Regime Based Portfolio Optimization: A Look at the South African Asset Market

---

Mr Nkosenhle S Mdluli

Supervisors:

Prof Ritesh Ajoodha

Mr Rudzani Mulaudzi

A research report submitted in partial fulfillment of the requirements for the degree of Master of Science in the field of e-Science

in the

School of Computer Science and Applied Mathematics

University of the Witwatersrand, Johannesburg

3 September 2023

Declaration

I, Mr Nkosenhle S Mdluli, declare that this report is my own, unaided work. It is being submitted for the degree of Master of Science in the field of e-Science at the University of the Witwatersrand, Johannesburg. It has not been submitted for any degree or examination at any other university.

Mr Nkosenhle S Mdluli

3 September 2023

Abstract

Financial markets change their properties (i.e mean, volatility, correlation, and distribution) with time. However, traditional portfolio optimization strategies seek to create static, all weather portfolios oblivious to this and current economic conditions. This produces portfolios that are unable to predict events with excessive skewness and kurtosis. This research investigated the difference in portfolio percentage return, of portfolios that incorporate regimes against one that does not. HMMs, binary segmentation, and PELT algorithms were used to identify regimes in 7 macro-economic features. These regimes, with regimes identified by the SARB, were incorporated into Markowitz's mean-variance optimization technique to optimize portfolios. The base portfolio, which did not incorporate regimes, produced the least return of 761% during the period under consideration. Portfolios using HMMs identified regimes, produced, on average, the highest returns, averaging 3211% whilst the portfolio using SARB identified regimes returned 1878% during the same period. This research, therefore, shows that incorporating regimes into portfolio optimization increases the percentage return of a portfolio. Moreover, it shows that, although HMMs, on average, produced the most profitable portfolio, portfolios using regimes based on data-driven techniques do not always outperform portfolios using the SARB identified regimes.

Acknowledgements

The support of the DSI-NICIS National e-Science Postgraduate Teaching and Training Platform (NEPTTP) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NEPTTP.

Contents

List of Figures

List of Tables

List of Abbreviations

ASEAN	**A**ssosciation of **S**outh **E**ast **A**sian **N**ations
CAC	**C**otation **A**ssistée en **C**ontinu
CAPE	**C**yclically **A**djusted **P**rice to **E**arnings ratio
CPI	**C**onsumer **P**rice **I**ndex
CVaR	**C**onditional **V**alue **a**t **R**isk
DAX	**D**eutscher **A**Aktieninde **X**
DJIA	**D**ow **J**ones **I**ndustrial **A**verage
FTSE	**F**inancial **T**imes **S**tock **E**xchange
GARCH	**G**eneralized **A**uto **R**egressive **C**onditional **H**eteroskedasticity
GA	**G**enetic **A**lgorithm
GDP	**G**ross **D**omestic **P**roduct
GNI	**G**ross **N**ational **I**ncome
GNP	**G**ross **N**ational **P**roduct
HMMs	**H**idden **M**arkov **M**odels
IIP	**I**ndex of **I**ndustrial **P**roduction
IP	**I**ndustrial **P**roduction
JSE	**J**ohannesburg **S**tock **E**xchange
LDA	**L**inear **D**iscrimant **A**nalysis
LIR	**L**ong-term **I**nterest **R**ate
LR	**L**ending **R**ate
MPT	**M**odern **P**ortfolio **T**heory
NEER	**N**ominal **E**ffective **E**xchange **R**ate
QDA	**Q**uadratic **D**iscrimant **A**nalysis
QMV	**Q**uadratic **M**ean **V**ariance
QVM	**Q**uadratic **V**ariance **M**inimisation
PELT	**P**runed **E**xact **L**inear **T**ime
REER	**R**eal **E**ffective **E**xchange **R**ate
SARB	**S**outh **A**frican **R**eserve **B**ank
SIR	**S**hort-term **I**nterest **R**ate
TIPS	**T**reasury **I**nflation **P**rotected **S**ecurity
VaR	**V**alue **a**t **R**isk

Chapter 1

Introduction

A portfolio is a collection of assets and investments held either by a person or an organization. This may include company equity, cash, bonds, and commodities. Before one can have a portfolio though, a decision must be made regarding how investment capital will be allocated amongst the different assets. This can be done using various allocation strategies namely:

- Constant weight asset allocation, this is when an investor keeps constant weight allocation between different assets, regardless of the prevailing market and economic conditions [13].
- Strategic asset allocation, this is when an investor allocates capital to different asset classes with a return target in mind and periodically re-balances the portfolio to ensure that targets are met [13].
- Dynamic asset allocation, this is when an investor constantly changes the asset allocation to factor in the prevailing market and economic conditions [13].
- Tactical asset allocation, this is similar to the dynamic asset allocation, but then changes are made on the basis of technical analysis in a bid to take advantage of potential opportunities[13].

All these strategies have their benefits and drawbacks. The main drawback though is that because correlations are not taken into account when allocating capital, the is a risk of having a highly correlated portfolio. This may result in asset returns moving in the same direction and consequently failing to hedge each others' risk.

Modern Portfolio Theory, introduced in the 1950s, seeks to incorporate correlation to maximize a portfolio's return whilst minimizing it's variance [24]. However, with time, it has been found to have flawed assumptions namely: assets' returns do not follow a constant, normal distribution, and their correlation matrix is not constant [33]. This is because asset returns show different behavior at different time periods. For instance, the South African equity index's average return distribution changes based on the historical period under consideration [23].

Many factors contribute to assets displaying different properties at different time periods. Chief amongst these are economic factors. This is because every economy in the world has changing economic conditions, as seen in the business cycle in figure 1.1. Any economy's business cycle can be divided into roughly 4 different regimes: boom, recession, depression, and recovery. Different economic conditions are prevalent in these regimes, and as such, the factors or conditions under which businesses operate change and so do their profits.

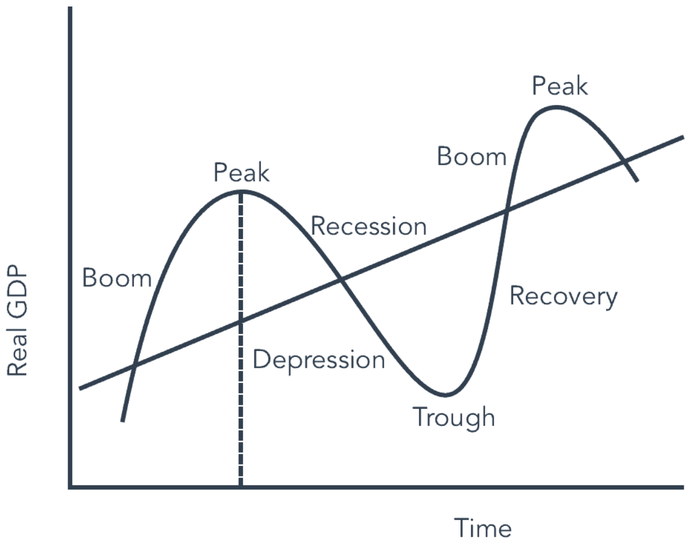


FIGURE 1.1: Business Cycle [1]

During the 2007 - 2008 global financial crisis, for instance, the was a large drawdown in asset prices. However, even though most assets had negative returns, gold and managed futures returned over 16.33% and 16.50%, respectively [11]. This shows that different asset classes are affected differently by the various economic conditions. This has increased the popularity of a dynamic asset allocation technique known as regime-based asset allocation (RBAA). This tries to factor in macroeconomic factors when building models for asset allocation [34].

The rest of this chapter will seek to discuss the problem statement and the research questions that this research set out to achieve. Moreover, it will discuss the contributions made by this research.

1.1 Problem Statement

Traditional portfolio allocation strategies seek to create static, all weather portfolios to meet their objective function across various economic conditions [34]. However, financial markets change their mean, volatility, and correlation patterns over time [3]. This is in response to changes in multiple factors, including economic factors. The South African financial markets are not different [23]. Empirical evidence has shown that returns of the various stocks and indices are neither normal nor constant [23]. As such, modeling the returns of the various stocks and indices as being normally distributed and having constant mean has various implications.

In the late 1980s, modeling regime change of non-stationary economic data was introduced [16]. However, it was not until the 2007-2008 financial crisis that RBAA gained traction [34]. While the literature indicates that considerable work has been done since, it is evident that very little work has been done in South Africa, more so in relation to data-driven techniques [23]. This research therefore sought to study the implications of incorporating regime changes to portfolio optimization with a particular focus on data-driven techniques.

1.2 Research Aim

The aim of this research was to study how percentage gross returns of portfolios incorporating economic regimes differ from that of one that does not incorporate regimes.

1.2.1 Research Objectives

In ensuring that the research aim was achieved, this research sought to achieve the following objectives:

- apply data-driven techniques, on macro-economic features, to identify regimes,
- incorporate identified regimes and SARB regimes into portfolio optimization using Markowitz's mean-variance optimization technique,
- reweigh the shares in the portfolio whenever there is a change in regimes to take advantage of the new regime's economic factors, and
- calculate the percentage gross return of the various portfolios and see how they differ from each other.

1.3 Research Questions

In addressing the above problem statement, and achieving the objectives, this research sought to answer the following research questions:

- How do regimes identified by data-driven techniques (HMMs, binary segmentation and PELT algorithms) differ, in terms of precision, recall, and F1-scores, from regimes identified by the SARB through a macro-economic theory based approach?
- Using portfolio percentage gross return as a metric, what is the impact of incorporating regimes (identified by the SARB as well as Hidden Markov models, binary segmentation and PELT algorithms) to portfolio optimization when compared to optimizing the portfolio without regimes, the base model?
- Using portfolio percentage gross return as a metric, how do portfolios using the SARB identified regimes, compare to portfolios using regimes identified through purely data-driven techniques (HMMs, binary segmentation and PELT algorithms)?

1.4 Research Contributions

This research contributes to the South African literature on using data driven techniques for regime-based portfolio optimization. This contribution comprises of:

- Using data driven techniques (i.e HMMs, binary segmentation and PELT) to identify economic regimes in South Africa. The SARB currently uses economic-theory based techniques to identify economic regimes. This research helps to expand on this by showing alternative ways of identifying regimes.
- Introducing the use of regimes identified through data driven techniques in portfolio optimization. Portfolio optimization generally uses tactical asset allocation. This research shows that incorporating regimes in portfolio optimization produces higher gross returns than when regimes are not incorporated.

1.5 Research Report Structure

This chapter gave an introduction to the research report. This included giving the problem and purpose statement, outlining the research questions and stating the research contributions. The rest of this report consists of 4 chapters. Chapter 2 gives a detailed breakdown of the background and related literature used in this research. This includes how the stock market is correlated to various macro-economic variables, regime detection techniques used by other researchers, and portfolio optimization techniques. Chapter 3 proceeds to discuss the techniques used in conducting this research, both regime detection techniques, and portfolio optimization techniques. The fourth chapter, chapter 4, discusses the results obtained, whilst chapter 5 summarizes the research and gives areas for potential future research.

Chapter 2

Background and Related Work

Asset management is a part of almost every person's life, be it directly or indirectly. This is because most people have savings and assets which they manage, be it in the formal or informal sector of the economy. As of June 2020, South Africa is said to have had R6.3 trillion assets under management by the country's asset managers [6]. This is only a fraction of the total assets held by South Africans as some are not under management by asset managers. As such, for an asset manager to maintain a competitive advantage over her competitors, it is important to ensure that returns are maximized. Individual investors alike should constantly look for better ways to optimize their asset allocation and maximize returns.

With the increase in the availability of historic data and the advances in artificial intelligence, there has been an increase in the application of artificial intelligence in portfolio optimization [17]. This application will likely increase with time, as the amount of data available increases. This chapter seeks to explore the literature relating to how regime detection and portfolio optimization have been done to date. This chapter will start by discussing the macro-economic factors correlated to the stock market, as discussed in the literature. It will proceed to look at the literature on regime detection and lastly, portfolio optimization.

2.1 Macro-Economic Variables and the Stock Market

An economy encompasses all activities relating to the production, trade, and consumption of goods and services in a unit. This unit may be a household, town, city, or country. Companies are crucial in the facilitation of such activities as they are involved in the production and consumption of goods and services. As such, factors affecting the economy also affect the financial performance of these companies whether directly or indirectly and consequently the stock market.

Economies differ based on the level of economic development. The World Bank classifies economies into 4 different categories based on the GNI per capita: low income, low-middle income, upper-middle income, and high income. This difference in income and economic development amongst countries results in differing correlations on their stock markets by the various macro-economic factors. For instance, in China, the stock market regulatory framework is still underdeveloped, resulting in low standards of corporate governance and unreliable company financial data [22].

Researchers have for some time been studying the relationship between the macro-economic variables of emerging economies and their corresponding stock markets. A summary of some of these literature can be found in table 2.1.

TABLE 2.1: A summary of reviewed literature on how macro-economic variables are related to the stock market.

Author(s)	Country(ies)	Feature(s)	Model(s)
Wongbangpo, Praphan and Sharma, Subhash C [38]	Indonesia, Malaysia, Singapore, Philippines and Thailand	GNP, CPI, money supply, interest rate and exchange rate.	Cointergration modelling.
Liu, Ming-Hua and Shrestha, Keshab M [22]	China	Exchange rate index, money supply, CPI, IP, SIR and LIR.	Heteroscedastic cointegration analysis using standard GARCH model.

Continued on next page

Table 2.1 – continued from previous page

Author(s)	Country(ies)	Feature(s)	Model(s)
Hsing, Yu [18]	South Africa	Real GDP, government deficit, money supply, domestic real interest rate, NEER, inflation rate and world interest rate.	Exponential GARCH.
Chandrashekar, Raghutla and Sakthivel, P and Sampath, T and Chittedi, Krishna Reddy [7]	India and Brazil	IIP proxy for GDP, CPI proxy for rate of inflation, LR proxy for interest rate and REER proxy for exchange rate	Cointergration modelling.

An increase in the GNP has a positive impact on the stock markets of Indonesia, Malaysia, Singapore, the Philippines, and Thailand, whilst CPI has the opposite effect [38]. This is because an increase in GNP increases cash-flow in the economy and consequently business revenue and profitability. An increase in CPI however increases business costs and may lead to a decrease in profitability in the short term, thus negatively impacting the stock market.

Money supply, interest rate, and the exchange rate affect ASEAN economies differently due to various reasons [38]. For instance, a high export economy will most likely have stock market performance being positively related to depreciation in the exchange rate against the U.S. dollar. This is because the U.S. dollar is the world's reserve currency; thus most trade happens using it. As such, it's depreciation renders goods sold to the world cheaper, thus increasing their demand and consequently the companies' revenue and profit. The opposite is true however for high import economies. The depreciation of the exchange rate makes imports more expensive, thus increasing business costs, which negatively affects revenue and profits.

Just as is the case with the ASEAN economies, the Chinese stock market is positively related to the economic output, measured through IP, whilst being negatively related to CPI [22]. With the Chinese economy being export oriented due to it's low manufacturing costs, it's stock market is positively related to depreciation in its exchange rate against the US dollar [22]. The interest rate, however, is negatively related to the stock market, whilst the money supply is positively related to the stock market [38].

