Mean-Variance Optimisation of A South African Index Based Portfolio Using Machine Learning

Katlego Makgoale

WITS Business School

Thesis presented in partial fulfilment for the degree of Master of Business Administration to the Faculty of Commerce, Law, and Management, University of the Witwatersrand

April 2021

DECLARATION

I Katlego Makgoale declare that this research report entitled 'Mean Variance Optimisation of A South African Index Based Portfolio Using Machine Learning' is my own unaided work. I have acknowledged, attributed, and referenced all ideas sourced elsewhere. I am hereby submitting it in partial fulfilment of the requirements of the degree of Master of Business Administration at the University of the Witwatersrand, Johannesburg. I have not submitted this report before for any other degree or examination to any other institution.

Katlego Makgoale

Signed at Port Elizabeth on 30th April 2021

Name of candidate	**Katlego Makgoale**
Student number	**0511394y**
Telephone numbers	**+27 68 013 6993**
Email address	**0511394y@students.wits.ac.za**
First year of registration	**2019**
Date of proposal submission	**15 January 2020**
Date of report submission	**30 April 2021**
Name of supervisor	**Sibanda Jakubose**

ABSTRACT

Author: Katlego Makgoale **Supervisor** Sibanda Jakubose

Title: Mean-Variance Optimisation of A South African Index Based Portfolio Using Machine Learning

This study embarked on a comparison of the effectiveness of the Markowitz Mean-Variance Portfolio Optimisation against utilising a Machine Learning Technique to construct an optimal portfolio. The study aimed to: Construct an optimal portfolio using the Mean-Variance Analysis Framework, Construct an optimal portfolio using a Machine Learning Technique (Support Vector Regression), Contrast the results of the Minimum-Variance Portfolio and the Machine Learning Portfolio. The stocks of the FTSE JSE FIN15 index were chosen to construct the portfolio. The historical returns of the stocks in the index were used to trained (December 2014 to June 2019) and test the models(June 2019 to December 2020). The Mean-Variance Analysis and Minimum-Variance Portfolio were constructed using Python code that the author compiled. Similarly, the Support Vector Regression model was built in Python. The weights for the Machine Learning portfolio were calculated using the pseudo-inverse matrix and the predicted value of the Regression Model.

It was found that the Minimum-Variance and Machine Learning portfolio produced different portfolios, but both containing fewer holdings than the original index. The performance of the Minimum-Variance Portfolio exceeded that of the index and the Machine Learning Portfolio with regards to relative(excess) returns and total returns in the out of sample period. It was found that the Machine Learning portfolio performs well at replicating the index returns but fails to exceed them and typically has a higher risk associated with it. It was concluded that the Minimum-Variance portfolio would be the most attractive to a risk-averse investor and the Machine Learning portfolio underperforms the Minimum variance and the index. Therefore confirming the effectiveness of Mean-variance Optimisation in a South African context against a Machine Learning Technique.

Keywords: Portfolio Optimisation, Machine Learning, Support Vector Regression, Johannesburg Stock Exchange, Python

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF TABLES AND FIGURES IN THE APPENDICES

Figure 1: Screenshot of Yahoo Finance accessed: 15 Jan 2021 54
Figure 2: Screenshot of EquityRT accessed via Wits proxy server access date: 1 Jan 2021 54

ACKNOWLEDGEMENTS

Dedicated to my Wife who held my hand through the doubts and late nights. To my Parents for their continuous support, oh and you too Kani and Special K.

Special thanks to my supervisor as well for accommodating me under exceptional circumstances.

DEFINITION OF KEY TERMS AND CONCEPTS

Term	Definition
Active Portfolio Strategy	An investment strategy where a portfolio manager actively seeks excess returns by transacting in financial instruments in order to 'beat the market'.
Python	A high-level programming language
Portfolio	An assembly of financial assets held by an individual, which is constructed in accordance with investment strategy objectives.
Efficient Frontier	A visual representation of the risk vs reward spectrum that a set of portfolios exists for which the highest expected return, at constant risk, can be found.
Optimal Portfolio	Also known as an efficient portfolio maximises an risk averse investor's returns with respect to return.
Passive Portfolio Strategy	A strategy whereby shares are bought and held in order to replicate the long term behaviour of an index.
Returns	The change in price, measured as a %, of a share relative to the purchase or reference price.
Portfolio Volatility	Degree of variation of returns of a portfolio, calculated using the standard deviation.
Index	A selection of financial instruments that is used as a tool to benchmark investments
Risk Free Rate	Rate of return of an investment with 'zero' risk. Normally inferred to be a government bond yield.
FTSE / JSE Fin 15	The largest 15 South African Financials ranked by net market cap, that are classified as SA Financials
Exchange Traded Fund	A tradeable financial asset that is designed to mimic/track an index, sector or commodity.

•

1. INTRODUCTION TO THE RESEARCH

1.1 Context and background to the study

More generally, this research evaluates the Mean-Variance Optimisation of a South African Index Based Portfolio Using Machine Learning. However, before getting to the research conceptualisation (Section 1.2), we briefly introduce the terms and concepts that we have used in conceptualising this research in Section 1.1 generally and broadly—while Chapter 2 has a more specific and detailed discussion on the research context. The research conceptualisation section provides for the research problem statement (Section 1.2.1) and consequently the purpose of this research (Section 1.2.2) as well as the research questions (Section 1.2.3). The delimitations and assumptions of the research study are in Section 1.3 while we discuss the significance of the research study in Section 1.4 and provide a preface to the research report in Section 1.5.

1.1.1 Quantitative Portfolio Management and Optimisation

The primary goal of investors is to earn a return on their investment. For portfolio managers, the goal is to achieve portfolio excess returns for a given level of risk to entice investors. Portfolio managers have long developed tools and methodologies to aid them in forecasting, analysing broad market trends, calculating valuations, and identifying mispriced securities (Lombard, 2015). There has been a continual drive to improve and formalise investment management techniques as applied by portfolio managers. One of the first formalisations was proposed by Benjamin Graham and David Dodd in their 1934 book Security Analysis, in which a quantitative approach to investing was introduced. Fundamental Investing promised safety of capital and 'an adequate return' (Graham, Dodd & Cottle, 1934).

Markowitz (1952) in his ground-breaking 1952 doctoral thesis, developed a methodology that differed and arguably gave improved result to that of Graham and Dodd. Markowitz developed a mathematical analysis framework that could be applied to investment management. The framework maximised a portfolios expected return for a given level of risk. The framework, developed by Markowitz, seeks to optimise a portfolio of risky assets (stocks) to achieve maximum portfolio returns for minimum risk. This method is called Mean-Variance Analysis Framework, and its output is the Minimum-Variance Portfolio. The Mean-Variance Analysis Framework has, over the decades, become one of the fundamental tools for Active Portfolio Managers. Although it has become popular over the decades, Mean-Variance Analysis Framework has limitations that impede universal application to all risky assets.

As portfolio managers seek to improve their performance, new techniques are being applied. Recently there have been advances and increasing interest in the application of Machine Learning in Finance. Machine Learning has seen practical application and use cases of late due to advances in computing power, the sharing of information, and the ability to rapidly extract, transform, and analyse vast amounts of data. Investigations of the applicability of Machine Learning continue to be made as the field of investing evolves. Particularly in use cases where Mean-Variance Analysis is weak (Emerson, Kennedy, O'Shea & Brien, 2019).

1.1.1 An introduction to formative evaluation

Quantitative approaches to investment management continue to evolve. It becomes essential to draw a comparison as to which technique (Mean-Variance Analysis Framework or Machine Learning) is best suited to any particular investment goal, such as maximising returns and outperforming an index. Although research has been conducted on the performance of Machine Learning against a 'classical' technique such as Mean-Variance Analysis, there is very little research conducted on comparing the two techniques on a South African index on the JSE. It is against this backdrop that this study will investigate the effectiveness of the Mean-Variance Analysis Framework against that of Machine Learning.

1.2 Research conceptualisation

1.2.1 The research problem statement

Portfolio Management is a challenging profession, even though many highly skilled professionals make up the industry. Investment risk and return can be defined as stochastic variables. Stochastic variables are best described by random probability distributions, such as the normal distribution (Bodie, Kane, & Marcus, 2014).

One problem within portfolio management is optimisation. Wessels and Krige (2005) confirmed that when compared against one another, Passive strategies (Index Tracking Exchange Traded Funds) outperformed the average of Active Portfolio strategies. They confirmed this even when transaction costs were factored into the comparison. Oldham and Kroeger (2005) found that only four funds of the twenty analysed were able to generate returns above the JSE All Share Index in the five year periods analysed. However, an investor seeking to achieve returns above the market return will find appeal in an Actively Managed Portfolio. The problem can therefore be phased as "It is very difficult to consistently beat the market".

Bodie et al (2014) explain that it is difficult to consistently outperform Passive Instruments (Index Tracking Funds) due to the Efficient Market Hypothesis. The Efficient Market Hypothesis states that the prices of risky securities fully reflect available information, both public and insider information. The Efficient Market Hypothesis implies that any attempts to gain excess returns from the market will effectively be a game of chance rather regardless of the investor's skills (Lombard, 2015; Zafar, 2012). Therefore, the Efficient Market and portfolio optimisation challenge prompts the portfolio manager to optimise their portfolio and continuously evaluate their investments effectively. The portfolio managers goals should be to identify and scrutinise underperforming portfolio assets that hinder their investment goals, such as beating the market. Inefficient portfolio optimisation can produce poor returns when underperforming assets are held too long.

Without a robust quantitative analysis technique such as the Mean-Variance Analysis Framework or Machine Learning, portfolio managers can fail to achieve their investment goals or objectives. The use of formal portfolios optimisation techniques such as the Mean-Variance Analysis Framework or Machine Learning also allows for efficient diversification. Without efficient diversification, investors rely on stocks in their portfolio that introduce higher than tolerable levels of risk and impact the portfolio's financial stability, leading to poor returns. This can be exacerbated during period of financial crisis or systematic shock (Cizauskas & Haslifah, 2019).

1.2.2 The Research Question

As quantitative investing and Machine Learning evolves, ease of mitigating or reducing the effect of the problems listed in Section 1.2.1(poor returns, instability of returns, inefficient diversification) become more of a pressing need to solve. It is with these problems in mind that the research questions of this study are:

- What is the optimal portfolio using the Mean-Variance Analysis Framework on the FTSE / JSE Financial 15?
- What is the optimal portfolio using the Machine Learning Technique, Support Vector Regression, on the FTSE /JSE Financial 15?
- How do the results from the Mean-Variance Analysis Framework compare to the results from the Support Vector Regression?

1.2.3 The Research Purpose (aim and objectives) statement

This study aims to compare the results of the Mean-Variance Analysis Framework and the Support Vector Regression in constructing optimal portfolios.

The objectives of the study have been framed into:

1. Construct an optimal portfolio using the Mean-Variance Analysis Framework.
2. Construct an optimal portfolio using a Machine Learning Technique (Support Vector Regression).
3. Contrast the results of the Minimum-Variance Portfolio and the Machine Learning Portfolio.

Two portfolio construction methods will be compared. The first will be that which Harry Markowitz initially devised (H. Markowitz, 1959). Arithmetic means of stock returns are used as expected returns to build a Minimum-Variance Portfolio. The second portfolio construction method will utilise the Support Vector Regression Model to compute the expected returns and then construct an optimal portfolio based on that prediction.

Both methods will seek to replicate or beat the benchmark index, the FTSE / JSE FIN15. The research will undertake an experimental line of inquiry. Price data will be collected from December 2014 to June 2019 to prepare and train the Mean-Variance Analysis Framework and the Support Vector Regression Model, respectively. Two portfolios will then be constructed from the stocks in the FTSE / JSE FIN 15. The performance of the Minimum-Variance Portfolio and the Machine Learning Portfolio will be compared against the FSTE /JSE FIN 15 from June 2019 to December 2020.

The author intends to evaluate the: total, excess returns, risk and return of the two portfolios. A conclusion will be drawn on the results of both portfolios and what would be most enticing to a Risk-Averse investor. The research is intended to contribute to the work of investment practitioners.

1.2.4 The Research Hypotheses

The research hypotheses can therefore be stated as follows:

> The Machine Learning Technique (Support Vector Regression) can outperform the Mean-Variance framework Analysis in constructing optimal portfolios from the FTSE / JSE Financial 15. The portfolios constructed using the Machine Learning Technique will provide greater returns and lower risk than those optimised using the Mean-Variance framework.

1.3 Delimitations and assumptions of the research study

A critique of Active Portfolio Management comes about when an empirical comparison of the performance of both techniques is conducted. The empirical research, which was famously conducted by Fama and French (2000) showed that consistently predicting financial markets is difficult. Wessels and Krige (2005) also conducted a study in which South African mutual funds could not outperform Exchange Traded Funds(ETFs). The likes of the (Fama & French, 2021) have empirically confirmed that it is difficult for an Active Management Strategy to beat passive instruments consistently.

The Efficient Market Hypothesis has not been validated outright. However, it has also not been refuted as there are instances where neither Active nor Passive Management techniques show superiority on a risk-adjusted basis. Modern Portfolio Theory also has assumptions that must hold in order for the framework to be applied. Some of the simplifying assumptions are:

- Future returns can be predicted based on historical returns.
- Investors are risk-averse and seek to maximise returns and minimise risk.
- Short positions are not considered, only long positions.

1.4 Significance of the research study

The research conducted to date on Machine Learning Techniques applied to portfolio optimisation of South African is scant (Hurwitz & Marwala, 2009). This study will contribute further to the research body of knowledge and to investment practitioners seeking to outperform the Mean-Variance Analysis Framework on the FTSE / JSE FIN15 index.

1.5 Preface to the research report

To this end, the report has six chapters. Following this introductory chapter, Chapter 2 provides a literature review covering the problem, the past studies, the explanatory framework and the conceptual framework. Chapter 3 discusses the research strategy, design, procedures, reliability and validity measures, and limitations. Chapter 4 and presents and discusses the findings, respectively, to interrogate our research questions, while Chapter 5 summarises and concludes the research.

2. LITERATURE REVIEW

This chapter has three broad objectives; namely to understand the research problem, to identify the knowledge gap, and to develop a framework for interpreting the research findings. Specifically, in Section 2.1, we detail the research problem. In Section 2.2, we review literature on studies that have attempted a similar study or research. With information arising from Section 2.2, we identify and detail qualitative attributes or quantitative variables that are key to this research in Section 2.3 as well as a framework that we will use to interpret our research findings in Section 2.4.

2.1 Research problem analysis

2.1.1 Modern Portfolio Theory (history and evolution)

In 1952 Harry Markowitz developed his now-famous theory, Modern Portfolio Theory, to select a portfolio best suited for the risk-averse investor (H. Markowitz, 1959). Within Modern Portfolio Theory, Markowitz described the Mean-Variance Analysis Framework as a method of constructing the optimum portfolio (Minimum-Variance Portfolio). A sufficiently diversified portfolio of assets seeks to maximise expected returns based on a level of tolerable risk. The volatility (standard deviation of expected returns) is used as a proxy for risk. The risk is an inherent aspect of an investment, and higher return premiums are generally proportional to higher risk. The risk-free rate is the return an investor can receive at zero risks (ZongMing, Koomson, & Guoping, 2017). Further contributions by William Sharpe in 1964, Fischer Black in 1972 and John Litner in 1965 have come to shape the way academics and practitioners think about average returns and risks of stocks(Fama & French, 2021). Modern Portfolio Theory is still widely applied today.

2.1.1 Mean-Variance Analysis

The most crucial goal of the Modern Portfolio Theory approach to investment management is to maximise returns on a risk-adjusted basis. The central tenant of Harry Markowitzs' approach to portfolio optimisation was simultaneously optimising a portfolios expected return (mean historical returns) and minimising the portfolios' investment risk. Markowitz presented Mean-Variance Analysis as the central framework of portfolio optimisation.

The popularity of Mean-Variance Analysis is partly due to the mathematical nature of the problem of portfolio optimisation. Portfolio Optimisation is essentially a mathematical optimisation problem and can be solved using Quadratic Programming or Convex Function optimisation. Convex optimisation problems are mathematical problems that are well understood and relatively easy to solve. A vast body of knowledge on solving convex optimisation problems and techniques has become more and more refined with advances in computing. This is why the Mean-Variance Analysis Framework is a near-universal method used by all portfolio managers (Martin, 2018; Perrin & Roncalli, 2020).

2.1.1 Mean-Variance Analysis Limitations

The critiques and limitations of Mean-Variance Analysis lie in the base assumptions at the heart of the technique. Mean-Variance Analysis estimates of risk are only appropriate when returns are normally distributed. Ilmanen (2011) has shown that traditional investments such as stocks or bonds exhibit returns that can be approximated with a normal distribution. However, more sophisticated financial instruments such as derivatives or hedge funds do not exhibit this characteristic. Mean-Variance as it was formulated only works when returns are assumed to be normally distributed.

Du Plessis and Ward (2009) explained in their study the criticisms and shortcomings of Mean-Variance even though it is such a popular technique of investment management. Although mathematically, the Mean-Variance Portfolio produces efficiently diversified portfolios, it has been criticised that the portfolios are unrealistic and may not be practical in their implementation. Mean-Variance Analysis also requires large volumes of data over distinctive periods to be provided to estimate the covariances. Finally, most investment companies are not organised in a manner in which the Mean-Variance Analysis Framework can be used.
Ivanova and Dospatliev (2017) also listed some further limiting assumptions of the Mean-Variance Analysis Framework discussed by Markowitz. The use of historical returns in estimating the probability distribution of possible future returns can be estimated by investors, meaning past performance can be used to predict future performance. Volatility (standard deviation) about the mean value of possible returns can be used to measure risk. Returns are desirable, and risks are to be avoided, and the financial market is perfectly liquid. These assumptions deviate from the behaviour of real financial markets.

In their study, Fama and French (2021) also added to the criticisms of Mean-Variance Analysis. Their results showed that stock price risks are multidimensional rather than only being a function of the dispersion of historical returns. They established that one dimension of risk could be described by the stocks market capitalisation (share price multiplied by outstanding shares). The second dimension of risk can be described by the ratio of a stocks book value (company value according to its reported balance sheet) and market capitalisation. Therefore, the use of dispersion of historical returns of a stock simplifies the possible actual extent of risk.

2.1.2 Machine Learning, the beginnings and a history

Machine learning is a subset of the field of Artificial Intelligence. Within the fields of Machine Learning are Supervised, Unsupervised and Reinforcement Learning. The discipline of Machine Learning has increased connectivity and has driven innovation and automation in the delivery of traditional tasks and services, which has benefited the investment industry. The development of easy to use programming languages, such as Python, has also enabled the introduction of Machine Learning into industrial applications such as Investment Management (Emerson et al., 2019).

Machine Learning has become a topic of interest within organisations seeking to leverage available data to gain a new understanding or leverage performance. Concerning investors and portfolio managers, Machine Learning, when used appropriately, promises to enable the investor to predict the future based on complex and hidden patterns in the data (Hurwitz & Kirsch, 2018).

2.1.3 Machine Learning and Portfolio Optimisation

Block (2017) examined how to perform an optimal diversification of a South African Portfolio consisting of equities from the Johannesburg Stock Exchange. A framework utilising a Machine Learning Simulation, specifically genetic programming and particle swarm optimisation, was applied. It was able be shown that the construction of an optimal portfolio is possible. The Machine Learning optimisation technique reduced the total number of shares in the portfolio down to only 15 Shares, which demonstrates efficient diversification. It was also noted that the portfolios that were indicated as being optimally diversified were able to attain market-beating returns. The equities analysed by Block (2017) consisted of a wide range of different sectors of the Johannesburg Stock Exchange. They were not focused on any particular Index or Sector, such as the Financial 15.

Hurwitz and Marwala (2009) studied the application of Machine Learning to optimise a portfolio with targeted risk/return. Their study used the Capital Asset Pricing Model developed by William Sharpe, Jack Treynor, John Littner and Jan Mossin as the portfolio optimisation methodology. The portfolio optimisation was executed by a Machine Learning algorithm (genetic programming) which managed the portfolio over time to achieve a targeted risk/return for the investor. The results showed that the developed system was capable of nominally efficient portfolio results but not truly efficient. Exploration of alternative optimisation techniques was recommended.

Ban, El Karoui, and Lim (2018) conducted a study. Two Machine Learning techniques were adapted (Performance Based Regularisation and Performance-based cross-validation) to apply Portfolio Optimisation. They introduced performance-based regularisation, a technique used to improve the stability of solutions to statistical problems, in their study, with the objective to constrain the quantities of portfolio risk and return. Performance-based regularisation was applied to the Mean-Variance and also the Mean-Conditional Value at Risk. They concluded that Performance-Based Regularisation showed some promise when applied to portfolio optimisation, which can be categorised as a decision problem. The effectiveness of their approach was limited to sample data and not confirmed against out-of-sample data, meaning that the results could not be generalised to out of sample data.

2.1.4 The Efficient Market

One of the fundamental tenants and challenges to a portfolio manager is the Efficient Market Hypothesis. First put forward by Fama (1960). It states that an investor will rarely find bargains or mispriced stocks in the markets and that stock prices will adjust quickly to the available information in the market. Studies have been conducted on the efficiency of the stock market. Vitali and Mollah (2010) tested the weak form of the Efficient Market Hypothesis on numerous South African indices. They found that the weak form of the Efficient Market Hypothesis could not be rejected outright for 2007-2009. Empirical confirmation of the Efficient Market Hypothesis has implications for the portfolio manager. Even with a portfolio optimisation technique at the portfolio managers fingertips (Mean-Variance Analysis or Machine learning), consistently outperforming passive instruments such as Exchange Traded Funds is more down to chance than the technical skill of the investor. It also means then a Passive Management strategy rather than an Active Management strategy would provide more consistent returns in the long run (Bodie et al., 2014). As such, in stock markets, stock prices usually represent all the information available to investors concerned. This generalise the problem in 'outperforming' the market for the portfolio manager.

