Meta-heuristic solution approaches to the portfolio optimisation problem

Bryce Evans Senekal

UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY

100
1918 · 2018

Thesis presented in partial fulfilment of the requirements for the degree of
Master of Commerce
in the Faculty of Economic and Management Sciences at Stellenbosch University

Supervisor: Dr Linke Potgieter March 2018

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 1, 2018

Abstract

The portfolio optimisation problem is a well documented and researched combinatorial problem in the financial and operations research fields. The problem definition is defined as having to decide on which financial assets to invest in so as to minimise the associated risk while still maintaining a desired level of return on the investment. To accomplish this, various models have been formulated to help find accurate methods of quantifying and then minimising this risk. One such model is the Markowitz Mean-Variance model, introduced in 1952 as the initial method of quantifying risk and beginning the renaissance of investing in a diversified portfolio. This paper attempts to solve three problems associated with investing in a diversified portfolio using the Markowitz model. These are; that the time taken to solve the models with traditional mathematical methods are unusable for real world dataset sizes; that the initial investment needed to purchase a fully diversified portfolio is large enough that the common investor may struggle to invest early enough in his or her lifespan; and that the Markowitz model is based on the assumption that financial assets expected returns are normally distributed.

These problems are solved in three parts. The first part is explaining what unit trusts are and how they can be used a tool for the average investor to use as an aid in efficiently investing in a diversified portfolio. The second is to overcome the estimations errors associated with Markowitz's assumption of normality issue by using a distribution free estimate of the variance-covariance matrix gained through shrinkage theory. An added effect of the shrinkage theory estimate is to attempt to correct the estimation errors that come with financial data due to the high dimensionality property it possesses. The third is to apply and compare a selection of meta-heuristics using the adjusted model on the portfolio optimisation problem to see if they provide a usable alternate technique for real world datasets. The meta-heuristics used in this paper are Simulated Annealing (SA), the Artificial Bee Colony (ABC), and the Pareto Envelop-based Selection Algorithm (PESA).

A collection of unit trusts were collected and evaluated, before using the listed meta-heuristics to find good solutions to the unit trusts selection problem. This would allow ordinary investors to have access to a diversified portfolio, whose risk may be lowered even further by diversifying between unit trusts. The shrinkage theory estimate was successfully applied to overcome the second problem and preliminary results indicate that there may be some benefit to using the new estimate as it may provide more accurate portfolio covariances and lead to more assured returns in the future. The solutions to two of the meta-heuristics, namely the ABC and the PESA, were found to be within an acceptable range of the true efficient set of portfolio for the data set, while the SA results were not successful. The solutions were all found within a relatively usable time period, namely 2 to 4 hours, and can be concluded to be of use for solving the portfolio optimisation problem for larger data sets in a usable time period.

Acknowledgements

The author wishes to acknowledge the following people for their various contributions towards the completion of this work:

- Firstly, I would like to thank my mother and father for providing me the opportunity to pursue this masters degree by providing me with both emotional and financial support. Despite the jokes about me studying forever, I know they are proud of the work I've done.
- And, of course, I would like thank my supervisor, Dr Linke Potgieter, for all it took to get this masters degree finished. For all her guidance, without which this thesis would have taken far longer, for all the patience in explaining the little details that contribute to an academic thesis that did not always make sense to me, and for her understanding when things became difficult.

Table of Contents

List of Reserved Symbols

Symbols in this thesis conform to the following font conventions:	
$\mathcal{A}$	Symbol denoting a **set**
A	Symbol denoting a **solution vector**
$\|\|a\|\|$	Symbol denoting the **frobenius norm**

Symbol	Meaning
$\cdot$	Symbol used to denote the multiplication operator

List of Acronyms

ACO: Ant Colony Optimisation

ABC: Artificial Bee Colony

ASISA: Association for Savings and Investment South Africa

ALSI: All Share Index

CCMV: Cardinally Constrained Mean Variance

CIS: Collective Investment Scheme

CISCA: Collective Investment Scheme Control Act

CVar: Conditional Value-at-Risk

FSB: Financial Services Board

FTSE: Financial Times Stock Exchange

GCLPM: Generalised Co-Lower Partial Moments

GRASP: Greedy Randomised Adaptive Search Procedure

GRG: Generalised Reduced Gradient

JSE: Johannesburg Stock Exchange

LPM: Lower Partial Moments

MAR: Minimum acceptable rate

MPT: Modern Portfolio Theory

PAES: Pareto Archived Evolutionary Strategy

PESA: Pareto Envelope-based Selection Algorithm

PSO: Particle Swarm Algorithm

SA: Simulated Annealing

SCM: Sample covariance matrix

SLP: Successive Linear Programming

SPEA: Strength Pareto Evolutionary Algorithm

SRA: Successive Regression Algorithm

UTCA: Unit Trust Control Act

VaR: Value-at-Risk

List of Figures

List of Tables

List of Algorithms

CHAPTER 1

Introduction

Contents

"...since you cannot successfully time the market or select individual stocks, asset allocation should be the major focus of your investment strategy, because it is the only factor affecting your investment risk that you can control." - William J. Bernstein. [6]

The quote by Bernstein summarises the ideology introduced by Harry Markowitz in 1952 when he challenged the notion of investing solely in the strongest of securities. Markowitz argued that in order to reduce his investment risk he needs to diversify his investment in a manner that will minimise the exposure to any one particular risk [50]. This may be done by investing in a variety of not perfectly correlated assets.

Diversification does, however, come with some of its own issues, two of which we focus on in this study. The first being that due to the sheer number of companies listed on financial markets, not to mention alternative investments such as futures or bonds, deciding on a suitable investment portfolio is a difficult task that requires a lot of expert knowledge. The second, once a preferred portfolio has been determined, the issue for most ordinary investors is that being able to purchase a diversified portfolio consisting of, for example, a number of securities, a few bonds, and commodity futures, requires a relatively large initial investment.

The second problem has been addressed by the establishment of collective investment schemes, including hedge funds, open-ended investment schemes, and unit trusts [17]. These schemes offer professional management, low initial investment amounts, diversification, and access to expensive shares for the ordinary investor. Unit trusts are therefore used as viable alternative to attempting to build a new portfolio for the investor, as they will allow an investor to begin investing in a fully diversified portfolio immediately due to the low minimum investment value. In order to help address the first problem, a number of mathematical models have been developed with the objective to optimise portfolio performance [51, 32, 33].

1.1 Unit trusts

Unit trusts are a pre-designed collective investment scheme, where groups of investors pool money together and invest in a number of securities, bonds, money market, and various other investments [17]. Each unit trust is made up to a set number of 'units' which represent the percentage share of ownership. In other words, if a unit trust is made up of a thousand 'units', each 'unit' would represent a tenth of a percent of ownership, and someone owning three hundred 'units' would be entitled to 30% of the capital in the unit trust.

To provide more information on unit trusts, a brief overview of the history of unit trusts in the South African context is provided in §1.1.1, before expanding on the different unit trust types and their structures in §1.1.2 and 1.1.3.

1.1.1 Historic overview

Unit trusts were first introduced to South Africa in 1965 in an advert in the Financial Mail proclaiming South Africa's first mutual fund, the term used for unit trusts in the United States at the time [17]. The growth of the unit trust industry took a sharp decline, however, due to the market crash in 1969, which resulted in regulatory authorities adopting a more cautious approach to marketing of unit trusts and affected investor confidence. The market did not reach the same level until 1983. Luckily, lessons were learnt from this crash and certain positive outcomes from the crash were gained. People gained experience in the volatility of financial markets and began to stagger investment amounts and not invest in lump sum values as well as learning that equity investments, particularly unit trusts, should be treated as long term investments. Industry also learnt that investors needed to be educated on the dangers of the market, especially the dangers of buying at high prices and selling at low ones.

Once the industry recovered, more funds started to be released, leading to 271 rand-denominated funds being present in the market by 1999. The reason behind the growth of unit trusts was the advancement of computing power in the 1990's. This allowed smaller companies to be able to launch unit trusts and larger companies were able to manage multiple unit trusts due to the fact that computers became more user friendly and data processing tools became more powerful. This growth trend continued exponentially, with more funds being created from 2000 to 2005 than all the funds created since the inception of unit trusts in 1965 (see Table 1.1 for a view of the growth pattern of unit trusts). If one observes the growth of the unit trusts in comparison to the economic growth, it is clear that unit trusts outperform the average economic growth, however, this can be attributed to unit trusts becoming a more attractive investment option as time progressed, rather than the actual performance of the unit trusts [52].

An important development in the history of unit trusts was the introduction of managed Prudential Funds in 1996. Prudential funds are funds that are managed according to asset management guidelines applicable to pension funds, more specifically of Regulation 28 of the Pension Funds Act. This allowed the unit trust market to move into territory previously reserved for the retirement funding industry by marketing the new balanced funds in addition to high risk-reward equity based unit trusts [52].

Year	Number of funds	Asset value (R '000 000)	Annual asset growth rate %	GDP (R '000 000)	Annual GDP growth rate %
1965	2	3		7 197	
1980	12	682.8	43.6	62 730	12.19
1990	36	7550.1	27.17	289 816	18.08
1995	88	33 675	34.86	548 100	13.59
2000	334	128 384.7	30.69	922 148	10.97
2005	617	415 131	26.46	1 529 658	10.65

TABLE 1.1: *A summary of the growth of the unit trust market compared to GDP growth for the years 1965 to 2005 [52].*

1.1.2 Structure of unit trusts

Every unit trust can be divided up into a few separate entities, as illustrated by Figure 1.1. The first is the set of investors who own the 'units'. The second is the portfolio of underlying assets. The portfolio is made up of the contributions of the individual investors which is then invested in a collection of securities, bonds, cash, and properties. The manager of the unit trust is responsible for the decision of which of these assets to purchase and when to purchase them. They also ensure that the fund functions on a daily basis. This includes different aspects such as the marketing of the fund, the creating and selling of 'units', and keeping records.

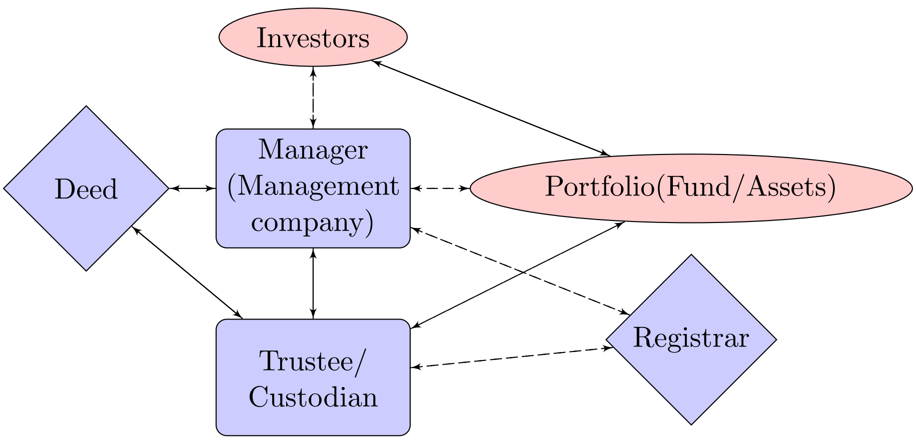


FIGURE 1.1: *An overview of the general structure of a collective investment scheme, depicting the relationships and interactions of the individual stakeholders.*

According to interviews from some fund managers from the top investment companies in South Africa; like Prudential [4, 61], Coronation [12], Investec [54], and Momentum [56], most agree with following certain strategies to avoid risk when creating a portfolio. These are to diversify the portfolio across geographical locations, currencies, asset classes, manager styles, and security types; to be disciplined when it comes to the actual purchase and not buy overvalued securities; to give a preference to companies with a proven track record of paying dividends to protect the income of the client; and to ensure that the securities purchased are relatively liquid to avoid losses made from not being able to sell the securities when the time comes.

These decisions are, however, governed by a deed, or a set of rules and regulations, which will fall under the authority of the Financial Services Board (FSB). This deed is enforced by a trustee, who is appointed on behalf of the fund, and is checked upon by a registrar, who is appointed by the FSB to ensure that the trustee and manager are performing their respective functions efficiently.

1.1.3 Unit trust types

There are many different types of unit trusts as illustrated by the classifications in Figure 1.2 [17]. It is possible, then, that two unit trusts who have very similar compositions, and therefore similar risk and return profiles, having very different characteristics, or visa versa, where two unit trusts with the same classification, may have very different risk and return profiles. This is due to the many permutations that the individual tiers of the classification process can have.

Tier one contains the domicile classification of the fund and is split into 4 subclasses. The subclass, the South African unit trust, is regulated by enforcing a minimum of 70% of the assets invested in the South African market at all times, and is allowed to then have the other 30% outside of South African under the caveat that 5% of the outside funds are still within Africa. Global funds, previously known as Foreign funds, are funds that contain at least 80% of the assets invested outside of South Africa but must be split across different locations, if not, the fund is then classified as a Regional fund, where the fund would be classified as a Japanese-Regional fund if 80% of the assets were located in Japan. The Worldwide unit trust contains no regulations on domicile allocation, other than those governing the investment in overseas funds. The regulations on investing overseas have changed throughout the years, but currently are restricted to allowing a Collective Investment Scheme (CIS) management company to invest no more than 35% of the total assets under management to an overseas fund. This relates to all funds under the management of the company, not only the foreign based ones and is designed to stop a company from only managing foreign-based unit trusts [17]. Another detail about the Regional funds is that due to focus on location, ranking becomes difficult as each fund is heavily influenced by local economic factors, and the Association for Savings and Investment South Africa (ASISA) discourages any formal attempt at ranking.

Tier two contains the asset allocation split of the unit trust. The asset allocation tier contains four distinct subsets, Equity funds, Interest-bearing funds, Property funds, and Multi-asset funds. Equity unit trusts require that the unit trust invests in a minimum of 80% towards securities on a stock exchange, the remainder may then be invested at the discretion of the manager. The Interest-bearing funds invest exclusively in bonds, money-market, or other interest earning securities. The Property unit trusts invest in property shares, real estate investment trusts, or other CIS in property. A point about true property unit trusts, are that they are close ended, meaning a buyer can only buy a 'unit' if there is a willing seller and vise versa. Finally, the Multi-asset unit trusts are a combination of all the above different classes. Multi-asset funds seek to maximise return over the long run throughout all investment options [17].