The South African stock market is also positively related to GDP growth, whilst being negatively related to inflation [18]. It is also negatively related to the interest rate whilst being positively related to depreciation in the exchange rate [18].

From the above analysis, it is clear that the macro-economic factors that affect the stock markets of emerging economies are very similar. Even though the effect that some factors have on the stock market varies from country to country, some almost always have the same effect on the stock market regardless of country. A summary of some of these factors as well as how they affect different economies is given in table 2.2.

	Macro-Economic Variables			
	Exchange Rate	**Inflation**	**Interest Rate**	**Output(GNP/GDP)**
Brazil	+	-	+	+
China	+	-	-	+
India	+	-	+	+
Indonesia	+	-	+	+
Malaysia	-	-	+	+
Philippines	+	-	-	+
Singapore	-	-	-	+
South Africa	+	-	-	+
Thailand	-	-	-	+

TABLE 2.2: A summary of macro-economic variables and their effect on different countries' stock market.

2.2 Regime and Change Point Detection

Regimes are defined as intervals of divergent behavior in financial and economic time-series [2]. Regime detection methods in asset management can broadly be grouped into 4 categories: macro-economic environments, fundamental equity valuations, technical indicators, and statistical regimes [23]. Macro-economic environments make use of macro-economic data to segment time periods by economic cycles. This cycle can have either two, three, or four periods depending on the requirements of the user. Regression techniques can then be used to forecast lagged observations of the economic variables to the corresponding target variable representing the economic cycles [23]. An example of such a variable can be the SARB quarterly recession indicator.

Fundamental equity valuations make use of fundamental equity data to try and determine whether markets are undervalued or overvalued [23]. The equity data is used to create value spreads in relation to historic norms. Consequently, based on the rules of the user, a market will be identified as undervalued or overvalued. A common example of such is the CAPE [23].

Technical indicators are informative signals obtained from the price, volume, or open interest of a security [9]. A wide range of technical indicators are available. However, the most successful indicators are well grounded theoretically and behaviorally with their predictive power incorporating the return distribution of the said security [23]. Multiple techniques can be used to predict regimes using technical indicators. These include probabilistic momentum and implied volatility. Probabilistic momentum converts monthly excess market returns for a given time period into a probability of outperforming the competing asset [23].

Which one of the above models one uses depends on the available data as well as the objective. Table 2.3 gives a summary of regime detection research for portfolio optimization.

TABLE 2.3: A summary of reviewed literature on regime detection.

Author	Features	Models	Significance
John M. Mulvey, Han Hao and Nongchao Li [30]	S&P 500 monthly adjusted closing prices and expected inflation (difference between TIPS return and US nominal government bond returns).	Trend filtering.	The regimes identified as well as their duration varies depending on the feature used.
Ruoyun Zhang, Chao Yi and Yixin Chen [39]	CSI 300 Index, CSI 500 Index, CSI Mid-High Credit Bond Index, Shanghai Securities Treasury Bond Index, South China Commodity Index and South China Gold index.	Hierarchical clustering.	Treasury bonds and credit bonds behave similarly across different regimes. Moreover, stocks and bonds tend to move in opposite directions.
Akioyamen, Peter and Tang, Yi Zhou and Hussien, Hussien [2]	Growth statistics, price and inflation indices, money supply statistics, interest rates, employment statistics, income and expenditure statistics, debt levels, and miscellaneous indicators which report housing stats and gross private domestic investment.	K-means clustering, LDA, QDA, logistic regression, decision tree classifier, adaptive boosting, and Naive Bayes classifier.	Supervised and unsupervised techniques can be combined to construct a regime detection framework.

2.3 Portfolio Optimisation

Investment portfolio building and diversification has long been a part of people's lives. Some of it's earliest traces date back to 935 B.C, which is when the book of Ecclesiastes is assumed to have been written. In it, the need for one to invest in multiple avenues for no one knows what risks lie ahead is touched upon [12]. It was not until the 1950s though that a modern understanding of portfolio selection was obtained [25]. This included the understanding of the implications of having assets with correlated risk, the clear distinction of efficient and inefficient portfolios, and the trade-off between risk and return for any portfolio [24].

Before that though, many investors were already involved in portfolio diversification of some form. One such method is what is now called naive diversification. This is when an investor uses simplistic techniques to divide a portfolio among different assets without the use of sophisticated mathematical models [5]. The $\frac{1}{N}$ is one of the most popular of naive diversification techniques. Here, an investor divides a portfolio equally among N different assets.

In the early twentieth century, although not explicitly stated in the literature, more advanced diversification techniques incorporating correlations were being utilized. This is seen in the top-down approach utilized at the time, which advocated for the world to be divided into 7 or 9 regions and separate grouping for companies operating throughout the world [5]. An investor would then divide a portfolio amongst companies within these different regions. The motivation behind this was the understanding that market risk was influenced by trading conditions in a particular region. As such, assets in the same region were exposed to the same risk.

In 1952 though, Harry Markowitz first published what is now popularly known as MPT. During this time, the maxim amongst investors was the need to maximise the discounted return for assets [24]. However, this failed to imply diversification in any form because an investor only needed to place all his/her capital into the asset with the greatest discounted value [24]. As such, this was quickly rejected. Markowitz instead provided the E-V formulation where given N assets, one seeks to maximise their expected return whilst minimizing the variance in their expected return.

Although portfolio theory was largely founded on variance as a measure of risk, with years gone by, various metrics have been developed to be used as risk measures. These can be used as complements or as substitutes since the motivation behind their use varies. One such risk metric is semi-variance. This metric assumes that investors are more concerned with a portfolio under performing rather than over performing. This is implicit in it's formulation since it quantifies the downward deviations of the portfolio returns from the mean or whatever target has been set.

The previously mentioned measures measure the dispersion/volatility of a portfolio's return overtime. However, other metrics do not measure dispersion. VAR, for instance, gives a summary of the worst loss that a portfolio will not exceed over a target horizon with a given level of confidence [19]. That is, it describes the percentile of the projected distribution of gains and losses over a given time-frame [19]. This helps asset managers to have better control of the risk exposure of a particular portfolio. The common methods used to calculate VaR include the historical method, variance-covariance method, and Monte Carlo simulation.

Having looked at the history of portfolio selection and risk measures, table 2.4 provides a summary of different authors' contributions to portfolio optimisation research.

TABLE 2.4: A summary of reviewed literature on portfolio optimization.

Author	Features	Models	Accuracy	Significance
Vasant, Jilten and Irgolic, Laurent and Rajaratnam, Kanshukan and Kruger, Ryan [37]	JSE daily stock returns.	20000 random weights were generated. The set creating a portfolio with minimum deviation was found numerically.	F-test and t-test are used.	Semi-variance has a better risk adjusted return compared to mean-variance. Also, it's best for portfolio sizes of 50 or less. Beyond 50, mean-variance works better as a risk metric.
Mulvey, John M and Hao, Han and Li, Nongchao [30]	Monthly adjusted closing prices of S&P 500, US equity, international equity, US treasury, corporate bond, real estate and commodities.	Trend filtering for regime detection as well as scenario based and mean-variance Markowitz models for portfolio optimisation.	Regime based mean-CVaR outperforms traditional mean-variance portfolio on all target returns bar at 6% target return.	Assets pose differential risk and returns based on economic regimes. As such, regimes help us better understand the dynamics of assets performance.

Continued on next page

Table 2.4 – continued from previous page

Author	Features	Models	Accuracy	Significance
Gilbert, Evan and Meiklejohn, Luke [15]	South African equities, bonds, cash, listed property returns and global equities and bond .	Models differ by risk measure: variance, omega ratio, CVaR and Wang transform.	The omega ratio produces the highest real returns for both strategic and dynamic asset allocation.	The risk metric used in your model is very important as it directly affects asset allocation and consequently portfolio return.
Zhang, Ruoyun and Yi, Chao and Chen, Yixin [39]	Macro-economic indicators, market technical indicators and historical asset prices.	Hierarchical clustering and Black-Litterman.	Rotational Black-Litterman: 22.53% return and 1.06 Sharpe ratio. Max Sharpe model: 1.34% return and 1.34 Sharpe ratio.	Altering the objective function of the Black-Litterman depending on the prevailing market conditions results in the highest return.

2.4 Conclusion

The previous sections discuss various elements involved in the economy and portfolio optimisation. Section 2.1 discusses how various macro-economic variables are correlated to stock market returns. Some of these have different correlations depending on the economic advancement of the economy. Section 2.2 discusses various regime detection techniques whilst section 2.3 discusses various portfolio optimisation techniques and how they have evolved with time. However, the is a gap in literature as to how these various sections can be combined to create and optimise a portfolio using the regimes identified from data, more so in relation to the South African stock market. There is an application of trend filtering techniques in detecting regimes, however, it does not utilise macro-economic data. Moreover, it is applied in the United States of America, and is not contrasted against non-regime portfolio optimisation. This thus highlights the need to expand literature on regime-based portfolio optimisation, especially in the South African context.

Chapter 3

Research Methodology

This research sought to study the impact of incorporating regimes in portfolio optimization. Although in some respects, this research took an exploratory approach, past research was also used as a guide.This was particularly the case with regards to selecting macro-economic variables to use for regime detection.

This chapter seeks to discuss the data and methodology used in conducting this research. Section 3.1 discusses the research design. This gives a high-level description of the research method used. Section 3.2 discusses the data used. This includes data sources, data type, and data pre-processing. Finally, in section 3.3, the methods used for regime detection and portfolio optimization are discussed.

3.1 Research Design

This research was carried out in 3 main stages: data collection and pre-processing, regime detection, and portfolio optimization. Within these stages, various processes were carried out. A summary of these stages is given below;

- Data collection and pre-processing:
  - data was sourced from Investing.com, GitHub and SARB,
  - data was visualized and analyzed,
  - missing data was imputed and,

  - data was converted to the necessary formats for use in proceeding processes.

- Regime detection. Models used were:
  - Hidden Markov models,
  - binary segmentation and,
  - PELT.
- Portfolio optimization using the mean-variance method.

3.2 Data

Two data sets were used to conduct this research: South African macro-economic data and JSE equity data. The macro-economic data was collected by the SARB and was sourced from [29]. The SARB compiles and keeps a public repository of various economic data that it uses when making decisions on the monetary policy. Only a subset of this data was used. Table 3.1 gives a brief overview of the macro-economic data used.

TABLE 3.1: An overview of macro-economic data collected by the SARB.

Feature	Description	Frequency	Duration
CPI headline	Consumer price indication used to report on headline inflation.	Monthly	1922 to 2020
Domestic output: all groups	This gives the monetary value of all the goods and service produced in a country during a given period.	Monthly	1970 to 2020
Leading business cycle indicator	Composite of indexes used to forecast and confirm the direction in which the economy is headed.	Monthly	1960 to 2020

Continued on next page

Table 3.1 – continued from previous page

Feature	Description	Frequency	Duration
Repo rate	Interest rate charged by the reserve bank to commercial banks when borrowing them money.	Daily	2001 to 2020

Data on JSE listed companies, FTSE SA all share, JSE top-40 and the Rand per US Dollar exchange rate was obtained on Investing.com [35]. This is a platform for financial markets that seeks to provide real time data on financial securities from 250 exchanges across the globe. An overview of this data is given in table 3.2.

TABLE 3.2: An overview of data obtained from Investing.com

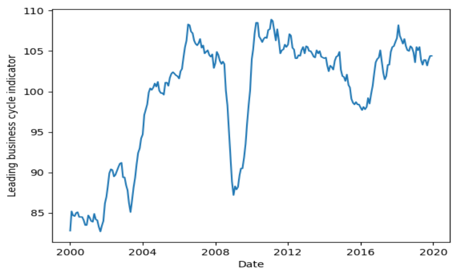
Feature	Description	Frequency	Duration
FTSE South Africa	An index tracking all shares in the JSE.	Daily	2000 to 2020
JSE listed companies	Keeps track of a company's share price	Daily	Varies
JSE top-40 index	An index tracking the top 40 stocks in the JSE by market capitalisation.	Daily	1995 to 2020
Rand per US Dollar exchange rate	The number of South African Rands needed to buy a dollar.	Daily	1970 to 2020

The data was used for two different yet complementary tasks. The macro-economic data, exchange rate data and equity indices data were all used for regime detection whilst individual equity data was used for portfolio optimization. Equity data comprised of historical stock prices, as assumed by the model used [26].

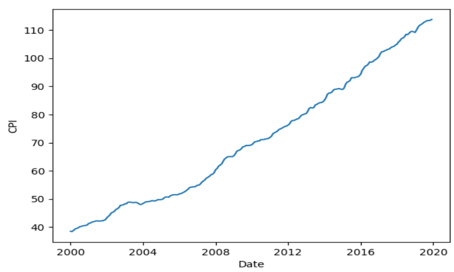
3.2.1 Data-Processing

Data on CPI headline, domestic output, and the leading business cycle indicator was obtained having had most of the cleaning and pre-processing been done [29].

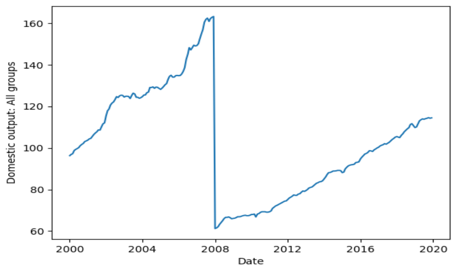
However, there were some repeating entries which were removed. This left the data in a clean and ready to use format. There is a drastic drop in domestic output in 2008 [29]. This is a data issue. However, when looking at the regime detection results, there does not seem to be an anomaly with other features, as such the data was maintained 4.3. Combining these features into a single data-frame produced a data frame with features spanning over 240 months. A graphical representation of this data is seen in figure 3.1.



(A) Leading business cycle indicator.



(B) CPI headline.



(C) Domestic output: all groups.

FIGURE 3.1: Graphs depicting the leading business cycle indicator, headline CPI and domestic output: all groups from from 2000 to 2020.

Data on the repo-rate, Rand per US Dollar, FTSE South Africa all share index and JSE top-40 index all covered varying time periods. This research considered the period from January 2000 to December 2019. This was because the macro-economic data covered this time period, as such, to ensure consistency, all the other data sets were truncated to cover this time period. Since these data sets were sourced from different sources, the was an inconsistency in the date entries. For instance, the repo-rate date had 5154 entries spanning from December 2001 to December 2019, whilst the Rand per US Dollar data set had 4997 entries spanning from January 2000 to December 2019. As such, since the date entries in the rand per US dollar set were consistent with the equity data to be optimized, these were then applied to the repo-rate and indices data sets to ensure consistency.

Combining the data set into one data-frame to ensure consistency in date entries resulted in missing data for some features. The proportions of missing data for each feature are given in table 3.3.

TABLE 3.3: A table showing the missing frequency of the FTSE SA all share index, JSE top-40 index, rand per US dollar exchange rate and repo-rate data sets.