2.2 Research knowledge gap analysis

As a reminder of the objectives of the study in Section 1.2.2, this study aims to:

1. Construct an optimal portfolio using Mean-Variance Analysis Framework
2. Construct an optimal portfolio using a Machine Learning Technique (Support Vector Regression)
3. Contrast the results of the Minimum-Variance Portfolio and the Machine Learning Portfolio.

The literature that has been examined has shown examples of Mean-Variance and Machine Learning applications as applied to Portfolio Optimisation, but no examples of the optimisation techniques contrasted against each other on a Johannesburg Stock Exchange Index.

In section 2.2, the author will elaborate on the extent of the knowledge gap that this report aims to contribute.

Beasley (2013) conducted research in which the Mean-Variance optimisation was applied to different financial scenarios. The study intended to define how the portfolio could be best optimised mathematically for each investment scenario. The most exciting investment scenario applied was the Index, tracking where the objective is to match the return achieved on a benchmark, here being the S&P500 index. The other financial scenario was enhanced indexation, where the goal was to exceed the return achieved on a benchmark index by providing returns relative (excess returns) to the benchmark. However, the index replication problem presented in the paper was computationally challenging to solve as the problem had evolved into a nonlinear mixed-integer optimisation problem. In the example where enhanced indexation was the objective Beasley (2013) provided a formulation to develop a portfolio that could provide returns over the Index. However, there was no control over the level of outperformance of the portfolio relative to the Index. The study also took place on stocks only available on the US stock markets (S&P500 and Russel 3000).

Regarding the most suitable Machine Learning Technique for the application of portfolio optimisation, there exist many differences of opinion from the literature. A comparison of two Machine Learning Techniques was made by Yan, Sewell, and Clack (2008), where a comparison was made of Genetic Programming and Support Vector Machines. The study results showed that Genetic Programming was more robust in optimising profits, and the Support Vector Machine sought to replicate monthly returns. The study was conducted on derivatives (Contracts for Difference) on the Malaysian stock market, not stocks.

Emerson, Kennedy, O'Shea, and O'Brien (2019) conducted an extensive review of the recent literature around quantitative finance and Machine Learning applications to this field, refer to Figure 1 below. Their study was a survey of the literature. It was found that Machine Learning and, more broadly, Artificial Intelligence are applied to three disciplines within Quantitative investment: Return Forecasting, Portfolio Construction and Risk Modelling. Deep learning (a form of Artificial Intelligence) appeared multiple times in the literature. Nevertheless, Support Vector Machines appeared more when applied to Return Forecasting with fewer examples of Support Vector Machine applied to portfolio construction.

Figure 1: Popular techniques featured in machine learning finance papers(Emerson et al., 2019)

	MLP	SVM	LSTM	GRU	RNN	CNN	RF	GPR	LR
Return Forecasting	7	5	4	2	-	1	2	-	-
Portfolio Construction	7	2	3	1	1	1	4	2	1
Risk Modelling	6	2	2	1	1	1	4	3	4

MLP	Multilayer Perceptron
SVM	Support Vector Machine
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
RNN	Recurrent Neural Network (basic)
CNN	Convolutional Neural Network
RF	Random Forests/Decision Trees
GPR	Gaussian Process Regression
LR	Logistic Regression

The literature examined shows very few examples of Mean-Variance and Machine Learning (Support Vector Regression) being applied to construct an optimal portfolio on a South African Exchange. It is against this gap in the knowledge base that the author wishes to embark on this study.

2.3 Variables key to the research

Market Indices are derived from stock prices and are considered a gauge for activity levels in the market (Bodie et al., 2014). Market indices give an overall picture of the sector, commodity or overall economy they represent, and their value is calculated daily using an assortment of methods/techniques (Patel et al., 2015). The oldest and most famous market index is the Dow Jones Industrial Average (DJIA). The DJIA was first computed in 1896 as an average of the 30 constituent stocks' prices (Bodie et al., 2014). The DOW is what is labelled a price-weighted average.

The JSE calculates and reports the index values for the many Market indices reported on the exchange. A famous and frequently reported index on the JSE is the 'Financials'. The FTSE/JSE Financial 15(Code: J212) is described as 'The top 15 companies which are constituents of the financial, economic group ranked by full market capitalisation' (Carte, 2009).

An investor can trade in the index through an Exchange Traded Fund, a fund that trades as security and seeks to mirror a particular index's behaviour and performance. An Example is the Satrix FINI ETF, which is managed by

Sanlam Investment Management (Pty) Ltd. The fund seeks to replicate the FTSE/JSE Financial 15 Index's performance by investing in the stocks of companies as per their weightings in the Index (Satrix, 2020).

Figure 2: FIN15 Close Price from December 2014 to December 2020

There are many examples in the literature where indices have been used to test or benchmark the performance of the Mean-Variance Analysis Framework and the Machine Learning Technique.

Lombard (2015) applied Mean-Variance analysis to the Johannesburg Stock Exchange, more specifically to test the effectiveness of Mean-Variance as a passive investment strategy concerning the FTSE / JSE Top benchmark 40. The study was conducted over data spanning 19.5 years. The historical returns of the stocks within the index were analysed using mathematical formulas and statistical models to explain the relationship between the variables. The Mean-Variance Analysis Framework was conducted using the expected return of stock prices, stock volatility, portfolio and risk, and return to conduct the analysis.

Regarding machine learning portfolio optimisation, Ma, Han, and Wang (2021) conducted their study over nine years of historical data on the components stocks of the China securities 100 index. A fusion of machine learning techniques (i.e. Random Forests and Support Vector Regression) used to construct an optimal portfolio. The daily returns of each stock in the index were used to train and then validate the machine learning technique, constructing the portfolio. The variables that were key to measuring the performance of the different models as we applied were the Mean Square Error, Mean Absolute Error. These metrics were used to show the models predictive ability.

A study conducted on the stocks Sao Paolo Stock Exchange, Paiva, Cardoso, Hanaoka, and Duarte (2019) used a fusion of Machine Learning techniques for portfolio selection. The Machine Learning Techniques were the Support Vector Machine and the Mean-Variance Analysis Framework. The results of the constructed portfolios were compared against the performance of the Ibovespa's, which is the Brazilian market index.

2.3.1 Risk-Free Rate

The Risk-Free Rate is defined as the rate of return that an investor with no risk can obtain. Since the Risk-Free Rate can be obtained with no risk, any risky investment must produce higher returns than the Risk-Free rate to entice investors. One of the leading assumptions of Modern Portfolio Theory is that investors are risk-averse. Therefore, a portfolio with Minimal Variance (risk) will entice a risk-averse investor. The return on short-dated government bonds is usually considered as a good proxy for the Risk-Free Rate. The RSA 5-year government bond yield will be the proxy for the Risk-Free Rate in this study. Domestically held short-dated government bonds are generally perceived as a good proxy for the risk-free rate.

2.3.2 Adjusted Closing Price

Prices in financial markets are based on the perception of value by investors. The close price quoted at the end of the day on the news comprises multiple factors that influence it, such as dividend pay-outs, rights offerings, stock splits, and corporate actions (Scott, 2020). The Adjusted Close Price is commonly used (Glabadanidis, 2020). The adjusted closing price is a calculation adjustment made to a stock's closing price and represents the stock's actual value.

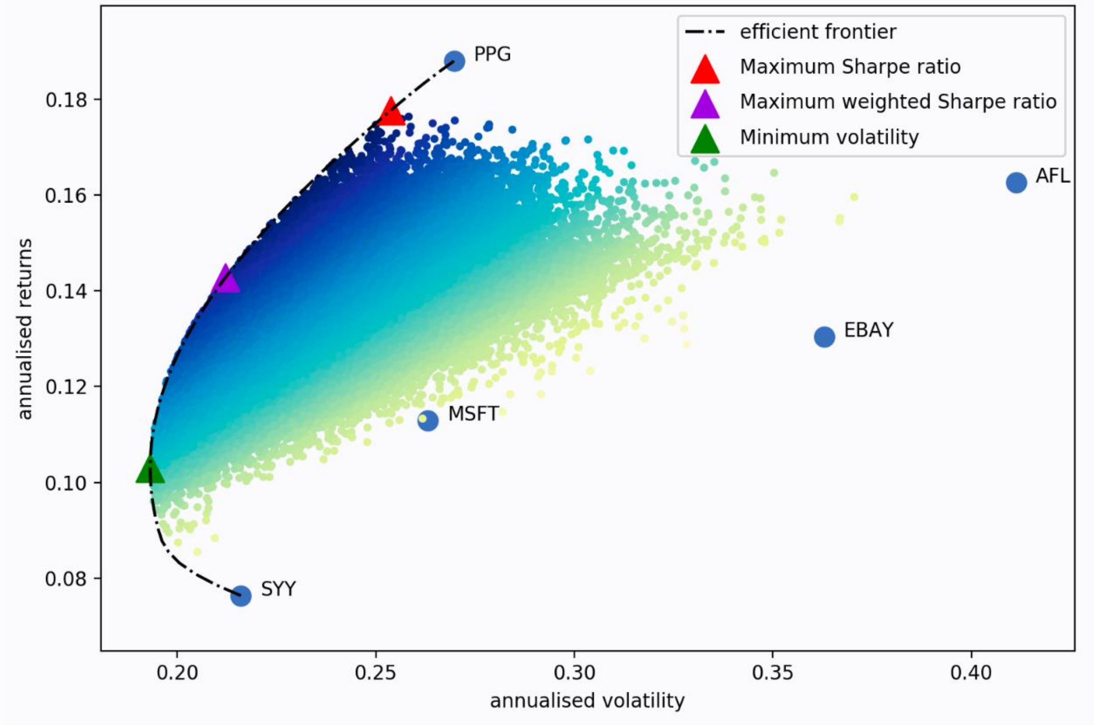
2.4 Framework(s) for interpreting research findings

2.4.1 Supervised Learning: Support Vector Machines

The Support Vector Machine is a classification technique that applies a transformation algorithm from the input space to a higher dimension to construct a boundary hyperplane. The Support Vector Machine helps classify problems by identifying extreme values and setting margins according to the extreme values of these data points. The decision boundary (boundary hyperplane). Support Vector Machines have been used in the financial industry for time-series prediction and financial analysis (Bahramy & Crone, 2013). Support Vector Machines support linear and nonlinear regression through Support Vector Regression. Support Vector Regression performs linear regression in higher dimensional spaces by trying to fit the error within the margin of the boundary. In Support Vector Regression, the goal is to ensure that the error does not exceed the threshold (Hastie, Tibshirani, & Friedman, 2009).

2.4.1 The Efficient Frontier

Figure 3: Representation of Efficient Frontier((Silva, 2020)



The risk-averse portfolio manager will take want to minimise the risk of their portfolio for their expected return. There are thousands of combinations of possible portfolios that can make up the range of possible investments to satisfy the portfolio managers investment goals. The Efficient Frontier, shown in figure 1, represents the spectrum of "possible " portfolios resulting from the collection of all possible combinations of securities available. The optimal or most "efficient" combination of securities falls along the efficient frontier. The Efficient Frontier uses two measures the expected return and the risk of the portfolio. In constructing the Efficient Frontier, volatility (standard deviation) indicates the amount of risk in the portfolio. Portfolios situated along the efficient frontier line represent the most "optimal" portfolio (Lombard, 2015).

2.4.2 Sharpe Ratio

In Investment Management the Sharpe Ratio is a measure of the reward to volatility of an stock or portfolio (Sharpe, 1994). The importance of the Sharpe Ratio suggests that the attraction of a portfolio can be measured so as to suggest to an investor which Investment is more in line with their investment objectives (Bodie et al., 2014; Jordan et al., 2015) of the trade-off between risk and reward

$$s_p = \frac{E[R_p - R_{rf}]}{\sigma_p}$$

Where R_p is the return of the portfolio.

Where R_{rf} is the Risk Free Rate.

Where $E[R_p - R_{rf}]$ is the expected value of the excess portfolio return.

Where σ_p is the standard deviation of the portfolio excess return.

2.4.1 Systematic and Firm Risk

The Beta value represents the amount of uncontrollable risk to which a stock is exposed. Beta is calculated as the ratio of stock and market covariance to the stock variance. The Beta value is dependent on a wide range of factors(Jordan, Miller, & Dolvin, 2015). Cyclical firms are prone to systematic risk due to stock performances moving with the changes in the market due to the business cycle (Van Rensburg, 2015). Variations in stock price due to the business cycle would typically imply a very high beta alongside companies that produce high priced consumer or industrial goods and luxury goods. Small market caps (less than R 1 Billion) typically would have significant Betas. However, the FTSE JSE FIN15 stocks would not apply here as all listed companies are medium to large caps and are the largest companies in the JSE's Financials category by market cap (Carte, 2009). A large market cap shields firms from high systematic risk. Firm-Specific Risks, or volatility are quantified as the standard deviations of the annualised return of each stock (the company).

2.5 Summary and conclusion

2.5.1 Summary of literature reviewed

What Chapter 2 has shown is that Mean-Variance Analysis Framework, as was developed by Markowitz, is still quite popular. However, there exists many variations and improvements to the fundamental theory. The shortcomings of Mean-Variance analysis come in the assumption of the normality of stock returns. Some assumptions of Mean-Variance Analysis Framework such as perfect liquidity of markets and how dispersion around the returns distribution can be used as an estimator of all of the stock risk when Fama and French (2021) showed with empire testing that risk appears to be multidimensional. As Machine Learning has advanced, it has been used in combination with the Mean-Variance Analysis Framework, as shown by Ban et al. (2018). There are recent examples of combinations of Machine Learning algorithms (Genetic Programming, Particle Swarm Optimisation) to portfolio optimisation.

However, the literature as was examined shows very few examples of Mean-Variance and Machine Learning applied to a South African index such as the Financial 15.

2.5.2 Proposed research strategy, design, procedure and methods arising from the literature reviewed

The research strategy to be employed in this study will be a quantitative study. The index returns will be evaluated against a Markowitz Minimum-Variance portfolio and a Machine Learning Portfolio based on a Support Vector Regression Model. The central research question will be evaluated against the empirical results generated over the study period. The research design will be that of an experimental study where the returns of the FIN15 are evaluated over five years. Research Data will be collected from the Yahoo Finance Application Programming Interface using the Python Programming Language. The data to be collected will be the daily financial information of the stocks for five years (Dec 2014 – Dec 2020). The target population will be the constituent shares in the FIN15.

3. RESEARCH STRATEGY, DESIGN, PROCEDURE AND METHODS

In Section 1.2.3, we have posed two questions that this research report intends to answer, that is, *'Can an optimised (Min Variance) and a Machine Learning portfolio produce returns above the FTSE/JSE Financial 15 index on which its stocks are based?'*,

We have since reviewed the literature and developed an interpretative and conceptual framework that will guide the techniques we will use. This chapter identifies and describes the research approach, design, and procedure and methods employed in this research to collect, process, and analyse empirical evidence.

Broadly, it has three objectives; namely, to identify and describe the research strategy (Section 3.1), the research design (Section 3.2), as well as the procedure and methods (Section 3.3). The chapter also describes the reliability and validity measures (Section 3.4) that this research applies to make it credible, as well as the technical and administrative limitations of the choices we make (Section 3.5).

3.1 Research strategy

Broadly the research strategy, also known as a research approach or a research paradigm, is the overall plan or structure for the methods the researcher will follow, the data collected, and the analyses conducted (Leedy & Ormrod, 2014). The research strategy can be categorised into three broad categories: qualitative, quantitative and mixed.

The strategy to be utilised in this study will employ quantitative research methodologies. The overall research strategy will use experimental methods that will identify relationships amongst the variables based on the experimental results and confirm or refute the research propositions stated in Section 1.2. The research will be conducted on share price information of the FIN15 constituents downloaded from Yahoo Finance databases. The indices' returns will be evaluated utilising the Mean-Variance Optimised portfolio. The expected returns for the Mean-Variance Optimisation process will be calculated using historical means and SVR. The research questions will be evaluated against the empirical results generated over the study period.

Yu, Chiou, Lee, and Lin (2020) recently conducted similar research in their journal submission to the International Review of Economics and Finance in 2020. The study is titled 'Portfolio Models with returns forecasting and transaction cost'. The research question being whether forecasting expected returns can improve the performance of a portfolio. The aim and objectives of the study were to examine and advance modern portfolio formulation methods such that expected and realised returns are maximised through the examination of different return forecasting techniques. The authors examine return forecasting techniques used widely in industry and academia, including Mean-Variance, Mean-Absolute Deviation, Downside Risk, Value-at-Risk, Conditional Value-At-Risk and Omega Ratio.

The benefit (Yu et al., 2020) from this approach to the research question was that the authors were able to directly address one of the critical concerns of Modern Portfolio Theory. One of the limiting assumptions of Modern Portfolio Theory being that future expected returns are based on historical returns. The out-of-sample performance of a portfolio can be affected by the weakness of the expected return prediction. A similar benefit of examining and comparing return forecasting applied to Modern Portfolio Theory is derived in this current research. The reason is that portfolio models are now being advanced by using machine learning techniques that analyse vast amounts of data to provide more robust expected return predictions.

3.2 Research design

A defining characteristic of academic research is the Research design. Research design is a structure and plan of action to conduct the research. With no step-by-step plan of action, the research author's thoughts and efforts cannot be laid down in a manner that will produce an output that contributes positively to the body of knowledge. Cooper, Schindler, and Sun (2003) define a research design as a "blueprint for collecting, measurement, and analysis of data". Bryman(2016) categorises research design into five generic types, which shall be explained briefly. A cross-sectional study requires collecting data on a sample at a particular moment to examine the data in connection with more variables and examine if relationships can be detected between the variables. A longitudinal study surveys a sample at multiple points in time, either as a panel or a cohort study. A case study involves the detailed scrutiny of a single case and examining the cases complicated and unique nature. A comparative study is a variation of a case study in that two contrasting cases are studied using identical methods. Finally, an experimental study gives the researcher power to produce effects in the variables under study. Fundamental is that a research design is a way that a researcher collects data.

This research study will utilise an experimental research design. The study will examine the effect of expected return prediction models (Arithmetic Mean vs Support Vector Regression) on the performance (Sharpe ratio and yield) of a portfolio of stocks from the FIN15. The comparison of performance will be made against the reference benchmark for each respective portfolio (FTSE/JSE Financial 15 index).

The following authors (Adams, Khan, Raeside, & White, 2007) state that an experimental study verifies or refutes knowledge through manipulation, testing and observation. Testing Modern Portfolio Theory on South African Financial instruments underpins the motivation to select an experimental research design. The justification for the selection of an experimental design is that the research questions must be tested. Mean-

Variance Optimisation and Support Vector Regression predictions for expected returns are applied to Modern Portfolio Theory. The effects are observed on the performance of the portfolios over the period.

A similar study was conducted by Ma, Han, and Wang (2021) in which the two machine learning models were utilised to enhance the mean-variance return prediction technique. The authors of the paper utilised the China securities 100 indexes as their sample dataset. They conducted experiments that tested the performance of the portfolio optimisation technique combined with the machine learning models. The reason for selecting the experimental research design is that the authors wished to test the performance of portfolio optimisation with machine learning for return prediction.

The study intends to test, observe and explain the results of the association between Machine Learning and Mean-Variance Analysis methods as applied an optimised portfolio of risky assets. An experimental research design will benefit from revealing the relationship between return forecasting accuracy and realised returns by observing changes in the expected return prediction technique utilised.

3.2.1 Markowitz Mean Variance Portfolio Optimisation

The most crucial goal of the Modern Portfolio approach to investment management is to maximise returns for a given risk. The central tenant of Harry Markowitzs' approach to portfolio optimisation was simultaneously optimising a portfolios expected return and minimising the portfolios' investment risk. Risk is quantified as the variance of returns of the stocks of the portfolio.

Markowitz presented portfolio optimisation as a mathematical framework, which at the centre of which lies a mathematical optimisation problem.

Optimisation is a challenging problem to solve in general. However, convex optimisation problems are well-understood, and portfolio optimisation problems are convex sets (Martin, 2018). A vast body of theory, as well as the refined solving routines, can therefore be applied.

A convex problem has the following form.

$$\begin{gathered} minimise\ x \quad f(\vec{x}) \\ subject\ to \ \ g_i(\vec{x}) \leq 0, i = 1, \ldots, m \\ \vec{A}x = \vec{b} \end{gathered}$$

Markowitzs' Mean-Variance Optimisation forms a spectrum of portfolios, the best of which is the portfolio that decreases the absolute risk for a predetermined expected return, the Minimum-Variance Portfolio (Zhang, Zohren, & Roberts, 2020). Markowitz was also able to realise that genuinely efficient diversification was only achieved when securities were found to have very little positive correlation or negative correlation to every other security in the portfolio and to the portfolio itself (Lombard, 2015).

3.3 Research procedure and methods

This section documents the actual procedure and the methods employed in this research to collect, collate, process, and analyse empirical evidence. Broadly, we detail the data and information collection instruments (Section 3.3.1), the target population and sampling of respondents (Section 3.3.2), the ethical considerations during the research process (Section 3.3.3), data and information collection process and storage (Section 3.3.4), data and information processing and analysis (Section 3.3.5)

3.3.1 Research data and information collection instrument(s)

An essential component of the research design process is in collating relevant and reliable research data. Research data must be gathered and analysed to test against the research question (Leedy & Ormrod, 2014). For this purpose, a data collection instrument must be of such nature that it allows for collating information specific to answering the research questions. The selection of the data collection instrument is an essential aspect of the research design. A correctly selected data collection tool will impact achieving the desired research result (Adams, Khan, Raeside, & White, 2007).