The third tier, namely the investment focus tier, expands on the second tier and provides more information on what types of investments the second tier assets will consist of. Simplified, this provides information as to which type of companies will be invested in if the unit trust has been classified as an Equity unit trust, and so on. To aid in summarising the third tier, the different focuses and a description of each has been provided in Tables 1.2, 1.3, and 1.4, while there is only one sub-category for Real Estate unit trusts, the general one.

Examples of the combinations of the different tiers include a South African-Equity-Mid Cap fund, a Global-Equity-Industrial fund, or a South African-Interest bearing-Variable term fund.

EQUITY

Type	Description
General Funds	Invest in shares across all sectors of the equity market and across all ranges of capitalisation.
Large Cap	Invest 80% in companies with a large market capitalisation, 100 of securities must be in this category.
Small & Mid Cap	Invest 80% in companies with a smaller market capitalisation, if a company becomes large cap, the fund is not obligated to sell if it constitutes less than 20% of the composition.
Resource	Invest 80% in companies listed in the metals, minerals, energy, chemical, or forestry industries on the JSE/FTSE. Up to 10% can be in other industries, provided they are related to the listed industries.
Financial	Invest 80% in companies whose primary business activities are in the Financial sector such as banks, insurers, and brokers. Up to 10% can be in other industries, provided they are related to the industry.
Industrial	Invest 80% in companies listed in the engineering, transportation, construction, electronic, telecommunications, or food producer and retail industries on the JSE/FTSE. Up to 10% can be in other industries, provided they are related to the listed industries.
Unclassified	Created to house funds whose listings can not be categorised in any other classifications.

TABLE 1.2: *A table containing the descriptions of the different requirements for each classification of equity typed unit trusts [17].*

MULTI-ASSET

Type	Description
Flexible	Invest in a combination of securities, bonds, money market, and property. Aggressively managed to use full advantage of the flexibility.
High Equity	Restricted to 75% maximum in equity, and 25% maximum in property.
Medium Equity	Restricted to 60% maximum in equity, and 25% maximum in property.
Low Equity	Restricted to 40% maximum in equity, and 25% maximum in property.
Income	Focussed on maximising income while preserving capital by investing predominantly in bonds, fixed deposits, and other income earning securities. Restricted to a maximum of 10% in equity and 25% in property.

TABLE 1.3: *A table containing the descriptions of the different requirements for each classification of multi-asset typed unit trusts [17].*

1.1.4 Cost structures

The costs associated with unit trusts were originally split into three different categories, namely initial fees, annual management service fees, and compulsory fees [52, 17]. Compulsory fees were a fixed charge charged every month to cover the costs associated with purchasing securities. Since 2002, however, the compulsory fee category was disbanded and renamed portfolio charges, so this section will focus on the initial and management fees [17].

Initial, or entry, fees are levied to cover the costs of the broker commission to buy the securities in the unit trusts and the costs of the administration of the management company of the unit

INTEREST-BEARING

Type	Description
Short-term	Invest in Interest-bearing assets with a weighted average of fixed maturity dates less than two years in the future.
Variable Term	Invest in Interest-bearing assets with varied maturity dates with no limit on the average weighted duration.
Money Market	Seek to maximise interest income with protecting overall income. The maturity dates of these investments are usually less than 13 months.

TABLE 1.4: *A table containing the descriptions of the different requirements for each classification of interest-bearing typed unit trusts [17].*

trust. This is a flat rate charged once off and deducted from the initial capital investment. The initial fee was capped under the Unit Trusts Control Act (UTCA), however, this was deregulated when the UTCA was replaced by the Collective Investment Schemes Control Act (CISCA), and can now be any value so long as it is disclosed to potential investors upfront [17]. One pitfall to avoid when evaluating initial fees is that the disclosed value only represents the administration costs, and not the broker fee, so investors should be aware of this when selecting a unit trust.

The annual management service fee is levied to cover the cost of business for the management company. These fees are also deregulated under CISCA, so once again investors need to be thorough in their research of costs. Within the annual fees are trailer fees paid to brokers to provide advice to clients and performance fees paid to fund managers to incentivise over-performance of the fund compared to a benchmark [52]. The choice of benchmark is then very important as it can affect the level of these performance fees. For the investor to decide on whether the performance fee is acceptable the following questions may be asked:

- Are the fees capped?

If not, the investor may end up losing a significant portion of the growth to the fees.

- Is the benchmark target reasonable?

If too low, the fund manager may still end up being rewarded for mediocre performance, while too high may cause the manager to take more risks to achieve the target.

- Is the benchmark period reasonable?

If too long, it causes investors to pay fees related to periods not associated with their investment in the fund, and if too short, allows the fund manager to follow riskier short-term strategies [17].

There are types of fees that do not fall within the description of the classifications, namely switching and exit fees. They are rarer fees that are only charged in certain circumstances. The switching fees are designed to protect management companies against clients switching between funds within the companies' set of funds. Most companies will either offer a limited number of free switches a year, or charge the difference between the current funds initial charge and the proposed funds initial charge. i.e. if the current fund has a 2% initial investment and the client wishes to move to one with a 5% initial investment, the company charges 3%. The exit fees are in turn designed to protect management companies from clients who wish to disinvest from funds before a given time period. While this is a rare charge, it can be used to protect management companies from initial periods of poor performance despite a long term growth [52].

An added consideration for investors seeking to diversify between unit trusts is how they will actually invest in the unit trusts once the portfolio has been decided upon. The investor may invest directly with the managing company of each unit trust and will incur the above costs for each separate investment or the investor may use a Linked Investment Service Provider (LISP). A LISP provides the investor with the option of investing in the chosen unit trusts in one location where the LISP has agreements with the managing companies of the unit trusts it offers [28, 17]. This offers convenience for the investor as they are able to switch between funds with ease when compared to directly investing, however, this will come with added fees from the LISP.

1.1.5 Unit trust performance

While unit trusts may solve the problem of investing in a fully diversified portfolio as discussed in the opening of this thesis, the question of whether they are actually able to beat the market average, and therefore worthwhile to invest in, is still to be answered.

Both Knight and Firer [40], and Biger and Page [8] found that when comparing risk-adjusted unit trust performance to that of the market average performance, no individual fund could be said to significantly under perform. Knight and Firer also showed that the same fund managers performed well by showing that the rankings for the funds were consistent at a 5% level [40].

Meyer-Pretorius and Wolmarans [52] tested the average performance of unit trusts during the period 1988 to 2005 against that of the JSE All Share Index (ALSI) for the same time period. In this study it was found that without costs, the unit trusts outperformed the average by approximately 1.5%, but when costs were incorporated the performance fell short of the ALSI by 5.6%. This highlights the discussion in §1.1.4 that investors need to be vigilant in understanding the cost structure of the unit trust they invest in.

Bertolis and Hayes [7] tested the general equity unit trust performance against that of the ALSI for the period 1994 to 2012. This was further subdivided into 6 periods so that each period would represent a different economic cycle period. The results of this study can be seen in Table 1.5.

Period	Economic Cycle	ALSI Growth	Unit Trust Growth
Jan 1994 – Dec 1996	Expansion	11.36% p.a	16.49% p.a
Jan 1997 – Aug 1999	Contraction	6.27% p.a	3.91% p.a
Sept 1999 – Dec 2003	Stable	13.47% p.a	13.62% p.a
Jan 2004 – Nov 2007	Expansion	20.72% p.a	19.11% p.a
Dec 2007 – Ayg 2009	Contraction	-3.85% p.a	-0.66% p.a
Sept 2009 – Dec 2012	Stable	16.4% p.a	14.07% p.a
Overall		14.44% p.a	14.01% p.a

TABLE 1.5: *A summary of the results presented by Bertolis and Hayes [7].*

Their results are inconclusive, only showing that unit trusts do not significantly underperform the market as a whole and that unit trust growth does not have an advantage over market growth in any period of the economic cycle.

The collection of the above mentioned results seem to indicate that on average the unit trust market performs on par with the general market. A good selection of unit trusts may therefore potentially perform as well as a good selection of securities.

1.2 Portfolio selection: An introduction and history

Portfolio selection is the construction of an investment portfolio consisting of risky assets, where risky assets are classified as investments in property holdings, securities in listed companies, and government bonds [75]. Risky assets are so called due to the inherent risk associated in purchasing them as the prices of these assets are linked to financial markets, which in turn are subject to supply and demand movements as well as socio-economic factors that cause prices to fluctuate. This risk being that by the time of maturity, i.e. when the money is due, the price may have dropped below that of the desired level of return. In addition, the companies that are invested in are not required to pay dividends, adding to the inherent risk of the price fluctuations.

The general process of portfolio selection is done in two steps wherein the portfolio is first divided into classes of risky assets such as the above mentioned securities, commodities, and bonds; and secondly each class is then constructed separately, while keeping the split between the classes themselves the same as in the first step. Consequently, a decision maker is required to find the best percentage split between having securities or property or bonds as well as which securities, commodities, and bonds are the best to include in the portfolio. This is why finding an optimal portfolio is not a trivial task.

Research into the portfolio optimisation problem began in 1952 when Harry Markowitz introduced the Mean-Variance model [50]. He challenged the conventional thinking of investing in the strongest security [51] and instead reasoned that a diversified portfolio of a few uncorrelated securities would give a safer, more assured return for an investor even if the strongest security performed badly. A simple formulation to quantify the risk of any given portfolio was presented, measured as the variance of the portfolio's return, along with a formulation for the expected return of a portfolio, i.e. how much a portfolio would be worth at the end of the investment period. For this model to function, Markowitz made the assumption of normally distributed returns to allow the variance be used as the measure of risk. The resulting model was therefore quadratic in nature. Markowitz later introduced the use of semi-variance as a more accurate quantification of risk.

One of the problems associated with the quadratic model is the computational time required to find the optimal solution for large datasets. Even more so, during the 1950's and 1960's the computing power available was not enough to solve the model for large datasets in a usable time. In order to address this, William Sharpe introduced a linear approximation of the Markowitz model in which he replaced the variance with a linear function of the general market risk associated with investing, and making the assumption that with adequate diversification, the missing risk information can be made very small, enough to render the missing information moot [68].

From the mid 1970's, new methods of quantifying risk were introduced in an attempt to provide alternatives to Markowitz' model, starting with Bawa's introduction of the Lower Partial Moments as a way of determining the semi-variance of the portfolio [5]. Following the collapse of a prominent bank due to misinformation by one of the traders, a simple version of downside risk was introduced called Value-at-risk, which showed the amount that stood to be lost in a certain percentage of the time under normal business cycles [33].

In addition to developing new quantifications of risk, new solution methodologies were developed to solve the portfolio optimisation problem in a usable time. These include alternate linearisation models, meta-heuristics, as well as stochastic programming techniques. A more detailed discussion is given in Chapter 2.

1.3 Financial advisors

A financial advisor at his or her core description is someone who assesses the financial needs of an investor and provides them with a plan for meeting these financial needs [71]. The responsibilities and services provided by many advisors include; meeting with clients personally to discuss their individual financial needs, explaining the different types of services available to the investor; making recommendations to the investor based on the individuals financial needs; help the investor with creating contingency plans for specific circumstances; and monitoring the investors portfolio and make the changes necessary to improve performance.

With the wide variety of funds available for an investor to choose from an investor may choose to go to an advisor for help in making this decision. Either an independent financial advisor or an associated financial advisor may be approached. The difference between the two is that an associated financial advisor will have to provide advice within the context of the products available at the firm for whom they work [58], in contrast to an independent advisor who is not restricted to a certain set of company products. As they are not linked to a particular company, they have a wider range of selections to help find a better investment for the investor. The investor also still has the benefit of investing in reputable companies, but retains a relatively unbiased source of advice. However, this does increase the costs somewhat as the independent advisor requires payment as well as the final investment fees [29]. From here the processes followed are similar in nature between the two types of advisors.

When an investor approaches an investment company for assistance in portfolio selection, the advisor assigned to the investor will conduct an interview with the investor to create a risk profile. This questionnaire will include determining the investor's goals for investing, i.e. to turn extra income into capital, to create a capital base for retirement, etc; determining what the earning potential of the investor; and any other concerns for the investor such as wanting to keep a certain percentage overseas or locally based. Once this is complete, the advisor will provide the investor with recommendations of the funds the company manage that best suit the investor, i.e an Equity fund, if the investor is young and willing to take chances with risk in order to potentially grow their capital base exponentially [41].

Currently, in the financial services industry, the process of selecting what to include in portfolios is done mainly by using expert knowledge gained from years of trial and error, combined with data analysis. This is due to the issues that arise from non-quantitative industry concerns like having to keep high investor confidence to avoid disinvestment, wanting to invest in companies that reflect good or preferable business practices such as charity work, or being outspoken against world issues, or perhaps wanting to support local businesses [75].

Theoretically, this is the process that should be followed, however, some research has been done into determining how objective investors are when it comes to making decisions in risky situations. Nofsinger and Varma performed a series of tests on a group of 100 financial advisors to determine how objective advisors are [62]. The tests done include the Cognitive Reflection Test, which determines how analytical or intuitive a person is, a risk aversion question combined with a reversed framing, and determining the utility of patience for each advisor. The reversed framing entails asking a question determining whether the advisor would choose an option to save 200 out of 600 people or have a chance at saving everyone and then to ask a similar question later except to phrase it as the option to have 400 out of the 600 people or to have a reduced chance for everyone to die. One would imagine that the same option should be chosen in both questions, however a third of the advisors chose a different option in the two questions. This is a worrying result for investors who would hope that the advice would remain constant. When compared to results from the general population, it indicates that financial advisors are more objective and rational than the general population. The choice of advisor, however, is important.

1.4 Problem description

When unit trusts were first introduced in 1965, growth was very slow with only 13 unit trusts being available to the public until 1982. However, with the improvement in technology coupled with a demand for unit trusts, this figure has steadily grown from the initial 13 to the 271 unit trusts at the close of the century and then to today's figures of over a thousand unit trusts [57]. The number of unit trusts available today make choosing the right combination of unit trusts to invest in a difficult choice.

The problem of selecting an optimal set of unit trusts may be modelled as a portfolio optimisation problem. However, the portfolio now consists of unit trusts, with their respective risk and return profiles, whereas in the original problem, the case of selecting different securities, bonds, or commodities is addressed. The unit trust selection model caters to investing in unit trusts as medium to long term investments, i.e. a 5 to 7 year holding period as so uses as much data available to the investor.

The Markowitz model may also be applied to the unit trust selection problem, as well as the other models mentioned in Chapter 2, to aid in this decision making process. However, using traditional solution methodologies to solve these models on large financial datasets typically take long to solve and makes them unsuitable for real world applications.