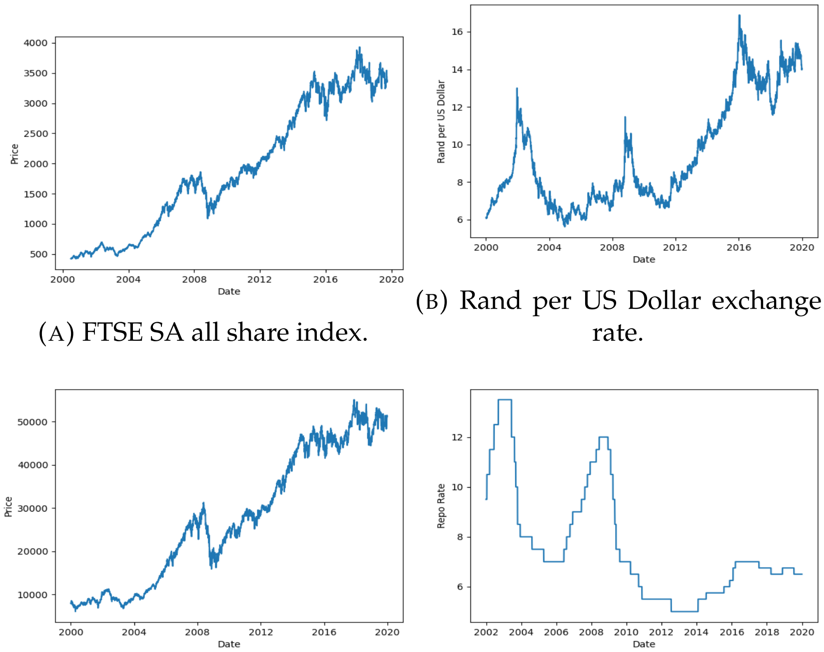
FTSE SA All Share	JSE Top-40	Rand/US Dollar	Repo-Rate
0.03882	0.00120	0.0	0.09986

Missing data was imputed using backward-fill and forward-fill missing data imputation techniques. Backward-fill imputes the missing data with the next valid data point, whilst forward-fill imputes the missing data with the last valid data point. Forward-fill was used because, with economic and financial data, the last value holds true up until a new value is recorded. Backward-fill, however was used because the FTSE SA all share index data set only had data from June 2000. As such, forward-fill was not feasible for some entries. A graphical representation of this data is seen in figure 3.2.

Although Investing.com has data on more than 200 JSE stocks, only 118 of these were used. This is because only stocks with data covering at least 95% of the period under consideration were selected. This is because when calculating the expected returns in portfolio optimization, historical data is used. As such, the more data available, the more accurate the expected returns. Each stock had the average, open, high, and low price. The average price was used and these were combined into a single data-frame for all stocks. To ensure consistency with the regime detection data, a similar index was used. Missing data was imputed using backward-fill and forward-fill.

3.2.2 Statistical Description of the Regime-Detection Data

Regime detection data exhibit various properties at different time periods, as can be seen in figures 3.1 and 3.2. This data was broken down into 4 different time



(A) FTSE SA all share index.

(B) Rand per US Dollar exchange rate.

(C) JSE top-40.

(D) Repo-rate.

FIGURE 3.2: Graphs depicting the FTSE SA all share index, Rand per US Dollar exchange rate, JSE top-40, and repo-rate.

periods namely: 2000 - 2004, 2005 - 2009, 2010 - 2014, and 2015 - 2019. Statistical properties(mean, standard-deviation, minimum value and maximum value) for all these time periods were obtained and can be seen in appendix A.

CPI headline, FTSE SA all share, and JSE top-40 all have means that increase across all 4 time periods. CPI headline is almost always positive since the SARB also has a positive inflation target. As such, they will consistently put in place policies that will drive inflation toward this target. FTSE SA all share and JSE top-40 are both JSE indexes. Since most companies in an economy tend to experience positive growth, this will be reflected in these indexes. The other features however, have fluctuating means as a response to various economic conditions.

Since the data under consideration is time series, it was tested for stationarity. This was done using the Augmented Dickey Fuller test. It's null hypothesis assumes the data to be stationary. For it to be rejected, the p-value must be greater than 0.05. All the data exhibited non-stationarity, meaning that they all had a p-value greater than 0.05, as can be seen in table 3.4. That is, none exhibited a constant mean and variance. Since stationarity was not a pre-requisite for the models used, the data was not transformed.

TABLE 3.4: Stationarity test results for features used in regime detection.

Feature	Test Statistic	p-value	Crit. Value(5%)	Stationarity
CPI headline	1.6206	0.9979	-2.8745	Not stationary.
Domestic output: All groups	-1.9627	0.3032	-2.8737	Not stationary.
FTSE South Africa	-0.5811	0.8752	-2.8621	Not stationary.
JSE top 40 index	-0.3717	0.9147	-2.8621	Not stationary.
Leading business cycle indicator	-2.4112	0.1386	-2.8744	Not stationary.
Rand per US Dollar exchange rate	-0.9538	0.7697	-2.8621	Not stationary.
Repo rate	-1.5861	0.4906	-2.8621	Not stationary.

3.2.3 Data Transformation

The binary segmentation and PELT algorithms used for change-point detection assume that the data has a variance of 1 when using mean as an indicator of change-point. As such, the data was transformed to have a variance of 1, whilst maintaining their mean, before using these algorithms. This was done by applying equation (1) to the data-set.

$$y_i = m_2 + (x_i - m_1)\frac{s_2}{s_1}, \tag{1}$$

where:

- y_i - transformed data point,
- m_2 - mean of new data set,
- x_i - old data point,
- m_1 - mean of old data set,
- s_2 - standard deviation of new data set, and
- s_1 - standard deviation of old data set.

3.3 Methods

Methods used in this research were divided into two broad categories: regime detection methods, which sought to partition the data into regimes, and portfolio optimization techniques, which sought to select the assets most likely to give the highest gross returns during the period under consideration. Both these are discussed in the sub-sections that follow.

3.3.1 Regime-Detection

Regime detection methods seek to define intervals of differing characteristics in a given data set. These characteristics include, the distribution, mean, variance, e.t.c. This research made use of 3 such methods: HMMs, binary segmentation, and PELT algorithms. The binary segmentation and PELT algorithms identified changes in the mean and variance.

Hidden-Markov Models

HMMs are probabilistic graphical models used to generate observable sequences from hidden information. They are a type of Markov chain and consequently exhibit the Markov property defined in theorem 3.3.1.

Theorem 3.3.1 (Markov Property [14]). *Given a discrete-time stochastic process* $\{X_t, n \geq 0\}$ *in a countable set S, the future state* $X_{t+1}, n \geq 0$ *in a countable set S is conditional only* *on the present state X_t and independent of the past states, $X_0, \cdots, X_{t-1}$. More specifically, $\forall n \geq 0$, and $\forall i, j, k, \cdots, m \in S$,*

$$P[X_{t+1} = j | X_t = i, X_{t-1} = k, \cdots, X_0 = m] = P[X_{t+1} = j | X_t = i] = p_{ij}.$$

They were initially used for speech recognition purposes in the 1970s but have since been applied in various other use cases such as analysing biological sequences. Besides their simplicity and malleability, their popularity stems from their ability to model an output of discrete states as a function of a hidden stochastic process. Formally, HMMs are defined in definition 3.3.1

Definition 3.3.1 (Hidden-Markov-Model [32])**.** This is a 5-tuple $(Q, \sum, \prod, A, B)$, where $Q = \{q_1, \cdots, q_N\}$ is a finite set of $\mathcal{N}$ states, $\sum = s_1, \cdots, s_N$ is the set of $\mathcal{M}$ possible symbols in the language, $\prod = \{\pi_i\}$ is the initial probability vector, $A = \{a_{ij}\}$ is the state transition probability matrix, and $B = \{b_i(v_k)\}$ is the emission probability matrix. The HMM can be denoted by $\lambda = (\prod, A, B)$, having the following constraints: The total transition property

- from the initial state to all hidden states is 1, i.e., $\sum_{i=1}^{N} \pi_i = 1$,
- from a hidden state q_i to all other hidden states is 1, i.e., $\forall i \in Q, \sum_{i=1}^{N} a_{ij} = 1$,
- from the hidden states to all observable states is 1, i.e., $\forall i \in Q, \sum_{i=1}^{M} b_i(v_k) = 1$.

Definition 3.3.1 and all the following discussions consider a case where the observations are discrete. For one to use HMMs in the real world, 3 problems must be solved: evaluation, decoding, and training.

Evaluation: Given $X = (x_0, x_1, \cdot, x_T)$, an observation sequence, and $\lambda = (\prod, A, B)$, a HMM, how can one efficiently calculate $P(X|\lambda)$, the probability of obtaining X given λ [20]? That is, how can one score just how well a given model λ matches a given sequence X. Solving this enables us to choose between different models.

Assume that we are given an observation sequence $X = (x_0, x_1, \cdot, x_T)$, state sequence $Q_1 = \{q_1, \cdots, q_T\}$, and HMM $\lambda = (\prod, A, B)$. We can thus calculate $P(X, Q_1|\lambda)$ using the fact that $P(X, Q_1|\lambda) = P(X|Q_1, \lambda)P(Q_1|\lambda)$ since these terms can be calculated using available data. Assuming the statistical independence of observations, we have:

$$P(X|Q_1,\lambda) = \prod_{t=1}^{T} P(x_t|q_t,\lambda)$$
$$= b_{q_1}(x_1)b_{q_2}(x_2)\ldots b_{q_T}(x_T). \tag{2}$$

Also, the probability of obtaining such a state sequence given the HMM can be calculated using equation (3).

$$P(Q_1|\lambda) = \pi_{q_1}a_{q_1q_2}\ldots aq_{T-1}q_T. \tag{3}$$

Combining equations (2) and (3), we obtain:

$$P(X,Q_1|\lambda) = P(X|Q_1,\lambda)P(Q_1|\lambda)$$
$$= (b_{q_1}(x_1)b_{q_2}(x_2)\ldots b_{q_T}(x_T))(\pi_{q_1}a_{q_1q_2}\ldots aq_{T-1}q_T)$$
$$= \pi b_{q_1}(x_1)a_{q_1q_2}b_{q_2}(x_2)\ldots aq_{T-1}q_T b_{q_T}(x_T). \tag{4}$$

However, equation (4) applies to only one sequence of states yet, to calculate $P(X|\lambda)$, we must calculate over all possible sequences of states of length T. As such, summing over all possible sequences of states we obtain:

$$P(X|\lambda) = \sum_{allQ} P(X|Q,\lambda)P(Q|\lambda). \tag{5}$$

The above calculation is rather computationally expensive thus infeasible since the possible number of sequences of states is exponential. As such, an alternative approach known as the backward-forward algorithm is used.

We define a forward variable which is an aggregate helper function by

$$\alpha_t i = P(x_1, x_2, \ldots, x_t, q_t = S_i|\lambda). \tag{6}$$

This is the probability of seeing observations $x_1, x_2, \ldots, x_t$ and ending up at state $q_t = S_i$ given the HMM λ. Since this accounts for all possible state sequence combinations before state $q_t = S_i$, summing it over all possible states solves $P(X|\lambda)$. $\alpha_t(i)$ can be solved inductively. Initializing $\alpha_t(i)$ for $t = 1$, we get

$$\alpha_1(i) = \pi_i b_i(x_1). \tag{7}$$

Inductively, we thus have:

$$\alpha_{t+1}(i) = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}] b_j(O_{t+1}) \text{ for } 1 \leq \text{t} \leq \text{T-1}, 1 \leq \text{j} \leq \text{N}. \tag{8}$$

As such, summing over (8), we have:

$$P(X|\lambda) = \sum_{i=1}^{N} \alpha_T(i). \tag{9}$$

Similarly, a backward variable, $\beta_t(i)$ can be solved for inductively. This is the probability of seeing observations $x_{t+1}, x_{t+2}, \ldots, x_T$ whilst starting at state $q_t = S_i$ given HMM λ. This is defined by

$$\beta_t(i) = P(x_{t+1}, x_{t+2}, \ldots, x_T, q_t = S_i|\lambda). \tag{10}$$

Solving (10) inductively, we obtain

$$\beta_T(i) = 1 \text{ for } 1 \leq \text{i} \leq \text{N}. \tag{11}$$

Inductively, we have:

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j O_{t+1} \beta_{j+1}(j) \text{ for } 1 \leq \text{i} \leq \text{N and t = T-1, T-2,} \ldots \text{,1.} \tag{12}$$

$\beta_t(i)$, although not used when solving the evaluation problem, is used when solving the next two problems to be discussed.

Decoding: Given $X = (x_0, x_1, \cdot, x_T)$, an observation sequence, and $\lambda = (\prod, A, B)$, a HMM, how can one choose the corresponding state sequence $Q = \{q_1, \cdots, q_N\}$? The are various criteria that can be used [32]. One such criteria seeks to find the individually most likely state q_t at time t. This is done by solving for each time step t:

$$\max_{1 \leq i \leq N} \gamma_t(i) = \max_{1 \leq i \leq N} P(q_t = S_i | X, \lambda). \tag{13}$$

Since $\alpha_t(i)$ gives us the probability of seeing observations $x_1, x_2, \ldots, x_t$ and ending up at state $q_t = S_i$ given the HMM λ, whilst $\beta_t(i)$ gives us the probability of seeing observations $x_{t+1}, x_{t+2}, \ldots, x_T$ whilst starting at state $q_t = S_i$ given HMM λ, both variables can be used to solve for (13). This is done as follows,

$$\begin{aligned} \gamma_t(i) &= \frac{\alpha_t(i)\beta_t(i)}{P(X|\lambda} \\ &= \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N} \alpha_t(i)\beta_t(i)}. \end{aligned} \tag{14}$$

Consequently, we get that the most likely state q_t at time t is

$$q_t = \arg\max_{1 \leq i \leq N} [\gamma_t(i)]. \tag{15}$$

This can be done for all time steps $1 \leq t \leq T$. However, this method does not guarantee that the path being traced out is feasible. That is, it does not ensure that there exists a path from q_i to q_{i+1}. As such, to solve this, an alternative approach which seeks to find a feasible state sequence Q that maximises $P(Q|X, \lambda)$ is used. This makes use of the Viterbi algorithm. It is also an inductive algorithm and is similar to the α variable used in the forward-backward algorithm. We first define

$$\delta_t(i) = \max_{q_1, q_2, \ldots, q_{t-1} = i} P(\{q_1, q_2, \ldots, q_t = s_i\}, \{x_0, x_1, \cdot, x_t\}|\lambda). \tag{16}$$