In Section 3.2. the research strategy of this study is explained as being a quantitative one and will be conducted as an experimental study. As such, the most appropriate data collection instrument will contain all the variables required to undertake the Markowitz Portfolio Optimisation Framework. Data is available for collation on numerous public financial databases. There are numerous Financial databases available, but Yahoo Finance is one of the world's most recognisable financial databases and contains an extensive repository of data on financial assets such as historical prices dates, the volume traded.

Their paper ZongMing et al. (2017) used Modern Portfolio Theory to examine the investment risk and returns of IBM and the Down Jones. The study aimed to analyse the use of Modern Portfolio Theory as an investment tool. The authors used the monthly returns of IBM from 1996 to 2015 to analyse Modern Portfolio Theory using the Dow Jones Industrial Index as a benchmark.

In their study, Yahoo Finance was utilised as the data collection instrument due to its ease of use as historical prices could be downloaded for analysis within Microsoft excel (ZongMing et al., 2017). What makes Yahoo Finance stand out against other financial databases is that many researchers use it. Yahoo Finance also has an easy to use Application Programming Interface (API), enabling the automation and rapid processing and collation of data when utilising Python. Yahoo Finance's API is well documented, and there are many functions within the

Python libraries(McKinney, 2012) that are built specifically to interface with Yahoo Finances' API. Tadlaoui (2018) also utilises Yahoo Finance in their study.

Images of the research data collections instruments are available on in the Appendix.

3.3.2 Portfolio Construction

3.3.2.1 Holding Period Returns

The returns are the changes in the price of a stock calculated as a percentage and are given by the following equation:

$$R_i = \left(\frac{P_{i+1} - P_i}{P_i}\right)(100)$$

Equation 1: Daily return calculation

Where P_i is the beginning price of the stock.

Where P_{i+1} is the ending price of the stock.

In calculating the daily returns of a stock, the log of daily returns is typically used. It is common practice in portfolio optimisation to take the log of adjusted close price returns. Log returns allow the covariance and correlation matrices to be calculated with ease(Hudson, 2015). Using the log of returns does raise some issues, which are discussed in Section 3.4.

3.3.2.2 Mean

In statistics the arithmetic mean or average is a measure of central tendency of a sample of numbers. The mean is calculated as follows:

$$r = \left(\frac{1}{n}\right)\left(\sum_{i=1}^{n} R_i\right)$$

Equation 2: arithmetic mean

Where R_i is the daily adjusted close price returns taken as a log.

Where n is the number of observations.

3.3.2.3 Variance

The variance of stock returns can be defined as a measure of uncertainty or risk of a stock. The risk is a measure of the dispersion of the stock's expected return (Bodie et al., 2014). It can be defined as follows:

$$\sigma_i{}^2 = \left(\frac{1}{n}\right)\left(\sum_{i=1}^{n} R_i - r_i\right)$$

Equation 3: Variance

Where R_i is the daily stock price.

Where r_i is the arithmetic mean of a stocks returns.

Where n is the number of observations.

The normality of returns of a stock will be evaluated in Section 5 of this study in further detail. To annualise, equation 3 must be multiplied by the number of trading days. It is common practice to multiply by 250, which is used as the number of trading days in a year (Rossi, 2018).

3.3.2.4 Volatility

The volatility is measured as the standard deviation of a stock's returns, and the standard deviation is calculated as the square root of the variance.

$$\sigma_i = \sqrt{\left(\frac{1}{n}\right)\left(\sum_{i=1}^{n} R_i - r_i\right)}$$

Equation 4: Volatility (Standard Deviation)

Where R_i is the daily price is returns taken as a log.

Where r_i is the arithmetic mean of a stocks returns.

Where n is the number of observations.

3.3.2.5 Correlation and Covariance

Covariance and Correlation are two critical factors in Modern Portfolio Theory used in determining the portfolio of risky assets. The Covariance measures the variability of the returns of two stocks and indicates their directional relationship (Hastie, 2009 ; Markowitz, 1991). The Covariance is a square matrix that gives the Covariance between each pair of stocks.

$$cov\left[\vec{R}_i, \vec{R}_j\right] = \left[\left(\vec{R}_i - \vec{r}_i\right)\left(\vec{R}_i - \vec{r}_j\right)^T\right] = \left[\vec{R}_i \vec{R}_j^T\right] - \vec{r}_i \vec{r}_j^T$$

Equation 5: Covariance matrix

Where $\vec{R}_i$and $\vec{R}_j$ are the returns vector of the stocks being compared

Where $\vec{r}_i$ and $\vec{r}_j$is the vector of mean returns of the stocks being compared

Correlation is a measure to which two stocks are linearly related and ranges between one and negative one. The Correlation matrix is a square matrix that gives the Correlation between two pairs of stocks.

$$corr\left[\vec{R}_i, \vec{R}_j\right] = \frac{cov\left[\vec{R}_i, \vec{R}_j\right]}{\sigma_i \sigma_j}$$

Equation 6:

Where $cov\left[\vec{R}_i, \vec{R}_j\right]$ is covariance matrix of the stocks being compared.

Where σ_i and $\sigma_i$is the volatility (standard deviation) of the stocks being compared.

3.3.2.6 Expected Returns

The Expected Return is the expected value of a random variable, which in Modern Portfolio Theory is stated to be the generalised product of the weighted average. Historical average returns are typically used to generalise the expected return. The idea being that Expected Returns in the long term will be a reasonable estimate of future returns (Ilmanen, 2011).

$$Ei[R] = \sum_{i=1}^{n} P_i R_i$$

Equation 7: Expected returns

Where R_i is the daily stock returns.

Where P_i is the probability for the return, this is often generalised to be the arithmetic mean of a stocks returns.

Where n is the number of observations.

3.3.2.7 Portfolio weights

The portfolio weights are the contribution of the stock in the total basket of assets. The simplified model of Modern Portfolio Theory assumes that the sum of the portfolio weights totals to one. This is in a version of Modern Portfolio theory that does not include shorting, where portfolio weights can be negative.

$$w_i + w_j + w_k = 1$$

Equation 8: Portfolio weights

3.3.2.8 Portfolio Return

The portfolio expected return is calculated

$$Ep[R] = \sum_{i=1}^{n} R_i w_i$$

Equation 9: Portfolio expected return

Where R_i is the expected return of the portfolio asset.

Where w_i is the weight of the portfolio asset as a percentage of the total portfolio.

For portfolios consisting of more than two assets, the equation for portfolio expected return can be calculated using linear algebra and is the dot product of the transposed returns vector and the portfolio weights. As such, the equation can be rewritten as

$$E_p[R] = \vec{R}^T \vec{w}$$

Equation 10: Portfolio expected return

Where $\vec{R}^T$ is vector of portfolio returns.

Where $\vec{w}$ is the vector of portfolio assets.

3.3.2.9 Portfolio Variance and Volatility

Similarly to that of a stock, the variance of a portfolio can be defined as a measure of uncertainty or dispersion around the central value(Expected Portfolio Return). In Finance, variance is a quantification of risk. The variance and subsequently the volatility of a portfolio are calculated as follows (Bodie et al., 2014).

$${\sigma_p}^2 = \vec{w}^T cov\left[\vec{R}_i, \vec{R}_j\right]\vec{w}$$

Equation 11: Portfolio variance

$$\sigma_p = \sqrt{\vec{w}^T cov\left[\vec{R}_i, \vec{R}_j\right]\vec{w}}$$

Equation 12: Portfolio volatility

3.3.3 Research target population

3.3.3.1 Research target population

The target population can generally be described as the grouped variables from which research is intended to be conducted, and conclusions are drawn. The population of relevance for this research would be the constituent shares listed on the FTSE/JSE Financial 15 Index (FIN15). The FIN15 consists of the top 15 companies as ranked by market capitalisation, which are constituents of the economic group classification of the Johannesburg Stock Exchange (FTSE/JSE Africa Index Series, 2020). As such, the portfolio weightings of the FIN15 constituents is calculated as part of the FTSE/JSE index methodology (FTSE/JSE Africa Index Series, 2020). The FIN15 was chosen for this study due to its long history on the JSE. The FIN15 dates back to 2002, when the Index was first originated. The FIN15 was also chosen due to its size. A more extensive Index such as the FTSE/JSE Top 40

Index would have introduced greater computational complexity and required more computing time to generate the model. This would not be suited for a Support Vector Regression model as it is not suited to working on datasets with a large number of features (equities) (Géron, 2019). The constituents and market capitalisation of the FIN15 Index are shown in

Table 1: FIN15 / STXFIN composition as at 22 December 2014

The FIN15 Index is rebalanced every quarter as market capitalisations of the constituent companies change daily due to fluctuations in price. The methodology utilised by the JSE in calculating the Index value will not be duplicated in this study. Instead, will utilise the Daily Adjusted Close Price information for the Exchange Traded Fund(ETF), "Fini ETF", created by Satrix Pty Ltd. The purpose of the "FINI ETF" is to track as closely as possible the value of the FIN15 Index. As such, STXFIN is registered as a Collective Investment Scheme and listed on the JSE (Satrix, 2020). As a Collective Investment Scheme, the fund must submit quarterly and annual statements to investors and report on the Stock Exchange News Service of significant transactions, refer to Table 4.

Every quarter when the FIN15 Index is rebalanced, duplicate transactions will take place on STXFIN. For this study, the rebalancing notices issued on the JSE will be used to source the portfolio weights data of the FIN15 Index. The STXFIN tracks the FIN15 Index so closely that the ETF Adjusted Close Price information and portfolio weights can be used in place of the FIN15 Index. From 2 February 2002 to 18 December 2020, the correlation coefficient between the value (price) of the FIN15 Index and STXFIN gives a value of 0.99976 for an observation window of 4715 observations.

Table 1: FIN15 / STXFIN composition as at 22 December 2014

No	Equities	Share Name	Weight
1	ABG	ABSA Group(Formerly Barclays Africa Group Ltd)	5.27%
2	CCO	Capital and Countries Properties Plc	1.58%
3	DSY	Discovery	2.99%
4	FSR	First Rand Ltd	13.70%
5	GRT	Growthpoint Properties Ltd	5.34%
6	INL	Investec Ltd	2.20%
7	INP	Investec Plc	5.27%
8	ITU	Intu Properties Plc	5.08%
9	MMI	MMI Holdings Ltd	2.82%
10	NED	Nedbank Group Ltd	4.37%
11	OML	Old Mutual Ltd	14.52%
12	REI	Reinet Investments SCA	3.41%
13	RMB	RMB Holdings Ltd	3.86%
14	RMI	Rand Merchant Insurance Holdings Ltd	2.65%
15	SBK	Standard bank Group Ltd	15.72%
16	SLM	Sanlam	11.22%
			100.00%

In their study (Huni & Sibindi, 2020) constructed Minimum-Variance Portfolios using Modern Portfolio Theory Framework utilising the tradable indices of the Johannesburg Stock Exchange, including the FIN15. The study supported the findings that Modern Portfolio Theory is an effective framework for the construction of optimal portfolios for the Johannesburg Stock Exchange.

The yield on the RSA 5-year government bond will be the proxy for the Risk Free Rate in this study. The Risk Free Rate that will be utilised 7.64%. The descriptive statistics of the Risk-Free Rate over the period 2016 to 2020 is shown below in Figure 3.

Figure 4: Histogram of Risk Free Rate returns (2016 – 2020)

Table 2: Risk Free Rate 24 June 2019to 21 December 2020

Start Date	24-Jun-19
End Date	11-Dec-20
Mean	7.64%
Standard Error	0.04%
Median	7.41%
Mode	7.26%
Standard Deviation	0.77%
Sample Variance	0.01%
Range	4.45%
Minimum	6.77%
Maximum	11.22%
Count	381

3.3.3.2 Sampling period

Table 3: Rebalancing Dates of FIN15 / STXFIN from December 2014 to December 2020

Period	Start Date	End Date	Trading Days	Cumulative Trading Days	Training or Testing
1	2014/12/22	2015/03/20	60	60	Training
2	2015/03/23	2015/06/19	59	119	Training
3	2015/06/22	2015/09/18	63	182	Training
4	2015/09/21	2015/12/18	62	244	Training
5	2015/12/21	2016/03/18	62	306	Training
6	2016/03/22	2016/06/17	58	364	Training
7	2016/06/20	2016/09/16	62	426	Training
8	2016/09/19	2016/12/15	63	489	Training
9	2016/12/19	2017/03/16	60	549	Training
10	2017/03/17	2017/06/15	59	608	Training
11	2017/06/21	2017/09/18	62	670	Training
12	2017/09/19	2017/12/19	64	734	Training
13	2017/12/20	2018/03/16	59	793	Training
14	2018/03/19	2018/06/19	61	854	Training
15	2018/06/20	2018/09/21	66	920	Training
16	2018/09/25	2018/12/21	62	982	Training
17	2018/12/24	2019/03/15	56	1038	Training
18	2019/03/18	2019/06/21	63	1101	Training
19	2019/06/24	2019/09/20	63	1164	Testing
20	2019/09/23	2020/06/19	184	1348	Testing
21	2020/06/22	2020/09/18	63	1411	Testing
22	2020/09/21	2020/12/21	62	1473	Testing

Chapter 2 the Expected Return model is a probability estimate and is utilised as the risk model. The risk model is developed based on the daily historical log-returns of the portfolio constituents. The arithmetic mean and the support vector regression model will be built on the training period data. The training period for the mean-variance and Support Vector Regression model is from period 1-18 as per Table 4

Although the FIN15/STXFIN contains 15 constituents at a time, due to listing and delisting over the period analysed, more than 15 shares will be present in the initial collation of data from the Stock Exchange News Service Notifications. Return Data on the Risk-Free Rate will only be utilised in the testing period of the study, Period 19 -22(24 June 2019 to 21 December 2020).

Support Vector Regression requires examples of data to train. Supervised Learning requires examples for testing the data set will be split into a period of training the model for expected return and testing the model of the expected return. The training period for the Mean-Variance Optimisation and Support Vector

Regression model is from period 1-18 as per Table 4. Block (2017); (Drue, 2018), in their thesis report, utilised a train test ratio of 80% to 20% of their data set.

3.3.4 Ethical considerations when collecting research data

It is a poorly held belief that ethical considerations have no place in financial research. Due to the study's pure quantitative nature, ethics is often an afterthought. The methods utilising machine learning in this research are being applied more and more within the investment discipline. An example is NMQRL research in South Africa (Zaidy, 2018). As portfolio managers continue to pursue excess returns for their clients, this will prompt innovation within the investment discipline, and failures may result. High profile failures of financial systems integrated with machine learning in the financial system have resulted in detrimental harm being inflicted on human beings who have livelihoods invested in the financial system. Recent examples include the 2010 flash crash, in which excessive high-frequency algorithmic and statistical investment transactions led to excessive price fluctuations on the US's financial market.

Whenever research or investigation is conducted with a possible focus on human beings or animals, ethics must be considered. Ethics in research can be categorised into four categories:

- Protection from harm
- Voluntary and informed participation
- Honesty with professional colleagues
- Right to privacy

The author of this research has been interested in Artificial Intelligence and the sub-discipline of Machine Learning techniques applied to their career within engineering. The authors' undergraduate individual research project was on Artificial intelligence as applied to manufacturing engineering.

As detailed in Section 3.3, no human or animal respondents focus on this investigation regarding the data collection and target population. No interactions that would bring harm to humans or animals will be conducted. All information is void of personally identifying information. All data that is collected is available on public financial databases (Yahoo Finance). Moreover, the Yahoo Finance database is a standard database that has been deemed suitable for research-related work.

3.3.5 Research data collection process

The research data collection method will be internet-based through accessing the Yahoo Finance API. The Yahoo Finance database is a standard database that has been deemed suitable for research-related work.

Research Data will be retrieved directly from the Yahoo Finance API via the Pandas Python Module. Yahoo Finance will be queried to collect information regarding the FIN15 / STXFIN index constituents from 22 December 2014 to 21 June 2019. Additional information will be collected to perform the portfolio optimisation. The information to be collected will include:

- Company name
- Stock Code
- Adjusted Daily Close Price
- FTSE/JSE FIN15 Daily Index Closing Value
- STXFIN Daily Adjusted Close Price
- Date
- Index rebalancing date
- Asset Weight before and after rebalancing dates

The JSE Stock Exchange News Service (SENS) contains information and details of the rebalancing of the STXFIN Exchange Traded Funds. Portfolio rebalancing notifications are issued every quarter by Satrix, the fund manager of the STXFIN Exchange Traded Fund. The rebalancing of the Exchange Traded Funds constituents ensures the fund continues to replicate and track the respective index value.

The Rebalancing dates, the new and previous weights are to be extracted from the rebalancing notifications issued on the SENS. Table 4 shows that between 22-12-2014 and 21-12-2020, 22 different rebalancing notices were issued. Meaning that 22 rebalances and modifications to the portfolio weights occurred. The rebalance dates, new and previous constituent weights are extracted from all 22 rebalance notifications using Excel Power Query and are shown in the Appendices.

The author's data will not need to be stored as it is accessed directly from the Yahoo Finance database via the API.

3.3.6 Research data and information processing and analysis

3.3.6.1 Research data and information processing

Data will be processed and analysed utilising Google Colab. Colab is a free cloud-based code development environment that utilises Python amongst other languages. The advantages of using Colab are that it does not require setup, supports the most popular Machine Learning libraries and is built for research and education purposes. The Python code was written to conduct the experimental study is available for viewing in the Appendices.

The steps of operationalising the data processing are similar to that presented by (George, Modern Portfolio Theory, Efficient Frontiers) in their presentation on Machine Learning for Finance in Python.

In order to conduct the data analysis, all the retrieved data will be loaded onto a folder on the Google Drive Cloud platform. The data that will be loaded for calling by the Python script will be kept on the Google Drive platform will be the rebalancing dates shown in Table 4, indicating the periods under which the constituents were kept constant.

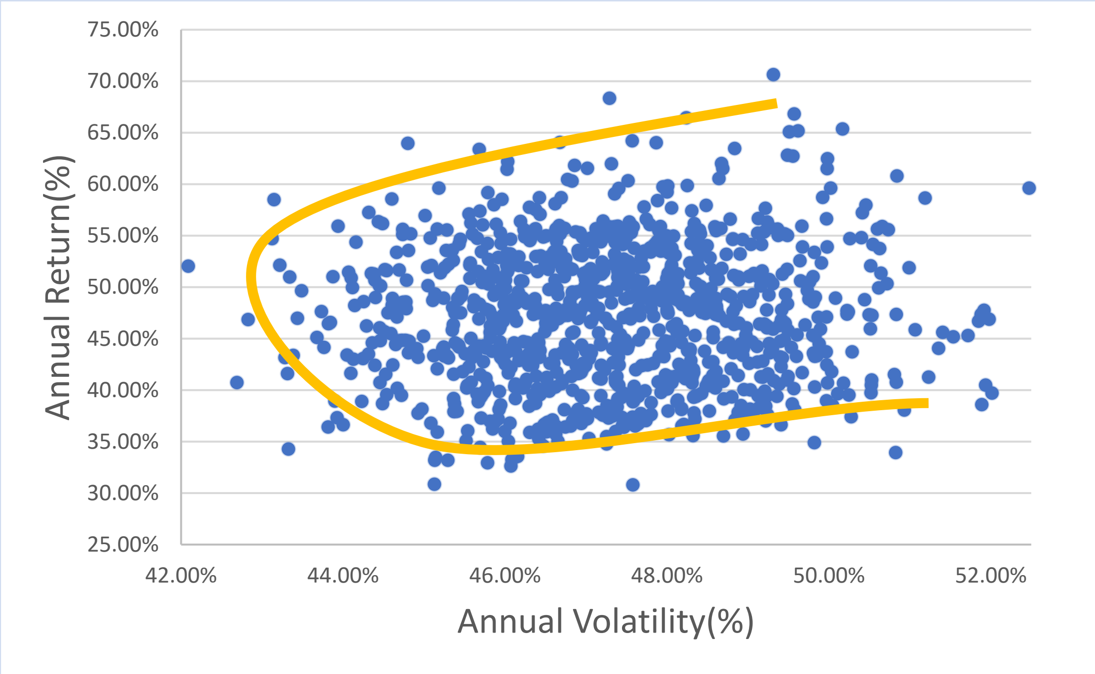
For the data analysis, the following modules must be loaded into the Google Colab script.

`Pandas` and for creating data structures and data analysis tools. `Numpy` for scientific computing, explicitly performing Matrix Algebra required for calculating Portfolio Volatility and Portfolio Returns. `Datetime`, a module for manipulating dates and times. `PyPortfolioOpt`, a module that can calculate the critical variables required in this study which include: returns, volatility, Sharpe ratio and efficient frontier of a portfolio of risky assets. `Matplotlib` for graphing and plotting. `Sklearn`, a machine learning library.

