root test, it was found that the indices used were not weak form efficient. On the other hand, when a linear root test was conducted to test for market efficiency, it was found to be efficient in the weak form. When further observing the results for the non-linear tests, Phiri (2015), noted that the ALSI, JSE Top 40, industrials and financial indices strongly rejected the weak form efficiency of the efficient market hypothesis. Considering that the results found by Phiri suggest that the ALSI follows a weak form inefficiency, the implication is that it may be possible to use a machine learning algorithm to identify trends in past price data information as a basis for future stock price prediction.

Although Phiri (2015) found that the All Share Index returns exhibit market inefficiency of the weak-form, Noakes and Rajaratnam (2016) find conflicting results. Noakes and Rajaratnam (2016) tested the efficiency of small (Small Cap index), mid (Mid Cap index) and large cap (JSE Top 40) indices on the JSE, using a random number generator test. It was found that there is a positive relationship between a company's size (where market capitalisation was used as a proxy for size) and company market efficiency. Price movement of large cap stocks were found to be random at an individual level, yet small cap stocks exhibit price movement that are non-random. A market state factor was also included in the analysis, and it was found that the JSE appeared to be less efficient during periods of financial uncertainty or crisis when compared to a more stable period. It is important to note that overall market efficiency may be similar to stocks, which have similar attributes. Considering that the JSE is a concentrated market, however, it is not necessarily correct to make an assumption on the entire exchange.

From a South African perspective, the review of the literature confirms that market efficiency has various points of view. There are a number of factors to be considered and as each factor is incorporated into research, different results are observed. Grater and Struweg (2015) concluded that the JSE is inefficient in the weak form, yet other studies suggest an inconsistency between stocks, data frequency, time periods, company size etc. While the results are mixed, at least a part of the literature does support the idea of using technical analysis methods in JSE stock price prediction.

2.2. Stock Price Prediction

Huang et al. (2005) described financial markets as behaving in a highly non-linear and dynamic manner. Furthermore, stock prices can be described as non-stationary, and are influenced by numerous macro-economic factors such as political climate, exchange rates, and interest rates, as well as micro-economic factors such as previous price movements and market behaviour, to mention but a few. These factors can be captured into the fundamental analysis and technical analysis that are performed on stocks (Patel et al., 2015; Pyo et al., 2017). Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit.

There have been vast amounts of literature that examines the possibilities of accurately predicting the stock price movements of numerous indices on several different stock markets around the world. The prediction methods have varied from traditional statistical methods such as k-means clustering, k-nearest neighbour and linear discriminant analysis, to the fairly new field of machine learning algorithms, such as artificial neural networks (ANN) and support vector machines (SVM). However, only a small percentage of the available literature is on stock price prediction using machine learning methods in emerging markets (Cao et al., 2005).

Fundamental Analysis

Fundamental analysis is the process of estimating the fair or intrinsic value of a stock. This is done by fundamental investment analysts who employ a valuation model. In order to determine this estimate, analysts evaluate past, present and future information pertaining to earnings of the company in order to produce a fair value (Wafi et al., 2015). The estimated value is then compared to the current market price to determine whether it is profitable to invest now based on the future expectation.

One of the most important principle of modern finance is that any asset value equals the present value of all expected future cash flows discounted at the required rate of return. However, the value of a company is based on its ability to achieve these cash flows, which are seen to be uncertain. Considering the complexity of stock valuation under different circumstances, a number of valuation techniques have been developed (Reilly and Brown, 2002, and Bodie, Kan, Marcus, 2009).

Table 2.1. Stock Valuation Methods

Stock Valuation Methods	
Dividend Discount Model (DDM)	
Multiples	**Earnings Multiplier Model (P/E ratio)**
	Price-to-Book Value Ratio (P/BV ratio)
	Price-to-Cash Flow Ratio (P/CF ratio)
	Price-to-Sales Ratio (P/S ratio)
Dicounted Cash Flow Models (DCFM)	
Residual Income Valuation Model (RI)	

Wafi et al. (2015) reviewed the theoretical foundations of each model (Shown in Table 2.1.), and concluded that each model has significant strengths and weaknesses, but each one's lack of accuracy can be accounted for based on the assumption of market efficiency. Wafi concluded that the best model that can be used to predict stock value is the Residual Income Model (RIM), as it does not require financial efficiency for its application. This model is credible for both developed and emerging markets.

Technical Analysis

Technical analysis is a trading technique used to evaluate potential investment opportunities by reviewing and analysing statistical trends and patterns from historic trading activity, such as past price movements and volume data. Dissimilar to a fundamental approach which attempts to determine an intrinsic value, technical analysis focuses on patterns of price movements, trading signals, and various other analytical and charting tools to determine the strength or weakness in investment opportunities.

Technical analysis is also known as charting and has been part of equity investment practice for many years. However, the concept of technical analysis has not received the same level of academic analysis as more common practices, such as fundamental analysis. One of the key reasons for this is due to the highly subjective nature associated with this method. Charts, which include shapes, trends and patterns are often in the eye of the beholder. Lo et al. (2002) proposed a systematic and automatic approach to technical pattern recognition using nonparametric kernel regression and applied this technique to a wide range of U.S. stocks from 1962 to 1996 to assess the efficiency of technical analysis. By evaluating the unconditional empirical distribution of daily stock returns to the conditional distribution which were conditioned on specific technical indicators, such as head-and-shoulder or double bottom patterns, it was concluded that over the 31-year sample period, a number of technical indicators do offer significant information and may have some real-world value.

Brown et al. (2015) concluded that technical analysis or the use of past prices to infer private information does in fact have value in a model. A two-period model was used to establish that rational investors use historic prices in creating their demands, and to determine the sensitivity of the value obtained when using technical analysis in conditions experiencing changes in exogenous parameters. The study concluded that investors have rational expectations of the relationship between prices and signals.

Machine Learning

Copious amounts of literature exist that examines the possibility of accurately predicting price movements of numerous indices on several different stock markets around the world. Attempted prediction methods have varied from traditional statistical methods such as k-means clustering, k-nearest neighbour, and linear discriminant analysis, to the fairly new field of machine learning algorithms, such as artificial neural networks (ANN) and support vector machines (SVM). However, only a small percentage of the available literature is on stock price prediction using machine learning methods in emerging markets (Cao et al., 2005).

With the increasing awareness of big data, together with improved technology, a stronger interest in machine learning has developed. The integration of statistical theory with programming has opened a new realm of possibilities in terms of prediction and classification, with the vast amounts of data available.

Two machine learning techniques of interest are neural networks (NN) and support vector machines (SVM). The creation of neural networks can be credited to McCulloch and Pitts (1943), whereas SVMs were founded on the principals of Vapnik (1999). Pyo et al. (2017) suggested that the main task of neural networks is to determine the optimal weights of the input variables, while Khemchandani et al (2009) suggested that support vector machines classify future data points by assigning each point to one of two disjointed half spaces within a pattern space, or higher dimensional feature space. The goal behind each technique remains constant in that it attempts to create a binary classifier or decision function from the available sample data, in order to classify a future sample with the lowest probability of misclassification (Kara et al., 2011).

Support Vector Machines

Huang et al. (2005) note that there are significant benefits to using Support Vector Machines (SVM) over neural networks. This enhanced performance can be attributed to the ability of SVM to not over-fit parameters when modelling, while always reaching a global optimal solution, which is not always guaranteed using neural networks. Hsu et al. (2009) mention that SVM methods are less complex than the Artificial Neural Network (ANN) method.

Ou and Wang (2009) used a variety of methods on daily closing prices of the Hang Seng index on the Hong Kong stock market for a period of seven years (January 2000 to December 2006). Ten methods were used in predicting the stock movements, namely tree-based classification, quadratic discrimination, k-nearest neighbour classification, linear discrimination analysis, logit models, naive Bayes, Bayesian classification with Gaussian process, SVM and least squares SVM. The best performing models were found to be the SVM and least squares SVM, as there were no assumptions about a priori probabilities required. Furthermore, the global optimum could be found, which was not achieved by the other methods.

Similarly, Huang et al. (2005) used price data from the Nikkei 225 index to predict the weekly price movement. The data for a period of January 1990 to December 2002 was used in the comparison of SVM models with models of random walks, linear discriminant analysis, quadratic analysis, Elman backpropagation neural networks and a combined SVM with other classification methods. SVM has the highest forecasting accuracy among the individual forecasting methods. One reason that SVM performs better than the earlier classification methods is that SVM is designed to minimise the structural risk, whereas the previous techniques are usually based on minimisation of the empirical risk. In other words, SVM seeks to minimise an upper bound of the generalisation error rather than minimising the training error, so SVM is less vulnerable to the over-fitting problem.

2.3. Modern Portfolio Theory and International Market Correlation

The next section reviews the second body of literature relevant to this study which is largely focused of international market correlation, modern portfolio theory and available statistical methodologies to determine the correlation between markets.

International correlations for stocks and bonds fluctuate widely over time. A number of studies exist that suggest that volatility appears to be contagious across markets. It is found that correlation between international market tends to increase in periods where there is high market volatility. Low correlation between international markets allows for diversification and is therefore at the heart of global portfolio diversification. By diversifying investments across international markets with low correlation allows for investors to manage the amount of total fund risk, with the assumption of not sacrificing or reducing performance. Modern Portfolio Theory or MPT considers the amount of correlation between a pool of assets in order to create the optimal weightings between assets to create the most efficient frontier. This allows a portfolio to optimise the expected return against the overall portfolio risk by including assets that have a low correlation with each other thereby offsetting losses in one asset with the gains of other assets. It is important to note however that correlation can change over time but can only be measured based on historic data. This means that assets that previously had a high level of correlation can become less correlated and begin to move in independently of each other. The research suggests that one of the major shortcomings of Modern Portfolio Theory is the fact that it assumes that correlation between assets are stable and does not fluctuate.

Markowitz (1952, 1959) and Tobin (1958) pioneered the concepts of Modern Portfolio Theory and Separation Theory, respectively. The idea stems from the efficient market hypothesis, and reinforces the idea that a rational investor will allocate assets in a portfolio in order to maximise the utility, which means weighting across assets that maximise returns and minimise risks.

Markowitz's Modern Portfolio Theory is used to determine a specific investor's ideal portfolio through the optimal blend of return to risk by using a mean-variance optimisation method. Standard deviation and variance are introduced as the risk measures used to determine a portfolio's risk (Markowitz, 1952). It was suggested that a rational investor would prefer higher return to lower return, but importantly the most preferred outcome is stable returns over time (Markowitz, 1952). The efficient frontier is defined as the most mean-variance efficient portfolios (or the optimal portfolio), which means the highest expected return for a given level of risk, or the lowest expected risk for a given level of return. Tobin (1958) built on Markowitz's theory with the creation of Separation Theory, by accounting for an investor's risk and return preference through a series of indifference curves. These two theories support the mean-variance optimisation for portfolio choice.

It is important to be aware that the assumptions underpinning these theories generally do not hold true for real world examples. Markowitz Theory assumes that returns are fully explained by the mean and variance. However other factors, such as skewness and kurtosis, are also significant. This means that the mean-variance optimisation strategy assumes that asset returns are normally distributed, and therefore that variance captures upside and downside risk equally. Fishburn (1977) concluded that MPT suggests investors equally weight negative and positive returns, as the variance accepts favourable upside deviation over the average return, and unfavourable downside deviation from the average as equivalent. However, in the real world, asset returns are usually not normally distributed and a rational investor does not view upside and downside risk symmetrically, but rather, they are more concerned with downside risk in order to limit losses in a portfolio (Leland, 1999).

Studies on the International Integration of Equity Markets

Many studies have tested the correlation relationships between stock markets. The review of the literature that follows starts by giving a brief overview of some of the most significant findings on international developed and emerging markets.

Grubel (1968) undertook a significant study which investigated the relationship between international stock markets. Grubel's study marked the beginning of substantial attention and research around the topic of the effects of international portfolio diversification. In his study, the stock markets of eleven geographically and economically varying countries was examined for volatility linkages, and it was concluded that through diversification, the return for a given level of risk can be significantly increased. It is widely understood that the above-mentioned benefits of diversification is one of the key reasons the topic is one of the most extensively researched and discussed topics in financial literature.

In one of the first long term studies on international market correlation, Longin and Solnik (1995) examined seven developed country major stock markets for the period 1960 – 1990. In this paper it was concluded that stock market integration increased over the period, and during periods of excessive global volatility, correlations tends to increase. Longin and Solnik concluded that by the 1990s, global stock market were closely correlated. A large amount of research exists around market correlation between USA, the UK and Australia. However, an increase in other developed and developing market is now on the rise as globalisation becomes ever present and stock markets of all countries in some way react to one another.

Arshanapalli and Dorkas (1993) found that the cointegration of international stock markets following the equities crisis in 1987 increased significantly. The study included data for the period 1980 to 1990, and found that cointegration before the crisis was relatively lower when compared to the correlation post the crisis. This observation was in some way contradictory to other studies, and one of the suggested reasons was the test for cointegration used. In a similar study, Rangvid (2001) used the idea that the number of European integrated markets was an indication of overall global market integration and concluded that during the 1980s and 1990s a significant increase in integrated markets were observed. The above studies clearly indicate that market correlation can vary in different economic conditions.

When looking at emerging markets, Raj and Dhal (2008) examined the correlation of the Indian stock market with global markets. With the use of the Augmented Dickey-Fuller test and the Johansen cointegration test, it was concluded that the Indian stock market has seen a steady increase in correlation since 2003. In a similar study, Zhang (2009) examines the relationship of Asian markets with the US. Using a unit root test and impulse response function, Zhang suggested that the US has a significant effect on the majority of Asian markets, with China being the least affected by movements in the US market.

The trend of international market integration was noted by Graham at al. (2012). A key point to note is that extensive research exists that determine the extent of market correlation, and how closely markets move in relation to each other. This study follows a similar objective with a key focus on South Africa's relationship to some of its key trading partners, which include both developed and emerging markets. In addition, an analysis of the BRICS countries is included.

Using GARCH to test Market Correlation

In a standard regression model, it is assumed that the residuals within the model are homoskedastic in nature, or simply constant over time. Bollerslev (1986) concluded that volatility in equity market prices evolves over time. Further, Crouhy and Rockinger (1998) and Sakthivel (2012) have shown that volatility clustering is persistent in stock price returns and finally, Change and Su (2009) conducted a study which concluded that volatility exhibits asymmetric behavior, with a tendency to increase when responding to bad news, and decrease when responding to good news. These studies imply that stock market price data is heteroscedastic in nature, which seems plausible considering that depending on market conditions, some periods will be more riskier than others, causing the standard errors to be larger. By testing for unit roots in the data, we can conclude the existence of share price return trends, and therefore heteroskedasticity of the residuals. This characteristic of data requires a model more appropriate than a simple regression, such as the Auto-Regressive Conditional Heteroskedasticity (ARCH) model. Goyal (2000) suggested that ARCH models perform improved volatility forecasts of stock price data and is appropriate when characteristic variance is expected in the time series data, as the ARCH model does not invalidate the standard ordinary least square inference. Ignoring the ARCH effects could result in skew results and weak beta coefficients, and for this reason the Generalised ARCH (or GARCH) model is used when this type of behaviour is observed.