$\delta_t(i)$ gives us the the highest probability of a single path along the given states, accounting for the first t observations and ending up at state $q_t = s_i$. Inductively, we thus have:

$$\delta_{t+1}(j) = [\max_i \delta_t(i) a_{ij}] b_j(O_{t+1}). \tag{17}$$

However, before we can move on to the next observation x_{t+1}, we need to keep track of the best state sequence for observations $x_0, x_1, \cdot, x_t$. This is done using the variable $\psi_t i$. Proceeding to solve for (16), we first define a base case,

$$\delta_1(i) = \pi b_i(O_1), \tag{18a}$$

$$\psi_1(i) = 0. \tag{18b}$$

Equation (18b) holds because the is no prior observation and thus no prior state. Inductively, we thus have:

$$\delta_t(j) = [\arg\max_{1\leq i\leq N} \delta_{t-1}(i)a_{ij}]b_j(O_t) \text{ for } 2 \leq \text{t} \leq \text{T}, 1 \leq \text{j} \leq \text{N}, \tag{19a}$$

$$\psi_t(j) = \arg\max_{1\leq i\leq N}[\delta_{t-1}(i)a_{ij}] \text{ for } 2 \leq \text{t} \leq \text{T}, 1 \leq \text{j} \leq \text{N}. \tag{19b}$$

Finally, to terminate our induction, we find the highest probability of a single path along the given states, accounting for all the observations and ending up at state T.This is given by:

$$P^* = \max_{1\leq i\leq N}[\delta_T(i)], \tag{20a}$$

$$q_T^* = \arg\max_{1\leq i\leq N}[\delta_T(i)]. \tag{20b}$$

Having performed induction, we can then recursively track back our results to reconstruct the sequence that has led us to q_T^*. This is done using $\psi_t(i)$ and is given by:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \text{ t=T-1, T-2, } \ldots, 1. \tag{21}$$

Training: Finally, given a set of observations $X = (x_0, x_1, \cdot, x_T)$, how do we learn the parameters of our HMM $\lambda = (\prod, A, B)$ to maximise $P(X|\lambda)$? This is regarded as the most difficult of the 3 problems, and at present, the is no analytical method known to solve it [32]. As such, an iterative method known as the Baum-Welch method is used. This is an iterative expectation-maximisation procedure, and as such, gives local optimum, which is not guaranteed to be a global optimum. We first define

$$
\begin{aligned}
\xi_t(i,j) &= P(q_t = s_i, q_{t+1} = s_j | X, \lambda) \\
&= \frac{P(q_t = s_i, q_{t+1} = s_j, X|\lambda)}{P(X|\lambda)} \\
&= \frac{\alpha_t(i) a_{ij} b_j O_{t+1} \beta_{t+1}(j)}{\sum\limits_{i=1}^{N} \sum\limits_{j=1}^{N} \alpha_t(i) a_{ij} b_j O_{t+1} \beta_{t+1}(j)}.
\end{aligned} \tag{22}
$$

This is the probability of being in state s_i at time t and state s_j at time $t+1$, given the observation sequence X and HMM λ. Since $\gamma_t(i)$ is the probability of being in state i at time t, summing ξ over j, we get

$$
\gamma_t(i) = \sum_{j=1}^{N} \xi_{i,j}. \tag{23}
$$

Summing both $\gamma_t(i)$ and $\xi_t(i,j)$ over t, we obtain the expected number of times state s_i is visited and the number of times transitions from state s_i to s_j are made. This can be used to update a HMM's parameters as follows,

$$
\begin{aligned}
\bar{\pi}_i &= \text{ expected number of times at state } s_i \text{ at time 1} \\
&= \gamma_1(i),
\end{aligned} \tag{24a}
$$

$$
\begin{aligned}
\bar{a_{ij}} &= \frac{\text{the expected number of times state } s_i \text{ is visited}}{\text{the number of times transitions from state } s_i \text{ to } s_j \text{ are made}} \\
&= \frac{\sum\limits_{t=1}^{T-1} \xi_t(i,j)}{\sum\limits_{t=1}^{T-1} \gamma_t(i)},
\end{aligned} \tag{24b}
$$

$$\bar{b_j}(k) = \frac{\text{expected number of times in state } j \text{ and observing symbol } v_k}{\text{expected number of times in state } j}$$

$$= \frac{\sum_{\substack{t=1 \\ s.t x_t = v_k}}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}. \tag{24c}$$

Using an initial model to calculate (24a), (24b), and (24c), a new model $\bar{\lambda}$ is obtained. This is repeated until $\bar{\lambda}$ converges to a critical point at which point we have reached our local optimum.

Mean and Variance

An arithmetic mean is the sum of given values divided by the total number of these values [27]. That is, given values $x_1, x_2, \ldots, x_N$, the mean is given by $\bar{x} = \frac{x_1+x_2+\cdots+x_N}{N}$. It is a measure of central tendency in a data set. Other types of mean include the geometric mean and harmonic mean.

Variance measures dispersion in a data set. Given a data set, it measures just how dispersed the data points are from the mean. This is given by the expected value of the squared deviations from the mean [36]. That is, given a data set $X = x_1, x_2, \ldots, x_N$, the variance is defined by:

$$Var(X) = E[(X - E(X))^2]. \tag{25}$$

Given a data set $y_1, y_2, \ldots, x_T$, mean and variance change point methods seek to find a set $\mathcal{T} = \{t_1, t_2, \ldots\} \subset \{1, 2, \ldots, T\}$ of index points where change points are located in the data set. This is done by minimizing the sum of a cost function $\mathcal{C}(\cdot)$ over all possible partitions in the data set, as seen in equation (26) [35].

$$\min_{\mathcal{T}} \mathcal{V}(\mathcal{T}, y) = \min_{\mathcal{T}} \sum_{k=0}^{K} \mathcal{C}(y_{t_k \ldots t_{k+1}}) + pen(\mathcal{T}) \tag{26}$$

If the number of change points is known prior, equation (26) extends to equation (27).

$$\min_{|\mathcal{T}|=K} \mathcal{V}(\mathcal{T},y) = \min_{\mathcal{T}} \sum_{k=0}^{K} \mathcal{C}(y_{t_k \dots t_{k+1}}) \tag{27}$$

Equations (26) and (27) can be solved in various ways. The different methods used can be classified according to cost function used, search criteria and constraint on the number of change points [35].

Since change point methods seeks to segment a given data set into multiple regions of homogeneity, cost functions help detect when there is a change in such homogeneity and thus mark that point as a change point. There are various choices of cost functions, and the choice on which to use is primarily dependant on the type of data at hand. For instance, when the data is independent and identically distributed, with piece-wise constant distribution, the cost function is given by equation (28) [35].

$$\mathcal{C}_{i.i.d}(y_{a,\dots,b}) = -\sup_{\theta} \sum_{t=a+1}^{b} logf(y_t|\theta) \tag{28}$$

In (26), $\theta \in \Theta \in \mathbb{R}^p$ whilst $f(\cdot|\theta)$ is a probability density function. θ is representative of any value of interest whose value changes abruptly at a change point. In instances though where the data has a Gaussian distribution with piece-wise constant mean, a quadratic error loss cost function is used. This is given by equation (29).

$$\mathcal{C}_{L_2}(y_{a,\dots,b}) = \sum_{t=a+1}^{b} ||y_t - \bar{y}_{a,\dots,b}||_2^2. \tag{29}$$

The above cost function can be extended to allow for a change in variance. This consequently results in equation (30).

$$C_{\sum}(y_{a,\dots,b}) = (b-a)logdet\bar{\sum}_{a,\dots,b} + \sum_{t=a+1}^{b}(y_t - \bar{y}_{a,\dots,b})' \sum_{a,\dots,b}^{-1}(y_t - \bar{y}_{a,\dots,b}). \tag{30}$$

The second defining criteria for regime detection techniques is the search criteria used. Search methods are generally classified into exact and approximate [35]. Exact search methods return the exact solutions to optimisation equations (26) and (27) whilst approximate methods only return approximate solutions. Examples of such search methods are given below. The PELT algorithm given in figure 2 is an exact search method whilst the binary segmentation algorithm shown in figure 1 is an approximate method.

3.3.2 Portfolio Optimisation

Having discussed what portfolio optimization is, as well as how it has evolved in chapter 2.3, this chapter will seek to discuss in detail, the key components and procedures involved in this. This research made use of mean-variance optimization. This is an optimization technique that uses variance as a risk measure and weighs it's trade-off against expected return [8]. Like most optimization techniques, mean variance has two critical components: expected returns and a risk model. These are covered in sections 3.3.2 and 3.3.2.

Having calculated the covariance matrix $\sum$ and the expected mean μ, the portfolio optimization problem reduces to an optimisation problem in which we seek to solve:

$$\begin{aligned} \min_{w} w^T \sum w&, \\ \text{subject to } w^T \mu \geq \mu^*&, \\ w^T \mathbf{1} = 1&, \\ w_i \geq 0&, \end{aligned}$$

where w is a vector representing the weights assigned to the different assets in the portfolio.

Expected Returns

Expected return is said to be the return, be it profit or loss, that an investor expects to make from an investment. Since this is expected, there is no guarantee that it will be realized. There are various ways in which this can be calculated: mean historical return and the capital asset pricing model (CAPM) estimate of returns are two such methods.

Mean historical return: This seeks to calculate the total percentage change in the assets value over the duration under consideration then average it out over the number of periods covered. This is seen in equation (30a). However, asset's value compounds in value and as such, the geometric mean is a more accurate measure of this rate and is given by equation (31b).

$$\text{AMHR} = \frac{\text{percentage change in price}}{100n}, \tag{31a}$$

$$\text{GMHR} = (1 + \frac{\text{percentage change in price}}{100})^{\frac{1}{n}} - 1, \tag{31b}$$

where n is the number of periods.

Capital asset pricing model (CAPM) estimate of returns: This model seeks to describe the general relationship between an assets expected return and it's systematic risk in relation to portfolio theory [31].

$$E_S = r_f + \beta(E_M - r_f), \tag{32}$$

where $E_s$is the expected return, r_f is the risk free rate, E_M is the market risk premium and β is the beta of the investment. The risk free rate is the theoretical return that an investor expects to get when investing in a risk free asset. This could be a highly rated government bond such a the US treasury bond. The market risk premium is reflective of the overall performance of of the stock market and is usually indicated by a stock index such as the JSE All Share.

Risk Models

This is a method or approach used to quantify an asset's risk. In the mean variance optimization procedure, the covariance matrix is used for this purpose. However, similar to the expected returns, we can only estimate the covariance matrix using past data since we do not have access to future data. Many models are used as a proxy for the covariance matrix, which includes sample covariance, semicovariance and shrunk covariance matrix.

Sample covariance: This makes use of sample data to calculate the covariance matrix. Given a T x N matrix of returns, R and a 1 x N vector of the average returns, $\bar{R}$, the sample covariance matrix can be calculated using equation (33).

$$\widehat{\sum} = \frac{1}{T-1} R'R - \frac{T}{T-1} \bar{R}'\bar{R}. \tag{33}$$

For equation (32) to hold, R must be non singular.

Semicovariance: It is plausible to assume that investors are more concerned with a portfolio under performing rather than over performing. As such, semi-variance was introduced as a risk metric as it quantifies the downward deviations of the portfolio returns from the mean or whatever target has been set. This can be calculated using equation (33).

$$\sum Cov_{ij} = \frac{1}{T}\sum_{t=1}^{T}[\min(R_{i,t} - B, 0) \cdot \min(R_{j,t} - B, 0)]. \tag{34}$$

Shrunk covariance: This seeks to calculate a weighted average matrix from the sample covariance and some target matrix with the hope of obtaining a more accurate covariance matrix [21]. The shrunk covariance matrix takes the form

$$\widehat{\sum}_{\alpha}^{*} = (1-\alpha)\widehat{\sum} + \alpha\sum, \tag{35a}$$

$$\alpha^{*} = \arg\max_{\alpha}\mathbb{E}[\|\sum - \widehat{\sum}_{\alpha}^{*}\|], \tag{35b}$$

where α is the shrinkage estimator [10].

3.3.3 Evaluation-Metrics

Regime detection is a classification problem; as such, classification metrics were used to evaluate it. The classification report, which is an extension of the confusion matrix was used. This is because it can factor in imbalances in class distributions. It calculates the following variables:

$$Precision = \frac{TP}{TP + FP}, \tag{36}$$

$$Recall = \frac{TP}{TP + FN}, \tag{37}$$

$$F1 - score = \frac{2 * (Precision * Recall)}{(Precision + Recall)}, \tag{38}$$

where:

- TP - true positive
- FP - false positive
- TN - true negative
- FN - false negative.

Of all the positive predictions made, which ones are actually positive? This is the question that precision answers. Recall, however, measures the proportion of true positive predictions made relative to the total number of positive predictions in the data set. The F1-score score consequently gives a summary of the precision and recall.

In evaluating a portfolio's performance, however, the total percentage increase in that portfolio's value was calculated. That is, given a portfolio of initial value x_0 and final value x_1, at the end of the period under consideration, the return by that portfolio is given by:

$$y_i = \frac{x_1 - x_0}{x_0} \cdot 100. \tag{39}$$

However, because this portfolio incorporates multiple regimes, covering different time periods, the cumulative return for the overall portfolio, inclusive of all time periods, is given by:

$$y = ((\prod_{i=1}^{N}(1 + \frac{y_i}{100})) - 1) \cdot 100. \tag{40}$$

Chapter 4

Results and Discussion

In a bid to achieve the objectives of this research, 2 groups of models were built: regime detection models and portfolio optimization models. In building these models, no results were preconceived; as such, the research used an exploratory approach. This chapter seeks to discuss the results of these models.

Section 4.1 discusses the results of the regime detection models. These were compared and contrasted against the business cycle phases identified by the SARB. The discussions were structured according to the regime detection model used. Section 4.2 proceeds to discuss the results of portfolio optimization. Portfolios based on regimes identified in section 4.1 are compared and contrasted against each other and against the base model, which does not make use of regime detection.

4.1 Regime-Detection

Regime detection was carried out using HMMs, binary segmentation and PELT algorithms. The ability of HMMs to model an output of discrete states as a function of hidden stochastic processes made it ideal for use in regime identification. With stocks' gross returns having a variable correlation matrix whilst also not following a constant, normal distribution, HMMs can model this variability of statistical properties as a sequence of discrete states [33].

Binary segmentation and PELT algorithms were used to detect changes in either the mean or variance. Since macro-economic factors change with time, so does their mean and variance. Such changes are usually synonymous with changes in economic conditions. For instance, during a period of high inflation, the implementation of deflationary monetary policies increases the repo-rate and consequently the mean value of the repo-rate for that period. As such, using the change in mean and variance as a measure of change points for regimes can detect plausible macro-economic regimes.

TABLE 4.1: A table showing the past business cycle regimes as identified by the SARB. [4].

Date	Trend	Duration
Jan 2000 - Nov 2007.	Upwards (1).	95 months.
Dec 2007 - Aug 2009.	Downwards (0).	21 months.
Sep 2009 - Nov 2013.	Upwards (1).	51 months.
Dec 2013 - Dec 2019.	Downwards (0).	73 months.