3.3.6.2 Generating the efficient frontier

The prices of the stocks in the index will be retrieved using the `pdr.get_data_yahoo` function in Pandas. The adjusted close price data is collected into a Pandas `dataframe`. The returns are calculated using the `expected_returns.returns_from_prices` function in the `PyPortfolioOpt` module. In the Markowitz Mean-Variance Optimisation, the log-returns are used rather than the simple returns. The Covariance and Correlation (`Corr` and `Corr` respectively) matrices are calculated for the stocks using the `Numpy` module. To create the hypothetical efficient frontier for the portfolio of assets, the weights of all assets are randomised. The random portfolio weights are used to calculate the portfolio return and volatility using equations 9 and 11. The process is repeated 999 times to create a visualisation of the efficient frontier. This is shown in **Error! Reference source not found.**

Figure 5: Visualisation of Efficient Frontier of Modified Index for December 2014 to June 2019(training period) with 1000 random portfolios, generated with Main Script Python code



3.3.6.3 Minimum Variance Portfolio

According to Markowitz, the Minimum-Variance portfolio is deemed to be the optimal portfolio. The features of the minimum variance portfolio are calculated using `ef.portfolio_performance` from the `PyPortfolioOpt` library. The `ef.save_weights_to_file` calculates the Minimum-Variance portfolio weights and saves them to the cloud. The weights are to be used in evaluating the performance of the Minimum-Variance portfolio over the testing period.

3.3.6.4 Machine Learning Portfolio(SVR)

Creating the Machine Learning Portfolio, stock data from December 2014 to June 2019 is used to train and fit the Support Vector Regression model. The periods for the training of the Regression Model are shown in Table 4. The dependent variable for the regression model is the matrix of stock returns. The size of the stock returns matrix is 1123 columns and 10 rows. The independent variable will be the Index's returns (FIN15 / STXFIN) over the training period. Before the Regression model can be trained, the returns data must be pre-processed and made the same size and shape by removing missing or irrelevant data such as returns on non-trading days(public holidays). Once the Index and assets have the same number of rows, any missing or blank data must be imputed using the median. The model is now ready to be trained and fitted. The `sklearn.preprocessing StandardScalar` is used to scale both the dependent and the independent variable to simplify processing and ensure that the Regression Model will converge. The regression model is created using the `svr_rbf.fit` and `svr_rbf.predict` functions from the `sklearn` module.

Equation 9 must be solved to determine what the weights for the Machine Learning Portfolio will be. Equation 9 is rewritten as Equation 13 follows such that it can be solved by substituting the mean returns for the testing period into the trained regression model. The pseudo-inverse of the returns matrix is calculated to determine the Machine Learning Portfolio weights. Where the dependant function for the regression model becomes the mean returns over the training period.

$$[\vec{R}].E_p[R] = \vec{w}$$

Equation 13: Weights of SVR Model

To conduct the information collection and analysis of Research Data, the author learned to code in the Python Programming language from scratch.

3.3.6.5 Research data and information analysis

The process of data analysis is the steps or methods performed to explain, describe detect outcomes, patterns that test the research hypothesis(Adams et al., 2007).

The Sharpe Ratio and the portfolio cumulative and total returns will be used to evaluate the results of the study and determine which Portfolio is would be more attractive for an Investor who fits the assumptions for the risk averse Modern Portfolio Theory investor.

3.4 Research strengthens—reliability and validity measures applied

3.4.1 Adjusted Close Price

The use of the adjusted close price in calculating the returns serves to improve the study. At the close of the trading day, the market price reflects the influence of corporate actions, dividend payouts, and stock buybacks. These affect the stock price and in a manner that has. Using the adjusted close price, the actual effect of a stock price and subsequent returns can be evaluated in isolation. Improving the efficacy of the Markowitz Optimisation framework.

3.4.1 Liquidity

The stocks and Index selected for the purposes of this study shows great liquidity. Liquidity of stocks is an important assumption within Modern Portfolio Theory and Mean-Variance Optimisation, as an illiquid stock will have the ability to affect price and therefore return. The stocks within the FIN15 are all medium to large cap companies, as shown in Table 1, with significant trading volumes. Where liquid stock is shown to have trading monthly trading volumes of 10 000(Lombard, 2015).

3.5 Research weaknesses—technical and administrative limitations

3.5.1 Technical - Log returns

The use of logarithmic returns simplifies the calculation of Covariance and therefore portfolio variance. The mean of a set of returns calculated using logarithmic returns is less than the mean calculated using simple returns. Higher portfolio variance will automatically reduce expected returns as a matter of algebra(Hudson & Gregoriou, 2015). Therefore, conclusions made in this study that use log returns cannot be immediately generalised to similar studies that have conducted optimisations using simple returns.

3.5.2 Technical – Optimised Support Vector Regression Kernel Parameters

Default parameters of the Support Vector Regressing kernels were used as it is beyond the scope of this research to conduct Kernal Parameter optimisation.

3.5.3 Administrative – Variable Portfolio Weights introduce error

The study's primary limitation is that companies enter and exit the Exchange Traded Fund / Index over a long period due to changes in market cap. Between each rebalancing event, the market cap of the individual stock varies over the period in review. The change in market cap will result in a change in portfolio weight in the index that is not a deliberate rebalancing to replicate the index. When the Minimum-Variance and Machine Learning portfolio weights have calculated, the portfolio's performance will be compared to an Index in which asset weightings can shift naturally. The influence of this shift in portfolio weights must be considered on the final result.

3.5.4 Administrative - Exclusion of instruments and introduction of bias

Throughout review (both testing and training), stocks move in and out of the index for various reasons. One reason that has caused a limitation on the study is the change of stock code. Due to changes in the stock codes of some stocks, the Pandas data reader function fails to retrieve a complete set of daily adjusted close prices for some stocks in the index. The Pandas data reader function returns NAN(Not A Number), see Figure 6. Markowitz Mean-Variance Optimisation cannot be performed on an incomplete dataset. As a result, the stocks with missing price data must be removed from the study and a "Modified Index" constructed containing the stocks with available price data for the period. Due to either delisting from the index, dropping out of the index or aquisition / split the following stocks have no longer been included in the training and testing period:

- Capital and Countries Properties – dropped out of index
- Intu Properties – delisted from exchange
- MMI Holdings – dropped out of index
- Old Mutual – split up
- Reinet Investments – dropped out of index
- RMB Holdings – unbundled
- Rand Merchant Insurance Holdings Ltd.

The Modified Index will be rebalance such that the sum of all weights remaining equals zero by equal distribution of the weights of the stocks that were removed.

Table 4: FIN15 / STXFIN and Modified Index weights composition as at 22 December 2014

	Code	Share Name	FIN15 / STXFIN weights	Modified Index Weights
1	ABG	ABSA Group(Formerly Barclays Africa Group Ltd)	5.27%	8.66%
2	CCO	Capital and Countries Properties Plc	1.58%	0.00%
3	CPI	Capitec	0.00%	3.39%
4	DSY	Discovery	2.99%	6.38%
5	FSR	Firstrand Ltd	13.70%	17.09%
6	GRT	Growthpoint Properties Ltd	5.34%	8.73%
7	INL	Investec Ltd	2.20%	5.59%
8	INP	Investec Plc	5.27%	8.66%
9	ITU	Intu Properties Plc	5.08%	0.00%
10	MMI	MMI Holdings Ltd	2.82%	0.00%
11	NED	Nedbank Group Ltd	4.37%	7.76%
12	OML	Old Mutual Ltd	14.52%	0.00%
13	REI	Reinet Investments SCA	3.41%	0.00%
14	RMB	RMB Holdings Ltd	3.86%	0.00%
15	RMI	Rand Merchant Insurance Holdings Ltd	2.65%	0.00%
16	SBK	Standard bank Group Ltd	15.72%	19.11%
17	SLM	Sanlam	11.22%	14.61%
			100%	100%

3.5.5 Covid impact on the market

In March of 2020 a global pandemic crisis impacted the entire world's financial systems. A contagious virus had spread from China through to the rest of the world and made its way into South Africa. The national government instituted a lockdown to slow the spread of the virus and prevent infection. The results of the lockdown lead to market volatility over the period of March 2020. The effects of the Covid19 pandemic are factored in the evaluation of the results of the study.

4. PRESENTATION AND DISCUSSION OF RESEARCH RESULTS

This section documents the results of the experimental study as conducted with the methodology detailed in Section 3.3.5 and the analyses conducted on the empirical evidence. The weights of the Modified Minimum-Variance Portfolio and the Machine Learning Portfolio are shown in Section 4.2. The portfolio performance (realised returns, excess returns and investment growth) of the portfolio constructed is evaluated in Section 4.2 using the realised returns of each portfolio constructed over the test period. Concluding remarks are with regards to the research question and the attractiveness of the portfolio to an investor.

4.1 Portfolio Construction

4.1.1 Portfolio Weights

The portfolio weights calculated by each of the Portfolio optimisation techniques employed are shown in the Table 6 below. The code in the appendices titled Main Script and SVR Predictions was used to calculate the portfolio weights. No significant relationship can be observed between the selection of weights between the Machine Learning and Minimum Variance portfolio. The only stocks selected by both portfolios were: Capitec, Growthpoint Properties and Nedbank. The top three holdings in the Minimum Variance portfolio were in Growthpoint Properties, Investec Plc and Capitec in descending order of weight. For the Machine Learning Portfolio, the top three holdings were in Capitec, Growthpoint Properties and Standard Bank Group.

Table 5: Portfolio weights calculated for modified index, Minimum Variance and Machine Learning Portfolio

Stock Code	Company Name	Modified Index	Minimum-Variance Portfolio	Machine Learning(SVR) Portfolio
ABG	ABSA Group(Formerly Barclays Africa Group Ltd)	8.66%	0.00%	0.00%
CPI	Capitec	3.39%	14.45%	53.60%
DSY	Discovery	6.38%	3.21%	0.00%
FSR	Firstrand Ltd	17.09%	0.00%	0.00%
GRT	Growthpoint Properties Ltd	8.73%	50.97%	27.67%
INL	Investec Ltd	5.59%	0.00%	0.00%
INP	Investec Plc	8.66%	30.00%	0.00%
NED	Nedbank Group Ltd	7.76%	1.37%	8.49%
SBK	Standard bank Group Ltd	19.11%	0.00%	10.24%
SLM	Sanlam	14.61%	0.00%	0.00%
		100.00%	100.00%	100.00%

4.1.1 Diversification

The calculation of the Minimum-Variance Portfolio has also led to a reduction in the number of stocks. From 10 stock now only 5 are held in the Minimum-Variance Portfolio. This is a demonstration of Efficient Diversification, which is explained as a reduction in the number of stocks which has led to a reduction in the covariance between assets (Markowitz 1952).

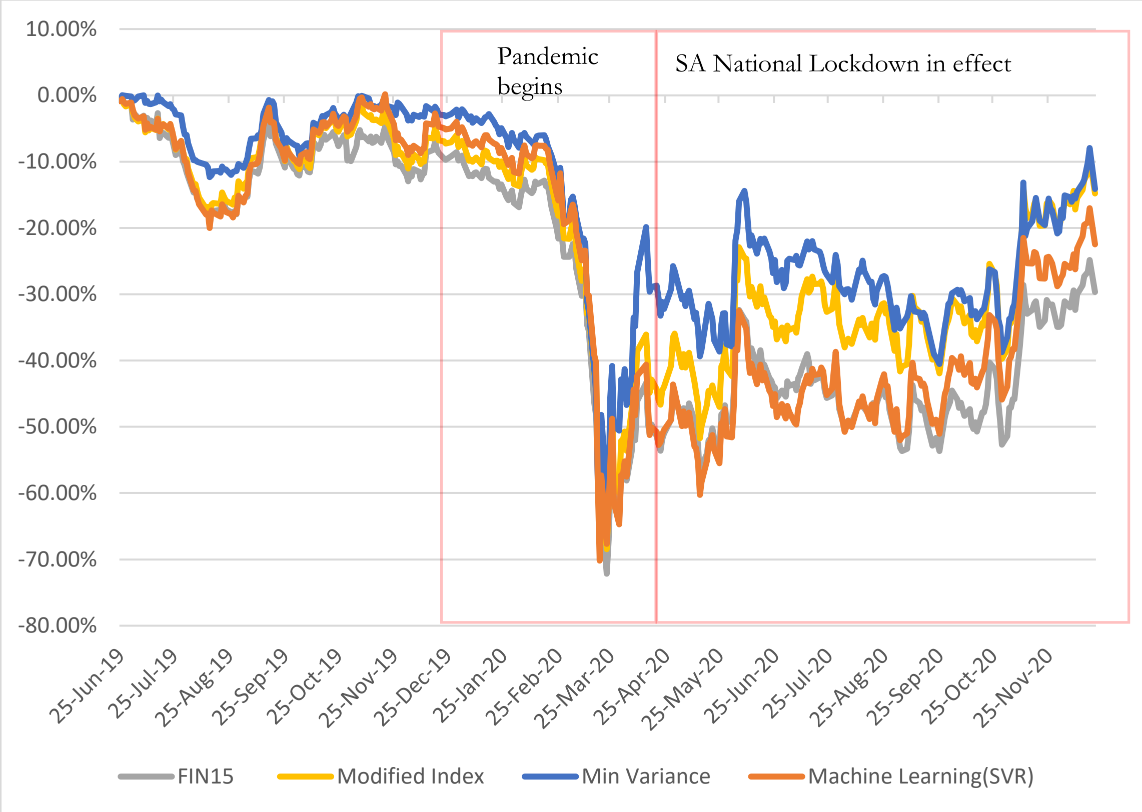
The Machine Learning Portfolio has also shown a decrease in holdings, from 10 to 4 stocks held. The support vector regression reduced the number of dimensions (features) but still managed to obtain a Rsquared value of 0.92, meaning that the stocks that were removed did not contribute in a statistically significant way to the expected returns prediction of the Support Vector Regression model

4.2 Portfolio Performance

4.2.1 Total Returns / Losses

The Systematic Shock effect of the Covid19 pandemic and the resulting lockdowns introduced a great deal of volatility into the JSE. It is essential to analyse the investment growth and returns over the period before starting the National Lockdown and the performance after the National Lockdown, as shown in Figure 7. The Minimum-Variance Portfolio generates the lowest losses compared to the Modified Index and the Machine Learning Portfolio for the entire period (pre and post lockdown). When benchmarked against the Modified Index, the Minimum-Variance outperforms the benchmark in total returns but only marginally. This is to be expected as many studies have empirically verified the effectiveness of the Markowitz Mean-Variance Optimisation to the construction of an optimal portfolio of risky assets (Huni & Sibindi, 2020). After the announcement of the National Lockdown in March 2020, the market and subsequently all portfolios recovered from the shock. The Minimum-Variance Portfolio generated the least losses from March 2020 to December 2020. Consistently providing returns above the Modified Index benchmark and the Machine Learning Portfolio. The Machine Learning Portfolio never beat the Modified Index, which is its benchmark. If the FIN15 is ignored in Figure 7, the Machine Learning Portfolio is the poorest performing portfolio.

Figure 6: Total Returns of portfolios over testing period: June 2019 – Dec 2020



This can be ascribed to the fact that the Machine Learning Portfolio only serves to replicate the Index. It is not constructed with any attribute that can minimise losses or maximises returns compared to the benchmark.

Table 6: Total returns at the end of testing period: June 2019 to December 2020

FIN15	Modified Index	Minimum-Variance	Machine Learning(SVR)
-29.68%	-14.81%	-14.13%	-22.49%

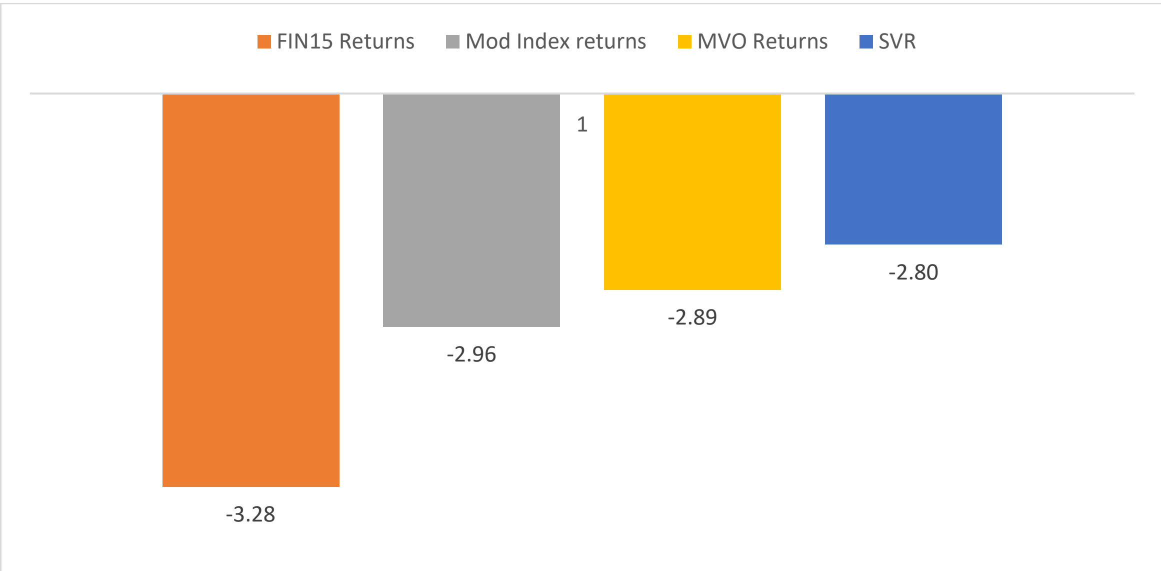
It can be seen from the correlation matrix in table 8 that the machine learning portfolio shows a higher correlation with the Modified Index than the min variance portfolio, therefore Although the returns of the FIN15 Index are included. The machine learning portfolio seeks to replicate rather than beat the Modified Index benchmark. This can be explained by the R^2 of the Machine Learning Portfolio over the test period. The confidence is calculated using the **`sklearn.svm.svr`** Python module. The R^2 value of the Machine Learning Portfolio is 0.921 over the training period and 0.95 over the test period which is higher than that of the Minimum-Variance when measured against the Modified index benchmark.

Table 7: Correlation matrix of constructed portfolio returns

	FIN15	Modified Index	Minimum-Variance	Machine Learning(SVR)
Index Returns	1			
Modified Index	0.97	1.00		
Minimum-Variance	0.84	0.87	1.00	
Machine Learning(SVR)	0.95	0.97	0.88	1.00

4.2.1 Sharpe Ratio

Figure 7: Sharpe ratios for all constructed portfolios and FIN15 index



Based on the individual Sharpe ratios alone, shown in Figure 8, an investor seeking returns would select the Machine Learning Portfolio due to its Sharpe Ratio being the least negative of all Constructed Portfolios. The Minimum-Variance Portfolio shows only a 0.09 difference to that of the machine learning portfolio. It has also been shown in Figure 7 and 10 that the Minimum-Variance portfolio provides returns over the Modified Index and the Machine Learning Portfolio. Considering all of this, an investor would be better off selecting the Minimum-Variance Portfolio.

Figure 8: Volatility vs Return of all Constructed Portfolios

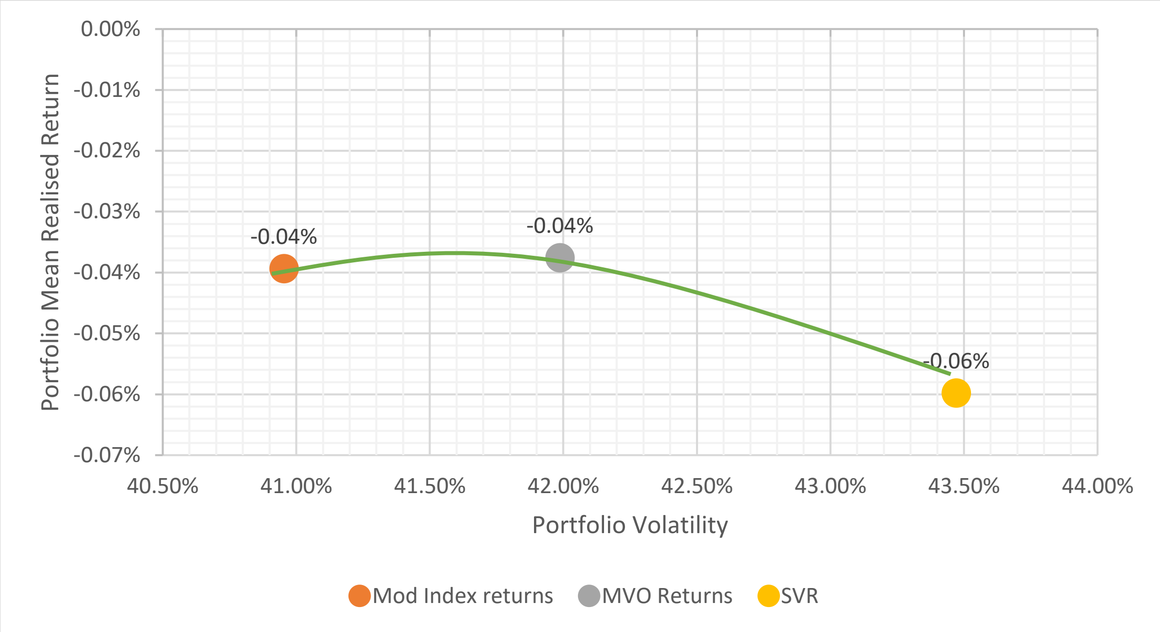
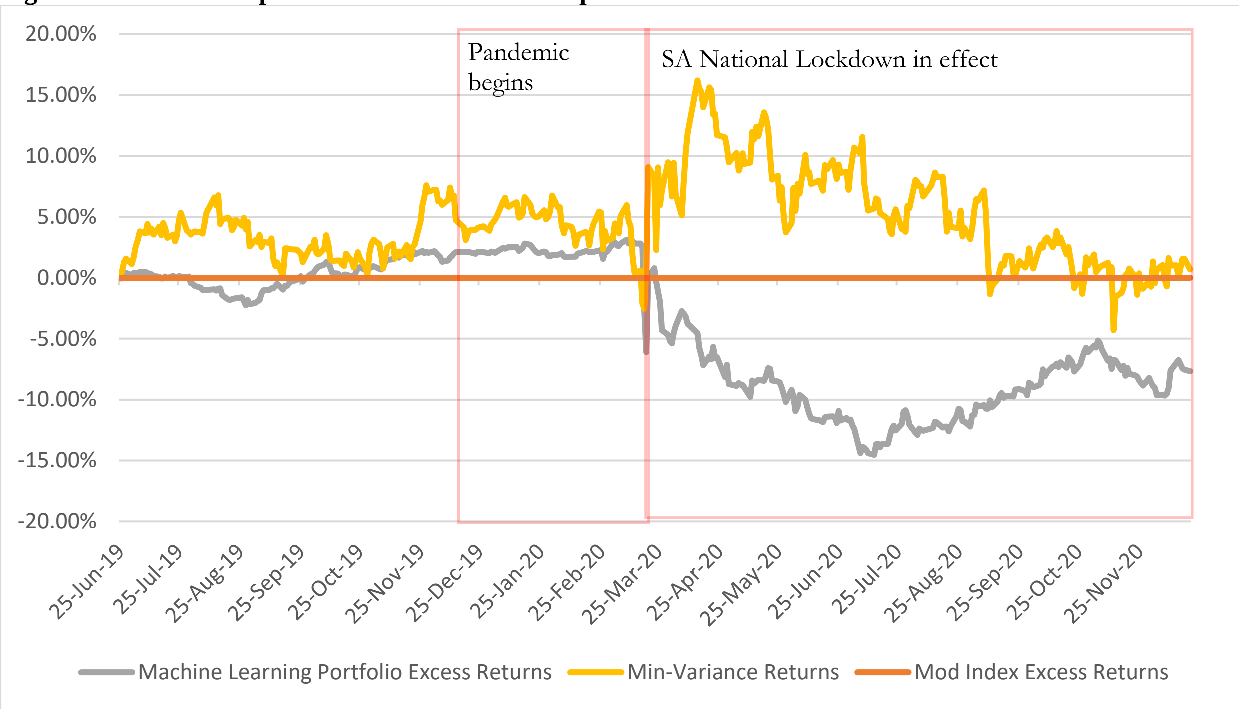


Figure 9 shows all the portfolios on their Risk(Volatility) vs Return plane. The optimum portfolio for an Investor as described in Modern Portfolio Theory being one that sits in the upper left had corner of the plane. Based on this evaluation the Machine Learning Portfolio(SVR) is the least desirable to an investor as the losses(negative returns) offered come at a higher risk, and are greater than losses presented by the Minimum Variance Portfolio and the Modified Index. An investor is disincentivised to select the Machine Learning Portfolio based on Figure 9.