Nwogugu (2006) suggested that the GARCH model is possibly inaccurate when residuals follow a normal distribution over the period being examined, resulting in overvaluation of the volatility. Nwogugu (2006) goes on to conclude that the most significant shortcomings of the GARCH model pertain to defining the parameters which result in negative results for the standard deviation, and variance which is unrealistic. In another study by Chang and Su (2009), inefficient estimation and a poor test for non-linear adjustment of volatility was highlighted as some of the key shortcomings of the GARCH model. Chinzara and Aziakpono (2009), and Samouilhan (2006), considered a univariate and bivariate GARCH model in an attempt to account for some of these shortcomings, and increase the testing power.

When using a GARCH model, a key requirement is a specific distribution of the error terms. The Student t-distribution is widely known to be the most appropriate, as it captures the observed kurtosis and asymmetries that are commonly found in financial data, including the effect of volatility in light of bad news and good news (Bollerslev, 1986). The Student t-distribution has "fatter tails" as found by Karlsson (2002), which was shown to be more accurate when compared to the Gaussian or normal distribution, which displays excessive kurtosis. Karlsson (2002) suggested that the error distribution fits the residuals more accurately when using a GARCH model but assumes stationarity in data.

As discussed in the previous section, there has been an increase in studies that examine the relationship between international stock markets. This increase is especially noticeable since the 1987 global equity crash and the 1997 Asian market crisis (Chinzara and Aziakpono, 2009).

Chinzara and Aziakpono (2009) conducted a study which is relevant in the context of this one, as it also considers the South African angle. A GARCH and multivariate autoregressive model was employed in order to examine the return linkages and volatility transmission between South Africa and some large developed and emerging markets across the globe. The results obtained from Chinzara and Aziakpono (2009) show that significant co-movement in stock markets existed between the local South African market and Australia, China and the US during the time period between 1999 and 2007.

A number of studies consider the volatility linkages between markets such as China, India, the UK and some emerging Eastern Europe countries with each of their major trading partners (see, for example, Allen et al., 2011; Chang and Su, 2009; Jeyanthi and Annapakiam, 2010; Samouilhan, 2006; Fedorova and Saleem, 2009). Chang and Su (2009) followed a similar approach as described by Chinzara and Aziakpono (2009), but included a threshold error correcting model (TECM) and a bivariate GARCH model, in which they concluded that volatility was fundamentally asymmetric in their investigation of the correlation between the stock markets of Vietnam market and its major trading partners. The approach was considered appropriate after it was found that the volatility of the Vietnamese market with its major trading partners was asymmetrical in nature. Samouihlan (2006) concludes that the asymmetric nature of the GARCH models accounts for the leverage effects observed in stock price returns. Samouihlan (2006) made use of the exponential GARCH or EGARCH variant when evaluating the relationship between global stock markets and the Johannesburg Stock Exchange. EGARCH has the advantage of using conditional variance to model the asymmetric directions of good and bad news based on past data.

2.4. Conclusion on Literature Findings

The literature reviewed in section 2.1 shows that large amount of research exists that indicate that the South African market is inefficient in the weak form, suggesting that stock price forecasting could be feasible. Price forecasting is an attempt to predict the direction and size of a future event, with the objective to allow investors to make better decisions based on future expectations. The literature examined in section 2.2 investigates two major approaches to forecasting, which fall under the categories of explanatory (causal) and time series methods. As the name suggests, explanatory forecasting considers inputs and outputs, and assumes a cause and effect relationship between the two, such as with fundamental analysis. Considering a constant relationship between the inputs and outputs, an explanatory approach will change the output as the inputs are toggled. On the other hand, time series forecasting treats the system as a black box and attempts to discover the factors affecting the behaviour. This is similar to a technical analysis approach.

There are two reasons why it is useful to treat the system as a black box. The first is that the system is likely to be highly complex and not understood. Even if it was understood, it is extremely difficult to measure the relationship between the factors that affect and govern the system. Secondly, the main concern is usually to forecast what is expected to happen and not necessarily why it happens. The time series approach is an expanding area of research, with new approaches being developed today. Considering that machine learning attempts to forecast outcomes based on historic data, this relatively new area of research falls under the time series approach (or broadly under technical analysis). Hence, the focus of this work is time series forecasting using machine learning, and in specific support vector machines which, as the research suggests, is well established as the most suitable model for stock price prediction.

Lastly, the understandings gained through section 2.3 of this literature review have molded the study in terms of choosing the GARCH model as an appropriate model for heteroskedastic data, as well as determining that the Student-t distribution best models financial data when performing a correlation analysis. In order to compliment the analysis of stock price prediction, an investigation into market correlation between South Africa and other developed and developing markets using a GARCH model, which has been shown to be one of the most suitable correlation measuring techniques, will be conducted. The correlation aspect could identify markets that are less correlated to South Africa, thereby suggesting possible safe havens when local stock markets are expected to decline.

Chapter 3: Data and Sampling

3.1. Data Information

The price index data used for both aspects of this study was obtained from Bloomberg's data feed station and includes weekly closing index price series of the composite or all share index of each country's market. Any missing data points were replaced with an average of the previous and following dates closing price as done by Longin and Solnik (1995). Furthermore, Longin and Solnik (1995) found that outliers are detrimental to the modelling process and were subsequently replaced by values that equal $\mu \pm 2\sigma$ where μ is the mean of the data set and σ is the standard deviation of the data set. The view that daily index prices encapsulates noise trading and monthly prices fails to capture dynamic markets relationships, was the basis for a decision to make use of weekly data in the models. Furthermore, in order to encapsulate a market's overall performance, composite indices were considered a good proxy of performance.

A wide range of developed and emerging markets were considered so as to incorporate markets that have some relation to South Africa. In economics, The BRIC countries is an acronym that refers to the countries of Brazil, Russia, India and China. These countries are deemed to be at a similar stage of newly advanced economic development and is widely linked with the transfer of international economic power to these fast-growing markets from the developed countries of the G7 economies which include the United States of America, the United Kingdom, Japan, Italy, France, Germany and Canada. South Africa (SA) became a member nation of the group and the acronym was renamed to BRICS in 2010. BRICS is the third largest trading group after Europe and Asia, and it is estimated that the group will overtake the G7 economies by 2027 (Forooha, 2009). These countries are considered to be at various stages of economic development and for this reason, this study includes a specific analysis of the BRIC nations.

Other countries that were chosen in this analysis include Nigeria, Hong Kong, Japan, Italy, France, Netherlands, Switzerland, Germany, Australia, Saudi Arabia, the United Kingdom (U.K.), the United States of America (U.S.) and Turkey. Some of these countries were chosen based on South

African economic data on the largest export and import partners over the past decade indicating a potential correlation between markets (South Africa Online, 2018).

Table 3.1. Composite equity indices of each country

BRICS		Trading Partners	
Country	**Index**	**Country**	**Index**
South Africa	ALSI	Italy	FTSE MIB
Brazil	BOVESPA	France	SBF 120
Russia	MICEX	Netherlands	AEX
India	NIFTY	Switzerland	SMI
China	Shanghai Composite	Germany	CDAX
		Nigeria	NGSE Index
		Australia	ASX 200
		Hong Kong	Hang Seng
		Saudi Arabia	Tadawul
		Japan	Nikkei 225
		U.S.	S&P 500
		Turkey	XU100
		U.K.	FTSE 100

3.2. Time Period Split

The total time frame that was used spans the period from the 1st of January 2000 to the 31st of December 2019. The data was then split into periods and classified with an overarching theme that resulted in recessionary and recovery periods in order to capture specific trends. The data was split into five separate periods for the following reasons. Firstly, Chatfield (2000) noted that classical regression methods work well when the variation is subjugated by a regular linear trends and/or regular seasonality. However, he also noted that they tend to work poorly when the trend and/or seasonal effects are changing through time or when successive values of the irregular fluctuations are correlated. Secondly, Loretan (2000) suggested that international correlation can change over time but can only be measured on a historic basis, therefore in order to effectively capture the correlation between the equity markets in each period, splitting the periods into smaller sub periods were considered to be the best approach. Finally, from a machine learning perspective, using excessive data for training has been shown to be detrimental to model fitting (Roman, 2018). Excessive data for model training results in the model being unable to identify specific trends and produce meaningful predictions. The above observations resulted in the decision to split the data into five subsets, where each period was characterised by some global event that possibly contributed to the state of the economy in the period. Table 3.2 below, indicates the five periods that were examined in this study.

Table 3.2. Time Period Analysis

	Dates	Number of Years	Number of Observations	Period Classification
Period 1	1 Jan 2000 - 31 Dec 2001	2	129	Recession
Period 2	1 Jan 2002 - 31 Dec 2003	2	105	Recession
Period 3	1 Jan 2004 - 31 Dec 2007	4	211	Recovery
Period 4	1 Jan 2008 - 31 Dec 2009	2	108	Recession
Period 5	1 Jan 2017 - 31 Dec 2018	2	107	Current

- **Period 1:**

  This period encapsulates the burst of the dot-com bubble which started in the late 1990's as a consequence of an investment trend within the internet and technology related fields sector (Whitefoot, 2017). The bubble finally reached its maximum in March 2000 which subsequently resulted in markets eventually crashing due to Microsoft being acknowledged as a monopoly in the market (Whitefoot, 2017). In conjunction with this, increasingly high interest rates and the attacks against America on 11 September 2001 also contributed to the poor market conditions during this time.

- **Period 2:**

  During this period a number of large developed markets including the U.S., Canada, Asia and Europe experienced a sharp and significant decline in stock prices. Analysts who studied this period consider the decline to be related to mean-reversion following the burst of the dot-com bubble (Whitefoot, 2017). Although classified as a recession, the underlying cause was vastly different and a split was considered the most efficient approach in order to identify different price trends in the modelling process.

- **Period 3:**

  Following four years of poor market conditions which saw significant stock price declines, markets finally took a positive turn and began to make steady improvements, considered to be a recovery from the previous recession. Importantly, this period was the longest period investigated.

- **Period 4:**

  This period is themed by the infamous sub-prime crisis that began in December 2007 and flagged the beginning of the greatest global recession since 1929's Great Depression (Amadeo, 2018). The crisis was the after effect caused by the housing bubble in the United States of America which saw U.S. mortgage backed securities being marketed and sold at extraordinarily low credit rates. This was worsened by the fact that large and reputable credit rating agencies were posting false ratings on credit providers. The bubble finally burst when the Federal Reserve Bank of America increased interest rates against the market expectation of interest rate cuts. Considering the scale of the U.S. economy, the effect of the crisis had a global impact, with many global markets facing considerable losses and bankruptcies.

- **Period 5:**

  Since the sub-prime crisis, no major global event was identified that significantly resulted in a continuous upward or downward trend in market prices. As a result a more recent period was used. In the current period, recession risks remain modest, however the period has seen both upside and downside in almost a head and shoulders pattern. A number of events in the past 2 years could have aided to this, namely the Trump inauguration and Brexit. On the upside global growth continues to pick up.

3.3. Testing Data Stationarity of the JSE All Share Index

By definition, a stationary time series is one that does not depend on time. This means that the statistical properties such as the mean, variance and covariance of the data set being tested are constant over time, and are adjusted to replicate the real values which are not impacted by trends or seasonality that the data could contain. On the other hand, a non-stationary time series is characterised as one which reflects a trend in the data, and therefore its statistical properties reflect a changing mean, variance and covariance. A unit root is a descriptive statistic that confirms whether a stochastic time series is stationary or non-stationary. If a time series is considered to contain a unit root it implies non-stationarity in the time series data. Models used for volatility linkages are dependent on the characteristics of the data so the existence of a unit root, or whether data is stationary or not, will provide an indication of the optimal model to use from a regression perspective.

3.4. The Augmented Dick-Fuller Test

Dickey and Fuller (1981) established a variety of tests to test whether a given time series contains a unit root. The original Dickey-Fuller (DF) undertakes that the error/white noise terms in the model are uncorrelated. However these researchers later suggested that this assumption is unrealistic. The subsequent development of the Augmented Dickey-Fuller (ADF) test consequently considered the correlation between the error terms in the model. Dickey and Fuller (1981) suggested that the ADF is best suited for large data sets, thus the ADF model is considered appropriate in this study considering that the smallest data set being tested contains 108 observation (Table 3.2. above shows the data used). The Augmented Dickey-Fuller test is a hypothesis test with the null hypothesis being that a unit root exists, and the alternative hypothesis being that a unit root does not exist. The test produces a test statistic that is compared relative to the asymptotic Chi-Squared distribution. All critical values of the ADF test are negative in which the smaller the value of the test statistic (the more negative), the stronger is the rejection of the null hypothesis. The critical values for the Augmented Dickey-Fuller test can be found in the appendices as appendix A.

3.5. Unit Root Testing

E-views is a statistical package, used for time series-oriented analysis. The data for the South African All Share Index (ALSI) used was tested for unit roots on the statistical software, E-Views. In order to circumvent scaling discrepancies, the data was log-transformed before being tested for the presence of unit roots in the ALSI data. The GARCH model assumes non-stationarity and heteroskedasticity in the dependent variable. As such this test was only conducted on the ALSI data set.

3.6. Results and Implications of the Augmented Dickey-Fuller Test

The results of the Augmented Dickey-Fuller test is reflected in Table 3.3. below. The results show that for each of the five periods of the test, the dependent variable (the ALSI), a test statistic greater than the critical value of the ADF- Fisher Chi-Squared statistic was generated.

Table 3.3. ADF test statistic results

	ADF - Fisher Chi-Square	**p-value**
Period 1	23.7487	0.9547
Period 2	12.7139	1
Period 3	6.89237	1
Period 4	24.8746	0.8879
Period 5	25.4434	0.1214

Based on the results obtained, we fail to reject the null hypothesis at the 5% confidence interval and therefore conclude that there is a unit root in each time period. As discussed above, this implies that the time series data in non-stationary and heteroskedastic in nature.

The GARCH model assumes that the data set used as the basis of correlation i.e. the dependent variable, should be non-stationary (Engle, 2000). The results of Chapter 3 conclude that the All Share Index data is characterised as non-stationary. A number of models exist which compliment non-stationary data, but due to its robust features this study makes use of a GARCH model which is discussed in Chapter 4.

Chapter 4: Identifying Volatility Linkages Using GARCH (1, 1)

The GARCH model has become a significant tool in the analysis of time series data, particularly within the financial industry to model complex financial applications. The model is especially useful when the purpose of the study is to analyse and compare the volatility of time series data, thereby giving an indication of the level of correlation between two separate sets of time series data.