As a result, meta-heuristics are explored as a possible solution methodology to overcome these limitations of the mathematical models for real world use. An additional problem for this paper is that most optimisation techniques are not usable for real world applications due to alternate concerns of real world investors beyond risk and return. Therefore the meta-heuristics will be created with the view of implementing them as part of a decision support system for a financial advisor or independent investor by adding some of the concerns as constraints, like keeping investor confidence high by ensuring that short term return is kept positive, as part of the model.

1.5 Scope and objectives of the project

Only a collection of South African domiciled unit trusts collected from the Bloomberg Terminal database will be considered for inclusion in the portfolios. Only the price history will be used to determine the expected return, other information, like the costs, will be excluded for this study as they can be incorporated later as an adjusted return. In order to determine a set of efficient portfolios, only meta-heuristics will be considered as a possible methodology to search through portfolio combinations, so as to determine the efficacy of meta-heuristics in determining the efficient set. Furthermore, the meta-heuristics will use an adjusted Markowitz Mean-Variance model to create portfolios that are theorised to be more accurate.

The following objectives are pursued in this study:

Objective I: Perform a literature review on

(i) the portfolio optimisation problem,
(ii) the development of meta-heuristics,
(iii) and the application of meta-heuristics to the portfolio optimisation problem.

Objective II: Evaluate unit trusts as an investment tool by

(i) Collecting and analysing data on a selection of South African domiciled unit trusts, and estimating the expected return of the funds and the covariance between the funds.

Objective III: Develop a methodology to solve the unit trust selection problem by

(i) Improving the covariance estimation used in the Markowitz model to address the inaccurate underlying assumption that the expected returns of all funds follow a normal distribution.

(ii) Formulating an adjusted Markowitz model that uses the improved covariance estimation obtained in Objective III(i) as a measurement of risk.

(iii) Applying a selection of meta-heuristics to the model in Objective IV(i) to find good solutions to the unit trust selection problem.

(iv) Comparing the results obtained from the different meta-heuristics by using suitable performance measures obtained in Objective I(iii).

Objective IV: Provide ideas of future research possibilities.

1.6 Thesis layout

Following this introduction chapter, the literature pertaining to modern portfolio theory, meta-heuristics, and their applications in portfolio selection is reviewed in Chapter 2. In Chapter 3, the data used in the project will be explored and discussed. Chapter 4 will explain the methodology of the algorithms and techniques used to achieve the objectives listed. In Chapter 5, all results will be shown and discussed, before concluding the thesis in Chapter 6.

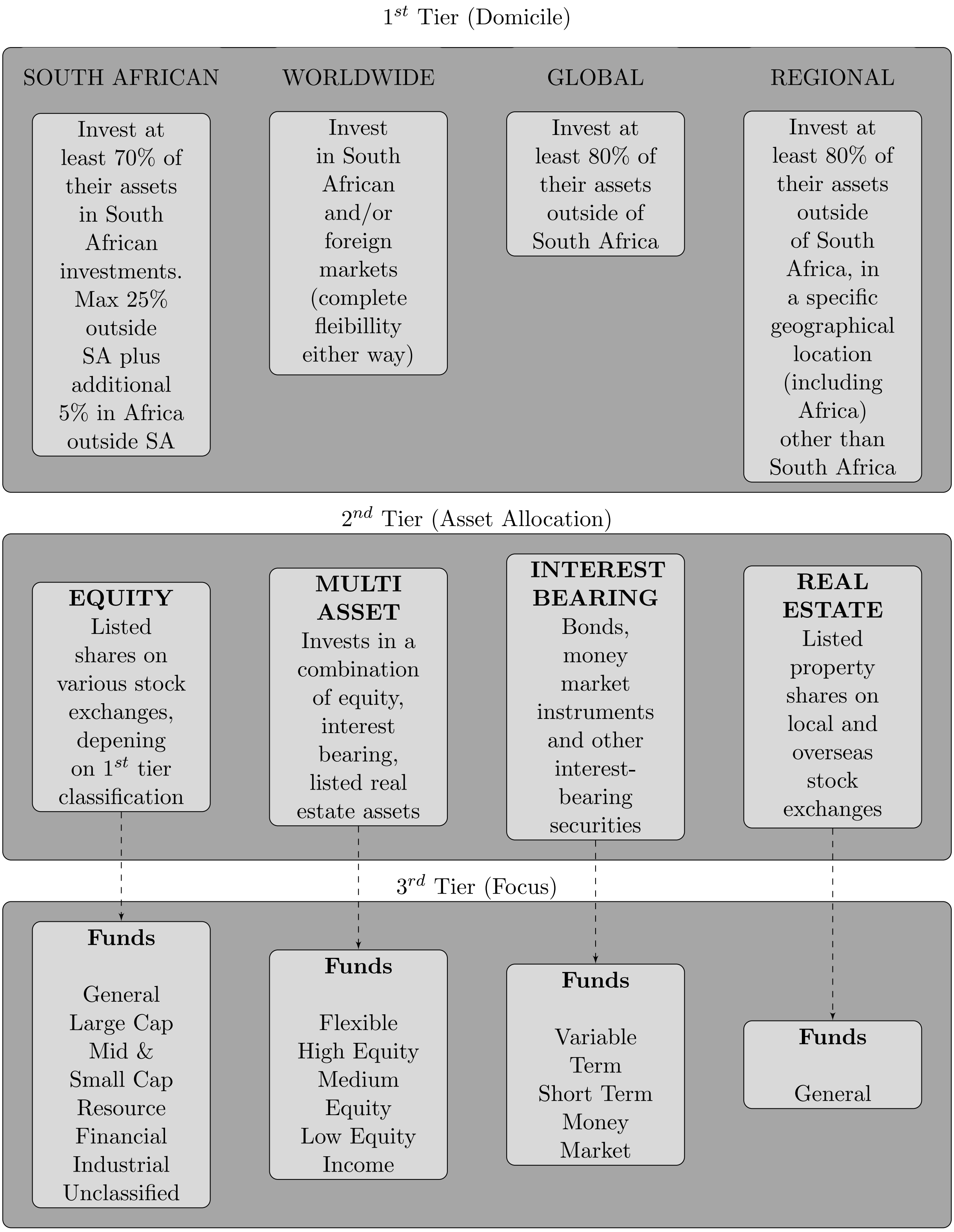


FIGURE 1.2: *An overview of the different permutations of unit trust classification types on the current financial market [17].*

CHAPTER 2

Literature review

Contents

This chapter will explore two sets of literature. The first being the works of the major contributors to Modern Portfolio Theory, Harry Markowitz and William Sharpe, as well as some additions that address the short-comings of their models. The second set will explain the workings of some of the meta-heuristics used in the past to attempt to find the efficient set of portfolios as well as discuss the results of the past attempts.

2.1 Portfolio optimisation

Portfolio optimisation is the process of combining various assets together in a manner that the proportions will ensure that there is not another combination that will be better according to some criterion. This criterion will combine the expected final value of the assets, the inherent risk associated with investing in the asset, and possibly other considerations not associated with financial betterment [35].

2.1.1 Historical perspective

In 1952, Harry Markowitz introduced the concept of portfolio optimisation as a mathematical alternative to portfolio selection and management techniques employed used at the time [50]. Markowitz's work is referred to as Modern Portfolio Theory (MPT). The historical view of an investor purely trying to maximise the anticipated (or expected) return of the securities is rejected by Markowitz as an explanation of investor behaviour in favour of the belief that an investor should also take into account the volatility of the security price. Large volatility is considered then as a negative attribute when choosing a security for a portfolio, whereas a large expected return gained from a security is considered a positive attribute.

From Markowitz' belief that investors should also try to minimise the variance of the securities chosen, and subsequently the variance of a portfolio, rose the expected return - variance of returns (E-V) rule. In order to explain the E-V rule, Markowitz presents a simple statistical example of expected value and calculations of variance and covariance.

Let R_{ij} denote the return on security i during time j, then the expected return, μ_i, of any security is given by $\sum_{j=1}^{n} \frac{R_{ij}}{n}$, where n is the number of observations of R_{ij}; the variance, σ_{ik}, is the covariance between securities i and k, given by $\frac{\sum_{i,k=1}^{n}(R_{ij}-\mu_i)(R_{kj}-\mu_k)}{n-1}$, where σ_{ii} is the variance of security i; and X_i is the proportion of the investor's portfolio allocated to security i. This would make the expected value, E, and variance, V, of the portfolio

$$E = \sum_{i=1}^{n} X_i \mu_i$$

and

$$V = \sum_{i}^{n} \sum_{j}^{n} \sigma_{ij} X_i X_j.$$

For a given selection of securities, the investor will have the option of combining them into a portfolio. A rational investor would want to select a portfolio that has the largest return for any given variance or the smallest variance for a given return. The set of portfolios with this characteristic is referred to as the *"efficient set"* of portfolios. Markowitz believed that for the E-V rule to be practical, two conditions would need to be met. Firstly, that the investor should want to use the E-V rule, thereby seeing the value of the efficient set of portfolios, and secondly that reasonable expected return and variance values must be available for the securities. The underlying assumption in the Markowitz model is that the securities are normally distributed to allow for the use of the sample covariance matrix as the measure of risk.

The Markowitz model does not imply that diversification is about holding a large number of different securities as some securities are highly correlated in their performance. Rather diversification is about minimising the covariance between the selected securities in a portfolio.

Criticisms of the Markowitz model include the fact that the assumption of normality is no longer a reasonable one, as research has proven that financial data is rarely normal. The use of the variance as the measure of risk is also not recommended, even Markowitz himself has said that a model using semi-variance would be preferred over the E-V model [51].

In 1963, William Sharpe provided two simple methods for analysing the efficient frontier defined in Markowitz's *Portfolio Selection* [69]. These were the Critical Line Method and the Diagonal Model. Of interest in this paper is the Critical Line method as the general outline of the steps create a heuristic that can be adapted for use by meta-heuristics. Sharpe's modifications are to rewrite the objective function as $\phi = \lambda \sum_i X_i E_i - \sum_i \sum_j X_i X_j \sigma_{ij}$, and to introduce corner portfolios. Corner portfolios are defined as being identical in composition other than the inclusion or exclusion of one security in one of the portfolios. Sharpe states that any portfolio on the efficient frontier can be determined as some combination of the corner portfolios between which it lies. The Critical Line method can be used to find the corner portfolios by following the steps outlined below:

1. The corner portfolio associated with $\lambda = \infty$ is determined. It is the portfolio consisting of one security with the highest return value. This is the starting point.

2. The relationships between the composition of the efficient portfolios and the λ value associated are calculated. The relationship between one section of the efficient frontier curve does not apply to another.

3. Using the relationships from step two, the securities are examined to determine at what level of λ a change in the portfolio will occur.

4. The next largest value of λ is chosen, this portfolio composition is determined and is considered the new corner portfolio.

5. As the relationships between the composition of the efficient portfolios and the associated λ values are only valid in one section, return to step two and repeat the process until $\lambda = 0$.

The issue with the Critical Line method is that it does not linearise the model, so as the number of securities being analysed increases so does the complexity in terms of number of comparisons and calculations, as the relationships need be recalculated at every corner portfolio and the new corner portfolio needs to be determined along with its composition. It will, however, speed up the process of finding the efficient frontier.

The next major development in portfolio selection is post modern portfolio theory, named to differentiate between the pioneering work of Markowitz and the research addressing the shortcomings of MPT. *Post Modern Portfolio Theory Comes of Age* is the article that coined the term Post Modern Portfolio Theory [65]. Written in 1993, authors Brian M. Rom and Kathleen W. Ferguson introduced the concept of Post Modern Portfolio Theory as the new preferred manner in which risk should be quantified. This new manner was that of differentiating between good variance and bad variance of a security's price. The authors argued that variance above the expected price was a good thing for an investor and should not be considered a penalty, in fact, during a bull market, where prices are on the rise, volatility may even be sought out. As a result, the downside risk model was introduced, in which a security's risk was defined as the magnitude of the drop below a specified rate of return given by the investor [13]. This rate of return is known as the Minimum Acceptable Rate (MAR). The MAR could range from the investors own choice to the actuarial rate needed to avoid having an underfunded pension plan. Due to the inclusion of the MAR, each result of the downside risk model now has a unique efficient frontier related to that MAR, as opposed to the general efficient frontier of the M-V model in which the investor will have to select either a risk or return and find portfolio that satisfies their needs. Another benefit of the downside risk model is that the risk statistic can be split into two parts, the downside probability and the downside magnitude. The downside probability is the likelihood of the security not meeting the MAR and the magnitude is the amount by which it will fail in the event that it fails. This gives a much better indication of the risk of a security beyond the variance of it's expected return.

2.1.2 Different measures of risk and modelling approaches

Through the years many different measures of risk have been tested due to the unpredictable nature of financial data and human nature wanting to minimising that risk. In this chapter, we focus on the Markowitz model and its extensions, the Lower Partial Moments (LPM) model, and the Value-at-Risk (VaR) model.

The articles that deal with the Markowitz, thereby using the variance as the measure of risk, use either the standard Markowitz model [73] or a variation of the model known as the Cardinality Constrained Mean Variance (CCMV) model [3]. The cardinality constrained model's objective is to

$$\text{maximise } \lambda \cdot r_P - (1-\lambda) \cdot \sigma_P \tag{2.1}$$

subject to

$$r_P = \sum_{i=1}^{N} w_i r_i \tag{2.2}$$

$$\sigma_P = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij}} \tag{2.3}$$

$$\sum_{i=1}^{N} w_i = 1 \tag{2.4}$$

$$\sum_{i=1}^{N} b_i \leq K \text{ where } b_i = \begin{cases} 1 \text{ if } i \ \in P, w_i \geq 1, \\ 0 \text{ if } i \ \notin P, w_i = 0. \end{cases} \tag{2.5}$$

The addition of equation (2.5), the cardinality constraint, allows the user to define a maximum number of investment options in the portfolio. This is of importance as over diversification of funds results in higher fees due to the investor needing to make more purchases. Cura [19], uses the CCMV model to achieve his goal of testing the Particle swarm optimization algorithm as a tool for mixed quadratic and integer programming problems. He notes that the traditional Markowitz model does not add in real world constraints such as the minimum lot purchases required for buying securities on the exchange market and so uses the CCMV model to restrict the number of assets in the portfolios. Dietmar and Kellerer [49] also provide another reason as to why the CCMV could be considered more attractive than the traditional Markowitz model. In practice investors tend to be uncomfortable investing in a wide variety of securities and will therefore want to limit their investment more than the traditional Markowitz model may suggest to do.