4.2.1 Excess Returns

Figure 9: Constructed portfolio excess returns compared to the Modified Index



When benchmarked against the Modified Index, the Minimum-Variance Portfolio outperforms the Machine Learning Portfolio in excess returns. Excess returns are those that are calculate relative to the index rather than in absolute terms(Beasley, 2013).

The Machine Learning Portfolio generated Excess Returns to the benchmark in September 2019. However, after the Systematic Shock event of the National Lockdown, the Machine Learning Excess Returns become Excess losses and remained that way for the remainder of the test period. From June of 2020, the Machine Learning Portfolio does begin to claw back and recover against the Modified Index Benchmark. Machine Learning Portfolio Excess Returns is insensitive primarily to short term shocks and volatility. Figure 10 shows that the ML Portfolio Excess Returns are smooth compared to that of the Min-Variance portfolio. The behaviour of the Machine Learning Portfolio follows that of the Modified Index as confirmed by the R^2 value.

The Minimum-Variance Portfolio exhibits less than 5% excess returns until March of 2020. From then, the excess returns generated by the Minimum-Variance portfolio spike but erode for the nine months to December 2020. Ultimately the Minimum variance Portfolio outperforms the Machine Learning portfolio in the metric of Excess returns.

4.3 Portfolio Analysis

4.3.1 Risk and Return, Efficient Frontier

The relationship between risk and return can best be visualised by comparing Figure 11 and Figure 5, which shows the visualisation of the Efficient Frontier by generating 1000 randomly weighted portfolios. During the testing period, all returns are negative, meaning that the Efficient frontier is below that shown in Figure 5. Portfolio return and volatility of the Minimum-Variance and Machine Learning Portfolio have been plotted along with the Modified Index in Figure 11. The Machine Learning and Minimum-Variance exhibit less volatility than that shown by the random portfolios generated by the code. The Minimum-Variance portfolio shows marginally higher returns than the benchmark index.

Figure 10: Visualisation of Efficient Frontier for June 2019 to Dec 2020(testing period) with 1000 random portfolios, generated with Main Script Python code

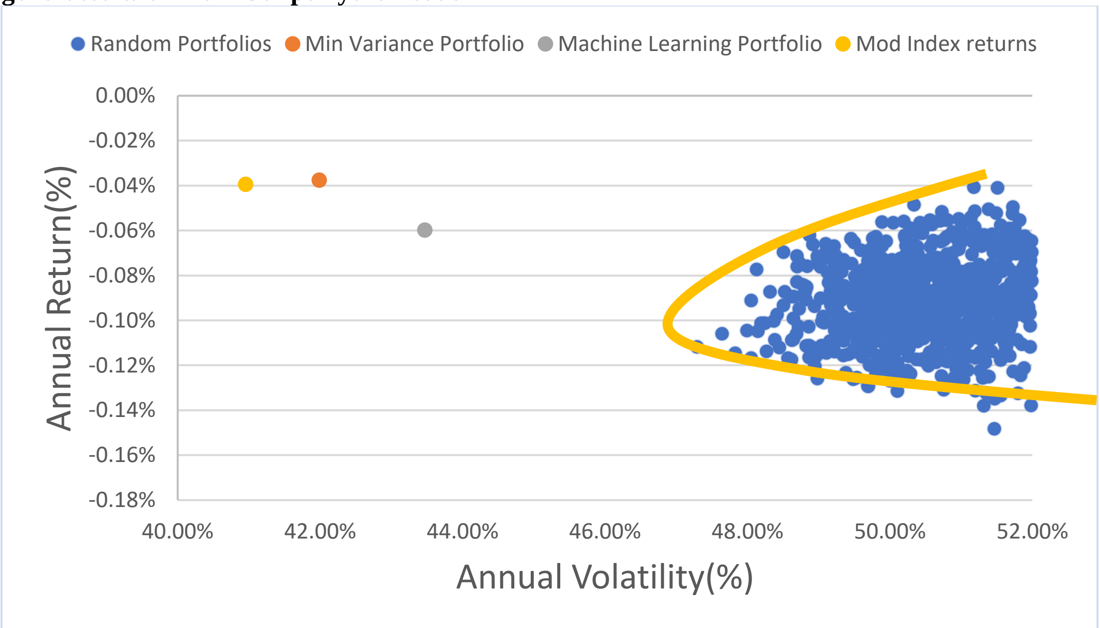


Table 8: Annual volatility of constructed portfolios: June 2019 - Dec 2020

FIN15	Modified Index	Minimum-Variance	Machine Learning(SVR)
37.22%	41.01%	42.04%	43.53%

4.3.1 Normality of Constructed Portfolio Returns

It can be seen in the returns histogram, figure 12, that the returns for all constructed portfolios appear to exhibit a normal distribution.

Figure 11: Histogram of portfolio returns

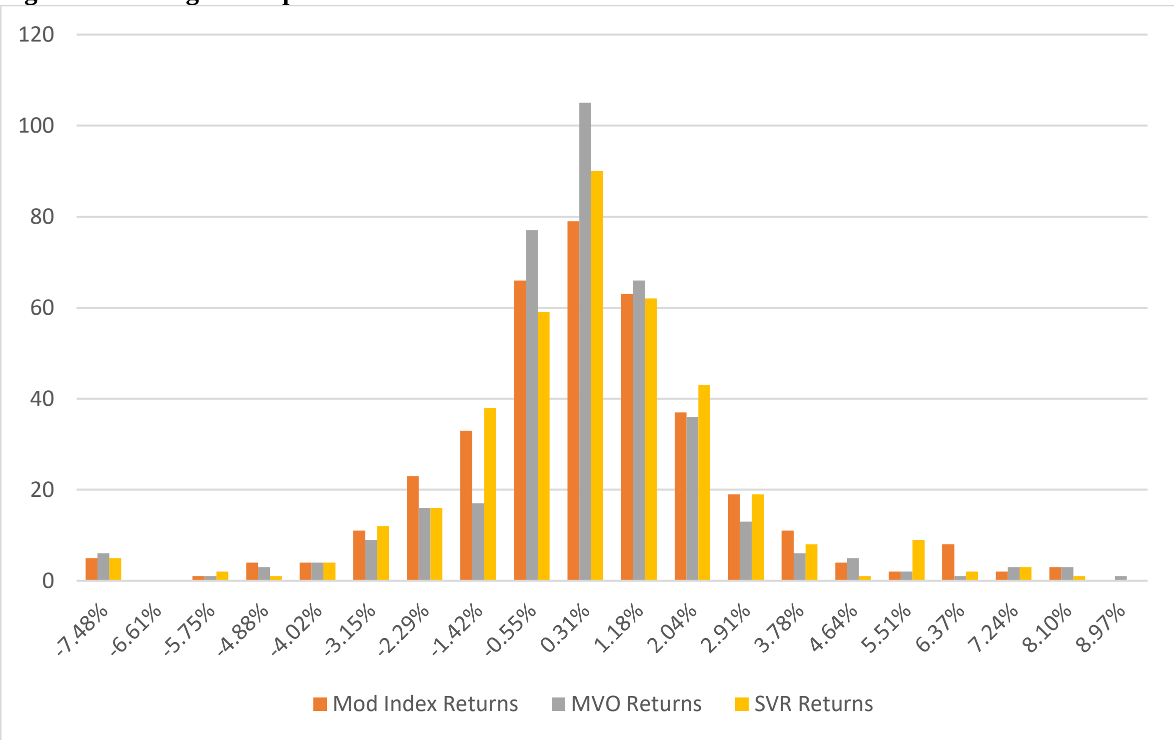


Table 10 shows the Shapiro Wilk normality test results. The Shapiro-Wilk test tests the null hypothesis that the data was drawn from a normal distribution. The p-value for all the constructed portfolios is well within the confidence value tolerance of 0.05. The normality of the constructed portfolios can be explained by the stocks in the portfolios all having Normal returns. A portfolio is a weighted linear sum of stock returns, and Normal distributions have been proven to be linear(Hastie et al., 2009). It can be concluded that the constructed portfolio will also exhibit normality in its returns distribution. The Shapiro-Wilk statistic and p-value for the matrix(377 rows X 10 columns) of stock returns have been calculated. The Shapiro value being 0.89 and the p-value is calculated for the respective Shapiro and p-value being 2.80e-45. The importance of testing the normality of the stock returns is because the Markowitz-Mean Variance optimisation assumes that the mean historical returns can be used as the expected returns. After all, the returns of stock portfolios are Normal. Had this not been so, the predictability of the mean to determine the expected returns would be weak. The normality of stock returns strengthens the argument for using mean historical returns as a value for expected returns.

Table 9: Shapiro-Wilk statistic and p-value for constructed portfolios

Index Returns Stat	Index Returns P-Value
0.923	0
Modified Index Returns Stat	Modified Index Returns P-Value 0
0.930	
Min-Variance Returns Stat	Min-Variance Returns P-Value
0.847	0
Machine Learning(SVR) Stat	Machine Learning(SVR) P-Value
0.889	0
	0

5. DISCUSSION OF REASEARCH RESULTS

5.1 Diversification

No significant relationship can be observed between the selection of weights between the Machine Learning and Minimum Variance portfolio. The only stocks selected by both portfolios were: Capitec, Growthpoint Properties and Nedbank. The top three holdings in the Minimum Variance Portfolio were in Growthpoint Properties, Investec Plc and Capitec in descending order of weight. As presented in section 4.1.1 both techniques showed an efficient diversification of the investment holding from the original index. The Minimum-Variance Portfolio reduced the number for stocks from 10 to 5, while the Machine Learning Portfolio reduced the number of holdings from 10 to 4, the returns generated were still wither on par or beat that of the index. The reduction in holding for both portfolio construction methods is for different reasons. For the Minimum-Variance Portfolio the reduction in number of holdings is due to a reduction in the overall covariance between assets. For the Machine Learning Portfolio, the remaining holdings are the ones that contributed in a statistically significant way to the expected returns prediction of the Support Vector Regression Model.

5.1 Back testing and Interpretation of Results

The Minimum-Variance Portfolio generated the least losses from March 2020 to December 2020. Consistently providing returns above the Modified Index benchmark and the Machine Learning Portfolio. The Machine Learning Portfolio never beat the Modified Index, which is its benchmark. If the FIN15 is ignored in Figure 7, the Machine Learning Portfolio is the poorest performing portfolio. This is because the Machine Learning Portfolio only serves to replicate the Modified Index. It is not constructed with any attribute that can minimise losses or maximises returns compared to its benchmark. From the correlation matrix in table 8 it can be seen that the Machine Learning Portfolio has a high correlation with the Modified index. The Machine Learning Portfolio seeks only to replicate the performance of the Modified Index, not beat it.

3.

5.1 Risk(Volatility) versus Reward(Returns/losses)

By assessing reward and volatility using measures such as the Sharp Ratio for the portfolios and plotting the Portfolio Return vs Portfolio Volatility as shown in Figure 9 An investor is disincentivised to select the Machine Learning Portfolio. This is because as the losses (negative returns) offered come at a higher risk, and are greater than losses presented by the Minimum Variance Portfolio and the Modified Index. Ultimately the Minimum variance Portfolio outperforms the Machine Learning portfolio in the metric of Excess returns as well. Meaning that it is a better choice in this metric as well.

6. SUMMARY, CONCLUSIONS, LIMITATIONS, AND RECOMMENDATIONS

6.1 Summary

We are reminded of the research question that was embarked on for the study. That being 'Can two portfolios, comprised of the FTSE/JSE Financial 15 constituents, be optimised using a combination of Mean-Variance Optimisation and Machine Learning to produce excess returns about their benchmark index? In the study, we embarked on an experimental study in which the two research questions were to be tested using stock data from the Yahoo Finance Database. The hypothesis that was tested being:

The Machine Learning Technique (Support Vector Regression) can outperform the Mean-Variance framework Analysis in constructing optimal portfolios from the FTSE / JSE Financial 15.

The Literature of Modern Portfolio Theory and Machine Learning were explored, and the conceptual framework with which to evaluate and test the hypothesis was decided to be Mean-Variance Analysis.
The expected returns and volatility of the portfolios were calculated, and the two portfolios constructed and compared with regards to performance, volatility, total returns, excess returns with respect to the index benchmark.

6.2 Conclusions

This study showed that a machine learning portfolio could not improve or outperform an optimised minimum variance portfolio. The machine learning portfolio seeks only to replicate the portfolio and not minimise volatility. Volatility is not used in the training and fitting of the regression model. This serves better as a means of replicating the index and not improving an investor's returns.

The portfolio weights calculated by the Minimum-Variance and the Machine Learning Portfolio are calculated using different objectives. The Minimum Variance Portfolio seeks to use convex optimisation to minimise the variance of the portfolio. Meanwhile, the Machine Learning Portfolio weights result from minimising the margin error in the Support Vector Regression hyperplane.
The Risk and Return relationship for both the training period and testing period differs radically in nature due to the systematic shock effect of the Covid19 Lockdown on South African Markets starting in March 2019. Regardless of this, the constructed portfolios all show lower risk and higher returns for the testing period of June 2019 to December 2020.

Regarding the performance of the Minimum-Variance Portfolio compared to the Modified Index Benchmark and the Machine Learning Portfolio, the findings of this study corroborate the findings of other studies in that they confirm that Mean-Variance Optimisation is an effective framework for the construction of an optimal portfolio of risky assets.

The Machine Learning Portfolio generates losses above those on the Modified Index benchmark from March 2020 and fails to climb up above the benchmark. Meanwhile, the Minimum Variance portfolio excess returns do erode concerning the benchmark from March 2020, but the Minimum Variance portfolio still exhibits positive returns compared to the benchmark.

Regarding the question(Section 1.2.3) as to whether a Minimum Variance portfolio produces returns above the FTSE/JSE Financial 15 index on which its stocks are based? The research hypothesis is confirmed when checked against the results shown in the plots showing the excess returns and total returns for the period.

Regarding the hypothesis which states, ' The Machine Learning Technique (Support Vector Regression) can outperform the Mean-Variance framework Analysis in constructing optimal portfolios from the FTSE / JSE Financial 15.'

The research hypothesis can be confidently rejected. This can be seen in the poor performance of the Machine Learning Portfolio with regards to excess and total returns compared to the benchmark. The Machine Learning Portfolio also performed poorly compared to the Minimum-variance portfolio. The Machine Learning Portfolio also showed high risk with a lower return for the period of the experiment(June 2019 – December 2020)

6.3 Recommendations

In future research the transaction costs of rebalancing of the portfolio must be taken into account to give a more realistic picture of the overall performance of an Index beating portfolio. The use of Support Vector Regression may also be more suited to the task of Return Forecasting, this must be compared with its suitability to Portfolio Construction as laid out in this paper.

6.4 Investment Management Implications

The aim of this study is to:

1. Construct an optimal portfolio using the Mean-Variance Analysis Framework.
2. Construct an optimal portfolio using a Machine Learning Technique (Support Vector Regression).
3. Contrast the results of the Minimum-Variance Portfolio and the Machine Learning Portfolio.

The results garnered from the listed objectives of the study will help investment practitioners who are seeking to outperform the Mean-Variance Analysis Framework on the FTSE / JSE FIN15 index. The Investment Management implications of this study will serve as a guide to making good Investment Management decisions. The framework within which these decisions can be made is generated by a comparison of the results of the Machine Learning Portfolio but also on the comparison to the recommendations made by Modern Portfolio theory generally. The management implications are summarised below.

Managerial Questions	ML Portfolio	MV Portfolio	Best Guidelines
Is the portfolio of holdings efficiently diversified?	Assets that do not contribute significantly to prediction of expected returns are removed from portfolio	Assets must not be perfectly positively correlated to one another.	**MV Portfolio:** The best portfolio is efficiently diversified when correlation between assets is minimised
What is the risk to the investor?	Highest when compared to the MV Portfolio and the Index	Lower than that of the ML Portfolio but higher than the index	**Index** The best portfolio has the highest reward but lowest risk
Which portfolio will give greater returns?	Losses are greater than that of the MV Portfolio and even the index	Losses are lower than the index and ML Portfolio	**MV Portfolio** The best portfolio has lowest losses

REFERENCES

Adams, J., Khan, H. T., Raeside, R., & White, D. I. (2007). Research methods for graduate business and social science students: SAGE publications India.

Bahramy, F., & Crone, S. F. (2013). Forecasting foreign exchange rates using Support Vector Regression. Paper presented at the 2013 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr).

Ban, G.-Y., El Karoui, N., & Lim, A. E. (2018). Machine learning and portfolio optimization. Management Science, 64(3), 1136-1154.

Block, A. E. (2017). The optimally diversified equity portfolio in South Africa: an artificial intelligence approach.

Bodie, Z., Kane, A., & Marcus, A. (2014). Investments (Tenth Glob). In: McGraw Hill Education, Bershire.

Bryman, A. (2016). Social research methods: Oxford university press.

Carte, D. (2009). The All-Share Index: investing 101. Personal Finance, 2009(341), 9-10.

Cizauskas, K., & Haslifah, D. (2019). Revisiting Markowitz: portfolio optimisation including higher moments of return.

Cooper, D. R., Schindler, P. S., & Sun, J. (2003). Business research methods.

Drue, S. (2018). A comparison of a factor-based investment strategy and machine learning for predicting excess returns on the JSE. Faculty of Science,

Emerson, S., Kennedy, R., O'Shea, L., & O'Brien, J. (2019). Trends and applications of machine learning in quantitative finance. Paper presented at the 8th International Conference on Economics and Finance Research (ICEFR 2019).

EquityRT. (2020). Retrieved from https://0-www-equityrt-com.innopac.wits.ac.za/

Fama, E. F. (1960). Efficient market hypothesis. Diss. PhD Thesis, Ph. D. dissertation.

Fama, E. F., & French, K. R. (2000). Forecasting profitability and earnings. The Journal of Business, 73(2), 161-175.

Fama, E. F., & French, K. R. (2021a). Common risk factors in the returns on stocks and bonds: University of Chicago Press.

Fama, E. F., & French, K. R. (2021b). The cross-section of expected stock returns: University of Chicago Press.

FTSE/JSE Africa Index Series. (2020). Retrieved from https://www.jse.co.za/sites/default/files/media/documents/2021-01/FTSE%20JSE%20Ground%20Rules%20v7%200%20Aug%202020.pdf

George, N. (Producer). (Modern Portfolio Theory, Efficient Frontiers). Retrieved from https://s3.amazonaws.com/assets.datacamp.com/production/course_6578/slides/chapter4.pdf

Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems: O'Reilly Media.

Glabadanidis, P. (2020). Portfolio strategies to track and outperform a benchmark. Journal of Risk and Financial Management, 13(8), 171.

Graham, B., Dodd, D. L. F., & Cottle, S. (1934). Security analysis: McGraw-Hill New York.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction: Springer Science & Business Media.

Hudson, R. S., & Gregoriou, A. (2015). Calculating and comparing security returns is harder than you think: A comparison between logarithmic and simple returns. International Review of Financial Analysis, 38, 151-162.