A data set is said to suffer from heteroscedasticity when the variances of the error terms are not equal, and the error terms are reasonably expected to be larger for some points or ranges than for others. As found in Chapter 3, the ALSI historic data used in this study is in fact non-stationary making the GARCH model a good fit as the basis of this study. The general warning is that when heteroscedasticity is found to be present in a data set, the regression coefficients for an ordinary least squares regression are considered to be unbiased. However, the standard errors and confidence intervals that were estimated by the conventional procedures will be too narrow for modelling, therefore providing an inaccurate sense of precision. Instead of treating this as a problem, the autoregressive conditional heteroscedasticity model (ARCH) and the generalised autoregressive conditional heteroscedasticity model (GARCH) treat heteroscedasticity as a variance to be modelled. This methodology results in the deficiencies of the three least squares to be corrected and a precision for the variance of each error term is computed.

4.1. GARCH Formulae

The generalised autoregressive conditional heteroscedasticity model (GARCH) contains two parameters which are referred to as p and q (p; q), where ‘p’ is the number of GARCH terms in the model and ‘q’ is the number of ARCH terms in the model. The GARCH(1, 1) model is the simplest and most robust of the family of volatility models and research has shown that GARCH (1:1) is a more reliable model in predicting correlation and volatility linkages between time series data and will therefore be the model of choice in this study (Hansen & Lunde, 2001).

GARCH (1, 1) is defined by the following equation:

$$R_t = \sqrt{h_t}.\varepsilon_t \tag{1}$$

$$\text{where } h_t = \alpha_0 + \alpha_1 R_{t-1}^2 + \beta_1 h_{t-1} \tag{2}$$

In Equation 1, R_t is referred to as the mean adjusted return of a share at time period t, and ε_t is referred to as a white noise variable such that ε Є~N (0, 1). In Equation 2, the parameters α_0, α_1 and β_1 are determined by a maximum likelihood estimation. Together, the equation in (1) and (2) implies that the variance today is a function of the previous days squared residual, the previous days variance and the weighted average long-term variance, such that the weights of α_1 and β_1 = 1.

One squared residual and one variance is captured at a time when using the GARCH(1, 1) model, while the ARCH model fluctuates the weights of parameters α and β until the best fit of the data is obtained. GARCH follows a similar procedure as ARCH. However, the difference is that in a GARCH model precedence is given to more recent data points in the time series.

Once all the inputs that are shown in Equation 2 above is obtained, the variance can be forecasted for the period under analysis using Equation 3 below:

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 R_t^2 + \beta_1 \sigma_t^2 \tag{3}$$

4.2. Maximum Likelihood Estimation and Log Likelihood

There are various methods to reach an estimation of the parameters used in a GARCH model. One such model is referred to as the maximum likelihood estimation (MLE), which is widely considered to be the most efficient (Bollerslev, 1986). As the name suggests, the MLE method maximises the log likelihood number through a series partial differentiation of the parameter estimates. The parameter figures that are obtained using this method are then substituted back into the log-likelihood function to obtain the log likelihood value.

The log likelihood value against a single data set is of little meaning as the likelihood value provides an indication of the overall quality fit of the model and should be interpreted in conjunction with other models in a data set so the fit relative to other data sets can be assessed. A smaller log-likelihood value indicates a more accurate fit of the model when comparing volatility linkages and therefore a positive correlation with the dependent variable. However, a large log likelihood value is evidence of a poor fitted comparison indicating negative correlation, which is why a relative comparison across models is key. To compliment the log likelihood value analysis, an Akaike Information Criterion method was considered appropriate to ensure the results are meaningful. This is explained below.

4.3. Akaike Information Criterion

Based on information theory, the Akaike Information Criterion (AIC) is considered to be the relative goodness of fit test for a GARCH model. The most efficient and optimal model when analysing the GARCH results is one which minimises the loss of information, and the Akaike Information Criterion (AIC) was used to provide a means of selecting the best fitting model from a set of models.

The Akaike Information Criterion (AIC) is defined by equation (4) below:

$$AIC = -2(loglikelihood) + 2K \qquad (4)$$

In Equation 4, K is considered to be the number of estimated parameters in the model. Estimated parameters are considered to be variables and the intercept. The absolute value of the AIC value must be considered when interpreting.

When interpreting the AIC value, a high absolute value implies that a higher level of information was lost in the modelling process whereas a lower absolute value implies that the model was able to capture most of the information with less lost information. As described for the log likelihood value, in isolation against a single model, the AIC value is of little meaning. The values should be captured and compared against a number of models in order to determine the efficiency relative to one another, where the model with the lowest AIC value is considered to be the most efficient.

In the context of this study, the GARCH model was used to determine the volatility linkages or correlation between South Africa and each country considered in Chapter 3. Notably, the GARCH model was run for each country and in each period as described in Chapter 3 of this document; therefore the results are obtained for each period separately. For each model the log likelihood value and the AIC value were obtained. The results were then interpreted in order to determine which country's index model lost the most information during the modelling process against the South African ALSI, therefore providing an indication of the overall model fit. Furthermore the AIC value can also provide an indication of correlation, as the greater the information loss in the modelling process, the less is the correlation between the price data sets. The results were interpreted in order to determine which of the countries are considered to be positively and negatively correlated to South Africa, based on the data input for each time period being analysed. A positive correlation is an indication that the country in question fluctuates in sync with the JSE, meaning that the general market trend (either upwards or downwards) of the South African market should be the same for positively correlated markets. On the other hand, a negative correlation is an indication that the country in question moves in the opposite direction to the JSE, and the general market trend would be reversed for negatively correlated markets. In order to represent a country's equity market, each country's largest stock index was used (refer to Chapter 3). Chen (2018) suggested that international market correlation is never constant but changes and adapts to prevailing market conditions. For this reason, the correlation in each period based on the prevailing market condition in each period being analysed is considered. In this chapter, the results pertaining to the correlation aspect is obtained and interpreted. In the later chapters this correlation is considered as a means for potential international diversification opportunities that investors could exploit in order to boost or stabilise portfolio returns.

4.4. GARCH Testing

Oxmetrics is an econometric statistical software providing an integrated solution for time series, forecasting, financial modelling and statistical analysis. It includes a front-end application called *Oxmetrics* which consists of a number of application modules, one of which is G@RCH. *Oxmetric's* G@RCH module was used to conduct all GARCH analysis. As previously explained, South Africa was considered the basis of each correlation and was therefore treated as the dependent variable when tested against each country's index using the GARCH model for each period.

Karlsson's (2002) suggestion of the distribution of the generalised error term in a GARCH model indicates that the Student-T error distribution is fitting. Furthermore Karlsson (2002) considered the Student-t distribution to be the most efficient for non-stationary data. The ALSI data (the dependent variable) in this study was found to be non-stationary as indicated in chapter 3. The Student-t distribution is also considered to be most appropriate for financial time series data. For each model the log likelihood, Akaike Information Criterion (AIC) and the parameters for the model were extracted and an in depth analysis was conducted to determine positive and negative correlations between the countries indices.

Using the South African ALSI as the dependent variable, the historic index price data for each time period was loaded into the Oxmetrics tool. Thereafter, the historic index price data of the independent variable was loaded into the software. The Student t-distribution is the default distribution when running a GARCH(1, 1) model. The parameter estimations, AIC values and log-likelihood values for each country's index as the independent variable was extracted against the South African ALSI as the dependent variable. It important to understand that the AIC and log-likelihood values for each country needs to be interpreted relative to each other in order to ascertain the level of correlation of each country's index relative to South Africa. A single value of one country is of little meaning.

Therefore, the AIC and log-likelihood values for each country, in each period was tabulated and compared in order to identify which country indices were positively or negatively correlated to the South African ALSI. For each period, a relatively high absolute AIC value constitutes that significant information was lost in the modelling process. A relatively low AIC value constitutes that the two data sets were a good fit and minimal information was lost in the modelling process. By relative this means relative to the other countries within that same period. Similarly, a relatively high log-likelihood value constitutes a higher quality model and positive correlation as compared to a low log-likelihood value.

The results for each period is discussed below and referred to in the following sections.

4.5. Results

The results obtained include the log likelihood value and the Akaike Information Criterion (AIC) for each period. The following key findings were obtained based on the analysis conducted.

Table 4.1. GARCH period 1 results

	AIC	Log-Likelihood
SA - Brazil	-4.799	320.583
SA - Russia	-3.691	252.664
SA - India	-3.791	252.762
SA - China	-3.884	267.433
SA - Nigeria	-3.796	255.087
SA - Hong Kong	-4.518	302.264
SA - Japan	-4.193	281.186
SA - Italy	-4.36	305.184
SA - France	-4.977	332.316
SA - Netherlands	-5.041	336.504
SA - Switzerland	-4.752	317.583
SA - Germany	-4.892	326.784
SA - Australia	-5.367	357.861
SA - Saudi Arabia	-3.792	247.157
SA - U.K.	-5.292	353.16
SA - U.S.	-4.72	315.534
SA - Turkey	-4.458	298.365

**Grey Highlight: Countries considered to have relatively negative correlation to South Africa*

**No Highlight: Countries considered to have relatively positive correlation to South Africa*

The first period that was analysed was characterised by the crash of the dot-com bubble which peaked and crashed in March 2000. The highest log-likelihood values were obtained by the relatively developed markets of Australia, the U.K., the Netherlands, France, Germany, Brazil, Switzerland, the U.S., Italy and Hong Kong in declining order. The corresponding AIC values also suggest significant information could not be used and was lost in the modelling process. This is an indication that these countries are considered to be negatively correlated to South Africa. It is interesting to find Brazil, considered to be one of the BRICS nations and an emerging market, in the list of negatively correlated markets. A number of developed markets were considered to be heavily impacted by the crash during this time as they embraced the technological age more than the emerging markets. The negative correlation with South Africa seen during this period is fitting as developed markets were more affected by this crash. On the other hand, Saudi Arabia, Russia, India, Nigeria, China, Japan and Turkey (in increasing order) reflected lower log likelihood and AIC values suggesting positive correlation with South Africa during this time. Three of the four BRICS nations were found in this result set dominated by emerging markets seen to be undergoing a similar market maturity. These countries may have also been less affected by the dot-com crash as the developed markets.

Table 4.2. GARCH period 2 results

	AIC	Log-Likelihood
SA - Brazil	-3.941	212.331
SA - Russia	-3.746	197.198
SA - India	-3.345	176.156
SA - China	-2.847	160.633
SA - Nigeria	-2.807	158.031
SA - Hong Kong	-3.557	187.165
SA - Japan	-3.309	174.248
SA - Italy	-4.138	216.421
SA - France	-4.565	239.586
SA - Netherlands	-4.578	240.362
SA - Switzerland	-3.882	214.464
SA - Germany	-4.682	245.774
SA - Australia	-3.736	206.892
SA - Saudi Arabia	-2.815	158.053
SA - U.K.	-4.364	229.12
SA - U.S.	-3.761	203.192
SA - Turkey	-3.663	192.684

**Grey Highlight: Countries considered to have relatively negative correlation to South Africa*

**No Highlight: Countries considered to have relatively positive correlation to South Africa*

Period 2 was characterised by a period of mean reversion, where markets showed a steady declining trend following the events of the dot-com bubble. Although the general market trend was considered to be recessionary in period 1 and 2, significant changes in international correlation was observed, which supports Chen’s (2008) suggestion of constant change in international market correlation and potential diversity opportunities. In declining order, Germany, the Netherlands, France, the U.K., Italy, Switzerland and Brazil was considered to reflect high log likelihood and AIC values and therefore suggest negative correlation to the JSE in this period. Of note, Australia narrowly dropped off the list of negatively correlated markets after being considered the most poorly correlated in period 1. The U.S. also dropped dramatically however, in light of the September 11, 2001 terrorist attacks in the U.S.A, investors may have been attracted towards safe haven markets with strong financial systems causing a stir in the U.S. market. The majority of negatively correlated markets were considered to be developed markets with strong financial systems. Largely in sync with period 1, Nigeria, Saudi Arabia, China, Japan, India, Hong Kong, Turkey, Russia, the U.S. and Australia in increasing order, was found to be positively correlated during period 2. Again, emerging markets dominated this list, suggesting that these markets moved more in sync with the South African market during this time. A comparison of period 1 and 2 clearly indicates how correlation can change in time, and how a countries individual circumstances can impact correlation during a period.

Table 4.3. GARCH period 3 results

	AIC	Log-Likelihood
SA - Brazil	-4.753	502.687
SA - Russia	-4.482	474.443
SA - India	-4.38	465.776
SA - China	-4.191	443.784
SA - Nigeria	-4.112	435.684
SA - Hong Kong	-4.553	481.818
SA - Japan	-4.466	472.753
SA - Italy	-4.865	514.444
SA - France	-5.012	529.518
SA - Netherlands	-4.85	512.733
SA - Switzerland	-4.727	498.068
SA - Germany	-4.882	516.206
SA - Australia	-4.881	518.125
SA - Saudi Arabia	-4.136	438.238
SA - U.K.	-5.082	539.018
SA - U.S.	-4.542	479.388
SA - Turkey	-4.542	479.635

**Grey Highlight: Countries considered to have relatively negative correlation to South Africa*
**No Highlight: Countries considered to have relatively positive correlation to South Africa*

Period 3 was characterised by a recovery period, after the markets finally took a positive turn following the previous years of recession. It is important to also note that this period accounted for a total of four years' worth of weekly data, which is the largest from all five periods. The fact that this period consisted of the largest data set may have contributed to a more robust GARCH model fitting. Furthermore, the time period overlapped with an increase in commodity prices, such as wheat and oil creating in increase in demand from the consumers in developed markets. This increase in demand also flagged the kick start for growth in export-orientated countries such as China, India and Russia. Although seen as a different market trend, this period's correlation results produced similar observations as period 2 in terms of ranking.

However, when analysing the results it was clear that no country actually showed an extremely strong positive correlation with South Africa as all countries observed high log likelihood values with corresponding high AIC values indicating a poor model fit and large loss of information in all models. The most negative correlation was seen in the U.K., France, Australia, Germany, Italy, Netherlands and Brazil in descending order. The most positively correlated markets were Nigeria and Saudi Arabia which have consistently showed positive correlations with the JSE in all periods thus far. However, this period did not observe very strong positive correlations, indicating that markets behaved largely independently during this period. The large data set in this period may have had an impact on results during this period.