Sing and Ong [27] and Harlow [32] discuss the efficacy and composition of the results of the LPM model in comparison to models such as the Markowitz model. The LPM for continuous distributions is given as

$$\text{LPM}_n(\tau, R_i) = \int_{-\infty}^{t} (\tau - R_i)^n \text{dF}(R_i), \tag{2.6}$$

where n is the type of moment, R_i is the return of asset i, $\text{dF}(R_i)$ is the probability density function of R_i, and τ is the desired rate of return. Harlow began by comparing the simplistic LPM model in its ability to determine frontiers given different target returns (MAR), before discussing the efficacy against the traditional Markowitz model. Harlow found that the LPM model was at least as efficient as the Markowitz model but offers the potential for more attractive investment portfolios due to not penalising over performance. Harlow also showed that the constructed portfolios will converge toward either end of the frontier, making the LPM model a better choice for investors seeking a more balanced portfolio. Sing and Ong's goal was also to extend the Markowitz by using the LPM as the measure of risk, however they build on Harlow's work by also extending the LPM model to the Generalised Co-Lower Partial Moments (GCLPM) model, a measure of co-semivariance, not just semi-variance. The GCLPM is defined as

$$\mathrm{GCLPM}_n(\tau, R_i, R_j) = \int_{-\infty}^{t} \int_{-\infty}^{\infty} (\tau - R_i)^{n-1}(\tau - R_j)\mathrm{dF}(R_i, R_j), \tag{2.7}$$

where

$$\mathrm{GCLPM}_n(\tau, R_i, R_j) \neq \mathrm{GCLPM}_n(\tau, R_j, R_i), \text{ and} \tag{2.8}$$

$$\mathrm{GCLPM}_n(\tau, R_i, R_i) = \mathrm{LPM}_n(\tau, R_i), \tag{2.9}$$

with $\mathrm{dF}(R_i, R_j)$ in equation (2.7) as the joint probability density function. As in Harlow's paper, the GCLPM was then compared the Markowitz model's efficient frontier. Sing and Ong found that the Markowitz model tends to over inflate the variances and thereby accidentally rule out funds that could be downside efficient, and the GCLPM improved on the LPM by elimating the assumption of symmetry through the diagonal of the co-LPM matrix.

Markowitz discussed the computational problems associated with his quadratic programming model [51], and many authors have attempted to linearise the model to overcome some of these problems. Mansini *et al.* [47] present an overview of some of the linearisation models and also show a computational comparison using real world data. Two of the linearisation models include, the Mean Absolute Deviations (MAD) model, where the risk is replaced by the sum of the downside deviations from the mean, and the Minimax model, in which the risk is defined as the difference between the mean of the portfolio and its minimum value.

Young [79], showed how the Minimax model can be used to constrain variables to integer or boolean values to allow for transaction costs to be incorporated. The results for the two minimax evaluations were that the minimax model tends to find portfolios that are more risk averse than risk neutral or -seeking. This is due to the fact that with the use of the minimum value being the reference point the model is more sensitive to outliers than other models.

Michalowski and Ogryczak [53] show an extended MAD model with downside risk aversion as a technique for solving the portfolio optimisation problem. They demonstrate that the MAD model can be used to incorporate downside risk while remaining linear. This is done by defining the MAD risk measure as

$$\delta(x) = E\{|R_x - \mu_x|\} = \int_{-\infty}^{\infty} |\mu_x - \xi| P_x(d\xi), \tag{2.10}$$

where P_x is a probability measure induced by R_x. The MAD risk measure is then twice the downside absolute semideviation

$$\bar{\delta(x)} = E\{\max\{\mu(x) - R_x, 0\}\} \tag{2.11}$$

$$= E\{\mu_x - R_x | R_x \leq \mu_x\} P\{R_x \leq \mu_x\} \tag{2.12}$$

$$= \int_{-\infty}^{\mu_x} |\mu_x - \xi| P_x(d\xi). \tag{2.13}$$

This model will be equivalent to the Mean Variance model if the historical returns used are normally distributed and can still be used if they are not, giving it an advantage over the Markowitz model.

Other alternate approaches include stochastic programming formulations such as the model presented by Künzi-Bay and Mayer [42]. In their paper, they described the Conditional Value-at-Risk model in terms of a two-stage recourse problem with a specialised L-shape method. The results of this paper were inconclusive, however the work was continued when Kolos Ágoston used the Successive Regression Algorithm (SRA) to solve the stochastic CVaR minimisation problem [1]. Ágoston found that the SRA was efficient for large sample sizes and not for small sample sizes, which is generally the opposite of most other methods that solve the portfolio selection problem. Comparison between the SRA and meta-heuristics is recommended as future work.

2.2 Meta-heuristics

The most comprehensive definition of meta-heuristics is "*an iterative master process that guides and modifies the operations of subordinate heuristics to efficiently produce high-quality solutions. It may manipulate a complete(or incomplete) single solution or a collection of solutions at each iteration. The subordinate heuristics may be high (or low) level procedures, or a simple local search, or just a construction method*" [76]. Simplified, this means a meta-heuristic is a set of steps that combine a search technique with an intelligent manner of searching in order to search a solution space in a manner that will find good solutions in a potentially short time period.

2.2.1 Classification of the different meta-heuristics

Meta-heuristics can be classified and divided into different subclasses. These subclasses are simple local search based meta-heuristics, evolutionary algorithms, swarm intelligence algorithms, and hybrids [10, 30, 76]. Each of these subclasses are defined as such due to the characteristics of the search techniques or the guiding process used.

The local search meta-heuristics are so called due to the fact that the exploration of the local neighbourhood in every iteration is the main characteristic of the algorithm. Algorithms in this category are; the Greedy Randomised Adaptive Search Procedure (GRASP), which operates by determining the neighbourhood around the current solution and choosing the option that will then optimise the chosen objective, this repeats until no better solution is found [26]; Simulated Annealing, in which a neighbour is chosen and evaluated and accepted according to a temperature parameter, if accepted the solution moves to the new neighbour. The acceptance via the temperature parameter allows the algorithm to move out of local minima [39]; and Tabu Search, which also selects the minimum of the neighbourhood, but will choose a minimum that is higher than the current solution. It then stores the previous solution in a tabu list, allowing it not be stuck in a loop at local minima [31].

Evolutionary algorithms are algorithms that mimic the evolutionary process by utilising the processes in nature that allow species to adapt to their environment [10]. Algorithms in this category utilise recombination, or crossovers, as well as mutations to create new solutions from the current set of solutions. Evolutionary algorithms include the Genetic Algorithm [21] and the Pareto Envelope-based Selection Algorithm (PESA) [16]. The main difference between these two algorithms are how they store the best solutions and how they select solutions for the recombination phase. The Genetic Algorithm selects from a parent population to create an equal sized child population which is then combined and sorted to select the best of the combined population until the new parent population is full. The PESA keeps the best solutions in an archive population and selects from this to create a new internal population, which can be of a different size to the archive.

Swarm intelligence algorithms require two characteristic requirements to be met in order for the algorithm to be classified as a swarm algorithm [36], namely, the concept of self-organisation, and the division of labour. Self-organisation is divided into four basic properties; positive feedback, negative feedback, fluctuations, and multiple interactions. Division of labour requires that certain jobs are performed by specialised individuals within the swarm [11, 36]. Popular algorithms in the swarm intelligence category include the Ant Colony Optimisation [24], the Artificial Bee Colony algorithm (ABC) [37], and the Particle Swarm Optimisation (PSO) [25]. The Ant Colony Algorithm works through the use of pheromone trails left by ants to find better food sources, where stronger pheromone levels lead to better solutions; the ABC works by utilising bees in a hive that will select a source to explore based on the dance performed by the bee that found the food source, the better the food source, the higher the chance of selection; and the PSO works by emulating the swarm mentality of certain groups of animals, i.e birds and fish, in their movements to avoid predators or to find food or new habitat locations. It does this by assigning each solution a velocity and position in the search space and tries to adjust it to reach the best known solution [70].

2.2.2 Applications of local search algorithms in portfolio optimisation

With regard to portfolio selection a number of different ideas considering simulated annealing have been applied. Maringer and Winkler [48], tested the effect of the changes to the Memetic algorithm in solving a portfolio optimisation problem with a VaR model extended with risk constraints. The changes made were to swap the Threshold Acceptance algorithm with Simulated Annealing and to evaluate how an elitist strategy would effect the outcomes. The elitist strategy is considered the return to the best solution after becoming stuck at a local minimum instead of just reheating the temperature. The results showed that replacing Threshold Acceptance with Simulated Annealing was favourable in cases where the solution space is computationally demanding, and that in the instances where simulated annealing was preferred over threshold acceptance, the simulated annealing with the elitist performed better than the normal simulated annealing method. This was due to the elitist's preventing poorer solutions from becoming locked at local minima as well as lessening the sensitivity of the algorithm to temperature changes.

Armañanzas and Lozano [3] also developed a simulated annealing approach in 2005 to solve the Cardinality Constrained Mean Variance model with minimum and maximum weightings, where instead of using one objective function, they used both risk and return individually to assess the portfolio. The modification of importance in Armañanzas and Lozano's paper is their neighbourhood selection. In place of a simple swap or exchange of weighting between securities, they move between solutions through the following steps:

1. Determine which securities in the current portfolio contributed the most towards the risk total.

2. Select which of the two adds the least towards profit. Define this as the pivot security.

3. The pivot securities weight is then shared amongst random securities, first selecting from the current solutions remaining securities before selecting from the empty weighted securities.

The results of their paper were inconclusive as they found that in some scenarios the simulated annealing method determined the majority of non-dominated solutions over other methods and in others was dominated fully by the same methods. This would mean that using this search move may produce unsatisfactory results and should be combined with another method.

2.2.3 Applications of evolutionary algorithms in portfolio optimisation

The Pareto Envelop-based Selection Algorithm (PESA) is tested by Corne *et al.* [16] , for its application in solving multi-objective problems. It was tested against two established evolutionary algorithms, the Strength Pareto Evolutionary Algorithm (SPEA) and the Pareto Archived Evolutionary Strategy (PAES) as the PESA combines elements from both. To keep the results fair the parameters that are shared between algorithms were kept identical. The algorithms were run through six test problems to test for their efficacy.

Test case 1 has a convex pareto front, but no other characteristics; test case 2 has a non-convex front; test case 3 has a large number of gaps in the front; test case 4 is multimodal and contains a large number of separate pareto fronts; test case 5 is deceptive; whereas test case 6 contains non-uniformly distributed solutions along the frontier. The results of the test cases are not completely conclusive but show that the PESA has a slight edge over the other two algorithms by dominating them completely in 3 of the test cases and being tied with them in case 4. To see which cases the algorithms perform better on, see Table 2.1. The authors however caution against a conclusion solely on test case results and recommend that further testing be performed on a wider variety of problems before concrete conclusions are made.

Test Case	Result
T1	PAES
T2	PESA
T3	PESA
T4	All three
T5	SPEA and PESA
T6	PESA

TABLE 2.1: *A summary of test case results for SPEA vs PAES vs PESA*

Dioşan [22], compared the use of the Pareto Envelop-based Selection Algorithm, the Non-dominated Sorting Genetic Algorithm, and the Strength Pareto Evolutionary Algorithm (SPEA II) in determining the efficient frontier of the Euronext Stock set of securities. For the study the traditional Markowitz model with no alterations was used. Three measures are used to test the final solutions of each algorithm; a S-metric, to determine how much of the solution space is dominated by the set of solutions; a Δ-metric, to determine how evenly the solutions points are spaced; and a C-metric, to determine the percentage of solutions in a set of solutions that are dominated by another set of solutions. The S-metric and Δ-metric will be adopted for use in this project. The results of the metrics can be interpreted as; a better S- and C-metric value corresponds to the algorithm with the higher value and a better Δ-metric is the algorithm with the lowest value. By comparing the results, the PESA algorithm wins two of the three categories and is concluded by Dioşan to be the stronger algorithm. As the parameters of the size of each population, the number of generations completed before termination, as well as the evolutionary operators are the same in all three algorithms, allowing for comparable results, Dioşan's claim that PESA is the stronger of the three algorithms is strengthened.

2.2.4 Applications of swarm intelligence algorithms in portfolio optimisation

Chen et al. [14] tested the ABC for its efficacy in determining the pareto fronts for the Hang Seng 31, S&P 100, DAX 100, and Nikkei 225 indices for the first time in 2012. As this was the first attempt to use the ABC on the portfolio optimisation problem, the goal was purely to test the efficacy of the algorithm. To view the efficacy, Chen *et al.*'s ABC optimisation algorithm was tested against a Simulated Annealing algorithm, a Tabu Search, and a Variable Neighbour Search using three different performance measures.

The first measure is the $D1_R$ index which calculates the average minimum distance from each solution in the approximated front determined by the algorithm and the reference front of the optimal solution. This measure is calculated as

$$D1_R = \frac{1}{|X_{\text{ref}}|} \sum_{x^* \in X_{\text{ref}}} \min\{d_{x^*x} | x \in X_{\text{app}}\},$$

where X_{ref} is the reference pareto front and X_{app} is the approximated pareto front and d_{x^*x} is given by

$$d_{x^*x} = \sqrt{(f_r(x^*) - f_r(x))^2 + (f_{er}(x^*) - f_{er}(x))^2},$$

where $f_r(\cdot)$ is the risk objective function and $f_{er}(\cdot)$ is the return objective function.

The second measure, the Δ measure, determines the diversity of the solutions in the pareto front in which a larger Δ means that the solution may not be well spread out over the solution space. The formula for the Δ measure is

$$\Delta = \frac{d_f + d_l + \sum_{i=1}^{|X_{\text{app}}|_c - 1} |d_i - \bar{d}}{d_f + d_l + (|X_{\text{app}}|_c - 1)\bar{d}},$$

where d_f and d_l are the Euclidean distances between the extreme points of the reference pareto front and the approximated front, d_i is the Euclidean distance between consecutive points in the approximated front, $\bar{d}$ is the averaged value of the d_i's, and $|X|$ is the number of solutions in the appropriate front.

Finally, the AE measure determines the ratio of run time to number of non-dominated solutions found. This measure can be calculated as

$$AE = \frac{T_{\text{app}}}{|X_{\text{app}}|},$$

where T_{app} is the run time taken to obtain the approximated pareto front. Chen *et al.*'s results show that the ABC algorithm outperforms the other non-population based algorithms in determining a pareto front for a portfolio selection problem and is a promising algorithm for use in this particular application. The metric for diversity and the $D1_R$ metric will be adapted for use in this project.