Huni, S., & Sibindi, A. B. (2020). An Application of The Markowitz's Mean-Variance Framework in Constructing Optimal Portfolios using the Johannesburg Securities Exchange Tradeable Indices. The Journal of Accounting and Management, 10(2).

Hurwitz, E., & Marwala, T. (2009). State of the Art Review for Applying Computational Intelligence and Machine Learning Techniques to Portfolio Optimisation. arXiv preprint arXiv:0910.2276.

Ilmanen, A. (2011). Expected returns: An investor's guide to harvesting market rewards: John Wiley & Sons.

Ivanova, M., & Dospatliev, L. (2017). Application of Markowitz portfolio optimization on Bulgarian stock market from 2013 to 2016. International Journal of Pure and Applied Mathematics, 117(2), 291-307.

Jordan, B. D., Miller, T. W., & Dolvin, S. D. (2015). Fundamentals of investments: valuation and management: McGraw-Hill Education.

JSE. (2020). Stock Exchange News Service. In.

Kotze, A. (2017). FTSE/JSE Top 40 Index Long-Term Returns. Available at SSRN 2978093.

Leedy, P. D., & Ormrod, J. E. (2014). Practical research: Planning and design: Pearson Education.

Lombard, C. F. (2015). Decision-making under uncertainty: Markowitz optimisation as a passive strategy on the JSE. North-West University (South Africa), Potchefstroom Campus,

Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. Expert Systems with Applications, 165, 113973.

Markowitz, H. (1959). Portfolio selection. In: Yale University Press New Haven.

Markowitz, H. M. (1991). Foundations of portfolio theory. The journal of Finance, 46(2), 469-477.

Martin, R. A. (2018). Mean-Variance Optimisation. Retrieved from https://pyportfolioopt.readthedocs.io/en/latest/MeanVariance.html

McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython: " O'Reilly Media, Inc.".

Oldham, G., & Kroeger, J.-A. (2005). Performance, persistence and benchmarks of selected South African unit trusts for the period 1998-2002. South African Journal of Business Management, 36(4), 81-90.

Paiva, F. D., Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2019). Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. Expert Systems with Applications, 115, 635-655.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. Expert Systems with Applications, 42(4), 2162-2172.

Perrin, S., & Roncalli, T. (2020). Machine learning optimization algorithms & portfolio allocation. Machine Learning for Asset Management: New Developments and Financial Applications, 261-328.

Rossi, A. G. (2018). Predicting stock market returns with machine learning. Georgetown University.

Satrix. (2020). Satrix FINI ETF Minimum Disclosure Document. Retrieved from https://satrix.co.za/media/51753?inline=true

Scott, G. (2020). Adjusted Closing Price. Retrieved from https://www.investopedia.com/terms/a/adjusted_closing_price.asp

Sharpe, W. F. (1994). The sharpe ratio. Journal of portfolio management, 21(1), 49-58.

Silva, V. (2020). The Science of Portfolio Optimization. Retrieved from https://towardsdatascience.com/the-science-of-portfolio-optimization-186607d30416

Tadlaoui, G. (2018). Intelligent Portfolio Construction: Machine-Learning enabled Mean-Variance Optimization.

Van Rensburg, J. J. (2015). Economics / Jan Janse van Rensburg, Campbell McConnell and Stanley Brue. In (pp. 0-2).

Vitali, F., & Mollah, S. (2010). Stock market efficiency in Africa: Evidence from random walk hypothesis. South Western Finance, 1-54.

Wessels, D. R., & Krige, J. D. (2005). Active versus passive investing I : the South African experience. Studies in Economics and Econometrics, 29(2), 1-33. Retrieved from https://journals.co.za/content/bersee/29/2/EJC21408

Yu, J.-R., Chiou, W.-J. P., Lee, W.-Y., & Lin, S.-J. (2020). Portfolio models with return forecasting and transaction costs. International Review of Economics & Finance, 66, 118-130.

Zaidy, H. (2018). Michael Jordaan stakes his retirement on machine learning at AI-powered fund. Business Day. Retrieved from https://www.businesslive.co.za/bd/companies/financial-services/2018-06-18-fintech-powered-nmrql-research-seeks-an-edge-with-machine-learning/

Zhang, Z., Zohren, S., & Roberts, S. (2020). Deep Learning for Portfolio Optimisation. arXiv preprint arXiv:2005.13665.

ZongMing, T., Koomson, P., & Guoping, D. (2017). Investment Risk and Returns: The Relationship Between A Stock and An Index Using the Modern Portfolio Theory. Available at SSRN 3014223.

APPENDICES

Appendix 1.1: Data collection instrument(s)

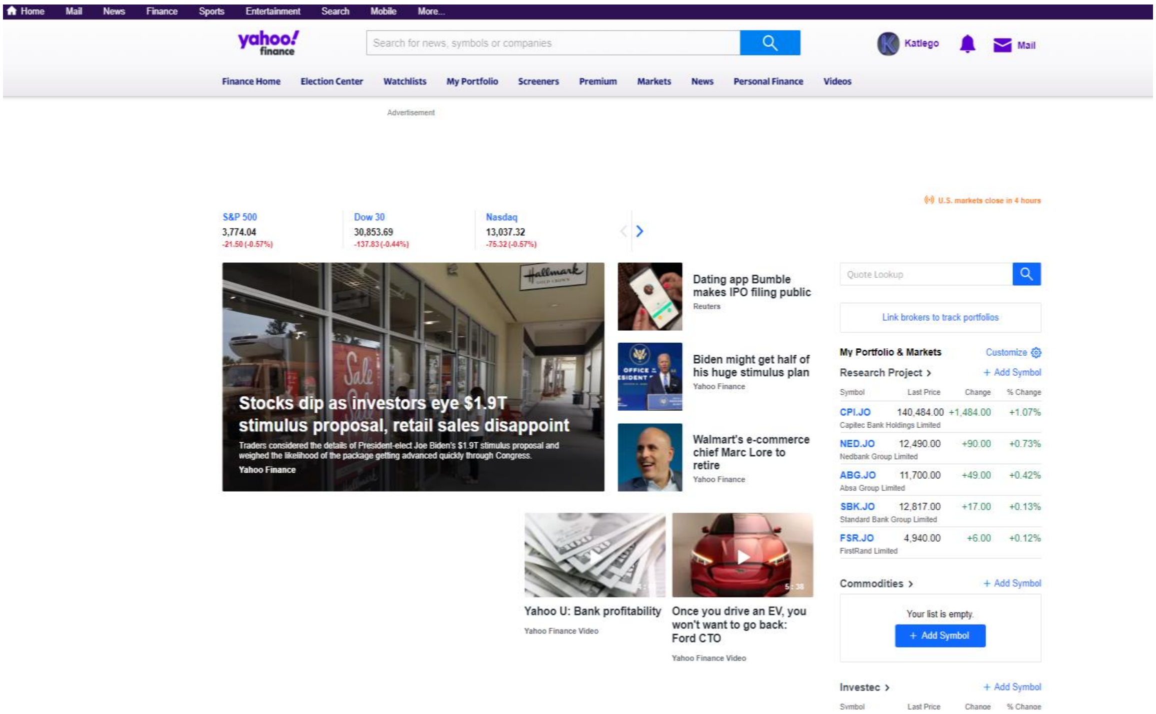


Figure 12: Screenshot of Yahoo Finance accessed: 15 Jan 2021

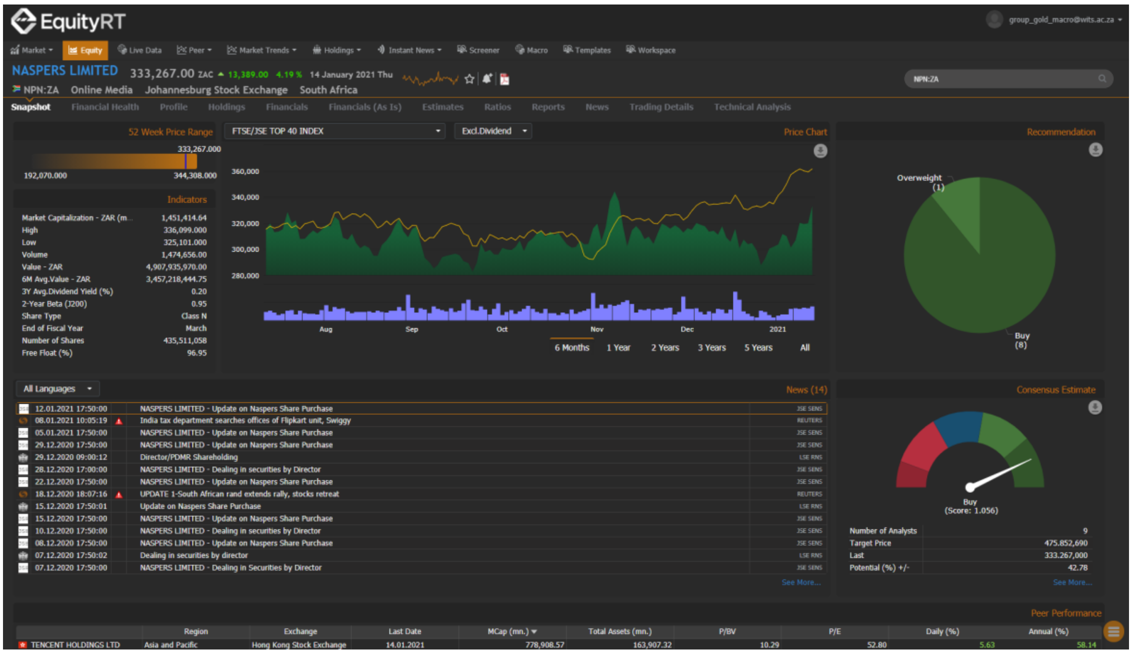


Figure 13: Screenshot of EquityRT accessed via Wits proxy server access date: 1 Jan 2021

Appendix 1.1: Data collection instrument(s) – Rebalancing of STXFIN

Table 10: Rebalance history of FIN15 / STXFIN for the period 22 December 2014 to 21 December 2015(JSE, 2020)

			22-Dec-14		23-Mar-15		22-Jun-15		21-Sep-15		21-Dec-15	
No	Company	Code	Preceding Weight	New Weight	Preceding Weight	New Weight	Preceding Weight	New Weight	Preceding Weight	New Weight	Preceding Weight	New Weight
1	ABSA(Formerly Barclays Africa, BGA)	ABG	5.24%	5.27%	4.89%	4.79%	4.95%	4.86%	4.82%	4.82%	4.82%	4.37%
2	Brait SE	BAT							0.00%	3.83%	3.83%	5.01%
3	Capital & Counties Properties plc	CCO	1.31%	1.58%	1.56%	1.85%	2.13%	2.12%	2.37%	2.34%	2.34%	2.80%
4	Capitec	CPI									0.00%	2.62%
5	Discovery	DSY	2.97%	2.99%	3.39%	3.31%	3.34%	3.35%	3.72%	3.74%	3.74%	3.94%
6	FirstRand	FSR	14.62%	13.70%	13.39%	13.51%	13.35%	13.43%	13.70%	13.47%	13.47%	12.11%
7	Fortress Income Fund Ltd A	FFA										
8	Fortress Income Fund Ltd B	FFB										
9	Growthpoint Properties Ltd	GRT	5.20%	5.34%	4.96%	5.05%	5.43%	5.57%	5.56%	5.62%	5.62%	5.33%
10	Intu Properties plc	ITU	5.05%	5.08%	4.86%	4.83%	4.80%	4.78%	5.24%	5.26%	5.26%	6.13%
11	Investec Ltd	INL	2.18%	2.20%	2.15%	2.13%	2.28%	2.25%	2.30%	2.30%	2.30%	2.49%
12	Investec plc	INP	5.24%	5.27%	5.25%	5.18%	5.58%	5.54%	5.64%	5.45%	5.45%	5.88%
13	MMI Holdings Ltd	MMI	2.80%	2.82%	2.80%	2.72%	2.60%	2.51%	2.26%	0.00%		
14	Nedbank Group Ltd	NED	4.34%	4.37%	3.95%	3.89%	3.92%	3.92%	3.81%	3.78%	3.78%	3.44%
15	NEPI Rockcastle PLC	NRP										
16	Old Mutual Limited(Previously Old Mutual Plc, OML)	OMU	14.43%	14.52%	15.98%	15.97%	16.02%	16.22%	16.60%	15.89%	15.89%	17.27%
18	PSG	PSG										
19	Quilter PLC	QLT										
20	Rand Merchant Insurance Holdings Ltd	RMI	2.63%	2.65%	2.60%	2.57%	2.51%	2.51%	2.57%	2.53%	2.53%	2.51%
23	Redefine Properties Ltd	RDF										
24	Reinet Investments SCA(Previously Reinet Inv Soc Anon, RNI)	REI	3.39%	3.41%	3.12%	3.07%	3.30%	3.29%	3.61%	3.56%	3.56%	0.00%
26	Remgro Ltd	REM										
27	RMB Holdings Ltd	RMH	3.83%	3.86%	3.69%	3.75%	3.60%	3.63%	3.83%	3.81%	3.81%	3.56%
28	Sanlam	SLM	11.15%	11.22%	11.40%	11.35%	10.21%	10.11%	9.23%	9.05%	9.05%	9.52%

No	Company	Code	22-Dec-14		23-Mar-15		22-Jun-15		21-Sep-15		21-Dec-15	
			Preceding Weight	New Weight	Preceding Weight	New Weight	Preceding Weight	New Weight	Preceding Weight	New Weight	Preceding Weight	New Weight
29	Standard Bank Group Ltd	SBK	15.62%	15.72%	16.01%	16.03%	15.98%	15.91%	14.74%	14.55%	14.55%	13.04%
Total Weight			100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.02%

Table 11: Rebalance history of FIN15 / STXFIN for the period 21 December 2015 to 19 December 2016(JSE, 2020)

			21-Dec-15		22-Mar-16		20-Jun-16		19-Sep-16		19-Dec-16	
No	**Company**	**Code**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**
1	ABSA(Formerly Barclays Africa, BGA)	ABG	4.82%	4.37%	4.37%	4.14%	4.14%	5.54%	5.54%	5.44%	5.85%	6.08%
2	Brait SE	BAT	3.83%	5.01%	5.01%	4.73%	4.73%	4.17%	4.17%	3.31%	2.45%	2.39%
3	Capital & Counties Properties plc	CCO	2.34%	2.80%	2.80%	1.96%	1.96%	1.98%	1.98%	0.00%		
4	Capitec	CPI	0.00%	2.62%	2.62%	2.63%	2.63%	2.77%				
5	Discovery	DSY	3.74%	3.94%	3.94%	3.44%	3.44%	3.56%	3.56%	3.33%	3.14%	3.13%
6	FirstRand	FSR	13.47%	12.11%	12.11%	12.37%	12.37%	12.33%	12.33%	12.52%	13.43%	13.75%
7	Fortress Income Fund Ltd A	FFA							0.00%	1.53%	1.55%	1.58%
8	Fortress Income Fund Ltd B	FFB							2.77%	2.03%	1.90%	2.06%
9	Growthpoint Properties Ltd	GRT	5.62%	5.33%	5.33%	5.89%	5.89%	6.26%	6.26%	5.92%	5.84%	5.77%
10	Intu Properties plc	ITU	5.26%	6.13%	6.13%	5.68%	5.68%	4.78%	4.78%	4.39%	3.88%	3.83%
11	Investec Ltd	INL	2.30%	2.49%	2.49%	2.45%	2.45%	2.07%	2.07%	2.02%	2.12%	2.15%
12	Investec plc	INP	5.45%	5.88%	5.88%	5.70%	5.70%	4.86%	4.86%	4.77%	5.03%	5.14%
13	MMI Holdings Ltd	MMI										
14	Nedbank Group Ltd	NED	3.78%	3.44%	3.44%	3.24%	3.24%	3.68%	3.68%	4.06%	4.33%	4.32%
15	NEPI Rockcastle PLC	NRP										
16	Old Mutual Limited(Previously Old Mutual Plc, OML)	OMU	15.89%	17.27%	17.27%	17.71%	17.71%	16.44%	16.44%	14.74%	13.73%	13.58%
18	PSG	PSG										
19	Quilter PLC	QLT										
20	Rand Merchant Insurance Holdings Ltd	RMI	2.53%	2.51%	2.51%	2.57%	2.57%	0.00%				
23	Redefine Properties Ltd	RDF							0.00%	4.46%	4.14%	4.07%
24	Reinet Investments SCA(Previously Reinet Inv Soc Anon, RNI)	REI	3.56%	0.00%			0.00%	4.67%	4.67%	3.94%	3.57%	3.39%
26	Remgro Ltd	REM										
27	RMB Holdings Ltd	RMH	3.81%	3.56%	3.56%	3.48%	3.48%	3.36%	3.36%	3.46%	3.63%	3.73%
28	Sanlam	SLM	9.05%	9.52%	9.52%	10.17%	10.17%	9.56%	9.56%	9.74%	9.49%	9.28%
29	Standard Bank Group Ltd	SBK	14.55%	13.04%	13.04%	13.84%	13.84%	13.99%	13.99%	14.35%	15.94%	15.76%
Total Weight			100.00%	100.00%	100.00%	100.02%	100.02%	100.00%	100.00%	100.02%	100.02%	100.01%

Table 12: Rebalance history of FIN15 / STXFIN for the period 19 December 2016 to 20 December 2017(JSE, 2020)

			19-Dec-16		17-Mar-17		21-Jun-17		19-Sep-17		20-Dec-17	
No	**Company**	**Code**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**
1	ABSA(Formerly Barclays Africa, BGA)	ABG	5.85%	6.08%	6.08%	5.34%	5.34%	7.98%	7.98%	6.89%	6.89%	8.16%
2	Brait SE	BAT	2.45%	2.39%	2.39%	0.00%						
3	Capital & Counties Properties plc	CCO										
4	Capitec	CPI			0.00%	3.34%	3.34%	3.35%	3.35%	3.62%	3.62%	3.65%
5	Discovery	DSY	3.14%	3.13%	3.13%	3.52%	3.52%	3.57%	3.57%	3.79%	3.79%	4.13%
6	FirstRand	FSR	13.43%	13.75%	13.75%	12.96%	12.96%	12.70%	12.70%	13.22%	13.22%	13.72%
7	Fortress Income Fund Ltd A	FFA	1.55%	1.58%	1.58%	1.59%	1.59%	1.59%	1.59%	1.51%	1.51%	1.38%
8	Fortress Income Fund Ltd B	FFB	1.90%	2.06%	2.06%	2.08%	2.08%	2.16%	2.16%	2.27%	2.27%	2.07%
9	Growthpoint Properties Ltd	GRT	5.84%	5.77%	5.77%	6.00%	6.00%	5.68%	5.68%	5.47%	5.47%	4.94%
10	Intu Properties plc	ITU	3.88%	3.83%	3.83%	3.40%	3.40%	3.28%	3.28%	0.00%		
11	Investec Ltd	INL	2.12%	2.15%	2.15%	2.09%	2.09%	2.17%	2.17%	2.15%	2.15%	1.61%
12	Investec plc	INP	5.03%	5.14%	5.14%	4.83%	4.83%	4.97%	4.97%	4.92%	4.92%	3.70%
13	MMI Holdings Ltd	MMI										
14	Nedbank Group Ltd	NED	4.33%	4.32%	4.32%	4.59%	4.59%	3.68%	3.68%	3.31%	3.31%	3.52%
15	NEPI Rockcastle PLC	NRP							0.00%	4.60%	4.60%	4.81%
16	Old Mutual Limited(Previously Old Mutual Plc, OML)	OMU	13.73%	13.58%	13.58%	13.34%	13.33%	13.33%	13.34%	13.06%	13.06%	12.09%
18	PSG	PSG										
19	Quilter PLC	QLT										
20	Rand Merchant Insurance Holdings Ltd	RMI										
23	Redefine Properties Ltd	RDF	4.14%	4.07%	4.07%	4.43%	4.43%	4.41%	4.41%	4.22%	4.22%	3.63%
24	Reinet Investments SCA(Previously Reinet Inv Soc Anon, RNI)	REI	3.57%	3.39%	3.39%	3.27%	3.27%	3.66%	3.66%	3.27%	3.27%	2.57%
26	Remgro Ltd	REM										
27	RMB Holdings Ltd	RMH	3.63%	3.73%	3.73%	3.51%	3.51%	3.25%	3.25%	3.35%	3.35%	3.30%
28	Sanlam	SLM	9.49%	9.28%	9.28%	10.20%	10.20%	9.11%	9.11%	8.83%	8.83%	10.24%
29	Standard Bank Group Ltd	SBK	15.94%	15.76%	15.76%	15.52%	15.52%	15.08%	15.08%	15.53%	15.53%	16.48%
Total Weight			100.02%	100.01%	100.01%	100.01%	100.00%	99.97%	99.98%	100.01%	100.01%	100.00%

Table 13: Rebalance history of FIN15 / STXFIN for the period 20 December 2017 to 24 December 2018(JSE, 2020)