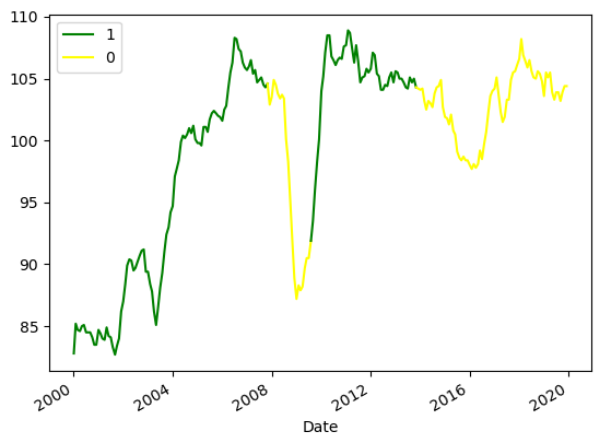


FIGURE 4.1: A graph showing the past business cycle regimes as identified by the SARB.

The regimes identified by the HMMs, binary segmentation, and PELT algorithms are compared and contrasted with those obtained by the SARB, as seen in table 4.1

and figure 4.1. Upward and downward trends are not measured purely by absolute increase in domestic output. Rather, the SARB uses the growth cycle definition of the business cycle. As such, it's regimes indicate fluctuations around the long-term trend of overall economic activity [4]. Therefore, periods with an increase in domestic output may be classified as being downwards and vice versa with periods with a decrease in domestic output.

According to the SARB, the South African economy was in an upward economic trend in 2000. As such, one key assumption was made regarding the results of the regimes of the various algorithms. The first regime identified by the HMMs was assumed to be an upward trend regime and was given a state of 1. However, with the binary segmentation and PELT algorithms, the first state was labelled as 0 on the graph, although representing an upward trend (1). In the data-set though, it was assigned 1, representing an upward trend. Proceeding states were then alternated between 0 and 1.

It is important to note that the data used for regime detection was of differing frequencies. Some features had a monthly frequency whilst some had a daily frequency, as seen in tables 3.1 and 3.2. These frequencies were maintained when the data was fed into regime detection models. However, because the JSE data which is used for portfolio optimisation is daily, the forward-fill data imputation technique was then used to ensure that the regime data was of a similar frequency to the JSE data.

4.1.1 Hidden-Markov Models

In constructing the HMMs for regime detection, 3 problems had to be solved: training, evaluation, and decoding. The first problem to be solved was the training problem because it is responsible for creating the HMM used in solving the other problems. However, before one can train an HMM, one must initialize it's parameters $(\prod, A, B)$. Since the Baum-Welch method used to solve the training problem is an expectation-maximisation procedure, it solves for the local optima. As such, the values assigned during initialization ultimately determine this local optima. In a bid to obtain a HMM that is as close as possible to the one that would be obtained at the global optima, each HMM was given 10 000 different initializations.

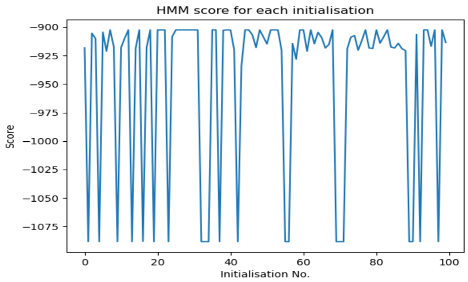
With 10 000 initializations being made for each HMM, each resultant HMM was evaluated and only the highest scoring HMM was selected for each feature. Since log-likelihood was used to score the HMM, the least-negative model was the best scoring model. A plot of the first 1 000 log-likelihood scores for each feature is shown in figure 4.2. As can be seen in figure 4.2, the HMMs for each feature have erratic log-likelihood scores, indicating different data fit for the HMMs. As such, when comparing the log-likelihoods of the different HMMs, the HMM with the largest log-likelihood score was selected for each feature.

After training and evaluating the HMMs, the Viterbi algorithm was used to infer the different regimes/hidden states for the different HMMs. Even though this research's regime detection was based on the business cycle, 2 regimes were assumed. This is in contrast to the 4 regimes usually implied by the business cycle. This is because, amongst the 4 regimes, 2 involve an upward trajectory of the economy whilst the remaining 2 represent a downward trajectory of the economy, albeit at varying rates. As such, these were grouped together to only have the upward and downward regimes, as is the case with the SARB.

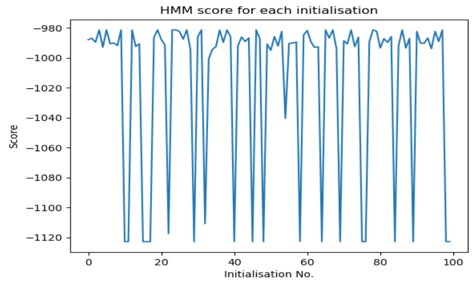
Figure 4.3 shows a graphical plot of the regimes identified by the HMM. Headline CPI, FTSE SA all share, and JSE top-40 all show one business cycle for the entire 20 year period under consideration. The regimes are of slightly different duration though. This is in contrast to the data from table 4.1, which shows the existence of two business cycles during the same period.

All the 3 stated features failed to identify the post financial crisis resurgent of economies, which saw double figure returns in stock market indices. Of these 3 however, only headline CPI correctly identified the start of the economic downturn of the subprime mortgage crisis. FTSE SA all share only identified economic downturn starting in 2015, long after it had started in 2013, whilst also missing the downturn in December 2007 to August 2009. The JSE top-40 on the other hand identified economic downturn starting in 2006, more than a year before the start of the December 2007 downturn.

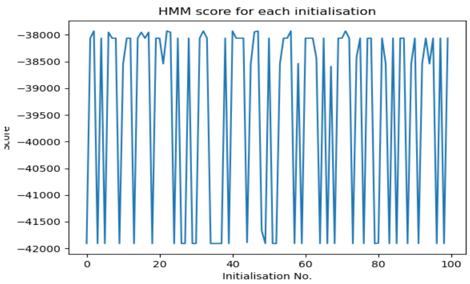
The leading business cycle indicator and repo-rate both identified 2 cycles during the 20 year period under consideration. This is similar to the data on table 4.1, which also identified 2 cycles. However, these cycles differ from each other and



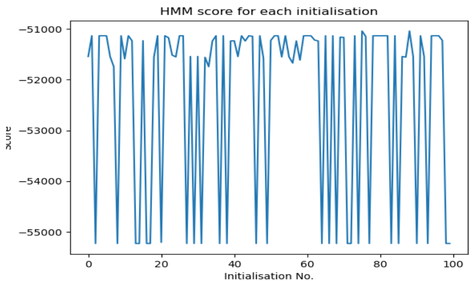
(A) CPI headline



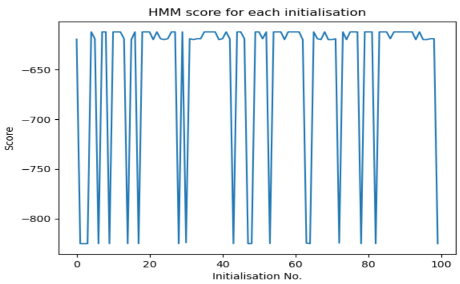
(B) Domestic output: all groups



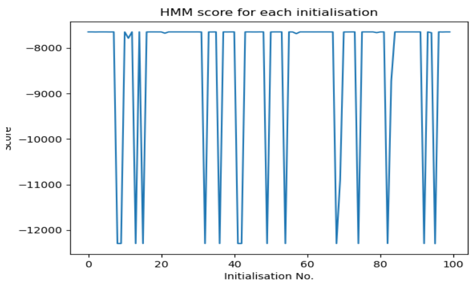
(C) FTSE SA all share



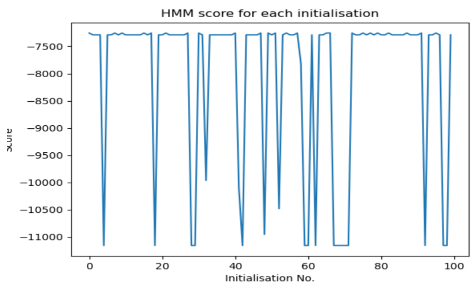
(D) JSE top-40



(E) Leading business cycle indicator



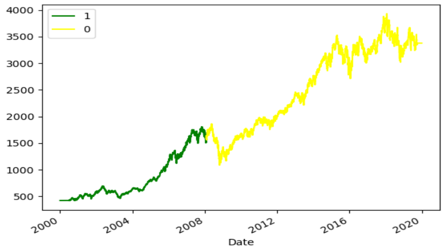
(F) Exchange Rate: Rand per US Dollar



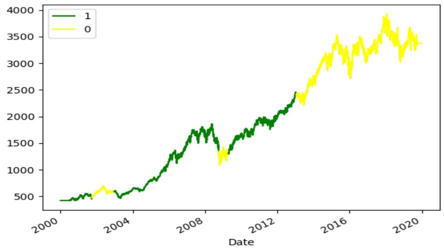
(G) Repo-rate

FIGURE 4.2: HMMs log-likelihood scores for the different features.

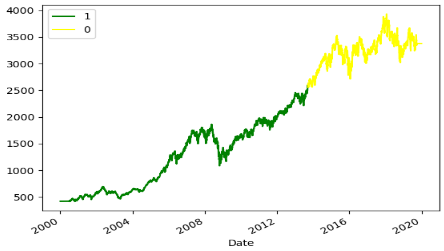
from the one identified in table 4.1. For instance, during the upward regime spanning Jan 2000 to Nov 2007, the leading business cycle indicator identified 2 different



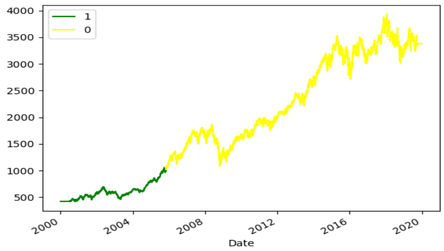
(A) CPI headline



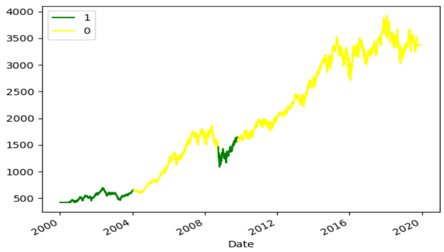
(B) Domestic output: all groups



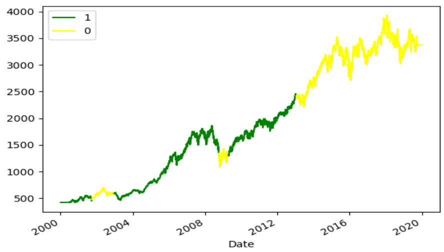
(C) FTSE SA all share



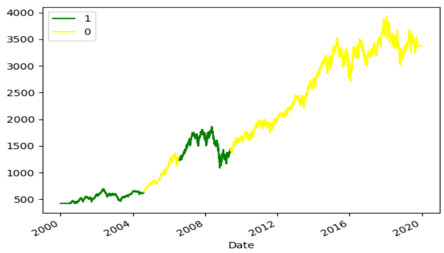
(D) JSE top-40



(E) Leading business cycle indicator



(F) Exchange rate: Rand per US Dollar



(G) Repo-rate

FIGURE 4.3: Regimes identified through the use of HMMs.

regimes, whilst the repo-rate identified 3 different regimes. This could be because of the repo-rate varying greatly during this period. It had a value ranging between

$7 - 13.5\%$ whilst having a standard deviation of 1.92% basis points. This greatly decreases post 2008 as the standard deviation becomes 1.06% basis points and consequently only 1 regime was identified.

The domestic output variable identified 3 regimes during the period under consideration. The first regime correctly corresponds to the first regime identified in table 4.1. However, the proceeding 2 regimes do not necessarily correspond to those identified in table 4.1. For instance, the downward trend identified stretches from 2007 to 2015 and the upward trend from 2015 to 2020. The final upward trend is in contrast to the SARB data in table 4.1, which labels the period as having a downward trend. The downward trend from 2007 to 2015 overlaps with an upward trend identified by the SARB.

Using data on the Rand/US Dollar, the HMM identified 3 cycles. Similar to the repo-rate, although of different durations, 3 regimes are identified between 2000 and 2008. This is in contrast to the data in table 4.1. However, it did correctly identify the upward trend stretching from 2009 to 2013 as well as the proceeding downward trend.

From figure 4.3, it is clear that the regime patterns identified by the HMMs differ from each other and from those of the SARB. This could be mainly attributed to HMMs modeling observations as being an output of hidden states. Since the data between the features varies from each other, in terms of statistical properties, as can be seen in figures 3.1 and 3.2, the sequence of the hidden states will also vary amongst the features. Table 4.2 shows the classification reports of the HMMs against the SARB regimes. This is to add a quantitative comparison between the regimes.

The HMM trained using the leading business cycle indicator data had the lowest F1-scores, 0.61 and 0.49, depending on which class label identified as positive, as seen in table 4.2. This is followed by the HMM trained using the repo-rate, which had F1-scores of 0.63 and 0.61. Both these models had a low precision score when the class label 0 was used as a positive label. Moreover, they both had a low recall score when the class label 1 was used as a positive label. This is because of the identified regimes not corresponding to the SARB regimes. For instance, during the time period of the first regime by the SARB, the leading business cycle indicator wrongly classified more than half the data, thus resulting in a high number of false positives and consequently, high precision.

TABLE 4.2: Classification reports for all regime detection features using the HMMs algorithm.

(A) CPI headline

	0	1
Precision	0.65	0.98
Recall	0.98	0.65
F1-score	0.78	0.78

(B) Domestic output: all groups

	0	1
Precision	0.77	0.89
Recall	0.85	0.83
F1-score	0.81	0.86

(C) FTSE SA all share

	0	1
Precision	0.98	0.87
Recall	0.78	0.99
F1-score	0.87	0.93

(D) JSE top-40

	0	1
Precision	0.56	1.00
Recall	1.00	0.48
F1-score	0.72	0.65

(E) Leading business cycle indicator

	0	1
Precision	0.47	0.82
Recall	0.88	0.35
F1-score	0.61	0.49

(F) Exchange rate: Rand per US Dollar

	0	1
Precision	0.77	0.89
Recall	0.85	0.83
F1-score	0.81	0.86

(G) Repo-rate

	0	1
Precision	0.51	0.80
Recall	0.81	0.49
F1-score	0.63	0.61

The HMM trained using the FTSE SA all share data had the highest F1-scores, 0.87 and 0.93, depending on which class label identified as positive, as seen in table 4.2. This can be seen in figure 4.3. Although the model missed the downward trend spanning from December 2007 to August 2009, it correctly identified the other regimes. Using the JSE top-40 data, the HMM trained has a precision of 1.0 when the positive class label is 1 and a recall of 1.0 when the positive class label is 0. This is because with the former case, the was no false positive whilst with the latter case, the was no false negative. The remaining HMMs had similar F1-scores.

4.1.2 Mean and Variance

In identifying change points using mean and variance, two search algorithms were used: binary segmentation and PELT. The PELT algorithm gives exact solutions, whilst the binary segmentation only gives approximate solutions. A large portion of their solution set intersected as a result.