While not focussed on the portfolio optimisation problem, Karaboga and Basturk [37] tested the ABC against a selection of other population based algorithms in the application of solving multi-objective problems. This was done by testing the algorithms efficacy in finding values that minimise five benchmark functions, each with a specific challenge for the algorithm to overcome. These are:

1. the Griewank function : $f(x) = \frac{1}{4000}(\sum_{i=1}^{D}(x_i)) - (\Pi_{i=1}^{D}\text{cos}(\frac{x_i}{\sqrt{i}})) + 1$,
2. the Rastrigin function : $f(x) = \sum_{i=1}^{D}(x_i^2 - 10\text{cos}(2\pi x_i) + 10)$,
3. the Rosenbrock function : $f(x) = \sum_{i=1}^{D} 100(x_i^2 - x_{i+1})^2 + (1 - x_i)^2$,
4. the Ackley function : $f(x) = 20 + e - 20e({}^{-0.2}\sqrt{\frac{1}{D}\sum_{i=1}^{D} x_i^2}) - e^{\frac{1}{D}\sum_{i=1}^{D}\cos(2\pi x_i)}$,
5. and the Schwefel function : $f(x) = D * 418.9829 + \sum_{i=1}^{D} -x_i\text{sin}\sqrt{|x_i|}$,

where D is the dimensionality.

The results of these benchmarks suggest that the ABC is more powerful than the other population based algorithms, namely the Genetic algorithm, the Particle Swarm Optimisation and a Particle Swarm Inspired Evolutionary Algorithm. The authors do, however, state that future works includes testing of the control parameters and the convergence of the algorithm.

Kendall and Su [38] use the PSO algorithm to find an efficient set of portfolios using the Markowitz model. No alterations are made to the Markowitz model other than testing the removal of the no short selling constraint, which would allow the securities to have negative weights. Short selling refers to the idea of selling shares before actually purchasing them, in the hopes that the price will drop before the shares are due to be transferred, allowing the seller to purchase the shares at a lower price than was promised. The algorithm was found to be more powerful than traditional mathematical solvers in the pure Markowitz model, however, the effect of the short selling was to exponentially increase the time needed to reach optimality.

Armañanzas and Lozano [3] tested the Ant Colony Optimisation (ACO) algorithm against a Simulated Annealing approach using the Markowitz model. The ACO is constructed by creating 3 colonies, each designated to solving a different objective. Colony 1 tries to minimise the risk objective; colony 2 minimises the negative of the return objective; and colony 3 attempts to balance the previous two objectives. This version of the ACO works by creating a connected network of all potential investments and placing an ant at each node every iteration. The ants then move up to K places, where K is the cardinality, selecting assets based on the current selection pheromone value of the arc ij. The weights,

$$w_i = \frac{\delta_i - \epsilon_i}{K},$$

of the K assets are then determined, where δ_i and ϵ_i are the maximum and minimum values for security i if it is included in the portfolio. Once the objective values have been calculated, the pheromone values are updated. After convergence of the ACO has occurred, a multi-objective greedy search is performed on the three final solutions from the colonies to find the pareto frontier. The results of this study show that the ACO is better suited to finding solutions in the upper portion of the pareto frontier and is less suited than the Simulated Annealing approach in finding the other solutions.

2.3 Chapter summary

This chapter contains the reviews of the literature concerning the portfolio optimisation problem, meta-heuristics and the application of meta-heuristics to the portfolio optimisation problem. The chapter begins with the major developments in the history of the portfolio optimisation problem, before discussing the applications of different measures of risk as well as the mathematical approaches used to solve the models. The chapter then moves onto a definition of meta-heuristics and the classification of the different types of meta-heuristics and their characteristics. This is then followed by a review of the different applications found of meta-heuristics that have been applied to the portfolio optimisation separated by classification.

CHAPTER 3

Data

Contents

This chapter will explore the dataset used for this project by providing an overview of the data in §3.1, an explanation of the method of calculating expected return and variance in §3.2, some descriptive statistics in §3.3, the results of the normality tests in §3.4, and a discussion over the issues that are associated with the dimensions of the data in §3.5.

3.1 Contents of the dataset

The price history of 832 unit trusts were obtained from the Bloomberg terminal database [9]. All the unit trusts included in the dataset are equity class unit trusts domiciled in South Africa with more than a year of price history data as per the classification of Chapter 1. Approximately 100 equity unit trusts domiciled in South Africa were excluded as they have less than a year of historical price data available.

From §1.1.3, these unit trusts include a minimum of 80% South African securities. The extracted price history from Bloomberg is 27 years worth of weekly data, from May 1990 to April 2017. Weekly data was used instead of monthly or annual data as most sources agree that the more data used the better the estimates [75].

3.2 Calculation of return and variance

To calculate the expected return, μ_i, of each unit trust i, we follow the original formulation introduced by Markowitz by 1952,

$$E[R_i] = \mu_i = \frac{\sum_{t=1}^{n} r_{it}}{n} \ ,$$

where r_{it} denotes the historic weekly growth percentage at time t and n denotes the total time periods in the observable dataset. While logic may dictate that the returns from a few years ago may not have an impact on tomorrow's returns, the patterns of growth may still be able to help predict the next set of patterns, since economic business cycles tend to repeat themselves.

The co-variance between unit trust i and unit trust j is calculated as the sample covariance of the dataset, where

$$\sigma_{ij} = \frac{\sum_{t=1}^{n}(r_{it} - \mu_i)(r_{jt} - \mu_j)}{n-1} \ .$$

Once again the entirety of the dataset is used to capture any repetitions of patterns in the dataset.

3.3 Descriptive statistics

A scatter plot of the unit trusts may be seen in Figure 3.1. By viewing the scatter plot, it is clear that the majority of the unit trusts are located near the bottom of the risk axis, but spread out along the return axis. This may be due to unit trusts already reducing risk by diversification in the construction of the fund. There are, however, a few outliers with an unusually high risk and a low return level. The general spread of the funds indicates that as return grows, so does the risk, a pattern that is consistent with theory. Due to the large number of funds with very similar risk and return levels, the scatter plot does not provide a clear picture of the number of funds within each return range.

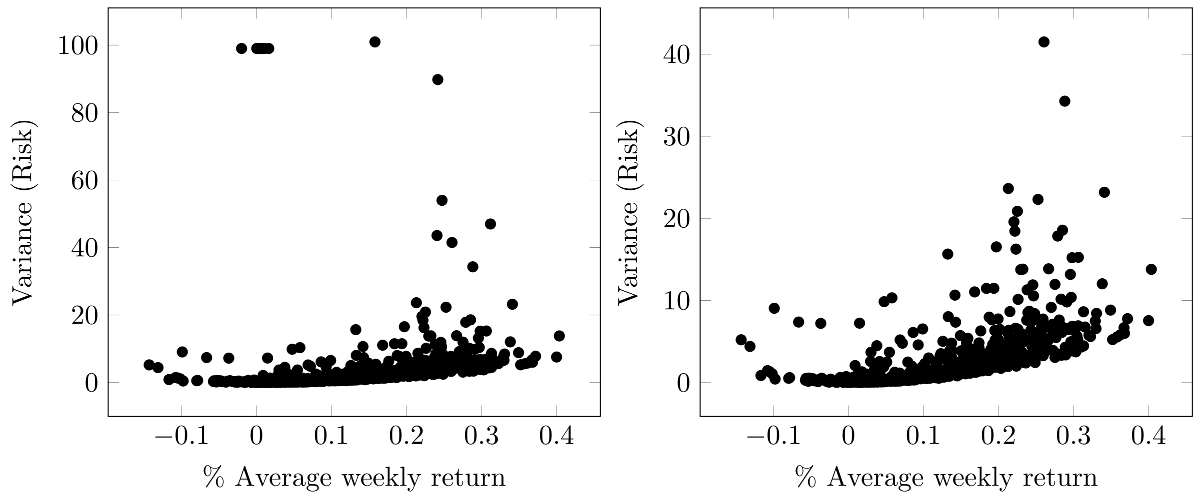


FIGURE 3.1: *Two scatterplots depicting the expected weekly return and risk, given as the variance, for each of the 711 unit trusts, shown with outliers(left) and without outliers(right).*

A histogram, given in Figure 3.2, of the number of funds within each average return range indicates a right-skewed distribution, with a median average weekly return of 0.117% or a 6.25% annual growth. When using the entirety of the dataset to gain the mean level of return, we gain a value of 0.124% weekly or converted, a 6.64% annual growth. When excluding unit trusts with an overall negative growth this figure changes to 0.137% weekly, which converted is a 7.38% annual growth. This indicates that at least half the funds are not able to outperform a simple savings account at any South African bank, emphasising the importance of choosing the right funds to invest in. The means for the periods prior to 2005 and 2012 are also compared to two studies for the periods 1988 to 2005 and 1994 to 2012 to see how this data set compares to previous studies. It was found that the mean annual return for the Bloomberg dataset before 2005 is 13.51% which is approximately in line with the findings for the period 1988 to 2005 with an annual return of 12.37% [52]. However the mean annual return for the Bloomberg data before 2012 is 9.19%, which is substantially lower than the findings for the period 1994 to 2012 at 14% [7]. One possible explanation for the difference in performance is that the data for the two studies are gained from different sources and could therefore contain different unit trusts. There is also therefore a decline in the average performance of unit trusts over time, which could suggest that the financial markets are currently unstable.

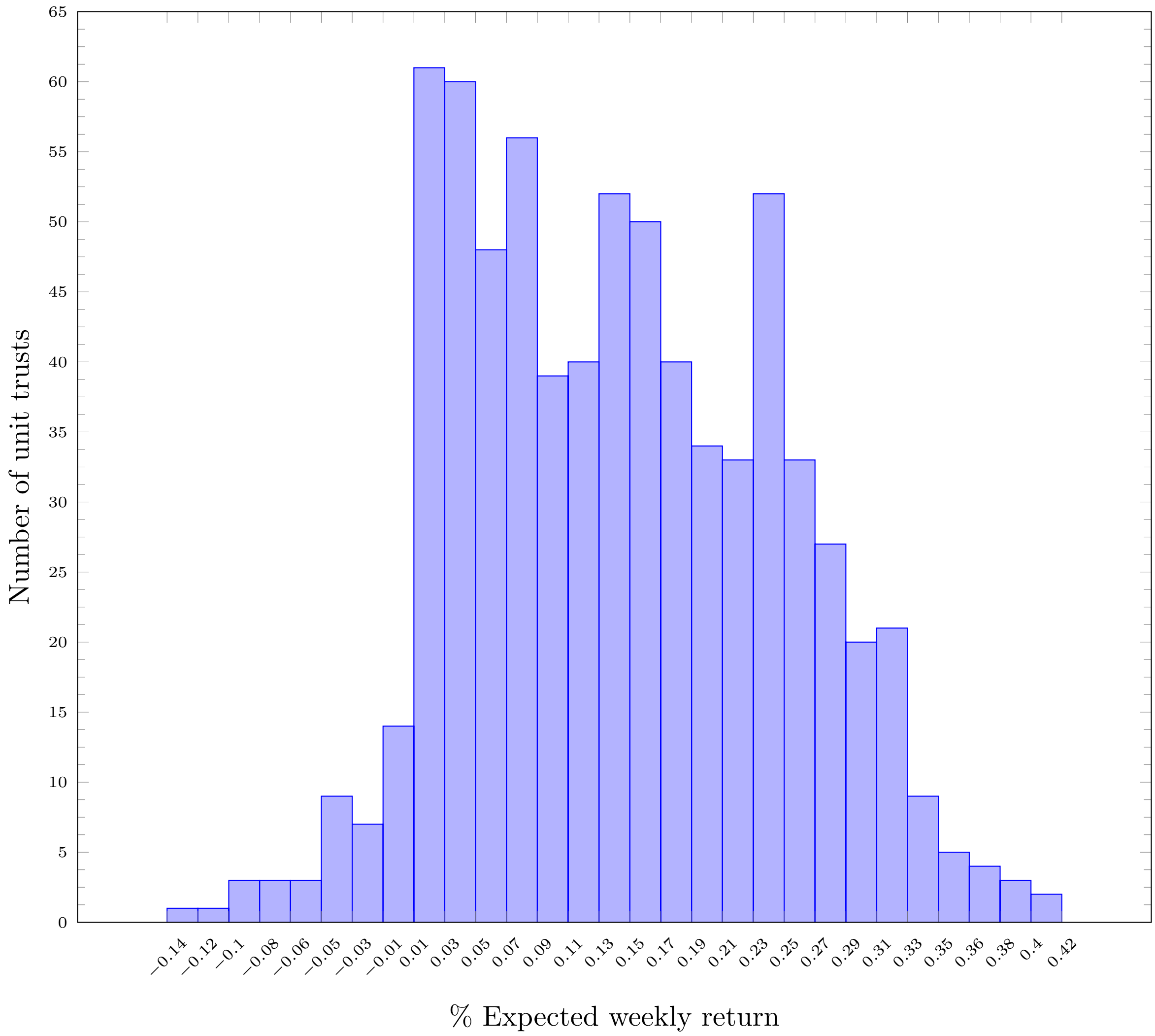


FIGURE 3.2: *A histogram of the frequency of different expected return categories from the dataset of 711 unit trusts.*

Two graphs, Figures 3.3 and 3.4, are provided to give a representation of the growth patterns associated with the categories of the histogram. It may be seen that even in the funds that represent better expected long term growths, there was an overall large drop around 2008. This was due to the 26% decline in the overall JSE/FTSE which created fears in investors that another recession was imminent. Including data from before, during, and after this 2008 crash will provide a better view of the overall performance of unit trusts.