			20-Dec-17		19-Mar-18		20-Jun-18		25-Sep-18		24-Dec-18	
No	**Company**	**Code**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**
1	ABSA(Formerly Barclays Africa, BGA)	ABG	6.89%	8.16%	8.16%	8.31%	8.31%	8.10%	8.10%	7.69%	7.69%	8.14%
2	Brait SE	BAT										
3	Capital & Counties Properties plc	CCO										
4	Capitec	CPI	3.62%	3.65%	3.65%	3.11%	3.11%	3.33%	3.33%	3.63%	3.63%	4.21%
5	Discovery	DSY	3.79%	4.13%	4.13%	3.94%	3.94%	3.54%	3.54%	4.19%	4.19%	3.91%
6	FirstRand	FSR	13.22%	13.72%	13.72%	14.17%	14.17%	13.35%	13.35%	15.87%	15.87%	14.95%
7	Fortress Income Fund Ltd A	FFA	1.51%	1.38%	1.38%	1.22%	1.22%	0.00%				
8	Fortress Income Fund Ltd B	FFB	2.27%	2.07%	2.07%	0.76%	0.76%	0.00%				
9	Growthpoint Properties Ltd	GRT	5.47%	4.94%	4.94%	5.29%	5.29%	5.41%	5.41%	4.66%	4.66%	4.97%
10	Intu Properties plc	ITU										
11	Investec Ltd	INL	2.15%	1.61%	1.61%	1.78%	1.78%	2.00%	2.00%	2.22%	2.22%	1.75%
12	Investec plc	INP	4.92%	3.70%	3.70%	4.09%	4.09%	4.68%	4.68%	4.83%	4.83%	3.82%
13	MMI Holdings Ltd	MMI										
14	Nedbank Group Ltd	NED	3.31%	3.52%	3.52%	3.95%	3.95%	4.05%	4.05%	4.13%	4.13%	7.98%
15	NEPI Rockcastle PLC	NRP	4.60%	4.81%	4.81%	2.92%	2.92%	3.04%	3.04%	3.22%	3.22%	2.67%
16	Old Mutual Limited(Previously Old Mutual Plc, OML)	OMU	13.06%	12.09%	12.09%	12.98%	12.98%	13.72%	13.72%	10.47%	10.47%	8.18%
18	PSG	PSG					0.00%	2.25%	2.25%	2.56%	2.56%	2.80%
19	Quilter PLC	QLT										
20	Rand Merchant Insurance Holdings Ltd	RMI										
23	Redefine Properties Ltd	RDF	4.22%	3.63%	3.63%	3.73%	3.73%	3.97%	3.97%	3.71%	3.71%	3.75%
24	Reinet Investments SCA(Previously Reinet Inv Soc Anon, RNI)	REI	3.27%	2.57%	2.57%	2.11%	2.11%	2.58%	2.58%	2.77%	2.77%	2.20%
26	Remgro Ltd	REM										
27	RMB Holdings Ltd	RMH	3.35%	3.30%	3.30%	3.45%	3.45%	3.41%	3.41%	3.80%	3.80%	4.28%
28	Sanlam	SLM	8.83%	10.24%	10.24%	10.01%	10.01%	9.02%	9.02%	9.82%	9.82%	9.78%
29	Standard Bank Group Ltd	SBK	15.53%	16.48%	16.48%	18.18%	18.18%	17.53%	17.53%	16.41%	16.41%	16.63%
Total Weight			100.01%	100.00%	100.00%	100.00%	100.00%	99.98%	99.98%	99.98%	99.98%	100.02%

Table 14: Rebalance history of FIN15 / STXFIN for the period 24 December 2018 to 23 September 2019(JSE, 2020)

No	Company	Code	24-Dec-18 Preceding Weight	24-Dec-18 New Weight	18-Mar-19 Preceding Weight	18-Mar-19 New Weight	24-Jun-19 Preceding Weight	24-Jun-19 New Weight	23-Sep-19 Preceding Weight	23-Sep-19 New Weight
1	ABSA(Formerly Barclays Africa, BGA)	ABG	7.69%	8.14%	8.14%	8.32%	8.32%	8.77%	8.77%	8.27%
2	Brait SE	BAT								
3	Capital & Counties Properties plc	CCO								
4	Capitec	CPI	3.63%	4.21%	4.21%	4.91%	4.91%	4.66%	4.66%	4.78%
5	Discovery	DSY	4.19%	3.91%	3.91%	3.48%	3.48%	3.72%	3.72%	3.02%
6	FirstRand	FSR	15.87%	14.95%	14.95%	14.49%	14.49%	14.99%	14.99%	14.44%
7	Fortress Income Fund Ltd A	FFA								
8	Fortress Income Fund Ltd B	FFB								
9	Growthpoint Properties Ltd	GRT	4.66%	4.97%	4.97%	4.97%	4.97%	4.78%	4.78%	4.83%
10	Intu Properties plc	ITU								
11	Investec Ltd	INL	2.22%	1.75%	1.75%	1.94%	1.94%	1.87%	1.87%	1.83%
12	Investec plc	INP	4.83%	3.82%	3.82%	4.19%	4.19%	4.05%	4.05%	3.94%
13	MMI Holdings Ltd	MMI								
14	Nedbank Group Ltd	NED	4.13%	7.98%	7.98%	7.36%	7.36%	7.01%	7.01%	6.62%
15	NEPI Rockcastle PLC	NRP	3.22%	2.67%	2.67%	2.93%	2.93%	3.10%	3.10%	3.35%
16	Old Mutual Limited(Previously Old Mutual Plc, OML)	OMU	10.47%	8.18%	8.18%	7.46%	7.46%	7.13%	7.13%	6.65%
18	PSG	PSG	2.56%	2.80%	2.80%	2.98%	2.98%	2.69%	2.69%	0.00%
19	Quilter PLC	QLT								
20	Rand Merchant Insurance Holdings Ltd	RMI								
23	Redefine Properties Ltd	RDF	3.71%	3.75%	3.75%	3.89%	3.89%	3.49%	3.49%	3.22%
24	Reinet Investments SCA(Previously Reinet Inv Soc Anon, RNI)	REI	2.77%	2.20%	2.20%	2.49%	2.49%	2.29%	2.29%	2.87%
26	Remgro Ltd	REM							0.00%	6.40%
27	RMB Holdings Ltd	RMH	3.80%	4.28%	4.28%	4.12%	4.12%	4.35%	4.35%	4.15%
28	Sanlam	SLM	9.82%	9.78%	9.78%	9.10%	9.10%	9.42%	9.42%	9.36%
29	Standard Bank Group Ltd	SBK	16.41%	16.63%	16.63%	17.37%	17.37%	17.67%	17.67%	16.25%
Total Weight			99.98%	100.02%	100.02%	100.00%	100.00%	99.99%	99.99%	99.98%

Table 15: Rebalance history of FIN15 / STXFIN for the period 23 September 2019 to 22 December 2020(JSE, 2020)

			23-Sep-19		22-Jun-20		21-Sep-20		21-Dec-20		22-Dec-20	
No	**Company**	**Code**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**	**Preceding Weight**	**New Weight**
1	ABSA(Formerly Barclays Africa, BGA)	ABG	8.77%	8.27%	7.60%	7.31%	7.31%	6.96%	6.96%	7.44%	7.44%	7.64%
2	Brait SE	BAT										
3	Capital & Counties Properties plc	CCO										
4	Capitec	CPI	4.66%	4.78%	6.44%	5.20%	5.20%	8.22%	8.22%	10.96%	10.96%	10.17%
5	Discovery	DSY	3.72%	3.02%	3.06%	4.35%	4.35%	5.39%	5.39%	5.31%	5.31%	4.86%
6	FirstRand	FSR	14.99%	14.44%	14.66%	13.91%	13.91%	20.50%	20.50%	22.22%	22.22%	22.16%
7	Fortress Income Fund Ltd A	FFA										
8	Fortress Income Fund Ltd B	FFB										
9	Growthpoint Properties Ltd	GRT	4.78%	4.83%	4.82%	4.67%	4.67%	4.10%	4.10%	3.80%	3.80%	3.85%
10	Intu Properties plc	ITU										
11	Investec Ltd	INL	1.87%	1.83%	1.13%	1.19%	1.19%	0.97%	0.97%	0.92%	0.92%	1.03%
12	Investec plc	INP	4.05%	3.94%	2.38%	2.66%	2.66%	2.23%	2.23%	2.26%	2.26%	2.56%
13	MMI Holdings Ltd	MMI										
14	Nedbank Group Ltd	NED	7.01%	6.62%	4.40%	4.29%	4.29%	3.91%	3.91%	3.95%	3.95%	3.88%
15	NEPI Rockcastle PLC	NRP	3.10%	3.35%	3.20%	3.72%	3.72%	2.77%	2.77%	3.42%	3.42%	3.28%
16	Old Mutual Limited(Previously Old Mutual Plc, OML)	OMU	7.13%	6.65%	6.17%	6.09%	6.09%	5.21%	5.21%	4.81%	4.81%	4.95%
18	PSG	PSG	2.69%	0.00%			0.00%	0.81%	0.81%	0.00%		
19	Quilter PLC	QLT			0.00%	3.16%	3.16%	3.22%	3.22%	2.38%	2.38%	2.64%
20	Rand Merchant Insurance Holdings Ltd	RMI							0.00%	2.13%	2.13%	1.98%
23	Redefine Properties Ltd	RDF	3.49%	3.22%	1.55%	0.00%						
24	Reinet Investments SCA(Previously Reinet Inv Soc Anon, RNI)	REI	2.29%	2.87%	4.93%	4.76%	4.76%	4.99%	4.99%	3.40%	3.40%	3.57%
26	Remgro Ltd	REM	0.00%	6.40%	7.73%	5.90%	5.90%	5.08%	5.08%	4.23%	4.23%	4.59%
27	RMB Holdings Ltd	RMH	4.35%	4.15%	4.31%	6.48%	6.48%	0.00%				
28	Sanlam	SLM	9.42%	9.36%	11.56%	11.53%	11.53%	9.99%	9.99%	8.30%	8.30%	8.79%
29	Standard Bank Group Ltd	SBK	17.67%	16.25%	16.06%	14.78%	14.78%	15.65%	15.65%	14.47%	14.47%	14.05%
Total Weight			99.99%	99.98%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Appendix 1.2: Data collection instrument(s) – Python Code compiled by author

1.2.1 Machine Learning Portfolio

```
# -*- coding: utf-8 -*-
"""SVR Predictions - Mean Variance Optimisation of A South African Index Based
Portfolio Using Machine Learning V3

Automatically generated by Colaboratory.

Original file is located at
    https://colab.research.google.com/drive/10Q8qhjfb6PssnYtYM0n1VX5SJVCQGBuh

**Preprocessing**
"""

##Mount Google drive so that datasets can be read from here.
from google.colab import drive
drive.mount('/content/drive')

#Import libraries
import numpy as np
import pandas as pd
from pandas_datareader import data as pdr
import datetime as dt
import matplotlib.pyplot as plt

##Portfolio optimisation libraries(pyPortopt)
!pip install PyPortfolioOpt
from pypfopt import expected_returns

#import machine learning libraries
from sklearn.svm import SVR
from sklearn.metrics import mean_squared_error

#ignore warnings
import warnings
warnings.filterwarnings('ignore')

#Create function to retrieve price data
def getPrices(assets, start, end):
      assetData = pdr.get_data_yahoo(assetsList, start=start, end=end)
      assetData = assetData['Adj Close']
      return assetData

#Define periods for training(22 December 2014 - 21 June 2019)
#Take note that training and testing vectors are not the same size, therefore
score function cannot be applied
StartDateTraining = '12/22/2014'
EndDateTraining = '06/21/2019'

"""**Extract, Clean, Transform and Load Training Data**"""

#Dependant variable[portfolio constituents]
#Extract price data for portfolio constituents
#Calculate log returns for training the SVR model
assetsList = ['ABG','CPI'     ,'DSY' ,'FSR' ,'GRT' ,'INL' ,'INP' ,'NED'
        ,'SBK', 'SLM']
assetsList = [stock + '.JO' for stock in assetsList]
```

```
AssetsPriceData_Training = getPrices(assetsList, StartDateTraining,
EndDateTraining)
AssetsReturnsData_Training =
expected_returns.returns_from_prices(AssetsPriceData_Training,
log_returns=False)

weightsFin15 = np.array([0.08662,     0.03392,       0.06382,       0.17092,
        0.08732,       0.05592,       0.08662,       0.07762,       0.19112,
        0.14612])

#Sort returns by date
AssetsReturnsData_Training = AssetsReturnsData_Training.sort_values('Date')

#Independant Variable[Index]
assetsList = ['STXFIN.JO']

#Calculate log returns for training and fitting, Dependant variable
IndexPriceData_Training = getPrices(assetsList, StartDateTraining,
EndDateTraining)
IndexReturnsData_Training =
expected_returns.returns_from_prices(IndexPriceData_Training, log_returns=False)

#Sort returns by date
IndexReturnsData_Training = IndexReturnsData_Training.sort_values('Date')

#Remove public holidays from assets list in the training data
#specify the index
Train_PublicHolidays2015 = ["12/25/2014","12/26/2014","1/1/2015","4/3/2015",
        "4/6/2015","4/27/2015","5/1/2015","6/16/2015","8/10/2015","9/24/2015","1
2/16/2015" ]
Train_PublicHolidays2016 =
["1/1/2016","3/21/2016","3/25/2016","3/28/2016","4/27/2016","5/2/2016","6/16/201
6","8/9/2016","9/24/2016","12/16/2016","12/26/2016","12/27/2016"]
Train_PublicHolidays2017 =
["1/2/2017","3/21/2017","4/14/2017","4/17/2017","4/27/2017","5/1/2017","6/16/201
7","8/9/2017","9/25/2017","12/16/2017","12/25/2017","12/26/2017"]
Train_PublicHolidays2018 =
["1/1/2018","3/21/2018","3/30/2018","4/2/2018","4/27/2018","5/1/2018","6/16/2018
","8/9/2018","9/24/2018","12/17/2018","12/25/2018","12/26/2018"]
Train_PublicHolidays2019 =
["1/1/2019","3/21/2019","4/19/2019","4/22/2019","4/27/2019","5/1/2019","6/17/201
9"]

Test_PublicHolidays2019 = ["8/9/2019","9/24/2019","12/16/2019","12/25/2019","
        12/26/2019"]
Test_PublicHolidays2020 =
["1/1/2020","3/21/2020","4/10/2020","4/13/2020","4/27/2020","5/1/2020","6/16/202
0","8/10/2020","9/24/2020","12/16/2020"]

Training_PublicHolidays = Train_PublicHolidays2015 + Train_PublicHolidays2016 +
Train_PublicHolidays2017 + Train_PublicHolidays2018 + Train_PublicHolidays2019

AssetsReturnsData_Training = AssetsReturnsData_Training.query(f'Date !=
@Training_PublicHolidays')

#Check the size of the portfolio constituents matrix compared to index returns
vector
print("Independant Data[Assets] Data Set Size:",
AssetsReturnsData_Training.shape)
print("Dependant Data[Index] Training Data Set
Size",IndexReturnsData_Training.shape)
```

```
#Resize dataframes to calculate regression model
#The values from the larger data set are dropped, due to the Asset Training Data
set being only 3 rows larger than the index, the effect of dropping the data can
be ignored as it is 0.35% of the number of observations


##calculate length of dataframes
Assets_TrainingDays = len(AssetsReturnsData_Training)
Index_TrainingDays = len(IndexReturnsData_Training)

#Make sure the dataframes are the same length, for the dependant variables.
#This is in order to calculate the score
if Assets_TrainingDays == Index_TrainingDays:
  pass
elif Assets_TrainingDays > Index_TrainingDays:
      AssetsReturnsData_Training =
AssetsReturnsData_Training[:len(AssetsReturnsData_Training)-(Assets_TrainingDays
- Index_TrainingDays)]
else:
      IndexReturnsData_Training =
IndexReturnsData_Training[:len(IndexReturnsData_Training)-(Index_TrainingDays -
Assets_TrainingDays)]

print("Dependant Training Data Set Size: ", IndexReturnsData_Training.shape)
print("Independant Training Data Set Size: ", AssetsReturnsData_Training.shape)

#Retrieve all the testing dates from the Index column
#Store the original dates for plotting the predicitons
#Convert series to dataframe
OriginalTrainingDates = IndexReturnsData_Training.reset_index()
OriginalTrainingDates = OriginalTrainingDates['Date']
#OriginalTrainingDates = OriginalTrainingDates.to_frame()

#Independant Data Training  set imputation
##Check for missing data in independant variables
#Impute with median and check for missing data again
AssetsReturnsData_Training.isnull().any()
AssetsReturnsData_Training = AssetsReturnsData_Training.fillna(method='ffill')

#Dependant Data Training  set imputation
##Check for missing data in dependant variables
#Impute with median and check for missing data again
IndexReturnsData_Training.isnull().any()
AssetsReturnsData_Training = AssetsReturnsData_Training.fillna(method='ffill')

"""**Extract, Clean, Transform and Load Testing Data**"""

#Define periods for testing(24 June 2019 - 12 December 2020)
#Take note that training and testing vectors are not the same size, therefore
score function cannot be applied
StartDateTesting = '06/24/2019'
EndDateTesting = '12/21/2020'

#Calculate returns data for testing, Dependant variable
assetsList = ['STXFIN.JO']

IndexPriceData_Testing = getPrices(assetsList, StartDateTesting ,EndDateTesting)
IndexReturnsData_Testing =
expected_returns.returns_from_prices(IndexPriceData_Testing, log_returns=False)

#Sort returns by date
IndexReturnsData_Testing = IndexReturnsData_Testing.sort_values('Date')
```

```
#Remove public holidays from assets list in the testing data
#specify the index
Test_PublicHolidays2019 = ["8/9/2019","9/24/2019","12/16/2019","12/25/2019","
        12/26/2019"]
Test_PublicHolidays2020 =
["1/1/2020","3/21/2020","4/10/2020","4/13/2020","4/27/2020","5/1/2020","6/16/202
0","8/10/2020","9/24/2020","12/16/2020"]

Testing_PublicHolidays = Test_PublicHolidays2019 + Test_PublicHolidays2020

IndexReturnsData_Testing = IndexReturnsData_Testing.query(f'Date !=
@Testing_PublicHolidays')

#Check the size of the portfolio constituents matrix compared to index returns
vector
print("Dependant Data(Index) Testing Data Set
Size",IndexReturnsData_Testing.shape)

# Retrieve all the testing dates from the Index column
#Store the original dates for plotting
#Convert series to dataframe
OriginalTestingDates = IndexReturnsData_Testing.reset_index()
OriginalTestingDates = OriginalTestingDates['Date']
OriginalTestingDates = OriginalTestingDates.to_frame()

"""**Training of Regression Model**"""

#Define dependent and independent variables
#Convert into numpy data structure
x = AssetsReturnsData_Training.to_numpy()
y = IndexReturnsData_Training.to_numpy()

#scale features to simplify processing
from sklearn.preprocessing import StandardScaler

sc_X = StandardScaler()
sc_y = StandardScaler()
X = sc_X.fit_transform(x)
Y = sc_y.fit_transform(y)

#Fit the index returns to the asset returns, first convert to numpy array
#Default kernal parameters will be used, kernel optimisation is a limitation of
the
svr_rbf = SVR(kernel = 'rbf', C=1e3, gamma = 0.1)
svr_rbf.fit(X,Y)

#Regression model
yPredict = svr_rbf.predict(AssetsReturnsData_Training.to_numpy())- 0.1676

"""**Plotting of models against training data**"""

#Plot training index versus regression model
plt.figure(figsize = (12,6))
plt.scatter(OriginalTrainingDates, IndexReturnsData_Training, s=5, color="blue",
label="index training data(FIN15/STXFIN)")
plt.plot(OriginalTrainingDates, yPredict, lw=2, color="red", label="regression
model")
plt.xlabel('Training Date')
plt.ylabel('Log Returns(%)')
plt.title('SVR Predictions to Training Data: 22 December 2014 - 21 June 2019')
plt.legend()
plt.show()
```

```
#scrap
#plt.plot(dates, prices, color= 'black', label= 'Training Data') #here you must
plt the next periods prices, not the training periods prices
#plt.plot(OriginalTrainingDates, svr_rbf.predict(Traindates), color= 'orange',
label= 'RBF model')

"""**Validation of Data**"""

# Testing Model: Score returns the accuracy of the prediction.
# The best possible score is 1.0
Confidence = svr_rbf.score(X,Y)
print("R-squared value of training data: ", Confidence)
print("MSE:", mean_squared_error(Y, yPredict))

"""**Export Data to Excel for graphing**"""

#Convert svr to dataframe
yPredict =  pd.DataFrame(yPredict, columns = ['SVR Returns Model'], index =
OriginalTrainingDates)

#Calculate modified index returns
#Convert to dataframe
ModifiedIndexReturnsData_Training = np.dot(AssetsReturnsData_Training,
np.transpose(weightsFin15))
ModifiedIndexReturnsData_Training =
pd.DataFrame(ModifiedIndexReturnsData_Training, columns = ['Mod Index Returns'],
index = OriginalTrainingDates)

#export all returns to excel
IndexReturnsData_Training.to_excel(r'/content/drive/MyDrive/Work in
Progress/ARP/Code/SVR Predictions/Index Returns.xlsx',sheet_name='STXFIN')
ModifiedIndexReturnsData_Training.to_excel(r'/content/drive/MyDrive/Work in
Progress/ARP/Code/SVR Predictions/Mod Index Returns.xlsx',sheet_name='Mod
Index')
yPredict.to_excel(r'/content/drive/MyDrive/Work in Progress/ARP/Code/SVR
Predictions/SVR Returns.xlsx',sheet_name='SVR')
```