Mean

The binary segmentation and PELT algorithms identified the same number of change-points in 3 of the 7 features. These were domestic output: all groups, leading business cycle indicator, and the exchange rate. The remaining features, bar the repo-rate, the difference in the number of change-points was minute. For instance, headline CPI and the FTSE SA all share, both had the number of change-points differing by 1. This shows that the binary segmentation and PELT algorithms, when using the same metric, generally identify the same number of change-points in a data set. This may be due to both algorithms using the same cost function, the modified Bayesian information criterion.

Although both algorithms generally detect the same number of change-points, the exact location of these change-points varies between algorithms. This can be seen in figures 4.4 and 4.5. This is most likely because of the binary segmentation being an approximate method, whilst the PELT algorithm is exact. This can be seen when looking at the features that had the same number of change-points. Their graphs are very similar, with the differences in location only being visible when looking at the actual data-set.

Only headline CPI, domestic output, and the repo-rate identified the start of the 2007 financial crisis, as can be seen in figures 4.4 and 4.5. Of these features, only domestic-output had a significant change in value during the same time. The repo-rate continued in an upward trend that had begun in 2006, and headline CPI continued unabated in it's upward trend as well, as can be seen in figures in 3.1 and 3.2.

Domestic output can show a drastic change during the start of the 2007 financial crisis because, unlike other macro-economic factors, the time lag between such an event and it's reflection in domestic output is minimum. This is because a financial crisis can quickly send an economy into recession due to lost jobs and output, as such, consumers are quick to scale back on consumption, due to uncertainty. Moreover, since the domestic output is recorded monthly, this decline in output is quickly reflected in SARB data.

TABLE 4.3: Classification reports for all regime detection features using the mean binary segmentation algorithm.

(A) CPI headline

	0	1
Precision	0.72	0.82
Recall	0.72	0.82
F1-score	0.72	0.82

(B) Domestic output: all groups

	0	1
Precision	0.70	0.83
Recall	0.76	0.79
F1-score	0.73	0.81

(C) FTSE SA all share

	0	1
Precision	0.33	0.57
Recall	0.28	0.63
F1-score	0.31	0.60

(D) JSE top-40

	0	1
Precision	0.52	0.70
Recall	0.57	0.66
F1-score	0.54	0.68

(E) Leading business cycle indicator

	0	1
Precision	0.50	1.00
Recall	1.00	0.36
F1-score	0.67	0.53

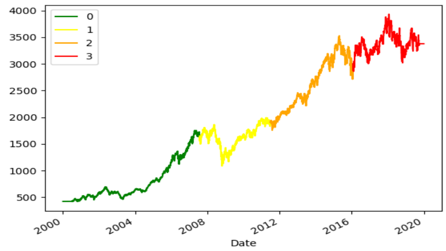
(F) Exchange rate: Rand per US Dollar

	0	1
Precision	0.64	0.74
Recall	0.58	0.79
F1-score	0.61	0.76

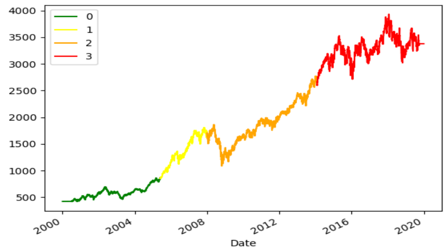
(G) Repo-rate

	0	1
Precision	0.72	0.80
Recall	0.68	0.83
F1-score	0.70	0.81

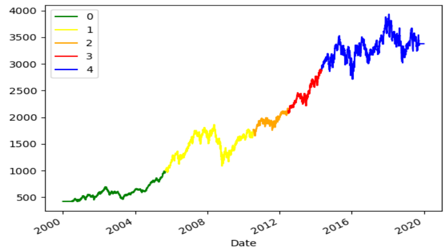
The leading business cycle indicator showed two changes, with both the binary segmentation and PELT algorithm, as seen in figures 4.4 and 4.5. This is in contradiction to the business cycle regime data from the SARB. This is seen as well in figure 3.1 as the business cycle indicator showed two significant dips in 2008 and 2016. However, because the algorithms search for segments with the minimum cost, they will at times not identify the exact change points if these do not result in minimum cost.



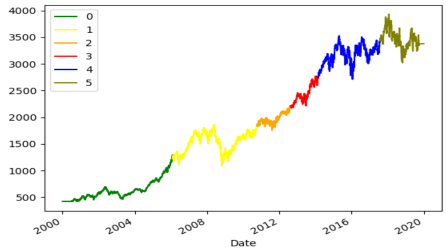
(A) CPI headline



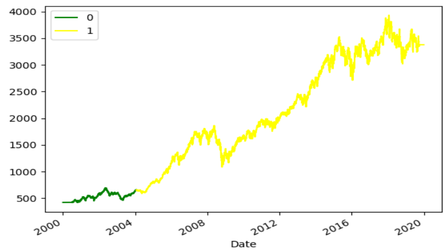
(B) Domestic output: all groups



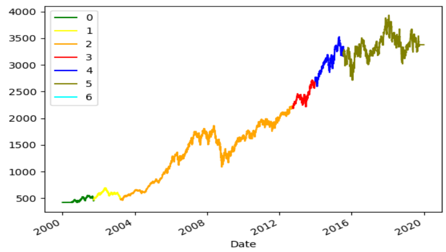
(C) FTSE SA all share



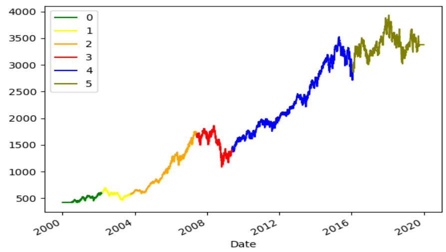
(D) JSE top-40



(E) Leading business cycle indicator

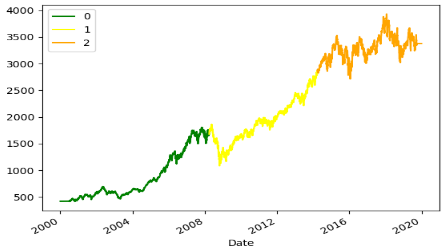


(F) Exchange rate: Rand per US Dollar

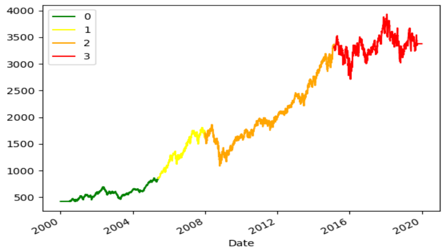


(G) Repo-rate

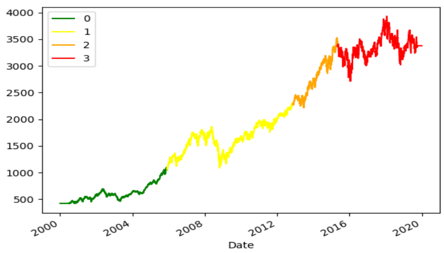
FIGURE 4.4: Regimes identified through the use of the binary segmentation algorithm, with the mean as a change point detector.



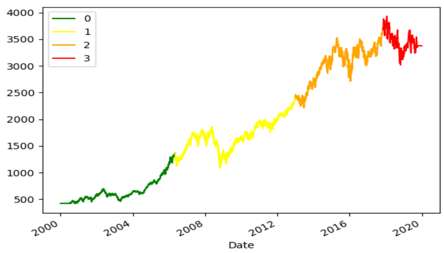
(A) CPI headline



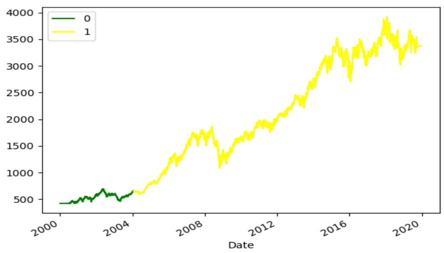
(B) Domestic output: all groups



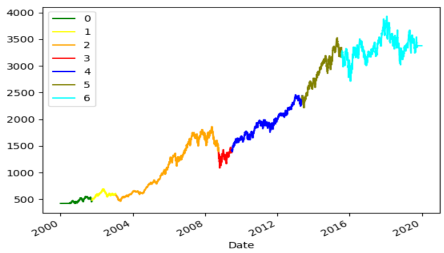
(C) FTSE SA all share



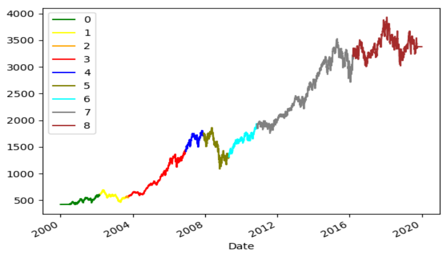
(D) JSE top-40



(E) Leading business cycle indicator



(F) Exchange rate: Rand per US Dollar



(G) Repo-rate

FIGURE 4.5: Regimes identified through the use of the PELT algorithm, with the mean as a change point detector.

TABLE 4.4: Classification reports for all regime detection features using the mean PELT algorithm.

(A) CPI headline

	0	1
Precision	0.30	0.57
Recall	0.23	0.65
F1-score	0.26	0.61

(B) Domestic output: all groups

	0	1
Precision	0.66	0.76
Recall	0.62	0.79
F1-score	0.64	0.78

(C) FTSE SA all share

	0	1
Precision	0.55	0.82
Recall	0.81	0.57
F1-score	0.65	0.67

(D) JSE top-40

	0	1
Precision	0.44	0.64
Recall	0.49	0.59
F1-score	0.46	0.62

(E) Leading business cycle indicator

	0	1
Precision	0.50	1.00
Recall	1.00	0.34
F1-score	0.66	0.50

(F) Exchange rate: Rand per US Dollar

	0	1
Precision	0.58	0.66
Recall	0.31	0.85
F1-score	0.40	0.74

(G) Repo-rate

	0	1
Precision	0.33	0.53
Recall	0.48	0.38
F1-score	0.39	0.44

CPI headline and domestic output: all groups had the highest F1-scores for regime detection using mean binary segmentation, regardless of which class label identified as positive, as seen in table 4.3. This is a consequence of having their change-points similar to those of the SARB, as seen in figure 4.4. The repo-rate had the next highest F1-scores. They are slightly lower than the first 2 due to the two regimes identified before 2004. However, post 2004, the regimes are similar to those of the SARB. The FTSE SA all share had the lowest F1-scores, 0.31 and 0.60, depending on which class label is labeled as positive.

For regime detection using mean PELT, domestic output: all groups and FTSE SA all share index had the highest F1-scores, as seen in table 4.4. They both had 4 regimes, as was the case with CPI headline and domestic output: all groups when using mean binary segmentation. This is similar to the SARB, which also identified 4 regimes. Headline CPI and the repo rate had the lowest F1-scores.

From tables 4.3 and 4.4, it can be seen that the binary segmentation produced regimes with higher F1-scores. This is despite the fact that binary segmentation is an approximate algorithm, whilst PELT is an exact algorithm.

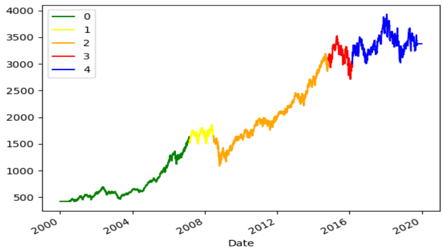
Variance

Using variance as a detector of change points, the change points detected were generally different, both in terms of location and number, from those detected by the mean in figures 4.4 and 4.5, regardless of the algorithm used. The variance tended to identify more features compared with the mean. This is because the mean and variance measure different attributes of data. The mean is a measure of central tendency, whilst variance is a measure of dispersion.

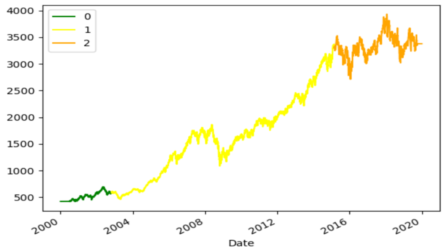
The PELT algorithm identified more change-points compared with the binary segmentation algorithm. In 4 of the 7 features used, it identified at least twice as many as the binary segmentation. This can be seen in figures 4.6 and 4.7. This is in contrast to the results obtained when using the mean for change-point detection. With the mean, both algorithms tended to detect the same equal number of change-points This may be due to the non-stationarity of the data since the variance is not constant.

Comparing these results to those produced by the SARB in table 4.1, the sequence and duration of the regimes identified were very different. Although, some features do have change-points similar to those seen in table 4.1. For instance, the domestic output, leading business cycle indicator, exchange rate, and repo rate all identified change points in 2004, which is when the economic boom that preceded the 2008 financial crisis started. Similar features, bar the leading business cycle indicator, could identify the 2016 change point, whilst domestic output, FTSE SA all share price and exchange rate were all able to identify the 2013 change point, depending on the algorithm used.

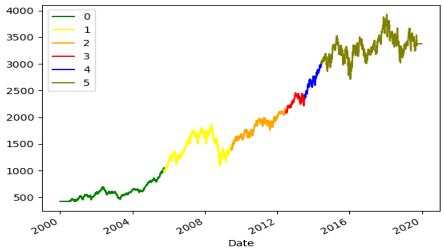
The leading business cycle indicator, as was the case with the mean change points, only identified one change point when using the binary segmentation algorithm. It failed to identify the financial crisis of 2008, which showed the largest decline in the leading business cycle indicator during the period under consideration. Moreover,



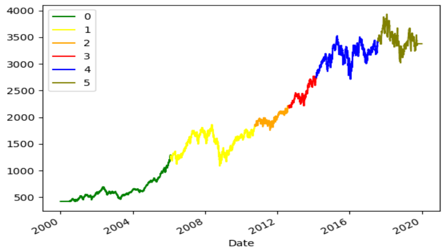
(A) CPI headline



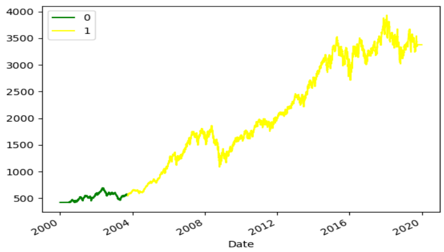
(B) Domestic output: all groups



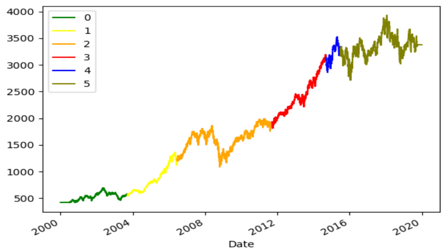
(C) FTSE SA all share



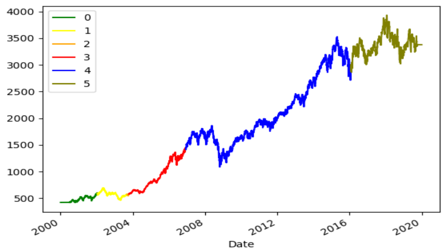
(D) JSE top-40



(E) Leading business cycle indicator

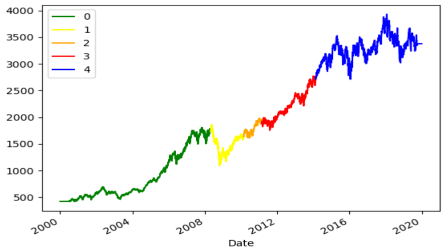


(F) Exchange rate: Rand per US Dollar

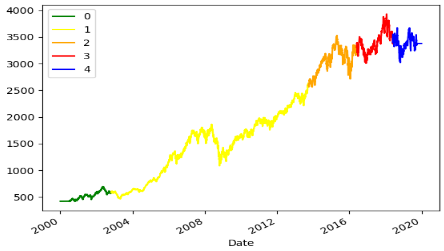


(G) Repo-rate

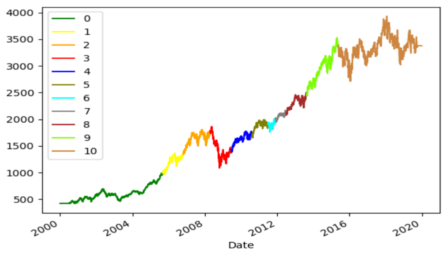
FIGURE 4.6: Regimes identified through the use of the binary segmentation algorithm, with the variance as a change point detector.