A view of the funds with the best expected return, variance, Sharpe ratio, or risk to return ratio, may be seen in Tables 3.1, 3.2, 3.3. The Sharpe ratio is calculated as $\frac{\mu_i - R}{\sigma_{ij}}$, where R is

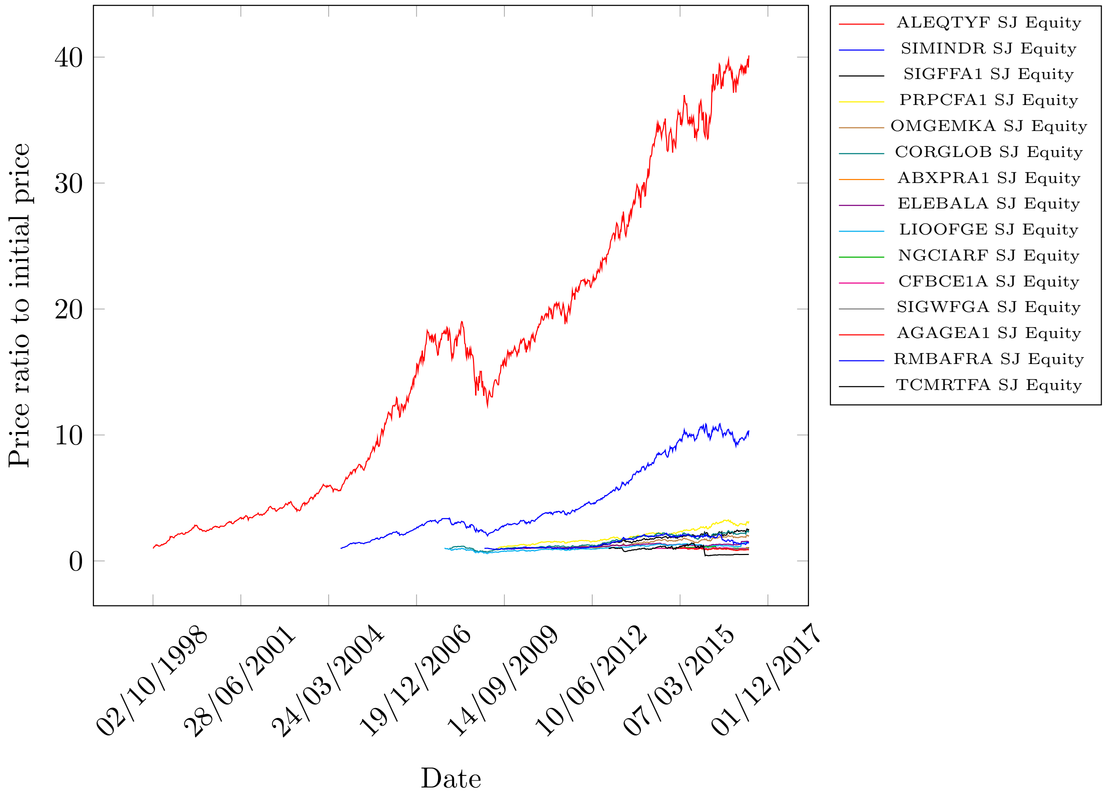


FIGURE 3.3: *Time series data of a selection of different unit trusts to represent the histogram categories.*

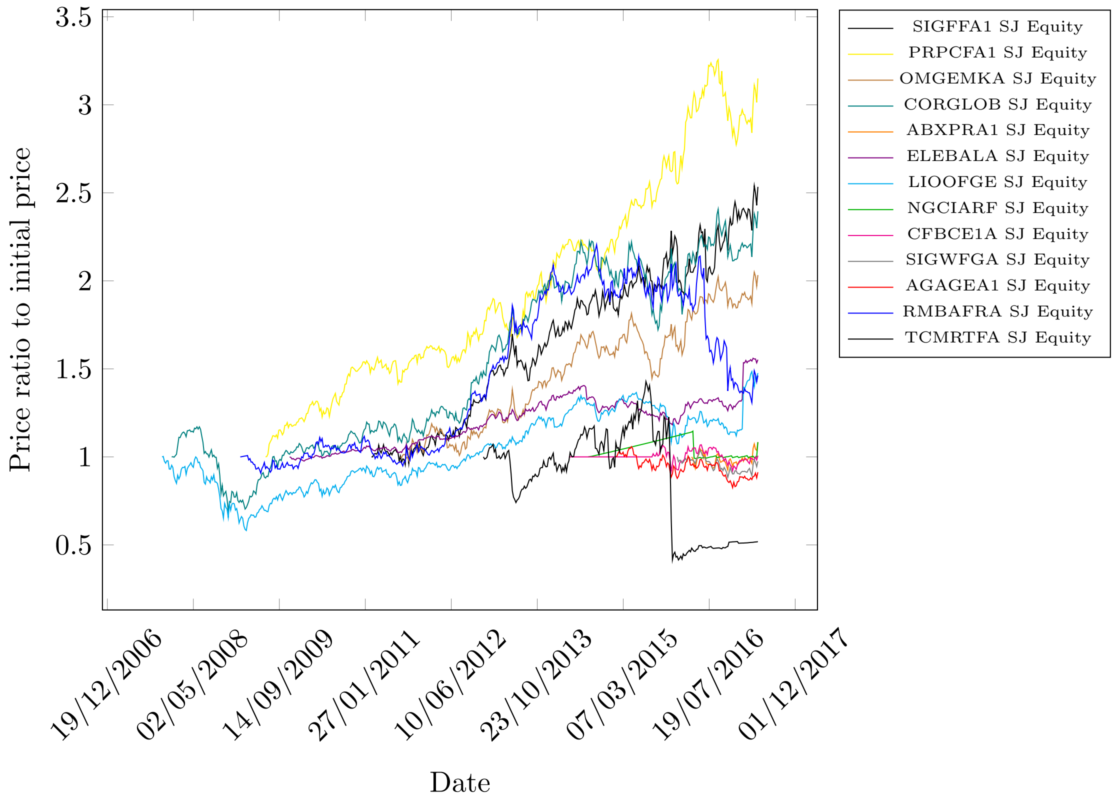


FIGURE 3.4: *Time series data of the unit trusts that represent the lower growth classes.*

a reference rate of return. Intuitively, a large Sharpe ratio then refers to a better fund. What may be seen from these tables is that no unit trust exists in more than one of the categories, reinforcing the trade off between the two performance measures of risk and return. Furthermore, it may be seen that while the changes within the top 10 unit trusts for risk and Sharpe ratio is not very high, the changes within the risk top 10 show that there is a significant difference in the stability of the unit trust performances.

Best Yearly Growth	
Fund Name	Value
Allan Grey Equity Fund	23.26%
Prescient China Balanced Feeder Fund	23.07%
SIM Industrial fund	21.29%
Element-Global Equity Fund	21.04%
Catalyst Global Real Estate Feeder	21.00%
Discovery Global Value Feeder Fund	20.66%
Old Mutual Global FTSE Rafi All World Index Feeder Fund	20.42%
MI-Plan IP - Global Macro Fund	20.05%
Coronation Industrial Fund	19.88%
Investec Commodity Fund	19.37%

TABLE 3.1: *A table depicting the top 10 funds in terms of expected annual growth.*

Best Sharpe ratio	
Fund Name	Ratio Value
ABSA Access BCI Stable Passive Fund	0.803
Sanlam Diversified Income Fund of Funds	0.779
Aureus Nobilis BCI Cautious Fund	0.695
Granate Multi Income Fund	0.686
Prescient Stable Income Fund	0.660
GCI MET Income Fund	0.593
Citadel SA Income H4 Fund	0.568
Prudential Namibian Enhanced Income Fund	0.557
PBi BCI Conservative Fund of Funds	0.551
Tresor Sanlam Collective Investments Income Fund	0.536

TABLE 3.2: *A table depicting the top 10 funds in terms of the Sharpe ratio.*

Best Risk	
Fund Name	Risk Value
Southchester IP Optimum Income Fund	0.0058
PSG Wealth Enhanced Interest Fund	0.0118
Old Mutual-Interest Plus Fund	0.0132
Prudential-High Interest Fund	0.0151
Absa Smart Alpha Income Fund	0.0155
Argon BCI Flexible Income Fund	0.0161
Prescient Stable Income Fund	0.0165
Tresor Sanlam Collective Investments Income Fund	0.0180
ABSA Access BCI Stable Passive Fund	0.0204
Gryphon Dividend Income Fund	0.0214

TABLE 3.3: *A table depicting the top 10 funds in terms of risk, or variance.*

3.4 Normality tests

A normality test can be defined as testing to see how well the normal distribution can model a given dataset and the probability that the dataset does in fact come from a normal distribution [23]. Normality tests were performed on the unit trusts to see whether the Markowitz assumption can be held for the chosen dataset. In Table 3.4, an example of the output of the normality test performed in SAS is given [66]. The non-normal result is gained by determining if the Shapiro-Wilk [67] and Kolmogorov-Smirnov [20] agree with each other. If not, the Cramer-von Mises [18] and Anderson-Darling tests [2] are also used. If the p-Value is less than 0.05, then the assumption of normality is rejected in favour of non-normality. We use 0.05 to represent 95% confidence level.

The results show that of the 711 unit trusts, only 22% of them may be statistically assumed as normally distributed, which means that the Markowitz Mean-Variance model is not applicable as it is described in literature.

Moments			
N	495	**Sum Weights**	495
Mean	0.00106921	**Sum Observations**	0.52925747
Std Deviation	0.02401812	**Variance**	0.00057687
Skewness	1.39521691	**Kurtosis**	12.9911484
Uncorrected SS	0.28553963	**Corrected SS**	0.28497374
Coeff Variation	2246.34858	**Std Error Mean**	0.00107953

Basic Statistical Measures			
Location		**Variability**	
Mean	0.001069	**Std Deviation**	0.02402
Median	0.001623	**Variance**	0.0005769
Mode	0.000000	**Range**	0.28904
		Interquartile Range	0.02272

Tests for Location $\mu_0 = 0$				
Test	**Statistic**		**p Value**	
Student's t	**t**	0.990434	**Pr** $>$ **|t|**	0.3224
Sign	**M**	14	**Pr** $>$ **|M|**	0.2235
Signed Rank	**S**	3989	**Pr** $>$ **|S|**	0.2065

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.88703	**Pr** $<$ **W**	$<$0.001
Kolmogorov-Smirnov	**D**	0.091503	**Pr** $>$ **D**	$<$0.0100
Cramer-von Mises	**W-Sq**	1.427584	**Pr** $>$ **W-Sq**	$<$0.0050
Anderson-Darling	**A-Sq**	8.485309	**Pr** $>$ **A-Sq**	$<$0.0050

Quantiles	
Level	**Quantile**
100% Max	0.0832308
99%	0.0622009
95%	0.0373712
90%	0.02823967
75% Q3	0.01578305
50% Median	0.00059628
25% Q1	-0.010015444

TABLE 3.4: *An example output for the Lion of Africa MET Equity Fund normality test as given as an SAS output.*

3.5 Dimension of the dataset

The number of observations (weekly returns) and variables (unit trusts) are 1404 and 711, respectively. This gives a parameter(p) to observation ratio(n) of $\frac{711}{1404} = 0.51$. Traditional covariance estimators are designed for a ratio where p/n is almost 0 and the high ratio in this dataset will create an unstable sample covariance matrix. This is known as the high dimensionality problem [15].

Due to the non-normality and high dimensionality an improved estimation for the covariance matrix is required.

3.6 Use of data for training and testing

A secondary data set containing 33 securities from a selection of list companies on the Johannesburg Stok Exchange, gained from a previous study at the University of Stellenbosch [72]. The three meta-heuristics were applied to this data set and the solutions gained were tested against the results from the aforementioned study to see if the algorithms could be successfully applied to another data set.

3.7 Chapter summary

This chapter contains a discussion of the data used in this study. Included in the chapter are; a description of how the expected returns and variances of the unit trusts are calculated; a cursory glance at the data via scatter plots and a histogram; descriptive statistics; top 10 tables of the unit trusts with the best expected growth, lowest variance, and the best risk to growth ratios; results of the normality tests; and a discussion of a problem associated with the size of the data set.

CHAPTER 4

Methodology

Contents

In this chapter, all techniques used to achieve Objective II(ii) and Objective III listed in Chapter 1 will be discussed. These include the improved estimate for the variance-covariance matrix using shrinkage theory estimation, the mathematical formulation for the model, all meta-heuristics used to find the efficient frontier and the measures used to evaluate the success of each meta-heuristic.

4.1 Problem formulation

For this project, the Markowitz mean-variance model is used, however, an additional constraint is added to ensure that solutions provide a positive return in the short term as well as using an improved estimate for the covariance instead of the sample covariance matrix as used in the original Markowitz mean-variance model. The reason for the additional constraint is that one of the main concerns for both investors and investment companies is consistent negative growth in a fund. If this is present in a fund, investors will generally disinvest in a fund, despite the fact that the fund has an overall positive long term growth.

The adjusted multi-objective model aims to

$$\text{maximise} \sum_{i=1}^{n} X_i \mu_i, \tag{4.1}$$

$$\text{minimise} \sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j \sigma_{ij}, \tag{4.2}$$

subject to

$$\sum_{i=1}^{n} X_i = 1000 \quad , \tag{4.3}$$

$$0 \leq X_i \leq 1000 \quad \forall\, i, \tag{4.4}$$

$$\sum_{i=1}^{n} X_i \mu_i^{st} > 0 \quad , \tag{4.5}$$

$$X_i \in \mathbb{Z} \quad \forall\, i, \tag{4.6}$$

where X_i is the decision variable describing the percentage of the portfolio invested in unit trust i, μ_i is the expected return of unit trust i, μ_i^{st} is the one year short-term expected return of unit trust i, σ_{ij} is the covariance of unit trusts i and j, and n is the total number of unit trusts. The multi-objective model can be changed to a single objective model with the objective to

$$\text{maximise } \lambda \sum_{i=1}^{n} X_i \mu_i - (1-\lambda) \sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j \sigma_{ij}, \tag{4.7}$$

where $\lambda \in \{0, 1\}$ is the relative ratio of risk to return and the constraints are kept the same.

Constraints (4.3), (4.4), and (4.6) ensure that the entire investment amount is invested, that the decision variable, X_i, is an integer value from 0 to 1000 to keep percentages tidy, and that the practice of short-selling is excluded. A value of 1000 is chosen as the full value of the portfolio as it allows the option of investing at a decimal percentage point, i.e. $X_i = 326$ being representative of investing 32.6% in that chosen fund. We use a tidy decimal value in this project as investment portfolios are advertised as tidy percentages, not as continuous variables. Constraint (4.5) ensures portfolios without a positive short term growth are excluded, with μ_i^{st} denoting the average return of unit trust i during the last year.

4.2 Shrinkage estimation

The Markowitz model is based on the use of the sample covariance matrix as defined in §3.2 to represent the risk of a set of financial assets. Unfortunately, as discussed in §3.5, the sample covariance matrix will be unstable and have large estimation errors [44, 77]. This results in extreme coefficients being developed and, as a result, any optimisation will have extreme reactions to these coefficients [63]. This, however, is not the only problem with using the sample covariance matrix. As mentioned in §3.4, the majority of the data used in this study is non-normal, which is in line with previous studies [78]. The non-normality is a problem as traditional estimators of covariance, like the sample covariance, are designed for Gaussian distributions. Even the majority of the field of robust covariance estimators designed for non Gaussian distributions require that the number of observations be greater than the number of parameters [78]. These two issues combined is why a better estimator is required when using the Markowitz model. In shrinkage theory, the covariance matrix estimate, Γ^*, is calculated that will minimise the loss $||\Gamma^* - \Gamma||$, where $||A|| = \sqrt{tr(AA^T/p)}$ is the Frobenius norm, and Γ is the true covariance matrix [44].