1.2.2 Comparison and Presentation of Results

```
# -*- coding: utf-8 -*-
"""Constituent Performance - Mean Variance Optimisation of A South African Index
Based Portfolio Using Machine Learning V1

Automatically generated by Colaboratory.

Original file is located at
    https://colab.research.google.com/drive/15mCzyaDwBTrzPEgGg8rvDuqiV9SbNCTF

**Preprocessing**
"""

##Mount Google drive so that datasets can be read from here.
from google.colab import drive
drive.mount('/content/drive')

#Load libraries
import pandas as pd
from pandas_datareader import data as pdr
from pandas import DataFrame
import numpy as np

##Portfolio optimisation libraries(pyPortopt)
!pip install PyPortfolioOpt
from pypfopt.efficient_frontier import EfficientFrontier
from pypfopt import expected_returns

## Graphing and plotting libraries
import matplotlib.pyplot as plt
import seaborn as sn

assetListFin15 = ['ABG','CPI','DSY','FSR','GRT','INL','INP','NED','SBK','SLM']
assetsFin15 = [stock + '.JO' for stock in assetListFin15]

##weight matrix to assign weights to the assets
weightsFin15 = np.array([0.08662,     0.03392,       0.06382,       0.17092,
        0.08732,       0.05592,       0.08662,       0.07762,       0.19112,
        0.14612])

##weight matrix to assign weights to the assets
#weightsMVO = np.array([0.00000,      0.14448,       0.03212,       0.00000,
        0.50971,       0.00000,       0.30002,       0.01366,       0.00000,
        0.00000])

##weight matrix to assign weights to the assets
#weightsSVR = np.array([0.10000, 0.10000, 0.10000, 0.10000, 0.10000, 0.10000,
0.10000, 0.10000, 0.10000,0.10000])

###Create a function to retrive the close price data of our listed financial
assets
def getPrices(assets, start, end):
      assetData = pdr.get_data_yahoo(assetsFin15, start=start, end=end)
      assetData = assetData['Adj Close']
      return assetData

"""**Calculate Performance Parameters**"""

#Real index returns for testing period
Prices_Index = pdr.get_data_yahoo('STXFIN.JO', '06/24/2019', '12/21/2020')
Prices_Index = Prices_Index['Adj Close']
```

```
Returns_Index = expected_returns.returns_from_prices(Prices_Index,
log_returns=False)

#export index returns to excel
Returns_Index.to_excel(r'/content/drive/MyDrive/Work in
Progress/ARP/Code/Constituent Performance Outputs/Real Index
Returns.xlsx',sheet_name='Real Index Returns')

#Modified index returns for testing period
Prices_Modified_Index = getPrices(assetsFin15, '06/24/2019', '12/21/2020')
Returns_Modified_Index_Assets =
expected_returns.returns_from_prices(Prices_Modified_Index, log_returns=False)
Returns_Modified_Index = np.dot(Returns_Modified_Index_Assets,
np.transpose(weightsFin15))

#Mean Variance Ooptimised Index returns for testing period
Prices_MVO = getPrices(assetsFin15, '06/24/2019', '12/21/2020')
Returns_MVO_Assets = expected_returns.returns_from_prices(Prices_MVO,
log_returns=False)
Returns_MVO = np.dot(Returns_MVO_Assets, np.transpose(weightsMVO))

#SVR Index returns
Prices_SVR = getPrices(assetsFin15, '06/24/2019', '12/21/2020')
Returns_SVR_Assets = expected_returns.returns_from_prices(Prices_SVR,
log_returns=True)
Returns_SVR = np.dot( Returns_SVR_Assets, np.transpose(weightsSVR))

#export all returns to excel sheet
#extract index
Dates = Returns_Modified_Index_Assets.reset_index()
Dates = Dates['Date']

#Convert arrays to dataframe
Computed_Returns = np.stack(( Returns_Modified_Index, Returns_MVO, Returns_SVR
), axis=-1)
Computed_Returns = pd.DataFrame(Computed_Returns, columns = ['Mod Index
Returns','MVO Returns','SVR Returns'], index = Dates)

#export all computed returns to excel
Computed_Returns.to_excel(r'/content/drive/MyDrive/Work in
Progress/ARP/Code/Constituent Performance Outputs/All
Returns.xlsx',sheet_name='All Returns')

#Calculate and export Shapiro-Wilk Normality tests
#Normality Test of index returns
from scipy.stats import shapiro

#Nomality test of Real index returns
Shapiro_Returns_Index = shapiro(Returns_Index)

#export shapiro tests to excel
Shapiro_Returns_Index = DataFrame(Shapiro_Returns_Index,index=['Index Returns
Stat', 'Index Returns P-Value'])
#Shapiro_Returns_Index.to_excel(r'/content/drive/MyDrive/Work in
Progress/ARP/Code/Constituent Performance Outputs/Shapiro Test.xlsx',
sheet_name='Index Returns')

#Normality Test of  modified index returns
Shapiro_Modified_Index = shapiro(Returns_Modified_Index)

Shapiro_Modified_Index = DataFrame(Shapiro_Modified_Index,index=['Mod Index
Returns Stat', 'Mod Index Returns P-Value'])
```

```
#Shapiro_Modified_Index.to_excel(r'/content/drive/MyDrive/Work in
Progress/ARP/Code/Constituent Performance Outputs/Shapiro Test.xlsx',
sheet_name='Mod Index Returns')

#Normality Test of  Mean Variance Ooptimised portfolio returns
Shapiro_MVO = shapiro(Returns_MVO)

Shapiro_MVO_Index = DataFrame(Shapiro_MVO,index=['MVO Returns Stat', 'MVO
Returns P-Value'])
#Shapiro_MVO_Index.to_excel(r'/content/drive/MyDrive/Work in
Progress/ARP/Code/Constituent Performance Outputs/Shapiro Test.xlsx',
sheet_name='MVO Index Returns')

#Normality Test of  svr index returns
Shapiro_SVR = shapiro(Returns_SVR)

Shapiro_SVR_Index = DataFrame(Shapiro_SVR,index=['SVR Returns Stat', 'SVR
Returns P-Value'])
#Shapiro_SVR_Index.to_excel(r'/content/drive/MyDrive/Work in
Progress/ARP/Code/Constituent Performance Outputs/Shapiro Test.xlsx',
sheet_name='SVR Index Returns')

Shapiro_Returns = [Shapiro_Returns_Index, Shapiro_Modified_Index,
Shapiro_MVO_Index, Shapiro_SVR_Index]
Shapiro_Export = pd.concat(Shapiro_Returns)
Shapiro_Export.to_excel(r'/content/drive/MyDrive/Work in
Progress/ARP/Code/Constituent Performance Outputs/Shapiro Test.xlsx',
sheet_name='Shapiro Tests')

#Correalation Matrix of financial assets in modified index
Modified_Index_Corr = Returns_Modified_Index_Assets.corr()
sn.heatmap(Modified_Index_Corr, annot=False)
plt.show()

#Descriptive Stats of Real index returnsindex
Summary_Returns_Index = Returns_Index.describe()
print("Index Returns Descriptives", Summary_Returns_Index)

#Descriptive Stats of Modified Index
Summary_Returns_Modified_Index = pd.DataFrame(Returns_Modified_Index, columns =
['Mod Index Returns'], index = Dates)
Summary_Returns_Modified_Index = Summary_Returns_Modified_Index.describe()
print("Modified Index Returns Descriptives", Summary_Returns_Modified_Index)

#Descriptive Stats of Mean Variance Optimised portfolio
Summary_Returns_MVO = pd.DataFrame(Returns_MVO, columns = ['MVO Returns'], index
= Dates)
Summary_Returns_MVO = Summary_Returns_MVO.describe()
print("MVO Returns Descriptives",Summary_Returns_MVO)

#Descriptive Stats SVR Portfolio
Summary_Returns_SVR = pd.DataFrame(Returns_SVR, columns = ['SVR Index Returns'],
index = Dates)
Summary_Returns_SVR = Summary_Returns_SVR.describe()
print("SVR Returns Descriptives",Summary_Returns_SVR)
```

1.2.3 Markowitz Mean Variance Portfolio

```
# -*- coding: utf-8 -*-
"""Main Script- Mean Variance Optimisation of A South African Index Based
Portfolio Using Machine Learning V2

Automatically generated by Colaboratory.

Original file is located at
    https://colab.research.google.com/drive/1427fKDt8potLrrs7DoxW0pwtuzECxT5a

**Description:**
This code serves to use the efficient frontier optimisation to optimise a
portfolio of securities.
"""

##Mount Google drive *from google.colab import drive
drive.mount('/content/drive')

"""**Preprocessing:**

"""

#Import libraries from python
##Data handling and manipulation libraries(pandas, numpy)
import pandas as pd
from pandas_datareader import data as pdr
import numpy as np
from numpy.linalg import multi_dot
import datetime as dt

##Portfolio optimisation libraries(pyPortopt)
!pip install PyPortfolioOpt
from pypfopt.efficient_frontier import EfficientFrontier
from pypfopt import expected_returns

## Graphing and plotting libraries
import matplotlib.pyplot as plt

#Retrieve portfolio and index information of Financial 15 portfolio
##Define the financial assets as a list
##Append the list so that the JSE API is referenced
assetListFin15 = ['ABG','CPI' ,'DSY' ,'FSR' ,'GRT' ,'INL' ,'INP' ,'NED'
        ,'SBK', 'SLM']
assetsFin15 = [stock + '.JO' for stock in assetListFin15]

##Create the weight matrix to assign weights to the assets

# weight of modified index
weightsFin15 = np.array([0.08662,     0.03392,       0.06382,       0.17092,
        0.08732,       0.05592,       0.08662,       0.07762,       0.19112,
        0.14612])

##Retrieve the portfolio information: prices, weights and dates
###Create a function to retrieve the adjusted close price data of the listed
stocks
def getPrices(assets, start, end):
      assetData = pdr.get_data_yahoo(assetsFin15, start=start, end=end)
      assetData = assetData['Adj Close']
      return assetData

##Reference a cloud based csv that holds the rebalancing dates
```

```
##Convert the data type to a list
dfRebalanceFin15 = pd.read_csv(r'/content/drive/My Drive/Work in
Progress/ARP/Code/Reference Files/RebalanceDatesFIN.csv')["Effective Date"]
list(dfRebalanceFin15)

##Create the adjusted close price data frame
Fin15 = getPrices(assetsFin15, dfRebalanceFin15[0], dfRebalanceFin15[23])
TradingDays = len(Fin15.index)

"""**Construction of Optimal Portfolio**"""

##Calculate the portfolio performance for the period under review
###Calculate the individual returns over the period
###calulate the mean returns
#### convert the mean returns to a dataframe
returnsFin15 = expected_returns.returns_from_prices(Fin15, log_returns=True)
MeanReturnsFin15 = returnsFin15.mean(axis=0)
MeanReturnsFin15 = MeanReturnsFin15.to_frame()
AnnualMeanReturnsFin15 = MeanReturnsFin15*TradingDays/250

###Calculate the variance of the daily stock returns financial assets over the
period in review
###Calculate the daily standard deviation of the returns
###Calculate the annual standard deviation of the returns
DailyVarFin15 = returnsFin15.var()
AnnualVarFin15 =DailyVarFin15*TradingDays

###Calculate the daily standard deviation, aka the daily volatility
###Calculate the annual standard deviation, aka the annual volatility[3]
DailyVolFin15 = np.sqrt(DailyVarFin15)
AnnualVolFin15 = np.sqrt(AnnualVarFin15)

###Calculate the covariance matrix of returns in the portfolio
###Calculate the correlation matrix of the returns in the portfolio
CovFin15 = returnsFin15.cov()
CovFin15 = np.array(CovFin15)
CorrFin15 = returnsFin15.corr()
CorrFin15 = np.array(CorrFin15)

##Calculate the variance of the portfolio
##Annualise the variance
PortVarFin15 = np.dot(weightsFin15,np.dot(CovFin15,weightsFin15))
AnnualPortVarFin15 = (PortVarFin15*TradingDays)

##Calculate the portfolio volatility
##Annualise the volatility
PortVolFin15 = np.sqrt(PortVarFin15)
AnnualPortVolFin15 = (PortVolFin15*np.sqrt(TradingDays))

##Calculate portfolio expected return using the mean of historical returns
PortExpReturnFin15 = np.dot(returnsFin15.mean(),weightsFin15)

##Plot the efficient frontier
###Concatenate the expected return and standard deviation
assetsPerformanceFin15 = pd.concat([AnnualMeanReturnsFin15, AnnualVolFin15],
axis=1) # Creating a table for visualising returns and volatility of assets
assetsPerformanceFin15.columns = ['Returns', 'Volatility']
assetsPerformanceFin15

##Define the number of random portfolios of the efficient frontier visualisation
numHoldings = len(assetsFin15)
numPort = 1000
```

```
###Define empty arrays for portfolio returns, volatility and asset weights
EffFront_PortWeightsFin15 = []
EffFront_PortReturnsFin15 = []
EffFront_PortVolFin15 = [] #this is only because of the multi_dot dunction which
outputs a 1X1 array

#Append the random portfolios to the lists generated above
 #Generate random weights
 #Make sure the random weights sum to one every time
 #append the returns
 ###Note: the order of the matrix multiplication in the last line of the loop
seems wrong, but it works, 25 Jan: no it dont 26 Jan: write this shit exactly as
you see it
for portfolio in range(numPort):
  EffFront_weightsFin15 = np.random.uniform( 0.0, 1.0,numHoldings)
  EffFront_weightsFin15 = EffFront_weightsFin15/np.sum(EffFront_weightsFin15)
  EffFront_PortWeightsFin15.append

  EffFront_ReturnsFin15 = (np.dot(EffFront_weightsFin15,
AnnualMeanReturnsFin15))
  EffFront_PortReturnsFin15.append(EffFront_ReturnsFin15)

  EffFront_VolFin15 =
np.sqrt(np.dot(np.transpose(EffFront_weightsFin15),np.dot(CovFin15,
EffFront_weightsFin15))*TradingDays)
  EffFront_PortVolFin15.append(EffFront_VolFin15)

#Create a dictionary to store the random portfolios
###Note, make sure this is daily or annual volatility
EffFront_Data = { ' Volatility': EffFront_PortVolFin15, 'Annual Returns':
EffFront_PortReturnsFin15,}
EffFront_portfolios = pd.DataFrame(EffFront_Data)

#Calculate the minimum volatility portfolio, maximum sharp ratio portfolio and
the efficient return based on a target of the risk free rate.
#Setup efficient frontier modules
AnnualMeanReturnsFin15 = AnnualMeanReturnsFin15.to_numpy()
ef = EfficientFrontier(AnnualMeanReturnsFin15,CovFin15)
Fin15MinVol = ef.min_volatility()
Fin15MinVol = pd.DataFrame.from_dict(Fin15MinVol,orient='index')
#You dont actually need the maximum sharpe ration portfolio
Fin15OptimalPerformance = ef.portfolio_performance(1,risk_free_rate=0.0764)

#Export  efficient frontier
EffFront_portfolios.to_csv(r'/content/drive/MyDrive/Work in
Progress/ARP/Code/Main Script Outputs/Efficient Frontier Plot.csv', index =
False)

#Export the optimal weights to a file
ef.save_weights_to_file(r'/content/drive/MyDrive/Work in Progress/ARP/Code/Main
Script Outputs/Min Volatility Portfolio Weights.csv')

"""**Seconday Code, not required to determine Optimal Portfolio**"""

#Only run this code when calculating index returns and performance
#This code module serves to only calculate the different returns of the index
for the period under review
#This code also serves to calculate the index parameters for the different index
Index_Prices = getPrices(assetsFin15 ,dfRebalanceFin15[0], dfRebalanceFin15[23])

Index_Mean_Returns = expected_returns.mean_historical_return(Index_Prices,
returns_data = False, compounding = False)
```

```
Index_CAGR_Returns = expected_returns.mean_historical_return(Index_Prices,
returns_data = False, compounding = True)
Index_EMA_Returns = expected_returns.ema_historical_return(Index_Prices,
returns_data = False, compounding = True)

Index_Performance = pd.concat([Index_Mean_Returns,
Index_CAGR_Returns,Index_EMA_Returns ], axis=1) # Creating a table for
visualising returns and volatility of assets
Index_Performance.columns = ['Mean', 'CAGR', 'EMA']
print(Index_Performance)

"""**USEFUL CODE STOPS HERE**"""

#Only run this code when calculating optimal returns and performance
#This code module serves to only calculate the different returns of the optimal
portfolio for the period under review
#This code also serves to calculate the optimal portfolio parameters

Asset_Prices = getPrices(assetsFin15 ,dfRebalanceFin15[0], dfRebalanceFin15[23])

Opt_Port_Mean_Returns = expected_returns.mean_historical_return(Asset_Prices,
returns_data = False, compounding = False)
Opt_Port_Mean_Returns =
np.sum(np.dot(Opt_Port_Mean_Returns.transpose(),Fin15MinVol))

Opt_Port_CAGR_Returns = expected_returns.mean_historical_return(Asset_Prices,
returns_data = False, compounding = True)
Opt_Port_CAGR_Returns = np.sum(np.dot( Opt_Port_CAGR_Returns.transpose(),
Fin15MinVol))

Opt_Port_EMA_Returns = expected_returns.ema_historical_return(Asset_Prices,
returns_data = False, compounding = True)
Opt_Port_EMA_Returns =
np.sum(np.dot(Opt_Port_EMA_Returns.transpose(),Fin15MinVol))

#Opt_Port_Performance = np.concatenate((Opt_Port_Mean_Returns,
Opt_Port_CAGR_Returns,Opt_Port_EMA_Returns), axis=1) # Creating a table for
visualising returns and volatility of assets
print('Mean = ', Opt_Port_Mean_Returns)
print('CAGR = ', Opt_Port_CAGR_Returns)
print('EMA = ', Opt_Port_EMA_Returns)

##Retrieve reference information
#Retrieve the 5 year government bond rate over the applicable dates [4]
#This will be the risk free rate used in the calculation of the optimal
portfolio
dfRiskFreeRate = pd.read_csv(r'/content/drive/MyDrive/Work in
Progress/ARP/Code/Reference Files/5Yr RSA Gov Bond Yield.csv')
##Notes: the reason for picking this index is that the rebalancing dates start
in 2014, so we will run our test over that date

dfRiskFreeRate.head()

"""References:
1. https://asxportfolio.com/shares-portfolio-analysis-EFP1
2. https://randerson112358.medium.com/python-for-finance-portfolio-optimization-
66882498847
3. finquest youtube
  https://www.machinelearningplus.com/machine-learning/portfolio-optimization-
python-example/#variance
4. Equity RT
"""
```

Table 16: Non Trading Days for FIN15 for period under review

Day Name	Day	Public Holiday
Thursday	25-Dec-14	Christmas Day
Friday	26-Dec-14	Day of Goodwill
Thursday	1-Jan-15	New Year's Day
Saturday	21-Mar-15	Human Rights Day
Friday	3-Apr-15	Good Friday
Monday	6-Apr-15	Family Day
Monday	27-Apr-15	Freedom Day
Friday	1-May-15	Workers' Day
Tuesday	16-Jun-15	Youth Day
Monday	10-Aug-15	National Women's Day
Thursday	24-Sep-15	Heritage Day
Wednesday	16-Dec-15	Day of Reconciliation
Friday	25-Dec-15	Christmas Day
Saturday	26-Dec-15	Day of Goodwill
Friday	1-Jan-16	New Year's Day
Monday	21-Mar-16	Human Rights Day
Friday	25-Mar-16	Good Friday
Monday	28-Mar-16	Family Day
Wednesday	27-Apr-16	Freedom Day
Monday	2-May-16	Workers' Day
Thursday	16-Jun-16	Youth Day
Tuesday	9-Aug-16	National Women's Day
Saturday	24-Sep-16	Heritage Day
Friday	16-Dec-16	Day of Reconciliation
Monday	26-Dec-16	Christmas Day
Tuesday	27-Dec-16	Day of Goodwill
Monday	2-Jan-17	New Year's Day
Tuesday	21-Mar-17	Human Rights Day
Friday	14-Apr-17	Good Friday
Monday	17-Apr-17	Family Day
Thursday	27-Apr-17	Freedom Day
Monday	1-May-17	Workers' Day
Friday	16-Jun-17	Youth Day
Wednesday	9-Aug-17	National Women's Day
Monday	25-Sep-17	Heritage Day
Saturday	16-Dec-17	Day of Reconciliation
Monday	25-Dec-17	Christmas Day
Tuesday	26-Dec-17	Day of Goodwill
Monday	1-Jan-18	New Year's Day
Wednesday	21-Mar-18	Human Rights Day
Friday	30-Mar-18	Good Friday
Monday	2-Apr-18	Family Day
Friday	27-Apr-18	Freedom Day
Tuesday	1-May-18	Workers' Day

Day Name	Day	Public Holiday
Saturday	16-Jun-18	Youth Day
Thursday	9-Aug-18	National Women's Day
Monday	24-Sep-18	Heritage Day
Monday	17-Dec-18	Day of Reconciliation
Tuesday	25-Dec-18	Christmas Day
Wednesday	26-Dec-18	Day of Goodwill
Tuesday	1-Jan-19	New Year's Day
Thursday	21-Mar-19	Human Rights Day
Friday	19-Apr-19	Good Friday
Monday	22-Apr-19	Family Day
Saturday	27-Apr-19	Freedom Day
Wednesday	1-May-19	Workers' Day
Monday	17-Jun-19	Youth Day
Friday	9-Aug-19	National Women's Day
Tuesday	24-Sep-19	Heritage Day
Monday	16-Dec-19	Day of Reconciliation
Wednesday	25-Dec-19	Christmas Day
Thursday	26-Dec-19	Day of Goodwill
Wednesday	1-Jan-20	New Year's Day
Saturday	21-Mar-20	Human Rights Day
Friday	10-Apr-20	Good Friday
Monday	13-Apr-20	Family Day
Monday	27-Apr-20	Freedom Day
Friday	1-May-20	Workers' Day
Tuesday	16-Jun-20	Youth Day
Monday	10-Aug-20	National Women's Day
Thursday	24-Sep-20	Heritage Day
Wednesday	16-Dec-20	Day of Reconciliation