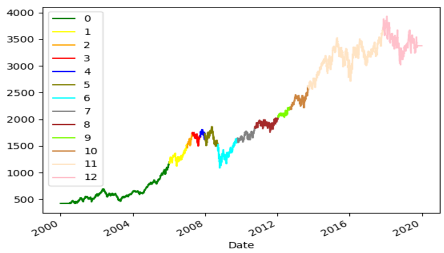
(A) CPI headline



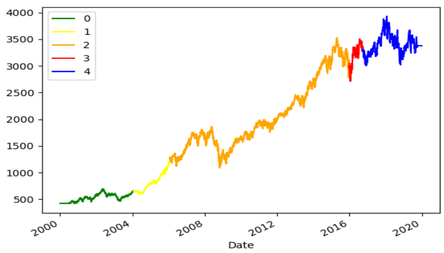
(B) Domestic output: all groups



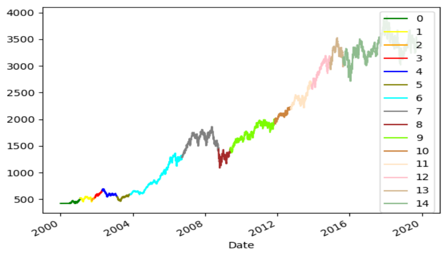
(C) FTSE SA all share



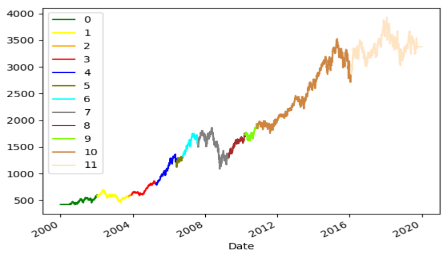
(D) JSE top-40



(E) Leading business cycle indicator



(F) Exchange rate: Rand per US Dollar



(G) Repo-rate

FIGURE 4.7: Regimes identified through the use of the PELT algorithm, with the variance as a change point detector.

TABLE 4.5: Classification reports for all regime detection features using the variance binary segmentation algorithm.

(A) CPI headline

	0	1
Precision	0.71	0.65
Recall	0.24	0.94
F1-score	0.36	0.77

(B) Domestic output: all groups

	0	1
Precision	0.25	0.37
Recall	0.39	0.23
F1-score	0.30	0.29

(C) FTSE SA all share

	0	1
Precision	0.69	0.92
Recall	0.90	0.74
F1-score	0.78	0.82

(D) JSE top-40

	0	1
Precision	0.76	0.95
Recall	0.94	0.81
F1-score	0.84	0.88

(E) Leading business cycle indicator

	0	1
Precision	0.49	1.00
Recall	1.00	0.31
F1-score	0.65	0.47

(F) Exchange rate: Rand per US Dollar

	0	1
Precision	0.52	0.74
Recall	0.68	0.60
F1-score	0.59	0.66

(G) Repo-rate

	0	1
Precision	0.45	0.65
Recall	0.50	0.60
F1-score	0.47	0.62

although of much less magnitude, it also missed the economic and political uncertainty of 2016, which saw South Africa's economic outlook being revised to negative by Fitch and Moody's. However, more change points were identified when using the PELT algorithm, resulting in a change point being identified in 2016.

The PELT algorithm identified 15 regimes when using the exchange rate data. For the repo-rate, JSE top-40 and FTSE SA all share, 13, 13, and 11 regimes were identified respectively. Assuming a 2 regime cycle, the PELT algorithm identified at least $5\frac{1}{2}$ cycles in 4 different features. However, according to the National Bureau of Economic Research, the average business cycle since 1945 has been 58 months [28]. This implies approximately 4 cycles during the period under research, thus contradicting the change points obtained by the PELT algorithm.

The JSE top-40 binary segmentation identified regimes had the highest F1-scores, 0.84 and 0.88, depending on class labels, as seen in table 4.5. They consisted of

TABLE 4.6: Classification reports for all regime detection features using the variance PELT algorithm.

(A) CPI headline

	0	1
Precision	0.34	0.59
Recall	0.21	0.74
F1-score	0.26	0.65

(B) Domestic output: all groups

	0	1
Precision	0.29	0.41
Recall	0.47	0.24
F1-score	0.36	0.31

(C) FTSE SA all share

	0	1
Precision	0.49	0.64
Recall	0.35	0.77
F1-score	0.41	0.70

(D) JSE top-40

	0	1
Precision	0.60	0.74
Recall	0.60	0.74
F1-score	0.60	0.74

(E) Leading business cycle indicator

	0	1
Precision	0.27	0.59
Recall	0.09	0.83
F1-score	0.14	0.69

(F) Exchange rate: Rand per US Dollar

	0	1
Precision	0.23	0.49
Recall	0.24	0.48
F1-score	0.24	0.49

(G) Repo-rate

	0	1
Precision	0.55	0.75
Recall	0.67	0.64
F1-score	0.60	0.69

6 regimes during the period under consideration. These were, in general, able to identify the first cycle corresponding to the SARB cycle.

The PELT algorithm generally produced regimes with low F1-scores, as can be seen in table 4.6. This could be a result of the PELT identifying much more regimes compared to the SARB. This consequently misplaced a lot of the data points, resulting in low precision and accuracy. As such, identification of more regimes doesn't necessarily lead to better results. Rather, identification of regimes that are as close as possible to the economic phenomena, as identified by the SARB does.

4.2 Portfolio Optimization

Having identified regimes using different features and methods in section 4.1, this research sought to optimize a portfolio of 118 stocks found on the JSE. This was done using 5 different models namely:

- a base model without regimes as seen in table 4.7,
- a model based on SARB identified regimes as seen in table 4.7,
- a model based on HMM identified regimes as seen in table 4.8,
- a model based on binary segmentation identified regimes as seen in table 4.9 and,
- a model based on PELT identified regimes as seen in table 4.10.

Results obtained from these models were compared against each other to see which models perform best as well as identify any trends.

The mean variance portfolio optimization technique was used to optimize these portfolios. This was done by first breaking down the time periods under consideration into the regimes identified in section 4.1. For instance, the SARB identified 2 distinct regimes, covering different time periods, as seen in table 4.1. As such, in this case, a portfolio was created, covering the upward trending regime, and another one covering the downward trending regime. The aggregate/cumulative gross returns for these portfolios were combined to obtain the aggregate/cumulative gross return for the SARB regime based portfolio. A similar method was followed for all the other regime identifying features.

Assuming that the average investor seeks to have minimum volatility in his or her portfolio, the portfolios were optimized to maximize gross returns whilst minimizing volatility. Since the models were regime incorporated, for each feature, the model optimized a portfolio for each regime. The aggregate/cumulative gross return of these regimes for each feature was then calculated to give the aggregate/cumulative gross return for that feature's regimes for the entire period under consideration.

The model did not factor in any costs related to trading such as brokerage fees. Moreover, the strategy used was a buy and hold strategy. That is, for each regime, one re-adjusted weights and bought those stocks suited for that regime according to the model. These were then held for the entire regime and cumulative gross returns calculated at the end of the time period.

The base model assumed the absence of regimes. Optimizing based on this gave a 761% gross return for the 20-year period under consideration. This averages to approximately 10.7% gross return per annum, as can be seen in table 4.7. This is generally higher than most fund investors gross returns. Moreover, with the SARB having an inflation target of $4 - 6\%$, this ensures an annual real gross return of at least 5.7%, before deducting costs. The SARB regimes' model had a gross percentage return of 1878% during the 20-year period under consideration, as can be seen in table 4.7. This is equivalent to an annual gross return of approximately 15.8%, which is much higher than 10.7% gross return by the base model.

TABLE 4.7: Portfolio optimization gross returns using no regimes and SARB detected regimes.

Detection Feature	No. of regimes.	No. of cycles.	Portfolio return.
None	0	0	761 %
SARB detected regimes	4	2	1 878 %

TABLE 4.8: Portfolio optimization gross returns using HMM regimes.

Feature	No. of regimes.	No. of cycles.	Portfolio return.
CPI headline	2	1	1 098.58 %
Domestic output: all groups	3	1.5	2 077.66 %
FTSE SA all share	2	1	1051.43 %
JSE top-40 index	2	1	976.92 %
Leading business cycle indicator	4	2	3413.71 %
Rand per US Dollar	6	3	11 221.63 %
Repo-rate	4	2	2 643.37 %

TABLE 4.9: Portfolio optimization gross returns using BinSeg regimes.

Feature	Change point type.	No. of regimes.	Portfolio return.
CPI headline	Mean	4	1 487 %

Continued on next page

Table 4.9 – continued from previous page

Feature	Change point type.	No. of regimes.	Portfolio return.
	Variance	5	1 643 %
Domestic output: all groups	Mean	4	1 275 %
	Variance	3	1 521 %
FTSE SA all share	Mean	5	587 %
	Variance	6	1 079 %
JSE top-40 index	Mean	6	1 056 %
	Variance	6	1 056 %
Leading business cycle indicator	Mean	2	1 121 %
	Variance	2	1 150 %
Rand per US Dollar	Mean	6	1 349 %
	Variance	6	1 349 %
Repo-rate	Mean	6	2 524 %
	Variance	6	2 011 %

TABLE 4.10: Portfolio optimization returns using PELT regimes.

Feature	Change point type.	No. of regimes.	Portfolio return.
CPI headline	Mean	3	1 281 %
	Variance	5	1 354 %
Domestic output: all groups	Mean	4	1 482 %
	Variance	5	2 360 %
FTSE SA all share	Mean	4	1 198 %
	Variance	11	1 270 %
JSE top-40 index	Mean	4	1 090 %
	Variance	13	1 411 %
Leading business cycle indicator	Mean	2	1 121 %
	Variance	5	1 201 %
Rand per US Dollar	Mean	7	2 168 %
	Variance	15	2 263 %
Repo-rate	Mean	9	4 976 %

	Variance	12	903 %

The exchange rate regimes identified by the HMM produced a portfolio with the highest return. It had a gross return of 11221% during the 20 year period under consideration, which when annualized, is 26.6%, as seen in table 4.8. This was more than double the annual gross return provided by the base model. This model's regimes had the second highest F1-scores of the regimes identified. However, it had 6 regimes and 3 cycles compared to the SARB's 4 regimes and 2 cycles. The next highest gross returning model only returned 4973%, which is less than half the highest returning model. This can be seen in table 4.10. It made use of the regimes identified using the mean, by the PELT algorithm.

The lowest performing model had a gross return of 587%, as seen in table 4.9, during the 20-years under consideration. This is equivalent to an annual gross return of approximately 9.2%. This portfolio made use of regimes detected by the binary mean segmentation algorithm, using FTSE SA all share data. The F1-scores for the regimes were 0.31 and 0.60, depending on which class is labeled as positive. This showed that it was unable to both capture-positive cases whilst being accurate with the classes it does not capture. However, although amongst the lowest, these were not the lowest F1-scores recorded.

TABLE 4.11: Average portfolio optimization gross returns by regime identification algorithm.

Algorithm	**Average Portfolio Return**
Base model	761%
Binary segmentation	1372%
HMMs	3211%
PELT	1719%
SARB	1878%

Portfolios using HMM detected regimes had, on average, the highest gross returns at 3211%, as seen in table 4.11. This is equivalent to an approximately 19.9% gross return annually. This was most likely because of HMMs being able to model an output of discrete states as a function of hidden stochastic processes. This creates multiple regimes based on stochastic properties. As such, a period where the expected gross returns are on average low, might be grouped as a single regime and vice versa when the expected gross returns are high. This ensures that regimes are created in a manner that maximizes expected gross returns since periods of low expected gross returns are not merged with periods of high expected gross returns, thereby reducing the expected gross returns and consequently, portfolio gross return.

The portfolio using SARB identified regimes produced the second highest average gross returns at 1872%, as can be seen in table 4.11. This is equivalent to an approximately 15.8% gross return annually. The SARB identifies regimes using fluctuations around the long-term trend of overall economic activity. The overall economic activity includes a combination of various factors, in various proportions, and direction. As such, the regimes identified by the SARB could be better suited to capture this, thus creating regimes that return higher results in the stock market.

Excluding the base model, the binary segmentation algorithm produced portfolios that on average, returned the least amount. They averaged 1372% during the 20 year period under consideration, which is approximately 14% per annum, as seen in table 4.11. This could be because the binary segmentation algorithm is an approximate search algorithm, rather than an exact algorithm. As such, the PELT algorithm, which is an exact search algorithm, produced portfolios with better gross returns compared to it.

Based of table 4.11, algorithms identifying more regimes tend to produce portfolios that on average, produce less gross returns. The binary segmentation and PELT algorithms had more regimes than the HMM and SARB, yet their portfolios returned less. This may be due to the identification of redundant regimes. For instance, the SARB only identified 4 regimes during the 20 year period under consideration, yet the PELT algorithm identified 15 when using variance changes on exchange rate data. This is much larger than what is identified by the SARB, thus could be identifying other regimes other than the economic regimes. Generally though, based on the discussions in section 4.2, it is clear that the introduction of regimes in portfolio optimization improves portfolio performance.

Chapter 5

Conclusions and Future Work

Through various models and experiments, this research showed that regime-based portfolios produced higher gross returns for an investor compared to traditional non regime-based portfolios. Moreover, it identified regimes based on different features and compare and contrast them against each other, thus successfully answering the research questions.

This chapter is the concluding chapter of the research and seeks to consolidate the findings of this research. Section 5.1 summarises the contributions made by each chapter, section 5.2 summarizes the findings of this research and how they answer the research questions, and section 5.3 identifies opportunities for future research

5.1 Conclusions

Asset management is a part of almost every person's life, be it directly or indirectly. Asset managers are constantly trying to gain a competitive advantage over their competitors, whilst individual investors are constantly looking for better ways to optimize their asset allocation and maximize returns.