Table 4.4. GARCH period 4 results

	AIC	Log-Likelihood
SA – Brazil	-4.657	248.646
SA – Russia	-4.663	249.133
SA – India	-4.751	253.533
SA – China	-4.626	247.023
SA - Nigeria	-4.644	248.231
SA - Hong Kong	-4.731	252.387
SA – Japan	-4.743	253.116
SA – Italy	-4.763	254.127
SA - France	-4.796	255.857
SA - Netherlands	-4.834	257.814
SA - Switzerland	-4.816	257.117
SA - Germany	-4.794	256.282

SA - Australia	-4.871	259.747
SA - Saudi Arabia	-4.632	247.22
SA - U.K.	-4.881	261.252
SA - U.S.	-4.734	252.642
SA - Turkey	-4.641	247.812

**Grey Highlight: Countries considered to have relatively negative correlation to South Africa*
**No Highlight: Countries considered to have relatively positive correlation to South Africa*

Period 4 was denoted by the sub-prime crisis and marked what is still considered the greatest global depression since 1929's great depression. The bubble was caused by the housing bubble in America and is further explained in Chapter 3. While the general market trend was recessionary in nature, a number of countries showed negative correlation with the JSE during this period indicating that every market was being impacted by this global crisis in a unique way based on their individual exposure to the crisis. In declining order, the U.K., Australia, the Netherlands, Switzerland, Germany, France, Italy, India, Japan, U.S. and Hong Kong all showed relatively high log likelihood values and corresponding AIC values, suggesting large information losses, implying a negative correlation with the JSE during this period. Interestingly, India and Japan, which was historically considered to be positively correlated to South Africa, now produced results which suggested the opposite. This is another example of how drastically international market correlation can change as markets adapt in time. In general, the majority of the markets showed strong negative correlation with South Africa during this time as markets were in a stir following the sub-prime crisis and the burst of the housing bubble. China, Saudi Arabia, Turkey and Nigeria showed what is considered the closest to positive correlation with the JSE during this time. With the exception of China, this list comprised of emerging markets. Of note, the results obtained during this period dispute the findings of Longin and Solnik (1995), who suggest that correlations among markets significantly increase during periods of increased volatility which was brought about by the sub-prime crisis. However, the market shock of a new experience may have counteracted this phenomenon.

Table 4.5. GARCH period 5 results

	AIC	Log-Likelihood
SA - Brazil	-4.215	221.586
SA - Russia	-4.288	225.491
SA - India	-4.311	226.561
SA - China	-4.128	217.151
SA - Nigeria	-4.158	218.684
SA - Hong Kong	-4.374	229.841
SA - Japan	-4.176	219.516
SA - Italy	-4.451	233.832
SA - France	-4.649	244.139
SA - Netherlands	-4.627	242.801
SA - Switzerland	-4.358	229.109
SA - Germany	-4.688	246.287
SA - Australia	-4.658	244.621
SA - Saudi Arabia	-4.153	218.334
SA - U.K.	-4.418	231.202
SA - U.S.	-4.633	243.308
SA - Turkey	-4.181	219.804

**Grey Highlight: Countries considered to have relatively negative correlation to South Africa*
**No Highlight: Countries considered to have relatively positive correlation to South Africa*

The current period (ended in 2018) was characterised by an uncertain market. Global events such as the Trump inauguration and Brexit have placed significant pressure. Closer to home, the fall of Jacob Zuma and rise of Cyril Ramaphosa has been considered as a potential turning point. In an uncertain future, the majority of the developed markets showed negative correlation with South Africa with Germany, Australia, France, U.S., Netherlands and U.K. reflecting the highest log likelihood values and corresponding AIC values. These countries have consistently appeared in the list of negatively correlated markets. On the other hand, China, Nigeria and Saudi Arabia observed the lowest log likelihood and AIC values. However, again, the positive correlated markets seem to be scarce as none of the markets showed extreme positive correlation.

It is important to understand the implications of correlation from a portfolio diversification perspective. A negatively correlated market can be seen as potential safe haven markets when local markets are declining. On the other hand, a positively correlated market should see similar returns as South Africa and the need for diversification is off little benefit during periods were the index price is expected to go up. The idea of investigating international correlation from a South African perspective in each period will provide insights of potential markets to exploits for local investors to offset losses and stabilise portfolio returns in varying market conditions, and results of these findings provide an indication of potential markets to exploit depending on predicted price movement of the South African All Share Index.

Chapter 5: Price Prediction using Support Vector Machines

Support vector machines are classified as one of numerous machine learning algorithms, where the foundations of support vector machines (SVM) have been developed by Vapnik (1998). The methodology has been gaining good popularity due to significant strides in technology and efficiency of implementation. Furthermore, it consists of multiple attractive features and promising empirical performance. Support vector machines are formulated using the Structural Risk Minimisation (SRM) principle which is considered to be superior to the concept of Empirical Risk Minimisation (ERM) which is employed by another conventional machine learning algorithm known as neural networks. The method of structural risk minimisation attempts to minimise the upper bound on the expected risk. On the other hand, Empirical Risk Minimisation (ERM) seeks to minimise the empirical error on the training data. The use of structural risk minimisation equips support vector machines with a greater ability to generalise which is the purpose of statistical learning. Historically, Support Vector Machines have been used for classification problems, but in recent years have been extended to the domain of regression problems in an adapted regression modelling technique referred to a Support Vector Regression.

5.1. Support Vector Regression

In the research area of machine learning, support vector machines are known to be supervised learning models with associated learning algorithms that analyze data that can be used for either classification or regression. When used in regression it is referred to as support vector regression or SVR analysis. In support vector machine modelling, when given a set of training data, each data point is marked to belong to one or another or two categories where the support vector machine training algorithm builds a model that assigns another new example to one of the categories making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples or training data points as separate points in space, plotted so that the examples of the separate categories that are explained above are divided by a distinct gap that is as wide as possible. New examples are then mapped into that same feature space and forecasted to belong to a category based on which side of the gap they fall.

The methodology behind support vector machines includes the construction of a hyperplane or set of hyperplanes in a high or infinite dimensional space which can then be used for classification, regression and other tasks such as outlier detection. Intuitively, an acceptable separation of the training data points is attained by the hyperplane that has the greatest distance to the nearest training data point of any class since in general the larger the margin, the lower the generalisation error of the classifier. This is referred to as the so-called functional margin.

When handling non-linear data, support vector machines can efficiently conduct a non-linear classification by using a method referred to as the kernel trick. This trick implicitly maps their inputs into a higher dimensional feature space. Considering the volatile nature of stock and index price data, one can reasonably assume that data associated with stock price time series data is in fact non-linear. Furthermore, in Chapter 3, it was established that the historic data of the All Share Index (ALSI) used in this study is non-stationary which confirms that the mean and variance of the data is constantly changing. Implicitly, this identifies the data as non-linear, and requires the data to be mapped into a higher dimensional feature space.

Support vector machines make use of 3 types of kernel functions depending on the type of time series data being used. These kernel functions are:

- Linear
- Polynomial
- Radial Basis Function (RBF)/Gaussian Kernel generally used for erratic non-linear data.

Although non-linearity can be reasonably assumed, part of the analysis conducted in this section confirmed this data characteristic by running the South African All Share Index price data plots (using an extract of the daily prices for the month 01 Dec 2018 – 31 Dec 2018) against each function using an open source programming language called Python. Being open-source, Python has the added advantage of having free libraries, which give access to pre-written code through a module called Scikit Learn. Using this module, this study included another test for linearity of the data points for the ALSI. The below graph depicts the closest relationship with the RBF kernel confirming non-linearity of the data. The Python code used in this implementation can be found in the appendices as appendix B.

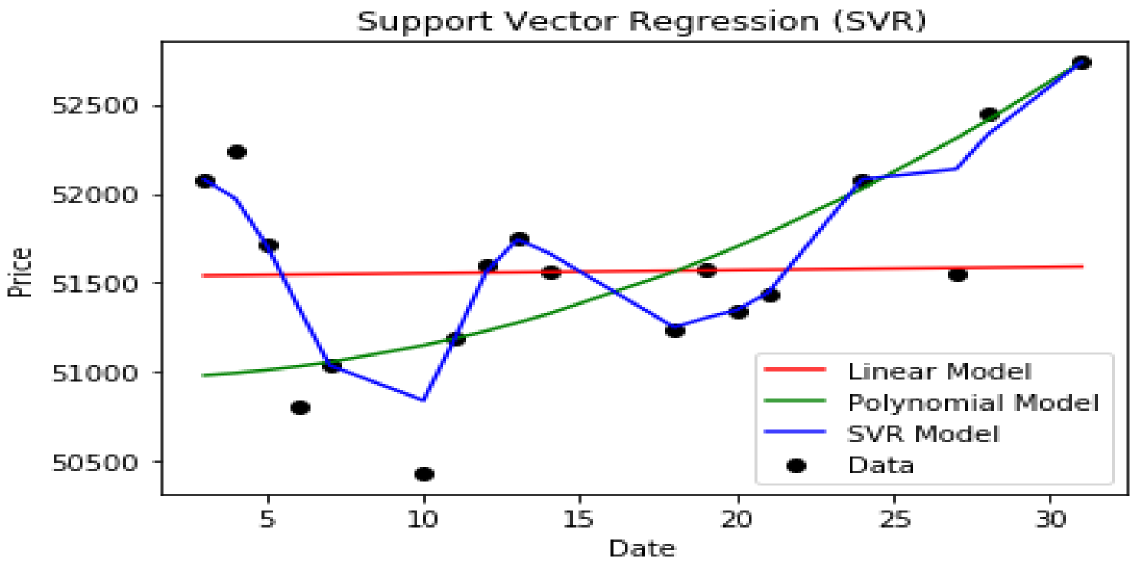


Figure 5.1. Comparison of data fit across SVM functions

5.2. Radial Basis Function

The Gaussian RBF (Radial Basis Function) is a popular Kernel method used in SVM models for non-linear data. The RBF kernel is a function whose value depends on the distance from the origin or from some point. The Gaussian Kernel is of the following format:

$$K(X_1, X_2) = exponent(-\gamma \|X_1 - X_2\|^2) \qquad (5)$$

Where: $||X_1 - X_2||$ = Euclidean distance between X_1 and X_2.

Using the distance in the original space, the dot product or similarity of X_1 & X_2, is calculated.

Parameters used in the radial basis function:

1. C is the inverse of the strength of regularisation.

   As the value of ‘C’ increases the model gets over fits.

   As the value of ‘C’ decreases the model under fits.

2. γ is the Gamma which is used only for RBF kernel.

   As the value of ‘ γ’ increases the model gets over fits.

   As the value of ‘ γ’ decreases the model under fits.

The RBF function is well suited for non-linear data as it maps the non-linear data into a higher dimension where it can then be separable. As seen in Figure 5.2 below, the data on the right pane is linearly inseparable in two dimensions:

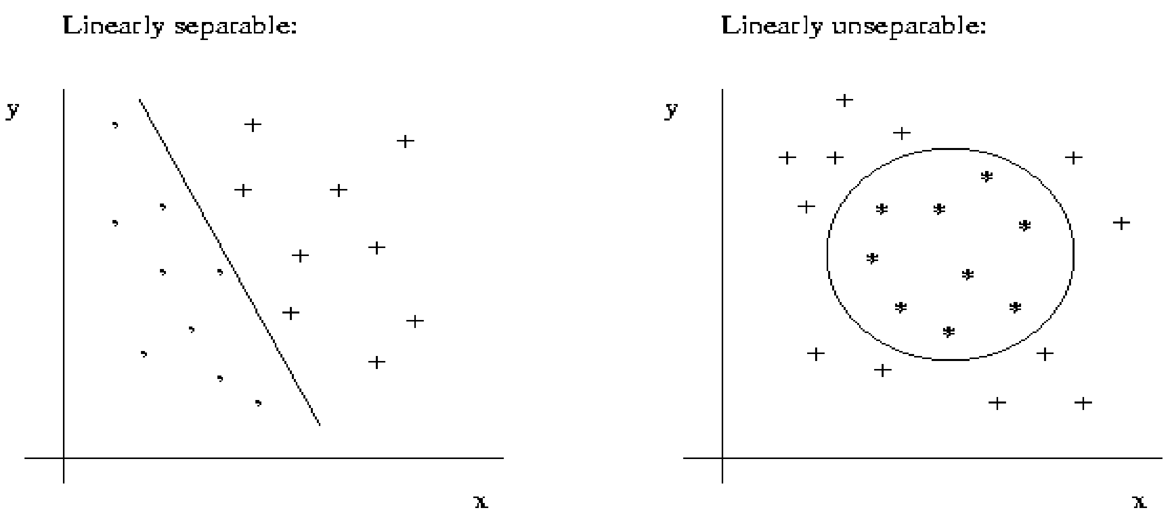


Figure 5.2. Linearly inseparable data points (Source: University of Finland, 2011)

However, using the RBF function, the data points are mapped into a higher dimensional feature space and becomes separable at a higher dimension. This is illustrated on the right pane in Figure 5.3.

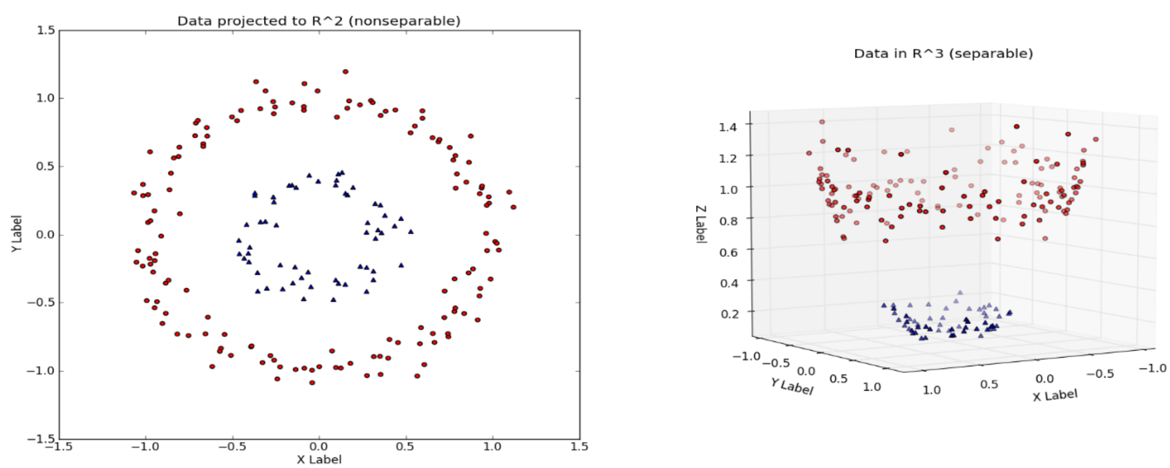


Figure 5.3. Data Separation at a higher dimensional space (Source: University of Finland, 2011)

5.3. Stock Price Prediction using SVR

Using conventional statistical theories and methodologies, it has been found that the ALSI data being worked with in this study is non-stationary and non-linear. Non-stationarity was implied following the tests conducted in Chapter 3 and non-linearity was confirmed above in this chapter after determining which kernel function will efficiently fit the data being used. In Chapter 4 the volatility linkages between South Africa and some of its major trading partners was investigated. Each period was analysed individually to determine whether current market conditions would play an impact in the correlation to South Africa. The correlation results which was unpacked in chapter 4 will then be interpreted along with the results obtained in the support vector machine price prediction trend and percentage accuracy.