To circumvent the issue of normality, Ledoit and Wolf's [43] estimate for large-dimensional covariance matrices is used as a plug-in for the shrinkage theory estimate by Chen *et al.* [15]. Ledoit and Wolf show that while Γ^* is not bona fide due to its reliance on four scalar functions of Γ, these four functions can be estimated to create S^*, an estimator for Γ^*. An advantage of using this method for estimation is that we do not need to assume any distribution for the data, just that there is only a finite forth moment, allowing its use despite not knowing the true distribution of the unit trust data. These scalars are $\mu = \langle \Gamma, I \rangle$, $\alpha^2 = ||\Gamma - \mu I||^2$, $\beta^2 = ||S - \Gamma||^2$, and $\delta^2 = ||S - \mu I||^2$, where $\langle A_1, A_2 \rangle = tr(A_1 A_2^T)/p$ is the inner product associated with the Frobenius norm, S is the sample covariance matrix and I is an identity matrix with dimensions equal to Γ. All of this is based on previous foresight of the true covariance matrix which means that estimates of these scalars will be needed to allow us to gain a workable covariance matrix despite not knowing the true covariance. In order to gain these estimates, the paper operate under the assumption of general asymptotes, denoted by the subscript n. The scalar estimates are

$$m_n = \langle S_n, I_n \rangle_n \,, \tag{4.8}$$

where $S_n = X_n X_n^T/n$ and $m_n - \mu_n \xrightarrow{q.m.} 0$,

$$d_n = ||S_n - m_n I_n||^2, \tag{4.9}$$

where $d_n - \delta_n \xrightarrow{q.m.} 0$,

$$\bar{b} = \frac{1}{n^2} \sum_{k=1}^{n} ||x^n_{.k}(x^n_{.k})^t - S_n||^2_n, \tag{4.10}$$

where $x^n_{.k}$ is the k^{th} row of the data matrix and then

$$b_n^2 = \min(\bar{b}^2, d_n^2), \tag{4.11}$$

where $b_n^2 - \beta_n^2 \xrightarrow{q.m.} 0$. This is done to ensure that the next estimator a_n^2 is non-negative, as a_n^2 is defined by

$$a_n^2 = d_n^2 - b_n^2, \tag{4.12}$$

where $a_n^2 - \alpha_n^2 \xrightarrow{q.m.} 0$. Replacing the unobservable parameters with the estimates yields the estimator for Γ^*:

$$S_n^* = \frac{b_n^2}{d_n^2} m_n I_n + \frac{a_n^2}{d_n^2} S_n. \tag{4.13}$$

Chen *et al.* then provide an outline for gaining a shrinkage estimate from any covariance matrix. All steps can be summarised by the fixed point iterations and the "clairvoyant estimator", or convergence, for the iterations. The fixed point iterations are

$$\widetilde{\Gamma}_{j+1} = (1-\rho)\frac{p}{n}\sum_{k=1}^{n}\frac{s_k s_k^H}{s_k^H \widehat{\Gamma}_j^{-1} s_k} + \rho I, \text{ and} \tag{4.14}$$

$$\widehat{\Gamma}_{j+1} = \frac{\widetilde{\Gamma}_{j+1}}{Tr(\widetilde{\Gamma}_{j+1})/p}, \tag{4.15}$$

where $s_k = \frac{x_k}{||x_k||_2}$ is the k^{th} normalized row of the data. With the correct value for ρ, these points converge to

$$\Gamma^* = (1-\rho)\frac{p}{n}\sum_{k=1}^{n}\frac{s_k s_k^H}{s_k^H \Gamma^{-1} s_k} + \rho I, \tag{4.16}$$

where the optimal value of ρ is given by:

$$\rho_O = \frac{p^2 - 1/pTr(\Gamma\Gamma^H)}{(p^2 - pn - n) + (n + (n-1)/p)Tr(\Gamma\Gamma^H)}, \tag{4.17}$$

when using S_n^* as the estimate for Γ.

The estimate Γ^* is then used as the variance-covariance matrix for the Markowitz model. This shrinkage estimate is designed to help with the estimation errors by shrinking the values in the new matrix towards a central value. This should, in theory, provide more accuracy as the larger estimation errors should be reduced the most [44, 77].

4.3 Solution methodology

A meta-heuristics approach is used in place of an exact solution approach, as the traditional methodologies for exact solutions take too long to be used in real world problems. The meta-heuristics chosen for use in this project are the Artificial Bee Colony (ABC), the Pareto Envelope-based Selection Algorithm (PESA), and Simulated Annealing (SA). These three meta-heuristics are chosen to represent each of the categories discussed in Chapter 2, due to their notable performance in literature.

4.3.1 Artificial Bee Colony

The ABC mimics the foraging behaviour of honeybees. The ABC hive consists of three different types of bees. Employed bees, located at every food source, that exploit their location for better food sources, where each food source represents a proposed portfolio to be used as a solution to the portfolio optimisation problem as formulated in equations (4.1) to (4.6). Onlooker bees, who exploit one of the solutions associated with an employed bee based on the probability associated with each employed bees solution. The probabilities are based on the dances that bees perform to the hive so that the best food sources can be chosen, and are calculated such that the better solutions have a higher probability of being selected, allowing for the algorithm to focus the search on better areas in the solution space. And finally, scout bees that will find food sources for employed bees to maintain in non-searched spaces once the employed bee has reached a local minimum. The parameters required by the ABC are the size of the hive, m, which will influence the number of each type of bee, the limit before abandoning a solution and moving to a scouted location, $\mathcal{C}$, and the stopping criteria, T.

The algorithm begins by initialising solutions for each employed bee and evaluating the initial solutions according to their fitness values, given by the objective function value for that solution. The exploitation phase begins by performing a local search around the food source associated with each employed bee, followed by the selection process and local search of the onlooker bees. If a better solution is found during this exploitation process the old solution will be abandoned for the new one. If a limit of non-improvement is reached, a local minimum is assumed and the employed bees who are assumed to be at a local minimum is moved to a scouted location [14, 36, 37]. This process is repeated until some stopping criteria is met. These steps are summarised in Algorithm 4.1.

Algorithm 4.1: ABC algorithm summary

Input : Size of the hive, criteria for move to scout location
Output: Population of solutions

```
initialise the employed bees solutions;
while stopping criteria not met do
    perform local search for each employed bee;
    select employed bees solutions and perform local search for each onlooker bee;
    update bee locations;
    move employed bees to scouted locations if necessary;
    store best solutions found;
end
```

4.3.1.1 Pseudo-code

The pseudo-code of the ABC is presented in Algorithm 4.2. The following subsections describe the ABC algorithm as it was designed and implemented to solve the portfolio optimisation problem as formulated in equations (4.1) to (4.6).

4.3.1.2 Initialisation

The algorithm begins by placing every bee at $\frac{m}{2}$ initial portfolio solutions $P_l = \{p_{l1}, p_{l2}, \ldots, p_{ln}\}$ for all $l = \{1, 2, \ldots, m/2\}$, where n is the number of unit trusts in the data set, m is the size of the hive, and p_{li} denotes the weight invested in unit trust i. This initial solution is generated by first identifying the unit trust, o, with the highest expected return from $\mu = \{\mu_1, \mu_2, \ldots, \mu_n\}$, and assigning $p_{lo} = 1000$ and $p_{li} = 0$ for all other unit trusts. Secondly, $m/2$ random weights, r_l, for $m/2$ randomly chosen unit trusts are generated, and each affected p_{li} is adjusted by

$$
\begin{aligned}
p_{lo} &= \frac{1000}{1 + r_l} && \text{if } \mu_o = \max(\mu),\\
p_{li} &= \frac{r_l}{1 + r_l} && \text{if } i \text{ randomly selected},\\
p_{li} &= 0 && \text{otherwise},
\end{aligned}
$$

for all $l \in \{1, 2, \ldots, m/2\}$ and all $i \in \{1, 2, \ldots, n\}$. This will result in $\frac{m}{2}$ initial portfolio solutions, each containing a maximum of 2 unit trusts with $p_{li} > 0$. Let $\mathcal{P} = \{P_1, P_2, \ldots, P_{m/2}\}$ and let the efficient frontier set, $\mathcal{P}^*$, be initialised by including only the portfolio containing unit trust o. This is contrary to literature where an independent random set of solutions is initialised for

Algorithm 4.2: ABC algorithm

Input : n, C_f^u, C_s^u,T,m.
Output: $\mathcal{P}^*$, the non-dominated set of portfolios.

initialise $\mathcal{P}$ and $\mathcal{P}^*$ as per §4.3.1.2;
create $\mathcal{C}^1 = \{c_1^1, c_2^1, \ldots, c_{m/2}^1\}$;
create $\mathcal{C}^2 = \{c_1^2, c_2^2, \ldots, c_{m/2}^2\}$;
while $t < T$ **do**
  determine candidate solution set for employed bees;
  for $l = [1, 2, \ldots, m/2]$ **do**
    determine $ER(V_l) = \sum_{i=1}^n v_{li}\mu_i$ and $V(V_l) = \sum_{i=1}^n \sum_{j=1}^n v_{li}v_{lj}\sigma_{ij}$;
    if $ER(V_l) \geq ER(P_l)$ ***and*** $V(V_l) \leq V(P_l)$ **then**
      $P_l \leftarrow V_l$;
      $c_l^1 \leftarrow 0$;
    end
    else
      $c_l^1 \leftarrow c_l^1 + 1$;
    end
  end
  determine candidate solution set for onlooker bees ;
  for $s = [1, 2, \ldots, m/3]$ **do**
    determine $ER(V_s) = \sum_{i=1}^n v_{si}\mu_i$ and $V(V_s) = \sum_{i=1}^n \sum_{j=1}^n v_{si}v_{si}\sigma_{ij}$;
    if $ER(V_s) \geq ER(P_l)$ ***and*** $V(V_s) \leq V(P_l)$ **then**
      $P_l \leftarrow V_s$;
      $c_l^1 \leftarrow 0$;
    end
    else
      $c_l^1 \leftarrow c_i^1 + 1$;
    end
  end
  for l *in* $\mathbf{c^1}$ **do**
    if $c_l^1 >= C_f^u$ **then**
      $P_l \leftarrow P_l^{new}$ as per §4.3.1.4;
      $c_l^2 \rightarrow c_l^2 + 1$;
      $c_l^1 \rightarrow 0$;
    end
    if $c_l^2 >= C_s^u$ **then**
      $P_l \leftarrow P_l^{\text{initial}}$;
      $c_l^1 \rightarrow 0$;
      $c_l^2 \rightarrow 0$;
    end
  end
  update $\mathcal{P}^*$ as per §4.3.1.5;
end

every employed bee [36] and despite the fixed starting point, the number of possible additions that can be made to the initial portfolio keeps the randomness of the searching. Every P_l is then checked for non-domination, where a non-dominated solution is considered to be a solution, P_l, where no other solution, P_{l^*} satisfies

$$\sum_{i=1}^n p_{li}\mu_i \leq \sum_{i=1}^n p_{l^*i}\mu_i$$

and

$$\sum_{i=1}^n \sum_{j=1}^n p_{li}p_{lj}\sigma_{ij} \geq \sum_{i=1}^n \sum_{j=1}^n p_{l^*i}p_{l^*j}\sigma_{ij}.$$

Any non-dominated P_i are then added to $\mathcal{P}^*$.

As part of the initialisation of the algorithm, the first set of fail counters for the employed bees, $\mathcal{C}^1 = \{c_1^1, c_2^1, \ldots, c_{m/2}^1\}$, and the second set, $\mathcal{C}^2 = \{c_1^2, c_2^2, \ldots, c_{m/2}^2\}$, along with their respective upper limits, C_f^u and C_s^u, are defined, as well as the stopping time, T.

4.3.1.3 Employed and onlooker bee phases

In the employed and onlooker bee phases, portfolio solutions or food sources are taken and used to find new solutions, in the hope that the new solutions are better. In order to generate candidate solutions, V_l, during the employed bee and onlooker bee phases, one of three operators are applied to the solutions in $\mathcal{P}$ with a random probability. The operator probabilities are determined by generating three random numbers, R_1, R_2, and R_3, where R_1 and R_2 are between 0 and 1, and R_3 is between 0 and R_2. If $R_1 \leq R_3$, then the first operator will be selected, if $R_3 \leq R_1 \leq R_2$, then the second operator will be selected, and if $R_1 > R_2$, then the third operator will be selected. The three operators are

1. A swap operator, where the weights of two randomly selected unit trusts in P_l are swapped.
2. A shuffle operator, where new weights are generated for all unit trusts in P_l with $p_{li} > 0$.
3. A selection operator, where the weight of the unit trust in P_l with the largest return over risk ratio, and $p_{li} > 0$, is increased by adding 10% of each other unit trust in P_l with $p_{li} > 0$.

Each candidate solution, V_l, is evaluated and compared to the associated solution, P_l, and if $\sum_{i=1}^n p_{li}\mu_i < \sum_{i=1}^n v_{li}\mu_i$ and $\sum_{i=1}^n \sum_{j=1}^n p_{li}p_{lj}\sigma_{ij} > \sum_{i=1}^n \sum_{j=1}^n v_{li}v_{lj}\sigma_{ij}$, then $P_l = V_l$ and the fail counter, $c_l^1 = 0$, otherwise $c_l^1 = c_l^1 + 1$.

The onlooker bees, on the other hand, select a solution from $\mathcal{P}$ to exploit based on the fitness of the employed bees solutions. For the selection process, the fitness is the return over risk ratio for each solution, namely

$$f_l = \frac{\sum_{i=1}^n p_{li}\mu_i}{\sum_{i=1}^n \sum_{j=1}^n p_{li}p_{lj}\sigma_{ij}}.$$

The probability of selecting a solution, P_l, is

$$P(P_l) = \frac{f_l}{\sum_{l=1}^{m/2} f_l}.$$

The onlooker bees then exploit the chosen portfolio from $\mathcal{P}$ by choosing one of the same three operators as the employed bees and determining a candidate solution V_s by applying the operator to the chosen portfolio. Each candidate solution, V_s, is evaluated and compared to the associated solution, P_l in $\mathcal{P}$. If $\sum_{i=1}^n p_{li}\mu_i < \sum_{i=1}^n v_{si}\mu_i$ and $\sum_{i=1}^n \sum_{j=1}^n p_{li}p_{lj}\sigma_{ij} > \sum_{i=1}^n \sum_{j=1}^n v_{si}v_{sj}\sigma_{ij}$, then $P_l = V_s$ and the fail counter, $c_l^1 = 0$, otherwise $c_l^1 = c_l^1 + 1$.