With financial markets constantly changing their mean, volatility, and correlation patterns over time, static, all weather portfolios may not be the best way to maximize returns [34], [3]. This research sought to explore the use of regimes as opposed to no regimes in portfolio optimization. In doing so, the research explored the macro-economic factors that best explain the movement in the stock market, how regimes identified using different features and by different models differed from each other and lastly, how portfolios constructed from these regimes differed from each other.

Chapter 1 gave an introduction to the research,outlining the problem statement, research aim, research objectives, and research contributions. This was proceeded by chapter 2, which provided a review of literature related to this research. This included literature on how macro-economic factors are related to the stock market, literature on how regimes and change points are identified in data and lastly, methods on portfolio optimisation. This, helped show the lack of literature combining all these concepts; regime detection, and portfolio optimisation, which is what this research sought to explore.

Chapter 3 discussed the data and methods used in this research. This included both regime identification methods and portfolio optimisation methods. Results of these methods were discussed in chapter 4. HMMs, binary segmentation, and PELT algorithms were used to identify regimes. Of these, HMMs produced the least number of regimes, and these regimes were the closest to those identified by the SARB. This was seen in HMM regimes having the largest F1-scores, on average, in their classification report. The PELT algorithm produced the most change points when using variance as an identifier of regimes.

Chapter 4 concluded by discussing the results of portfolio optimisation. Regime-based portfolios were found to have much higher gross returns compared to the base model without regimes. Portfolios using regimes identified by the HMMs had the highest average gross return, at 3211%, whilst portfolios using binary segmentation identified regimes had the lowest average gross returns, at 1372%.

5.2 Research Contribution

This research sought to explore how percentage gross returns of portfolios incorporating economic regimes differ from that of one that does not incorporate regimes. By so doing, it would expand the South African literature on using data driven techniques for regime-based portfolio optimization. This was successfully done as discussed below.

This research was able to use data driven techniques (i.e HMMs, binary segmentation and PELT) to identify economic regimes in South Africa. HMMs produced the least number of regimes, and these were the closest to those identified by the SARB. The PELT algorithm however identified the most regimes, when using variance as a measure of regime detection.

Moreover, this research was able to show that in optimising portfolios, regime-based portfolios produced higher percentage gross returns compared to portfolios without regimes. Portfolios using regimes identified by the HMMs had the highest average gross return, at 3211%, whilst portfolios using binary segmentation identified regimes had the lowest average gross returns, at 1372%.

In conclusion, this research helps add to the body of work on portfolio optimization in South Africa. Previous research focuses mostly on economic theory to identify regimes as well as the different portfolio optimization techniques available to investors without incorporating regimes. This research helps to illustrate the results of data driven techniques in regime detection. Moreover, it helps to incorporate the use of regimes in portfolio optimization techniques and seeing how approximate and exact numerical search algorithms differ in results.

5.3 Future Work

This research explored the use of unsupervised techniques for regime identification in portfolio optimization. Potential areas of research exist to expand on this. This includes:

- This research used offline analysis to identify change points. That is, it first evaluated the entire data-set before deciding on the change points. Although useful in some areas, it unfortunately provides very little use in real-time investing. As such, exploring the use of online techniques to identify could provide real-time portfolio selection frame-work.
- The inclusion of alternate asset classes and South African regulations involved in asset management. For instance, a portfolio built by an asset manager for retirement purposes must adhere to certain regulations different from other portfolios. This ultimately affect the allocation process.

- This research made use of JSE data that covered the entire period under consideration. However, this automatically eliminates companies that delisted from the JSE during this time period hence introducing survival bias. As such, one could explore how the portfolio optimisation could incorporate such companies and automatically re-allocate capital once such companies delist.

Appendix A

Statistical Properties of Regime-Detection Data

TABLE A.1: Statistical properties of CPI headline from 2000 to 2019.

Period	Mean	Std. Dev.	Min. Value	Max. value
2000 - 2004	45.07	3.80	38.40	49.80
2005 - 2009	58.60	6.42	49.90	69.10
2010 - 2014	79.01	6.21	69.80	89.20
2015 - 2019	102.79	7.17	89.40	113.80

TABLE A.2: Statistical properties of domestic output: all groups from 2000 to 2019.

Period	Mean	Std. Dev.	Min. Value	Max. value
2000 - 2004	116.43	11.08	96.30	129.40
2005 - 2009	112.70	40.16	61.20	163.30
2010 - 2014	77.87	7.05	66.80	89.20
2015 - 2019	102.82	7.74	88.40	114.60

TABLE A.3: Statistical properties of FTSE SA all share from 2000 to 2019.

Period	Mean	Std. Dev.	Min. Value	Max. value
2000 - 2004	556.92	89.61	420.16	814.13
2005 - 2009	1 378.16	286.92	779.70	1 861.84
2010 - 2014	2 248.59	443.10	1 571.16	3 194.60
2015 - 2019	3 345.79	193.06	2 713.05	3 931.29

TABLE A.4: Statistical properties of JSE top 40 index from 2000 to 2019.

Period	Mean	Std. Dev.	Min. Value	Max. value
2000 - 2004	8 887.22	1 149.03	6 076.33	11 429.85
2005 - 2009	20 811.88	5 024.80	11 223.82	31 315.34
2010 - 2014	33 213.25	6 861.00	23 066.67	47 080.38
2015 - 2019	47 878.80	2 866.32	41 543.66	55 065.37

TABLE A.5: Statistical properties of leading business cycle indicator from 2000 to 2019.

Period	Mean	Std. Dev.	Min. Value	Max. value
2000 - 2004	89.38	5.80	82.70	101.20
2005 - 2009	100.76	5.95	87.20	108.30
2010 - 2014	105.41	1.56	102.50	108.90
2015 - 2019	102.81	2.91	97.70	108.20

TABLE A.6: Statistical properties of rand per US dollar exchange rate from 2000 to 2019.

Period	Mean	Std. Dev.	Min. Value	Max. value
2000 - 2004	8.01	1.61	5.61	13.00
2005 - 2009	7.38	1.09	5.67	11.47
2010 - 2014	8.66	1.46	6.60	11.74
2015 - 2019	13.70	1.12	11.30	16.90

TABLE A.7: Statistical properties of repo-rate from 2000 to 2019.

Period	Mean	Std. Dev.	Min. Value	Max. value
2000 - 2004	10.11	1.90	7.50	13.50
2005 - 2009	8.88	1.82	7.00	12.00
2010 - 2014	5.55	0.53	5.00	7.00
2015 - 2019	6.58	0.40	5.75	7.00

Appendix B

Change-Point Algorithms

Algorithm 1: BinSeg Algorithm [35]

Input: signal $\{y\}_{t=1}^{T}$, cost function $\mathcal{C}(\cdot)$, stopping criteria.
Initialize $L[]$ <- $\{\}$.
repeat
  $k < -|L|$, number of breakpoints;
  $t_0 < -0$ and $t_{k+1} < -T$;
  if *k > 0* **then**
    Denote by $t_i (i = 1, \ldots, k)$ the elements (in ascending order) of L, i.e $L = \{t_1, \ldots, t_k\}$.;
  else
    .
  end
  Initialize G a $(k+1)-$long array. ;
  for $i = 1, 2, \ldots, k$ **do**
    $G[i] < -\mathcal{C}(y_{t_i \ldots t_{i+1}}) - \min_{t_i < t < t_{i+1}} [\mathcal{C}(y_{t_i \ldots t}) + \mathcal{C}(y_{t \ldots t_{i+1}})]$.;
  end
  $\bar{i} < -\arg\max_{i} G[i]$;
  $\bar{t} < -\arg\min_{t_i < t < t_{i+1}} [\mathcal{C}(y_{t_i \ldots t}) + \mathcal{C}(y_{t \ldots t_{i+1}})]$;
  $L < -L \cup \bar{t}$;
until *stopping criterion is met.*;
Output: set $L[T]$ of estimated breakpoints indexes.

Algorithm 2: PELT Algorithm [35]

Input: signal $\{y\}_{t=1}^{T}$, cost function $\mathcal{C}(\cdot)$, penalty value β.
Initialize Z, a (T+1) long array; $Z[0]$ <- $-\beta$.
Initialize $L[0]$ <- θ.
Initialize χ <- $\{0\}$.
for $t = 1, 2, \ldots, T$ **do**
    $\bar{t} < -\arg\min_{s \in \chi}[Z[s] + \mathcal{C}(y_{s \ldots t}) + \beta]$.;
    $Z[t] < -[Z[\bar{t}] + \mathcal{C}(y_{\bar{t} \ldots t}) + \beta]$.;
    $L[t] < -L[\bar{t}] \cup \{\bar{t}\}$.;
    $\chi < -\{s \in \chi : Z[s] + \mathcal{C}(y_s \ldots t) \leq Z[t] \cup \{t\}$;
end
Output: set $L[T]$ of estimated breakpoints indexes.

Bibliography

[1] Prateek Agarwal. “Business Cycle”. In: *Intelligent Economist* (2022).

[2] Peter Akioyamen, Yi Zhou Tang, and Hussien Hussien. “A hybrid learning approach to detecting regime switches in financial markets”. In: 2020, pp. 1–7.

[3] Andrew Ang and Allan Timmermann. “Regime changes and financial markets”. In: *Annu. Rev. Financ. Econ.* 4.1 (2012), pp. 313–337.

[4] South African Reserve Bank. “Quarterly Economic Review”. In: *South African Reserve Bank* (2020).

[5] Brian J. Bloch. “Naive Diversification vs. Optimization”. In: *Investopedia* (June 2019).

[6] Londiwe Buthelezi. “These were South Africa’s biggest asset managers in 2020”. In: *News24* (Mar. 30, 2021).

[7] Raghutla Chandrashekar et al. “Macroeconomic variables and stock prices in emerging economies: A panel analysis.” In: *Theoretical & Applied Economics* 25.3 (2018).

[8] James Chen. “Mean-Variance Analysis”. In: *Investopedia* (2021).

[9] James Chen. “Technical Indicator: Definition, Analyst Uses, Types and Examples”. In: *Investopedia* (2021).

[10] GUILLAUME Coqueret and VINCENT Milhau. “Estimating covariance matrices for portfolio optimization”. In: *ERI Scientific Beta White Paper* (2014).

[11] Jeff Desjardins. “Crisis Investing: How 14 Different Asset Classes Performed in Times of Distress”. In: *Visual Capitalist* (2015).

[12] *Ecclesiastes 11:2 NLT*.

[13] Camille Gaines. “Types of Asset Allocation”. In: *Retire Certain* (2022).

[14] Robert G Gallager. *Stochastic processes: theory for applications*. Cambridge University Press, 2013.

[15] Evan Gilbert and Luke Meiklejohn. "A comparative analysis of risk measures: A portfolio optimisation approach". In: *Investment Analysts Journal* 48.3 (2019), pp. 223–239.

[16] James D Hamilton. "A new approach to the economic analysis of nonstationary time series and the business cycle". In: *Econometrica: Journal of the econometric society* (1989), pp. 357–384.

[17] James B Heaton, Nick G Polson, and Jan Hendrik Witte. "Deep learning for finance: deep portfolios". In: *Applied Stochastic Models in Business and Industry* 33.1 (2017), pp. 3–12.

[18] Yu Hsing. "The stock market and macroeconomic variables in a BRICS country and policy implications". In: *International Journal of Economics and Financial Issues* 1.1 (2011), pp. 12–18.

[19] Philippe Jorion. *Value at risk-the new benchmark for managing financial risk*. 3. McGraw-Hill Education, 2007.

[20] Magnus Karlsson. "Hidden Markov Models". In: *Erişim Tarihi* 20 (2004), p. 2017.

[21] Clarence CY Kwan. "An introduction to shrinkage estimation of the covariance matrix: A pedagogic illustration". In: *Spreadsheets in Education (eJSiE)* 4.3 (2011), p. 6.

[22] Ming-Hua Liu and Keshab M Shrestha. "Analysis of the long-term relationship between macro-economic variables and the Chinese stock market using heteroscedastic cointegration". In: *Managerial Finance* (2008).

[23] E Maré and E Flint. "Regime-based tactical allocation for equity factors and balanced portfolios". In: *South African Actuarial Journal* 19.1 (2019), pp. 27–52.

[24] Harry M Markowitz. "Portfolio Selection". In: *The Journal of Finance* 7.1 (1952), pp. 77–91. ISSN: 00221082, 15406261.

[25] Harry M Markowitz. "The early history of portfolio theory: 1600–1960". In: *Financial analysts journal* 55.4 (1999), pp. 5–16.

[26] Robert Andrew Martin. "PyPortfolioOpt: portfolio optimization in Python". In: *Journal of Open Source Software* 6.61 (2021), p. 3066. DOI: 10.21105/joss.03066. URL: https://doi.org/10.21105/joss.03066.

[27] "Mean". In: *Wikipedia* (). URL: "https://en.wikipedia.org/wiki/Mean".

[28] Ardavan Mobasheri. "A history of economic cycles going back to the 1850s suggests a recession is near". In: *CNBC* (2017).

[29] Rudzani Mulaudzi. "South African Reserve Bank Cleaned Economic Data". In: *GitHub* (2021). URL: https://github.com/rudzanimulaudzi/sarb_feature_set.

[30] John M Mulvey, Han Hao, and Nongchao Li. "Machine learning, economic regimes and portfolio optimisation". In: *International Journal of Financial Engineering and Risk Management* 2.4 (2018), pp. 260–282.

[31] André F Perold. "The capital asset pricing model". In: *Journal of economic perspectives* 18.3 (2004), pp. 3–24.

[32] Lawrence Rabiner and Biinghwang Juang. "An introduction to hidden Markov models". In: *ieee assp magazine* 3.1 (1986), pp. 4–16.

[33] Immanuel Seidl. "Markowitz versus regime switching: an empirical approach". In: *The review of finance and banking* 4.1 (2012), pp. 33–43.

[34] Bruce Stewart. "Introduction to Regime Based Asset Allocation". In: *Investments & Wealth Institute* (2020).

[35] Charles Truong, Laurent Oudre, and Nicolas Vayatis. "Selective review of offline change point detection methods". In: *Signal Processing* 167 (2020), p. 107299.

[36] "Variance". In: *Wikipedia* (). URL: "https://en.wikipedia.org/wiki/Variance".

[37] Jilten Vasant et al. "A Comparison Of Mean-Variance And Mean-Semivariance Optimisation On The JSE". In: *The Journal of Applied Business Research* (2014).

[38] Praphan Wongbangpo and Subhash C Sharma. "Stock market and macroeconomic fundamental dynamic interactions: ASEAN-5 countries". In: *Journal of asian Economics* 13.1 (2002), pp. 27–51.

[39] Ruoyun Zhang, Chao Yi, and Yixin Chen. "Explainable Machine Learning for Regime-Based Asset Allocation". In: *2020 IEEE International Conference on Big Data (Big Data)*. IEEE. 2020, pp. 5480–5485.