For the remainder of this chapter, a support vector regression (SVR) technique using a radial basis function was employed to attempt to predict the weekly index price for the final 20 weeks of each time period that was described in Chapter 3. A 20-week prediction window was selected (i.e. 5 months) as most data sets consisted of 2-year periods and a suitable training/testing split is considered 75-80% training and 20-25% testing (H20.ai, 2017). Therefore, most periods consisted of 19 months of training data and 5 months of testing data, with the exception of period 3 which totalled 4 years of data meaning 43 months were used as training and 5 months as testing. The 20 week prediction was left constant in period 3 in order to make meaningful comparisons.

It is important to understand that each period's price prediction was done within the time period i.e. in a two-year period window, the last 20 weeks of the period were predicted and tested against the actual value. An accuracy measure known as Mean Relative Error (MRE) was used to assess the accuracy of the model (Marwala, 2017). The MRE is an accuracy measure which evaluates the average change in the error term. The accuracy measure is determined using the following formulae:

$$\text{MRE} = (\text{average of } (1 - \frac{Predicted\ Price - Actual\ Price}{\text{Actual Price}})) * 100 \qquad (6)$$

The sensitive nature of index prices should be considered when interpreting this accuracy value. Even though a price could be predicted considerably off the actual price, this measure could still produce a relatively high accuracy measure. A good accuracy measure is considered to be between 95 to 105, with 100 being absolutely perfect (Marwala, 2017). Notably, this measure can exceed 100 if the model consistently produces values greater than the actual price. This can be noted when thoroughly examining the inputs into Equation 6 above.

The prediction model was run against the South African All Share Index in order to determine the upcoming trend predicted by the model. Once the trend was determined, either being a forecasted increase in the index price or a forecasted decrease in index price, the correlation results provide insights into which international equity markets to exploit in each scenario and in each period. It is important to note that although the general market condition could be in a recessionary or recovery stage, the predicted index prices for the final 20 weeks could be expected to increase. Considering that this study focuses on a short-term prediction window, this study evaluates the diversification opportunities based on a short-term trend analysis.

While the focus of this study is around diversification opportunities from a South African perspective, most of the analysis focuses on the prediction trend of the South African All Share Index. However, the prediction model was also run against each country in order to answer the final research question, which is to determine if market correlation translates to prediction power. This is discussed in Chapter 6 of this study. Furthermore, by considering the prediction of each market we are able to confirm if the diversification benefits hold any value when implementing this investment strategy. The assumption is that a negatively correlated market will follow an opposite trend to South Africa and a positively correlated market will follow a similar trend in index price to South Africa.

The RBF function explained above is used for all models that were ran. All prediction models were run in an open source programming language called Python. As previously explained, Python is an open source programming language which allows access to free libraries which include key functions that were used in this study. Furthermore, Python makes use of Jupyter Notebook which is an open-source web application that allows users to create and save live code. This is advantageous as a user does not need to install an independent development environment (IDE) in order to write code but rather have the flexibility of programming directly off a web interface. The following Python Libraries were used:

- SciKit-learn is a free software machine learning library for the python programming language. It includes features such as classification, regression and clustering algorithms which are used in support vector machines, random forests, gradient boosting, k-means and DBSCAN. The library is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy. While the library can be used for a number of modelling techniques, the specific function of support vector regression is used in this study.
- Numpy is a fundamental library for scientific computing within the Python programming language. It includes a powerful N-dimensional array object with sophisticated functions. The package is commonly used for linear algebra and random number capabilities which fits in nicely with the use case in this study.
- Matplotlib is a Python 2-dimensional plotting library which produces figures that are off publication quality in a variety of formats and interactive environments across platforms. Using this library a user can generate plots, histograms, bar charts, error charts, line graphs etc. In this study, this library was used in order to graph predicted versus actual index price plots.

Anaconda is a data science platform allowing a user access to these libraries via the web application. Once installed the following process flow was conducted in the Python code.

1. In order to make use of the libraries described above, the necessary libraries were imported into the session.
2. In order to make use of the relevant data needed in this model, a empty list was defined in order to eventually import the time series price data.
3. Once the list was defined, a function was written in order to pull the data into the session. Furthermore, cleaning of the data was necessary as follows:

   a. Part of this step involves identifying missing data points and making adjustments to cater for them.
   b. Missing dates and data points were replaced with an average of the previous and following dates closing share prices (Longin and Solnik, 1995).
   c. Outliers were detected and replaced with $\mu \pm 2\ \sigma$, depending on whether the observation was on the lower or upper bound, respectively (Longin and Solnik, 1995).
4. Once all libraries and functions to cater for data was complete. A function was written to define the implementation of the support vector regression model. As described above, this was specifically defined for the radial basis function. The two parameters (see Section 5.2 above) required for the radial basis function was set with C = 1e3 and $\gamma = 0.1$.
   a. Part of this step involves looping the data through each date in order to identify a trend and learn from previous movements.
5. A final function was written using the Numpy library in order to compute the model accuracy. This required considerable computing power as it needed to loop through each data point and each predicted value.
6. Finally using the Matplotlib library the predicted values were plotted against the actual values.

The basis of all machine learning algorithms involves a set of training data in order for the machine to identify a trend in data and learn from historic data. For this reason, the data was split into training and testing data points. This was done for each time period and each index. Within the function described in step 4 of the process flow, a sub function available through the Scikit-learn library was used to process the data split. This function is referred to as "train_test_split". In each data set, the splitting was done in order to ensure that the final 20 data points (in weekly intervals) were predicted based on the support vector machine algorithm. This means that all the data prior to the final 20 weeks were used as training data. Considering that each period has varying weekly data points, the split between training and testing in each period was not consistent in each model. Therefore, period 3, which included the largest data set was trained using the most amount of data. In each period the final 20 points was the test data and a trend and prediction accuracy was determined. The Python code used for this modelling can be found in the appendices as Appendix C.

In the section below the results of the SVR model, in conjunction with the correlation results obtained in Chapter 4, are discussed.

5.4. Results

From the results obtained, an examination of the graphed results of the predicted prices and the actual prices for the South African All Share Index was conducted. A point worth noting is that although a period may have been denoted by a recession or a recovery, the graphs reflect only the 20 weeks that were predicted and therefore may not necessary be reflective of the overall market condition. Since the prediction window is only 20 weeks, there is the possibility of a 20 week general increase in prices during a recession and vice versa. Based on this observation, the impact of the correlation results are considered for each 20 week period. The table below shows the prediction accuracy value (MRE value in Equation 6 above) that was computed for each period and each country. Using these results an indication of how accurately the model fitted the data and how that translated into prediction power can be determined. Furthermore, the expected market trend for the prediction period can be assessed and finally based on the correlation results, a suggestion relating to international portfolio diversification opportunities can be explored.

Table 5.1. Model accuracy (MRE) using SRV prediction

	Period 1	Period 2	Period 3	Period 4	Period 5
AUSTRALIA	99.42	99.77	99.15	99.11	99.26
BRAZIL	97.74	94.54	64.54	95.41	95.22
CHINA	99.41	99.57	97.93	98.86	99.05
FRANCE	98.39	99.35	99.31	99	99.41
GERMANY	98.98	99.13	99.57	99.26	99.37
HONG KONG	97.19	99.17	81.18	98.24	98.83
INDIA	99.34	99.38	98.1	98.7	99.33
ITALY	84.22	99.25	98.59	98.7	98.79
JAPAN	97.63	98.82	98.76	98.76	98.75

NETHERLANDS	98.8	99.1	99.51	99.32	99.45
NIGERIA	99.02	96.65	61.99	85.37	99.21
RUSSIA	98.99	98.81	99.37	98.42	99.31
SA	98.74	99.14	79.35	99.04	98.79
SAUDI ARABIA	99.65	99.42	99.16	99.08	99.36
SWITZERLAND	98.28	99.31	99.18	99.48	99.34
TURKEY	96.18	98.01	74.57	91.71	94.88
UK	98.95	99.5	99.3	99.13	99.5
US	98.94	99.61	99.37	99.37	99.16

Period 1:

In period 1, which was characterised by a recession caused by the burst of the dot-com bubble found strong prediction power over most markets.

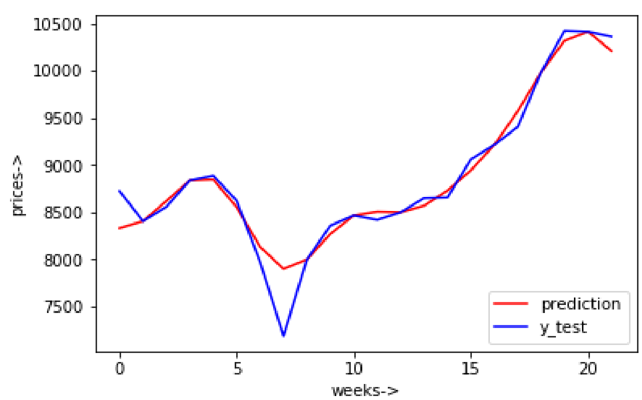


Figure 5.4. South African All Share Index prediction (Period 1)

** y_test refers to actual price data points*

Considering the price fluctuation during period 1, there is limited opportunity for international diversification as the predicted price ended the 20 week window at a price higher than the start of the 20-week window. The graph above shows a significant downward trend in the ALSI index price between week 4 and 7. Therefore, a rational investor looking to offset potential local losses during this period will consider investment opportunities that will reduce the overall portfolio effect of the ALSI decline during week 4 – 7. Based on the correlation results during period 1 as per Table 4.1 in Chapter 4, the top three countries that had equity markets most negatively correlated to South Africa included, Australia, the U.K. and Netherlands. All these countries produced a relatively high prediction accuracy observed in Table 5.1 above. By considering negatively correlated markets, an investor may be able to offset local losses with foreign gains in markets that are expected to be negatively correlated to the South African equity market. When investigating the graphed trend analysis of each country, all of the above countries experienced a similar downward trend in the respective index price, therefore suggesting that diversification benefits were not profitable and would not offset local losses during the first period. For period one it can be concluded that the suggested investment approach is of little value. The over-arching theme characterised by period one (the burst of the dot-com bubble) may have had an impact on all markets during these weeks making historic correlation a poor reflection on current price movement in the period.

The ALSI index price experienced an overall upward trend in period one, during periods of a predicted increase in price, a rational investor should reduce or minimize international diversification as positively correlated markets should experience similar gains and the value of diversification is nullified. The prediction graphs for Australia, the U.K. and Netherlands can be found in the appendices as appendix D.

Period 2:

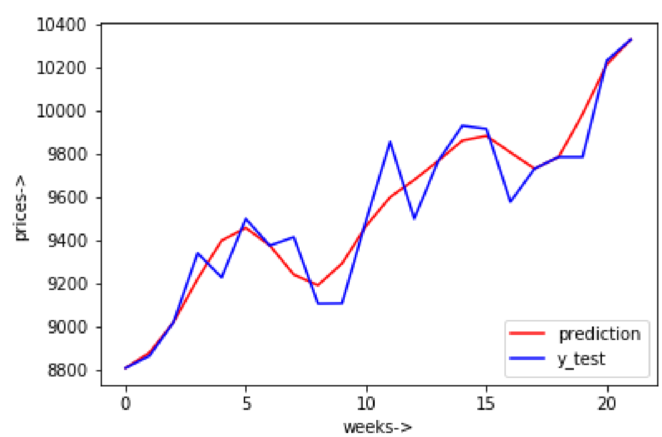


Figure 5.5. South African All Share Index prediction (Period 2)

** y_test refers to actual price data points*

In period 2, which was also classified as a recessionary period due to mean reversion following the dot-com bubble crash, produced similar prediction accuracy levels as period 1. In period 2, the South African ALSI produced a prediction accuracy (MRE) of 99.12, which is considered relatively high. Again, the fact that the predicted values failed to capture large price jumps is concerning, and is likely to be a reason that machine learning algorithms is not appropriate for active trading. It is important to note that the machine learning algorithm does well to predict a general price direction move. The predicted index price trend again accurately forecasted the fluctuation during the 20 week period. In the short time period, the predicted index price observed two noticeable price increases and decreases, where the declining index price behavior was observed between week 5-8 and 15-17. These weeks provide potential opportunities where a rational investor would seek to offset these price declines and portfolio losses by considering investment opportunities in negatively correlated international markets.

Based on the correlation results for period 2 (Table 4.2 in Chapter 4), the top three countries that had equity markets most negatively correlated to South Africa included Germany, the Netherlands and France. All these countries produced a relatively high prediction accuracy as shown in Table 5.1 above. When investigating the graphed trend analysis of each country, all three countries saw significantly more fluctuation in their respective index price when compared to the South African ALSI which contributed to the negatively correlated equity market relation. The first and second ALSI price dips coincided with corresponding index price increases in each of the three equity markets that were considered to be the most negatively correlated to the South African ALSI.

Therefore, in period 2, it can be concluded that the use of international diversification using the equity markets of Germany, the Netherlands and France in a portfolio would have been beneficial. The two step approach would allow an investor to first consider the positive and negative correlations between countries within a time period and then make rational decisions based on the outcome of the predicted prices for that period. An investor can then consider negatively correlated markets when local index price is expected to decline, thereby offsetting potential future losses. The prediction graphs for Germany, the Netherlands and France can be found in the appendices as Appendix E.

Period 3:

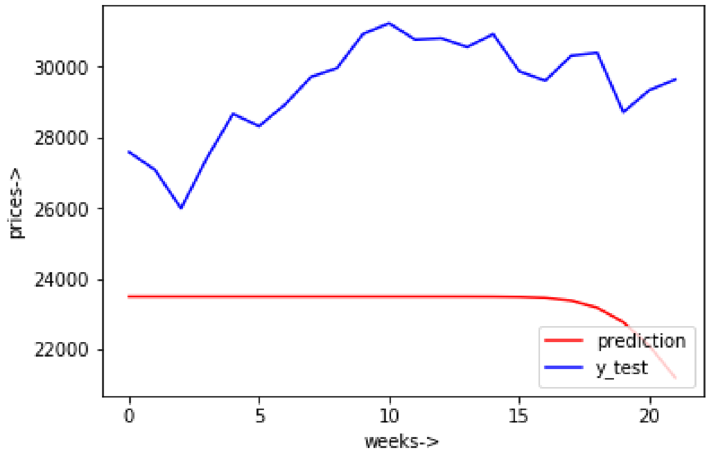


Figure 5.6. South African All Share Index prediction (Period 3)

** y_test refers to actual price data points*

In period 3, which was considered to be a recovery period where markets finally took a positive turn following a number of recession periods, the price prediction accuracy values (MRE) for a majority of the countries were poor as a number of country's showed values less than 95 (see table 5.1 above). Period 3 constituted the largest data set spanning 4 years. Based on this, one would expect strong prediction power as the prediction model is able to learn from a larger data set and include varying price movements into the algorithm. On the contrary, this period observed the worst overall accuracy results, with South Africa producing a model accuracy of 79.35 which is considered poor in the context of this study and how the accuracy value is computed (Equation 6). The results found in this period may also suggest that when using large data sets that include large upward and downward fluctuation, over fitting results and therefore leading to poor prediction power, as the algorithm is unable to identify a consistent trend in the data movement. It is difficult to consider international diversification opportunities in this period as the machine learning algorithm was unable to capture the dynamic market relationships and produce meaningful results. It would be inaccurate to say that the South African ALSI was expected to decline (the red line in figure 5.6.), as the actual price slightly increased during the period. However, the correlation results in Chapter 4 (Table 4.3) did not find any significantly strong positive correlations during this period, and the majority of markets were considered negatively correlated. If the price was expected to decline based on predicted values during this period, a rational investor would consider negatively correlated markets to offset potential losses. In this period, the U.K., France and Australia were considered to be the most negatively correlated and also produced relatively accurate prediction accuracy. On the other hand, Nigeria and Saudi Arabia was considered to be the most positively correlated.

Period 4:

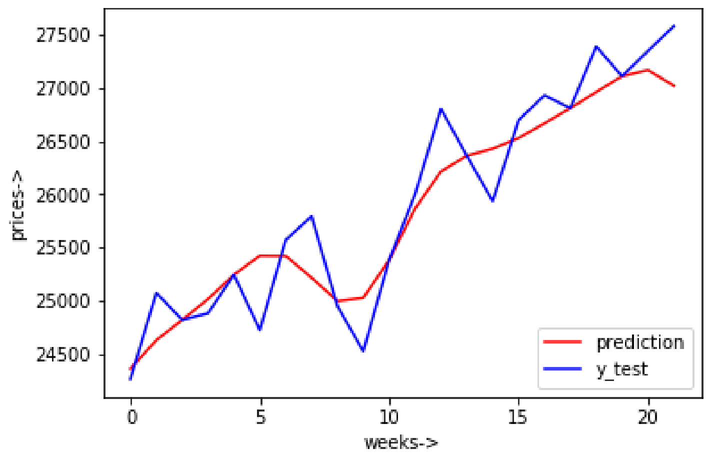


Figure 5.7. South African All Share Index prediction (Period 4)

** y_test refers to actual price data points*

Period 4 was denoted by the sub-prime crisis and, considering the severity of this crisis, markets behaved extraordinarily strange during this period. However, the results seem to be consistent with period 1 and 2. All markets observed strong prediction power, with the South African All Share Index computing an index price prediction accuracy (MRE) of 99.04 which is considered a good accuracy. Again, although the period being analysed was denoted by a recession, the 20 week period that was predicted, forecasted an incoming increase in prices. As was the case in previous periods, the prediction power fails to capture sharp spikes and dips but does well to produce an overall market trend as shown in Figure 5.7 above. In this case, the shorter data set allowed for the algorithm to identify a stand out trend in the price movement and produce good prediction accuracy. The predicted ALSI price (denoted by the red line in figure 5.7.) experienced a clear dip in index price after week 5, continuing to about week 8. This period of local price decline, provides an opportunity for international portfolio diversification. A rational investor should seek to minimise the effect of the ALSI decline by investing in negatively correlated international markets. The idea is that a negatively correlated equity market would not necessarily experience the same decline, but would likely experience index price increases, thereby offsetting local losses and stabilising overall portfolio returns. Based on the correlation results during period 4 (Table 4.4 in Chapter 4), the top three countries that had equity markets most negatively correlated to South Africa were the U.K., Australia and Netherlands. All these countries produced a relatively high prediction accuracy, as shown in Table 5.1. When investigating the graphed trend analysis of each country, there were encouraging opportunities to offset the ALSI price decline by considering investment opportunities in the U.K., Australia and Netherlands. The predicted price decline of the South African ALSI during week 5-8 largely coincided with corresponding index price increases in each of the three countries that were investigated. In period 4, it can be concluded that the suggested investment strategy could potentially prove beneficial to an investor who would consider international equity markets to diversify their portfolio, thereby offsetting predicted local losses. The prediction graphs for the U.K., Australia and Netherlands can be found in the appendices as Appendix F.

Apart from the weeks between 5 and 8, The ALSI price in period 4 largely predicted increases throughout the 20-week window. International diversification is of little value during predicted price increases, as positively correlated markets should experience similar gains and negatively correlated markets should decline in theory, thereby providing little to no value for a rational investor seeking stable portfolio returns.

Period 5:

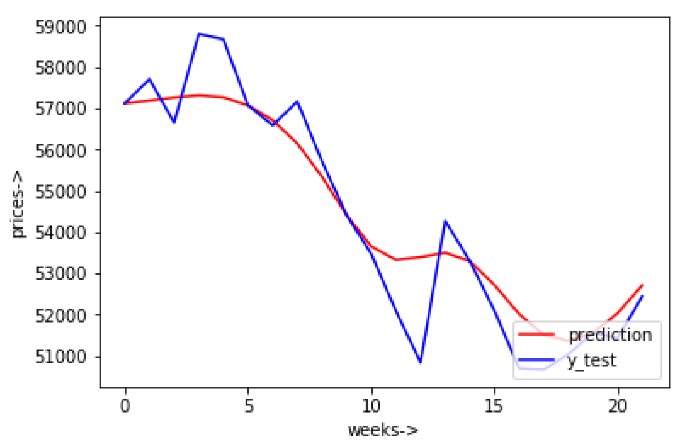


Figure 5.8. South African All Share Index prediction (Period 5)

** y_test refers to actual price data points*

The time gap between periods 4 and 5 is considerable (eight years), and this period may have indicated a shift in local market efficiency. The current period was characterised by uncertainty, and considering historic upward and downward fluctuation, which generally is not conducive for a machine learning algorithm attempting to pick up a trend in time series data, a poor predictive power result is expected. Contrary to this expectation, all markets observed high predicted accuracy values, with South Africa producing a price prediction accuracy value (MRE) of 98.79. An important consideration in all results and in all periods, is that while the prediction power generally shows a high level of overall accuracy, the model fails to capture important price fluctuation spikes which is a significant part of any investment strategy, especially in active portfolio management and trading. Period 5 experienced overall downward ALSI price prediction, where the predicted price ended the 20-week window considerably lower than the start of the period. This price trend provides significant opportunity for international diversification.

A rational investor looking to offset local predicted losses in an international portfolio would likely consider investments in countries that are negatively correlated to South Africa. Based on the correlation results during period 5, the top three countries that had equity markets most negatively correlated to South Africa included Germany, Australia and France. All these countries produced a relatively high prediction accuracy. When investigating the graphed trend analysis of each country, all three countries although considerably negatively correlated to South Africa, unexpectedly produced an index price that was significantly lower than the start of the 20 week prediction window, therefore suggesting that diversification benefits with all three countries was not profitable and would reduce overall portfolio return. The results observed in this period challenges the concept being considered in this study. Based on the negative correlation relationship, a rational investor would attempt to offset their local losses by investing abroad.

However this theory is not useful in this period as all markets are expected to decrease. The end of 2018 saw significant global events such as the Trump inauguration and Brexit that had a large impact on all developing and developed nations. These factors could have impacted the latter part of 2018, causing all markets to decline. The prediction graphs for Germany, Australia and France can be found in the appendices as Appendix G.

Chapter 6: Conclusion

6.1. Key Findings

This study set out to address three research objectives. The first centred on market correlation and determining the level of correlation, either positive or negative, between the JSE in South Africa and other major developing and developed nations. Secondly, this study tested the concept of the efficient market hypothesis (EMH) on the JSE, which states that all public and private information that could affect the current stock price value is accounted for by the market before the public can trade on it. This means that it would be impossible to forecast future prices as the current price is already reflective of everything that is known of the stock. The efficient market hypothesis was tested by investigating whether a machine learning algorithm known as support vector machines could predict future prices within a 20-week time period with a reasonable accuracy. The third question, merged the results of the first two, by attempting to determine potential international markets to exploit within each period analysed, based on the predicted price trend of the South African All Share Index. An investment strategy was investigated from the perspective of a rational investor looking to achieve enhanced or stable portfolio returns, by diversifying into international markets based on the future predicted prices of the South African ALSI. If ALSI prices were expected to decline, a rational investor would consider negatively correlated international markets in order to offset predicted losses and achieve stable returns in an internationally diversified portfolio. Alternatively, if the ALSI price were expected to increase, the value of diversification is nullified as positively correlated markets should experience similar gains and negatively correlated markets should decline in theory, thereby providing little to no value for a rational investor seeking stable portfolio returns.

In order to determine the best method to determine market correlation between the chosen countries, as described in Chapter 3, the collected data of the All Share Index (ALSI) was subject to the unit root test known as the Augmented Dickey-Fuller Test. The ALSI was considered the dependent variable in Chapter 4. The results obtained observed high critical values and it was concluded that the null hypothesis failed to be rejected and therefore the ALSI data being worked with is non-stationary and heteroskedastic in nature. Considering the nature of the data and guided by the literature on the topic, it was concluded that the GARCH(1, 1) model was the most efficient and appropriate correlation modelling methodology to use.

In Chapter 4, the application of the GARCH(1, 1) model, which was considered the most appropriate model, was used to identify volatility linkages between each country's major listed equity index and the South African ALSI. The model was run against each country in each time period, as defined in Chapter 3, and separate results for each period was obtained. Varying findings were made for each period but the equity markets in Germany, the U.K., Australia, France and the U.S. were most consistently found to appear in a list of countries that were considered to be negatively correlated to South Africa. Conversely, the equity markets of Saudi Arabia, Nigeria, India, Turkey, China and Russia were most consistently found to appear in a list of countries that were considered to be positively correlated to South Africa. Interestingly, a number of BRICS nations appeared in the results of positive correlation as well as Nigeria which shares our continent.

Finally, in Chapter 5, a machine learning algorithm known as Support Vector Machines or SVM, was used to predict the final 20 weeks price data points for each time period. The data was split into a training period for the machine, and the final 20 price points were used as a testing period and for computing a model accuracy for each country in each time period. Again, varying results were found in each period, but it was concluded that the machine learning algorithm in general does well to capture a general market trend and direction, but fails to capture sharp spikes and dips in index prices. In some scenarios, however, the model was unable to fit an accurate model, resulting in poor prediction accuracy results. However, the overall prediction accuracy in each model was generally high and a tentative conclusion that it is potentially feasible to predict a market trend using machine learning was made. The recession periods generally provided stronger predictive power relative to the results obtained from the recovery period examined. The recovery period was characterised by the largest data set and the prediction algorithm failed to produce a meaningful fit. In addition, in Chapter 5, a predicted trend for the final 20 weeks (of each period) of the South African All Share Index was conducted. International diversification opportunities were explored based on whether the price trend in each period was expected to increase or decrease over the 20 week period. The results in each period varied drastically and suggested that this method of international diversification based on historical correlations could be profitable in some instances. However, in a large number of cases, this method fell apart, and would have likely been detrimental to overall returns. This was especially apparent in period 5, which was the most recent time period analyzed. A conclusion is therefore, that using this methodology to enhance predicted index returns from a South African perspective produces inconsistent and unreliable results. While it is a reasonable approach, in practice the results are considered inconsistent and either a tuning of the methodology or another methodology should be considered to achieve the desired result. Finally, it is important to understand that the correlation and prediction results were run for each period as described in Chapters 4 and 5.

It should also be noted that for every data set used, the predicted prices failed to capture large spikes and dips in price data but was able to efficiently capture an overall trend in the price. Although the model accuracy value (MRE value) showed strong predictive power as an average of the predicted prices, the failure of the model to capture sharp price changes is a major limitation to its practical value.

In all periods, the equity markets of Australia, the Netherlands, the UK, France and Germany all consistently showed the least correlation with South Africa's JSE, and therefore appear to be the best alternative markets to offset local losses when a declining local market is expected. Potential investors can use these international markets when future prediction in SA is expected to decline, with the hope to keep overall portfolio performance steady. Considering a low correlation to South Africa, investors can anticipate that the market trajectory of these countries will counter the negative effects of local downward trending market performance. On the other hand, the equity markets of Nigeria, Saudi Arabia, China, India and Russia consistently showed the highest correlation to South Africa. However, positively correlated equity markets were not considered in the suggested investment strategy as it was deemed to add little to no value in enhancing portfolio returns.

6.2. Limitations and Future Research

In terms of future research from the market correlation and international diversification perspective, it would be interesting to note the possible volatility linkages between countries within the African continent. With the exception of Nigeria, no other African country was investigated in this study. Considering the close relationship that South Africa has with many African countries, this will be a useful investigation. However, the problem of data access poses a significant hurdle when African countries are considered as there is a lack of resources in obtaining share and index price information for most African countries. Furthermore, if this data is available, it is generally available in monthly intervals as found on the Bloomberg station. As discussed in Chapter 3, monthly data was considered inappropriate for this sort of investigation as it would restrict the robustness of the statistical models used in this paper. However, monthly data will allow for more African countries to be included.

In order to confirm the diversification opportunities that could be present in any given period, an additional test for co-integration between the countries may be beneficial to solidify the relationships between stock returns from each country with South Africa. Furthermore, the composites and indices that were selected in this study varied significantly in market size and also contained a largely different combination of equity instruments from a number of industries and sectors. In order to reduce this risk, it would be useful to ascertain which index or set of shares is the most accurate representation of the equity market of each country by conducting a sensitivity analysis of the most effective shares within an index.

From a machine learning perspective, and in order to more efficiently investigate the efficient market hypothesis of each country, the work on index price prediction can be extended by increasing the number of parameters in the feature space. By including other technical indicators such as the moving average, inflation rates, volume traded etc. the model could be made more robust and provide more accurate predictions. These factors, together with the index prices of the market to train and test the different models, could be a way of improving the method investigated in this study. This would reveal whether the prediction accuracy of the support vector machine algorithm can be improved. On the other hand, this increases the risk of over fitting the model which may be detrimental to the results. Intensive research must be conducted in order to confirm the most efficient parameters to use in the model.

From a different viewpoint, this study limited the future prediction points to the last 20 weeks within the dataset. An interesting result will be to determine the prediction power in longer periods. In longer time periods, markets can fluctuate immensely and it will be insightful to determine if support vector machines can capture these fluctuations. Furthermore, the 20-week prediction window resulted in limited or inconsistent diversification opportunities when the correlation results were interpreted together. By extending the prediction window used in this study, greater insights can be gained into whether such a strategy could be practically feasible. A further advancement to this study could be the input of individual stocks into the model in order to create a paper portfolio over a set period to determine how well a model will perform in terms of returns, when compared to the overall market with actual returns.

The use of other macroeconomic variables such as interest rates, CPI and GDP as inputs into the forecasting model could also prove worthwhile in order to determine if these variables have any predictive power over asset prices. By allowing the model to determine a relationship between these variables and the asset price, investors will be able to understand the affect if these variables are expected to increase or decrease in the future.

Finally, Support Vector Machines were used as the machine learning technique of choice in the study as previous research identifies this as the most applicable technique. However, new models and algorithms are continuously being developed. Other techniques include Artificial Neural Networks, Decision trees and Bayesian Networks which are some available techniques to be explored further from a financial point of view.

References

Allen, D., Amram, R., & McAlee, M. (2013, August). Volatility Spillovers from the Chinese Stock Market to Economic Neighbours. *Australian Research Council, 94*, 238-257.

Amadeo, K. (2018). *Stock Market Crash of 2008*. (T. Balance, Editor) Retrieved April 2019, from The Balance: https://www.thebalance.com/stock-market-crash-of-2008-3305535

Arshanapalli, B., & Doukas, J. (1993). International stock market linkages: evidence from the pre- and post- October 1987 period. *Journal of Banking & Finance, 17(1)*, 193-208.

Bodie, Z., Kan, A., & Marcus, J. (2009). *Investments.* 8th Ed. McGraw-Hill Irwin.

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

Brown, D., & Jennings, R. (2015). On Technical analysis. *The Review of Financial Studies, Volume 2*(Issue 4), 527 – 551.

Cao, Q., Leggio, K., & Schniederjans, M. (2005). A comparison between fama and french's model and artifcial neural networks in predicting the chinese stock market. *Computers & Operations Research, 32(10)*, 2499-2512.

Chang, H., & Su, C. (2009). The relationship between the Vietnam stock market and its major trading partners – TECM with bivariate asymmetric GARCH model. *Applied Economics Letters, Vol 17*, 1279-1283.

Chatfield, C. (2000). Times Series Forecasting. *University of Bath, England: Chapman and Hall/CRC Press*.

Chinzara, Z., & Aziakpono, M. (2009, January). Dynamic Returns Linkages and Volatility Transmission between South African and World Major Stock Markets. *Grahamstown: Rhodes University, 33*, 69-94.

Crouhy, M., & Rockinger, M. (1998, August). Volatility Clustering, Asymmetry and Hysteresis in Stock Returns: International Evidence. *Social Science Research Network, G14, G15*, 1-35.

Dickey, D., & Fuller , W. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica, Vol. 49 (No. 4)*, 1057-1072.

Engle, R. (2000). GARCH 101: An Introduction to the Use of ARCH/HARCH models in Applied Econometrics. *Stern School of Business, New York University*. Retrieved from https://www.stern.nyu.edu/rengle/GARCH101.PDF

Eriksson, P., Forsberg, T., & Gustavsson, N. (2011). *Fundamental Stock Analysis.* Jonkoping International Business School, Jonkoping University.

Fama, E. (1965). The Behaviour of Stock Market Prices. *The Journal of Businesses, 38*, 34-105.

Fedorova, E., & Saleem, K. (2009, August). Volatility Spillovers Between Stock And Currency Markets:Evidence From Emerging Eastern Europe. *School of Business, Lappeenranta University of Technology, 22nd Australasian Finance and Banking Conference 2009*.

Fishburn, P. (1977). Mean-Risk Analysis with Risk Associated with Below-Target Returns. *The American Economic Review, 67 (2)*, 116-126.

Forooha, R. (2009). *Power Up.* The Daily Beast. Retrieved April 29, 2019, from http://www.thedailybeast.com/newsweek/2009/03/20/power-up.html

Giot, P. (2005). Relationships Between Implied Volatility Indices and Stock Index Returns. *Journal of Portfolio Management, 31*, 92 - 100.

Goyal, A. (2000, May). *Predictability of Stock Return Volatility from GARCH Models.* University of Los Angeles, Anderson Graduate School of Management,.

Graham, M., Kiviaho, J., & Nikkinen, J. (2012). Integration of 22 emerging stock markets: a three-dimensional analysis. *Global Finance Journal, 23(1)*, 34-47.

Grater, E., & Struweg, J. (2015). Testing Weak Form Efficiency in the South African Market. *Journal of Economic and Financial Sciences, 8*, 621 - 632.

Griffoen, G. A. (2003). Technical Analysis in Financial Markets. *Journal of Finance*. Retrieved from https://ssrn.com/abstract=566882 or http://dx.doi.org/10.2139/ssrn.566882

Grubel, H. (1968). Internationally diversified portfolios: welfare gains and capital flows. *The, 58(5)*, 1299-1314.

H2O.ai. (2017, June 9). *Splitting Datasets into Training/Testing/Validating*. Retrieved from H2O.ai: https://h2o-release.s3.amazonaws.com/h2o/rel-vajda/1/docs-website/h2o-docs/data-munging/splitting-datasets.html

Hansen, P., & Lunde, A. (2001). A Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)? *United States: Brown University*.

Huang, W., Nakamori, Y., & Wang, S. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research, 32(10)*, 2513-2522.

Jeyanthi, B., & Annapakiam, J. (2010). Interdependence And Volatility Spillovers Under Market Reforms: The Case Of National Stock Exchange. *International Business & Economics Research Journal, 9*, 77-87.

Kara, Y., Boyacioglu, M., & Baykan, O. (2011). Predicting direction of stock price index movement using artifcial neural networks and support vector machines: Thesample of the istanbul stock exchange. *Expert systems with Applications, 38(5)*, 5311-5319.

Karlsson, L. (2002). Theoretical Survey. *Model Implementation and Robustness Analysis*.

Khemchandani, R., Chandra, S., & Et al. (2009). Knowledge based proximal support vector machines. *European Journal of Operational Research, 195(3)*, 914-923.

Kruger, R. (2011). Evidence of Return Predictability on the Johannesburg Stock Exchange (Doctoral Dissertation). *University of Cape Town, Cape Town*.

Leland, H. (1999). Beyond Mean-Variance: Performance Measurement in a Nonsymmetrical World. *Financial Analyst Journal, January/February*, 29-36.

Lo, A., Mamaysky, H., & Wang, J. (2002). Foundations of Technical Analysis: Computational Algorithms, Statistical Inference and Empirical Implementation. *The journal of Finance, Volume 55*(Issue 4).

Longin, F., & Solnik, B. (1995). Is the correlation in international equity returns constant: 1960 - 1990? *Journal of international money and finance, 14(1)*, 3 - 26.

Loretan, M. (2000, June). Speacial Feature: Evaluating changes in correlations during periods of high market volatility. *BIS Quarterly Review, Board of Governors of the Federal Reserve System International Finance Working Paper No. 658*.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance, 7(1)*, 77 - 91.

Markowitz, H. (1959). Portfolio Selection: efficient diversification of investments. *New York: John Wiley & Sons Inc.*

Markowitz, H. (1991). Foundations of Portfolio Theory. *The Journal of Finance, The Journal of American Finance Association*.

Marwala, L. R. (2017). *Forecasting the Stock Market Index Using Artificial Intelligience Techniques.* University of the Witwatersrand, Johannesburg.

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

Noakes, M., & Rajaratnam, K. (2016). Testing Market Efficiency on the Johannesburg Stock Exchange using the Overlapping Serial Test. *Annals of Operations Research, 243*, 273 - 300.

Nwogugu, M. (2006). Further Critique of GARCH Stochastic-Volatility Models and related approaches. *Applied Mathematics and Computation, 182(2)*, 1735-1748.

Ou, P., & Wang, H. (2009). Prediction of stock market index movement by ten data mining techniques. *Modern Applied Science, 3(12)*, 28.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machines learning techniques. *Expert Systems with Applications, 42(1)*, 259-268.

Phiri, A. (2015). Efficient Market Hypothesis in South Africa: Evidence from Linear and Non-Linear Unit Root Tests. *Managing Global Transitions, 13*, 369.

Pyo, S., Lee, J., Cha, M., & Jang, H. (2017). Predictability of Machine Learning Techniques to Forecast the Trends of Market Index Prices: Hypothesis Testing for the Korean Markets. *Plos ONE 12(11): e0188107*. Retrieved from https://doi.org/10.1371/journal.pone.0188107

Raj, J., & Dhal, S. (2008). Integration of India's stock market with global and major regional markets. *Press & Communications CH 4002 Basel, Switzerland: 202*.

Rangvid, J. (2001). Increasing convergence among European stock markets? A recursive. *Economics Letters, 71(3)*, 383-389.

Reilly, F., & Brown, K. (2002). *Investment Analysis and Portfolio management.* 7th Ed. South-Western.

Roman, V. (2018, December 23). *How To Develop a Machine Learning Model From Scratch*. Retrieved from Towards Data Science: https://towardsdatascience.com/machine-learning-general-process-8f1b510bd8af

Sakthivel, P., Bodkhe, N., & Kamaiah, B. (2012). Correlation and Volatility Transmission across International Stock Markets: A Bivariate GARCH Analysis. *International Journal of Economics and Finance, 4(3)*, 253-264.

Samouilhan, N. (2006). The Relationship Between International Equity Market Behaviour and the JSE. *University of Cape Town, Working Paper number 42*.

Seetharam , Y. (2016). The Dynamics of Market Efficiency: Testing the Adaptive Market Hypothesis in South Africa.

South Africa Online. (2018, February 26). *About South Africa; Economy*. Retrieved April 2019, from South Africa Online: http://www.southafrica.co.za/about-south-africa/economy/

Tobin, J. (1958). Liquidity Preference as Behavior Towards Risk. *The Review of Economic Studies, 25(2)*, 65-86.

Vapnik, V. (1998). Statistical Learning Theory. *networks, IEEE Transactions on neural, Wiley, New York, 10*(5).

Vapnik, V. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks, 10(5)*, 988-999.

Wafi, A., Hassan, H., & Mabrouk, A. (2015). Fundamental Analysis Models in Financial Markets – Review Study. *Procedia Economics and Finance*.

Whitefoot, J. (2017). *The Dotcom Bubble Crash Explained in a Timeline*. Retrieved April 2019, from The Lombardi Letter: https://www.lombardiletter.com/the-dotcom-bubble-crash-a-timeline/9251/

Zhang, Y. (2009). Linkages of Stock Prices in Major Asian Markets and the United States. *International Conference on Econometrics and the World Economy, Fukuoka University, Fukuoka, Japan, Center for Advanced Economic Study*. Retrieved April 2019, from http://www3.econ.fukuokau.ac.jp/english/researchcenter/paper24/11_Zhang.pdf

Appendices

Appendix A

Critical values for the Augmented Dickey-Fuller test

	With Out Trend		With Trend	
Sample Size	**1% Confidence**	**5% Confidence**	**1% Confidence**	**5% Confidence**
T = 25	-3.75	-3	-4.38	-3.6
T = 50	-3.58	-2.93	-4.15	-3.5
T = 100	-3.51	-2.89	-4.04	-3.45
T = 250	-3.46	-2.88	-3.99	-3.43
T = 500	-3.44	-2.87	-3.98	-3.42
T = ∞	-3.43	-2.86	-3.96	-3.41

Appendix B

Python code used to identify the most efficient support vector kernel to use. The RBF kernel was chosen.

```
import csv
import numpy as np
from sklearn.svm import SVR
import matplotlib.pyplot as plt
dates = []
prices = []
def get_data(filename):
    with open(filename, 'r') as csvfile:
        data = csv.reader(csvfile)
        next(data)
        for row in data:
            dates.append(int(row[0].split('/')[2]))
            prices.append(float(row[4]))
    return
def predict_price(dates, prices, x):
    # convert the matrix to [n][1]
    dates = np.reshape(dates, (len(dates), 1))
    svr_lin = SVR(kernel = 'linear', C = 1e3)
    svr_poly = SVR(kernel = 'poly', C = 1e3, degree = 2)
    svr_rbf = SVR(kernel = 'rbf', C = 1e3, gamma = 0.1)
    svr_lin.fit(dates, prices)
    svr_poly.fit(dates, prices)
    svr_rbf.fit(dates, prices)
    plt.scatter(dates, prices, color = 'black', label = 'Data')
    plt.plot(dates, svr_lin.predict(dates), color = 'red', label = 'Linear Model')
    plt.plot(dates, svr_poly.predict(dates), color = 'green', label = 'Polynomial Model')
    plt.plot(dates, svr_rbf.predict(dates), color = 'blue', label = 'SVR Model')
    plt.xlabel('Date')
    plt.ylabel("Price")
    plt.title("Support Vector Regression (SVR)")
    plt.legend()
    plt.show()
    return svr_lin.predict(x)[0], svr_poly.predict(x)[0], svr_rbf.predict(x)[0]
get_data('ALSI.csv')
predicted_price = predict_price(dates, prices, 20)
```

Appendix C

Python code used to clean and train the data, predict future prices, commute the model accuracy and graph the necessary results.

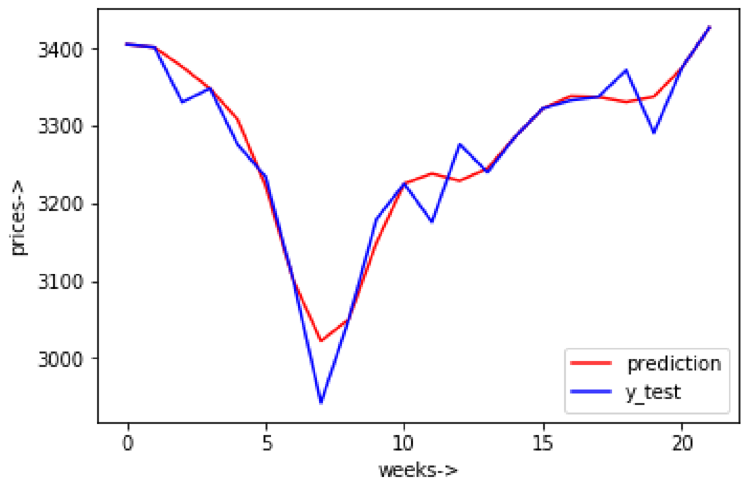
```
import csv
import numpy as np
from sklearn.svm import SVR
import matplotlib.pyplot as plt
sopening = []
sclose = []
sx=[]
sa=[]
sb=[]
count=0
original=0
def get_data(filename):
        with open(filename, 'r') as csvfile:
                csvFileReader = csv.reader(csvfile,delimiter=';')
                next(csvFileReader)
                for row in csvFileReader:
                        sopening.append(float(row[1]))
                        sclose.append(float(row[4]))
        return
get_data('ALSI.csv')
def predict_price(sopening,sclose, y):
  sopening = np.reshape(sopening,(len(sopening), 1))
  svr_rbf = SVR(kernel= 'rbf', C= 1e3, gamma= 0.1)
  sx=[y]
  sx=np.reshape(sx,(len(sx),1))
  svr_rbf.fit(sopening,sclose)
  return svr_rbf.predict(sx)
for u in range(100):
  sa.append(predict_price(sopening,sclose,sopening[(99-u)]))
for u in range(100):
  sb.append(sclose[99-u])
for i in range(len(sa)):
  if (sa[i]>sb[i]):
    count=((sa[i]-sb[i])/sb[i])
    original=original+count
  if (sa[i]<sb[i]):
    count=((sb[i]-sa[i])/sb[i])
    original=original+count
t=100-((original/len(sb))*100)
print("Accuracy",t)
import matplotlib.pyplot as plt2
plt2.plot(sa,color='red', label='prediction')
```

```
plt2.plot(sb,color='blue', label='y_test')
plt2.xlabel('days->')
plt2.ylabel('prices->')
plt2.legend(loc='upper left')
plt2.show()
```

Appendix D

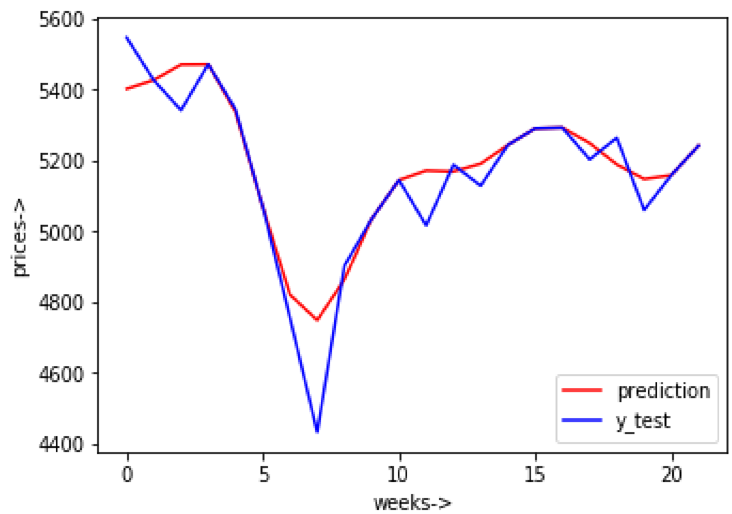
The prediction graphs of Australia, The U.K. and Netherlands in period 1.

Australian ASX 200 prediction (Period 1)



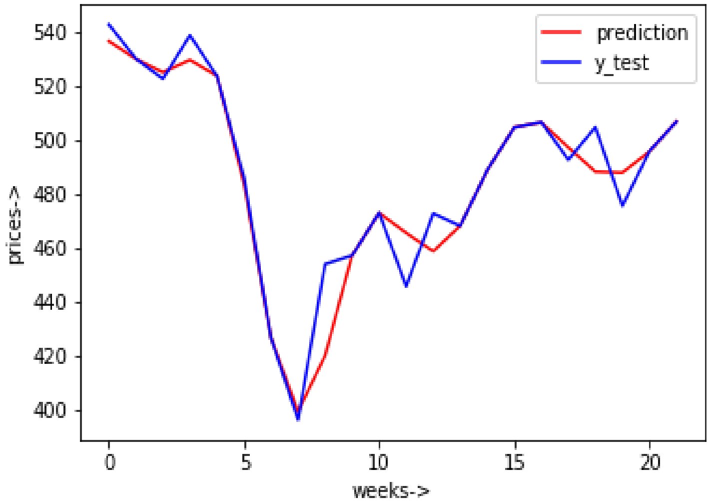
*** y_test refers to actual price data points*

The U.K. FTSE 100 prediction (Period 1)



*** y_test refers to actual price data points*

The Netherlands AEX prediction (Period 1)

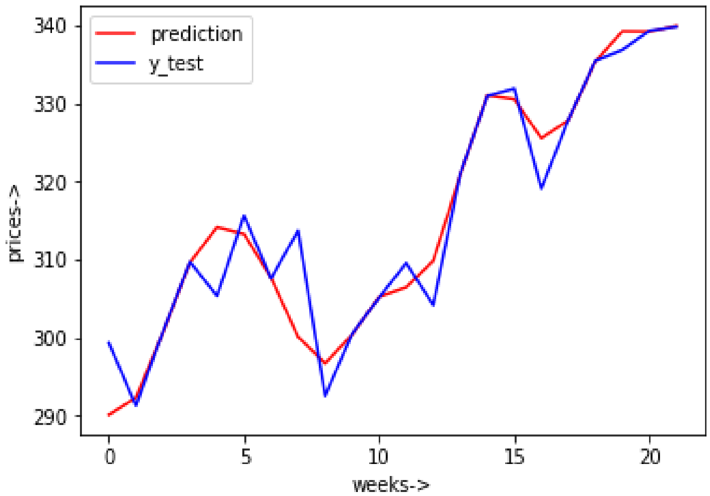


*** y_test refers to actual price data points*

Appendix E

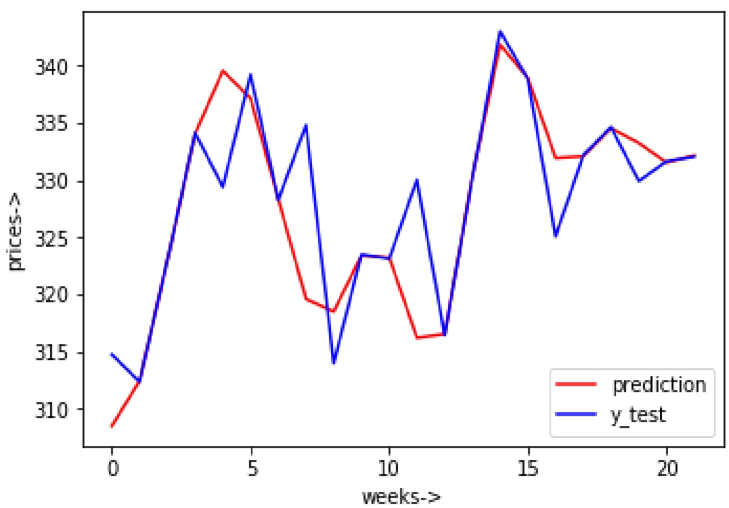
The prediction graphs of Germany, the Netherlands and France in period 2.

German CDAX prediction (Period 2)



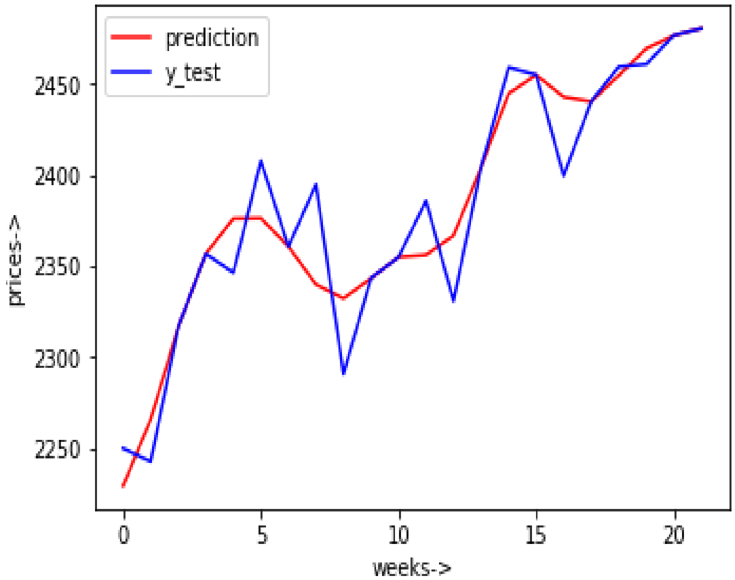
*** y_test refers to actual price data points*

The Netherlands AEX prediction (Period 2)



*** y_test refers to actual price data points*

French SBF 120 prediction (Period 2)

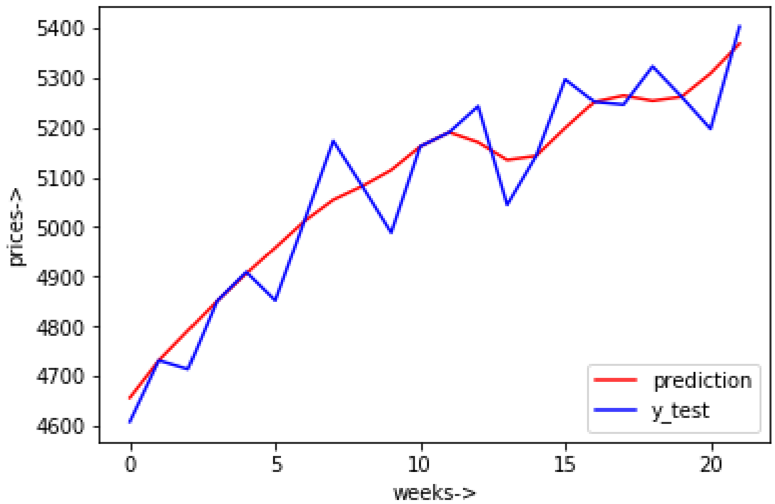


*** y_test refers to actual price data points*

Appendix F

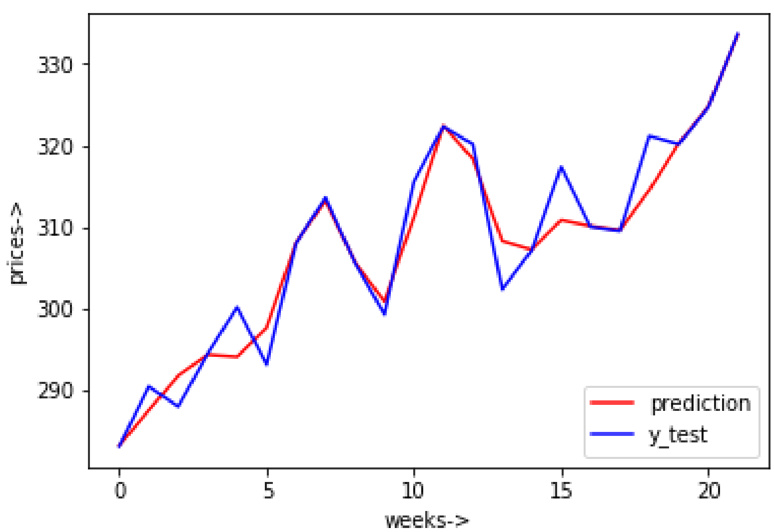
The prediction graphs of the U.K., the Netherlands and Australia in period 4.

The U.K. FTSE 100 prediction (Period 4)



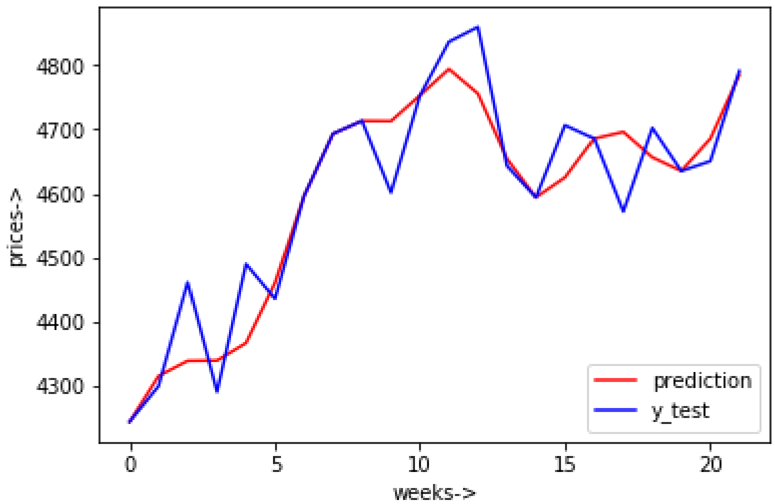
*** y_test refers to actual price data points*

The Netherlands AEX prediction (Period 4)



*** y_test refers to actual price data points*

Australian ASX 200 prediction (Period 4)

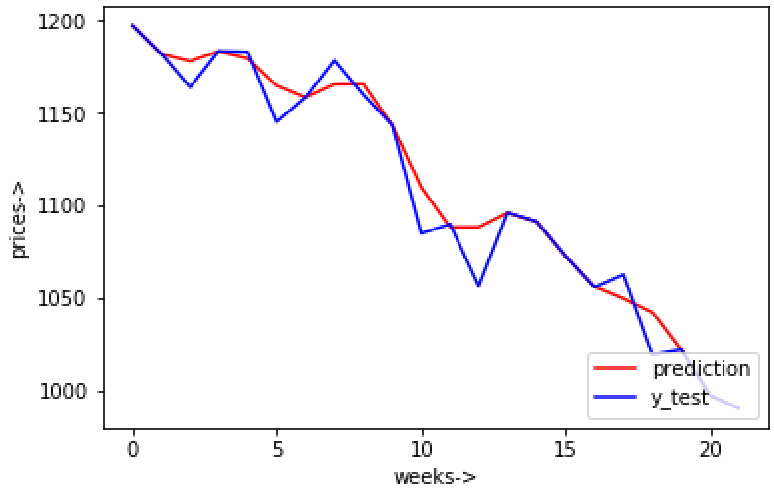


*** y_test refers to actual price data points*

Appendix G

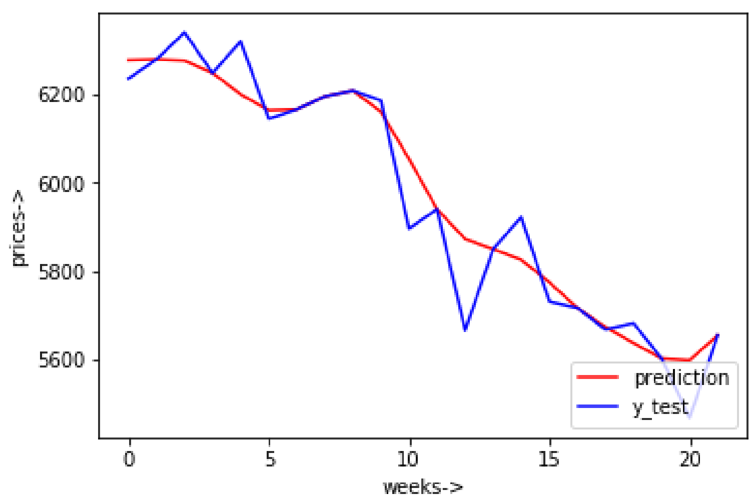
The prediction graphs of Germany, Australia and France in period 5.

German CDAX prediction (Period 5)



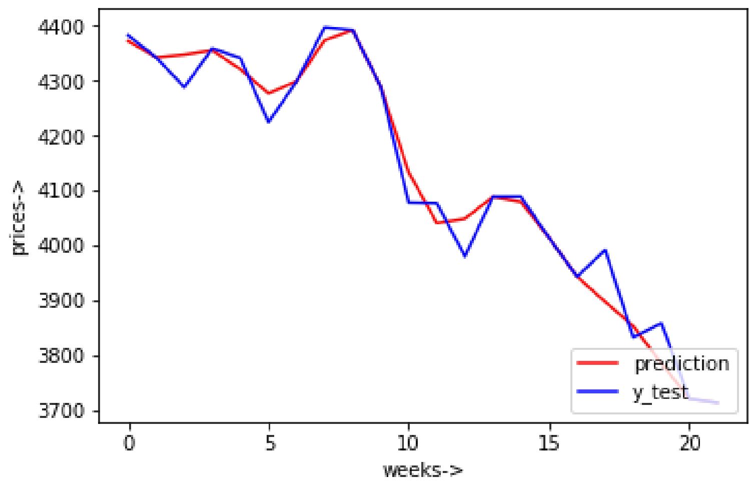
*** y_test refers to actual price data points*

Australian ASX 200 prediction (Period 5)



*** y_test refers to actual price data points*

French SBF 120 prediction (Period 5)



*** y_test refers to actual price data points*