4.3.1.4 Scout phase

If any counter, c_l^1, has reached a user defined upper limit, gained through experimentation via parameter calibration, C_f^u, the solution, P_l, is changed to a new scouted solution, where a randomly selected unit trust with $p_{li} = 0$ is added to the solution with a random weight, or a randomly selected unit trust with $p_{li} > 0$ is removed. These two modifications are selected with equal probability, however, the algorithm will only choose the increase in unit trusts if the portfolio currently has less than three unit trusts in it. The second fail counter is then increased ($c_l^2 = c_l^2 + 1$) and the first fail counter is reset ($c_l^1 = 0$).

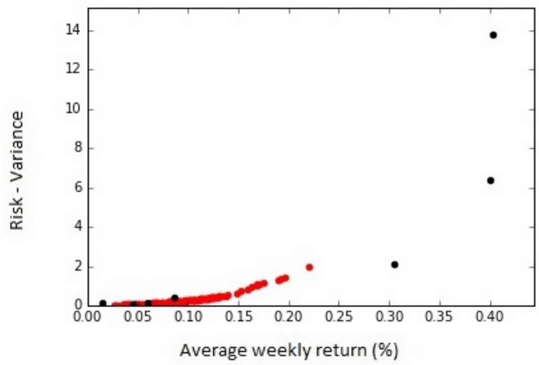
If any c_l^2 reaches its upper limit, C_s^u, the solution is not moved to the scouted location, but instead are moved back to the initial solution consisting of the unit trust, o, with the highest expected return from $\mu = \{\mu_1, \mu_2, \ldots, \mu_n\}$.

4.3.1.5 Frontier determination

The final step to the algorithm is to update the efficient set of portfolios, $\mathcal{P}^*$ by comparing the solutions in $\mathcal{P}$ to see if any of the solutions in $\mathcal{P}$ dominate any solution in $\mathcal{P}^*$. If any P_l is non-dominated it is added to $\mathcal{P}^*$, and if any solution, P_l^* is dominated, it is removed from $\mathcal{P}^*$.

4.3.1.6 Algorithm and parameter discussion

The modified initialisation and search phases are all designed to partly mimic the corner portfolio by Sharpe as mentioned in §2.1.1, by trying to find the best combination of a portfolio containing certain unit trusts before moving onto the next set of unit trusts, rather than moving randomly between different sized portfolios. This modification was found to improve the solution quality as may be seen in Figure 4.1.



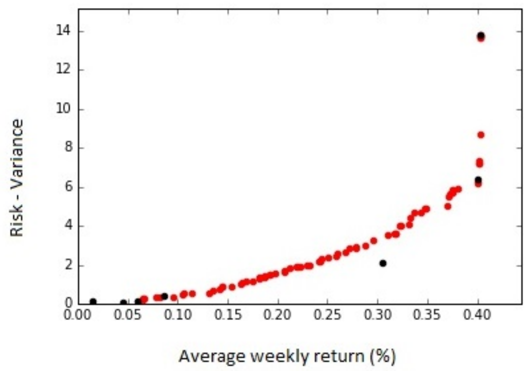


FIGURE 4.1: *Two scatter plots depicting the ABC at 2 hours with a fully random initialisation (left) vs ABC at 2 hours with the fixed start (right) in which the red points represent potential non-dominated solutions for the problem and the black points represent select points on a frontier gained using optimisation software.*

The effects of the parameters have an important role in deciding on the final value for the parameters. The size of the hive, if too small, means that too little of the solution space is being evaluated every iteration resulting in more iterations are possibly required, which results in more time being needed as other tasks are performed every iteration. However, too large of a hive size results in each iteration taking a significant time to completion, therefore control over the timing of the solution, and the stopping criteria in turn, becomes erratic. The hive size was set to 5000 as this was large enough to search the solution space effectively, but small enough to allow the algorithm to stop near to the specified target within an acceptable range. The limit to the scout phase decides how long the algorithm will stay at the assumed local minimum. Due to the large solution space and the fact that a small swap will make a massive difference, the limits are kept relative small. The first limit is kept to 500 to give the algorithm time to explore the percentages with the new securities in the portfolio and the second is kept to 10 to allow for quicker resets to a new portfolio. All parameter values for the ABC are gained through experimentation via a parameter calibration test.

4.3.2 Pareto Envelope-based Selection Algorithm

The PESA is an evolutionary algorithm that operates through crossing over sections of two parent solutions to form a child solution in the hopes that by combining two good parents the child produced will be better than the parents. In addition to crossovers, a parent can be mutated to create a new child, by changing one of the sections in the parent. As opposed to the genetic algorithm, which combines the parents and children into one set before choosing the strongest from the combined set to keep, the PESA keeps an archive of the best solutions and only one population. An outline to the procedure is describe in Algorithm 4.3.

Algorithm 4.3: PESA algorithm summary

Input : Size of the internal population, crossover and mutation rates
Output: Population of solutions

create initial internal population;
determine archive set from internal population;
while *stopping criteria not met* **do**
  empty internal population;
  while *internal population not full* **do**
    choose two parents from archive set;
    determine new child solution by applying crossovers/mutations;
    add child solution to internal population;
  end
  evaluate internal population for fitness;
  determine if any solutions in the internal population are non-dominated, place those solutions into the archive;
  remove dominated solutions from the archive;
end

The parameters used in the PESA are the internal population size, m, the mutation rate, r_m, the crossover rate, $r_c = 1 - r_m$, and the stopping criteria, T.

4.3.2.1 Pseudo-code

The pseudo-code of the PESA is presented in Algorithm 4.4 and will be discussed in the following subsections as it was implemented in this project.

Algorithm 4.4: Pareto Envelope-based Selection Algorithm

Input : $m, r_m,$ and T.
Output: $\mathcal{P}^*$, the non-dominated set of portfolios.

initialise $\mathcal{P}$ and $\mathcal{P}^*$ as per §4.3.2.2;
while $t < T$ **do**
  $\mathcal{P} \leftarrow$ **null**;
  create new $\mathcal{P}$ using the crossovers and mutations with rates $\frac{1-r_m}{3}$ and $\frac{r_m}{2}$;
  update $\mathcal{P}^*$ as per §4.3.2.4;
end

4.3.2.2 Initialisation

The algorithm begins by creating a set of randomly generated initial solutions, $\mathcal{P} = \{P_1, P_2, \ldots, P_m\}$, where $P_l = \{p_{l1}, p_{l2}, \ldots, p_{ln}\}$, with m being the desired size of the internal population, n being the number of unit trusts in the dataset, and p_{li} denoting the weight invested in unit trust i. One of the portfolios in this initial set is the portfolio consisting solely of the unit trust o, with the highest expected return from $\mu = \{\mu_1, \mu_2, \ldots, \mu_n\}$. The initial solutions are then evaluated for non-dominance as described in §4.3.1.2, and if they are non-dominated they are added to the archive, $\mathcal{P}^*$. Part of the initialisation is to define a stopping time, T. The initial set, $\mathcal{P}$ is then emptied to be used for the child solutions.

4.3.2.3 Crossovers and mutations

The iterative process of the PESA is the selection of and the application of crossovers and mutations to the archived solutions, known as parent solutions. The selection criteria chosen for this study is the tournament selection process, in which two parents are selected from the archive with an equal random chance, and compared using the return to risk ratio. The winning parent is then the parent with the highest ratio. Two parents will be chosen to create a child, and this will be repeated until the desired internal population size is achieved by creating the required number of children. In this study, we utilise three crossovers and two mutations to create children. Crossovers are chosen with probability $\frac{r_c}{3}$ for each of the three crossovers, and the mutation with probability $\frac{r_m}{2}$ for the two mutations. The crossovers are

1. Crossing over the weights of two parent solutions at the halfway point, so that the child consists of the first half of the first parent and the second half of the second parent. As an example, suppose the dataset consists of 8 unit trusts, this crossover would be performed as
$$P_1^* + P_2^* \rightarrow P_1 = \{p_{11}, p_{12}, p_{13}, p_{14}, p_{25}, p_{26}, p_{27}, p_{28}\}.$$
2. Crossing over the weights of two parent solutions which have been divided into quarters, so that the child solution consists of the first and third quarters of the first parent solution, and the second and fourth quarters of the second parent solution. As an example, suppose the dataset consists of 8 unit trusts, this crossover would be performed as
$$P_1^* + P_2^* \rightarrow P_1 = \{p_{11}, p_{12}, p_{23}, p_{24}, p_{15}, p_{16}, p_{27}, p_{28}\}.$$
3. Crossing over the weights of two parent solutions at a random point, so that the child solution consists of the first parent solution until the random point and the second parent solution from the unit trust after the random point to the final unit trust. As an example, suppose the dataset consists of 8 unit trusts, this crossover could be performed, amongst others, as either
$$P_1^* + P_2^* \rightarrow P_1 = \{p_{11}, p_{22}, p_{23}, p_{24}, p_{25}, p_{26}, p_{27}, p_{28}\},$$
$$P_1^* + P_2^* \rightarrow P_1 = \{p_{11}, p_{12}, p_{13}, p_{24}, p_{25}, p_{26}, p_{27}, p_{28}\}.$$

The above crossovers were the only crossovers described in the literature explored by the author and, to keep the PESA simple in its use, any other possible crossovers are excluded from this project.

The mutations are

4. To remove a randomly selected unit trust with $p_{li} > 0$ and rebalance the solution.

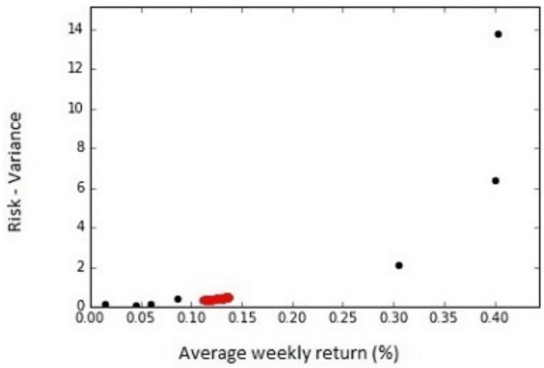
5. To perform the swap operator, where the weights of two randomly selected unit trusts with $p_{li} > 0$ are swapped.

4.3.2.4 Frontier determination

Each P_l is tested for non-dominance against every other P_l and against every P_l^* and added to $\mathcal{P}^*$ if they are non-dominated, and any dominated P_l^* are removed from $\mathcal{P}^*$. Once the domination check is completed, $\mathcal{P}$ is emptied for use in the next iteration.

4.3.2.5 Algorithm and parameter discussion

The addition of the fixed portfolio in the initialisation is not standard to literature [55], however, it was found to significantly improve the solution quality. This is evident from Figure 4.2 (right), clearly showing a vast number of additional points on the efficient frontier.



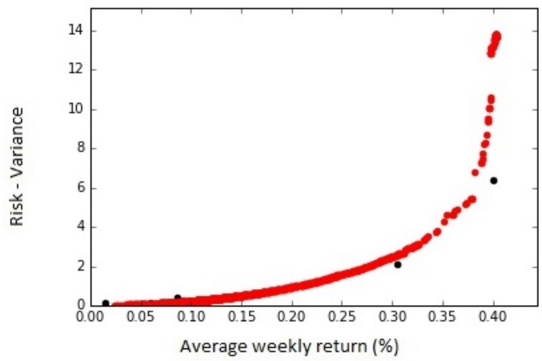


FIGURE 4.2: *Two scatter plots depicting the ABC at 2 hours with a fully random initialisation (left) vs ABC at 2 hours with the fixed insertion at the start (right) in which the red points represent potential non-dominated solutions for the problem and the black points represent select points on a frontier gained using optimisation software.*

The parameter testing for PESA yields that the internal population size acts the same as the hive size in the ABC, in that too small means that not enough solution space can be explored at every iteration and too large affects the control over the stopping criteria. A population size of 5000 is therefore used as it provided the best results when the parameter calibration was performed. The mutation rate is kept low, at 5%, in order to give more priority to crossovers as the mutations can be seen as similar to search moves from the ABC and SA.

4.3.3 Simulated Annealing

SA mimics the method of heating and cooling of controlled metals in order to have the metal set at its strongest configuration. To achieve this, an initial temperature is chosen, which determines if new solutions are accepted as a new current solution. Initially, the temperature will allow solutions that are worse than the current solution to be accepted, but as time progress the temperature will be lowered to enforce a hill climbing effect. Due to the acceptance function, SA allows the solution to climb out of local minima more effectively than most other algorithms, and is a good choice for solution spaces with many local minima [39].

The outline of SA is found in :

```
Input  :
Output:
set initial temperature;
determine initial random solution;
while stopping criteria not met do
    perform search on current solution;
    determine acceptance;
    memorise best solution;
    update temperature;
end
```

For use in the project, multiple SA solutions are determined in parallel, each associated with a different λ value from (4.7). The parameters used in the SA algorithm are the initial temperature, t_p, the cooling factor, S^c, and the limit on non-acceptance to increase the temperature, $\mathcal{C}$.

4.3.3.1 Pseudo-code

The pseudo-code of SA is presented in Algorithm 4.5. The following subsections will describe SA as it was implemented for use in this project.

4.3.3.2 Initialisation

The SA algorithm begins similarly to the ABC, by placing every annealing crystal at m identical initial portfolio solutions $P_l = \{p_{l1}, p_{l2}, \ldots, p_{ln}\}$ for all $l = \{1, 2, \ldots, m\}$, where n is the number of unit trusts in the data set, m is the number of λ values to be evaluated, and p_{li} denotes the weight invested in unit trust i. This initial solution is generated by first identifying the unit trust, o, with the highest expected return from $\mu = \{\mu_1, \mu_2, \ldots, \mu_n\}$, and assigning $p_{lo} = 1000$ and $p_{li} = 0$ for all other unit trusts. Secondly, m random weights, r_l, for m randomly chosen unit trusts are generated, and each affected p_{li} is adjusted by

$$\begin{aligned}
p_{lo} &= \frac{1000}{1 + r_l} && \text{if } \mu_o = \max(\mu),\\
p_{li} &= \frac{r_l}{1 + r_l} && \text{if } i \text{ randomly selected},\\
p_{li} &= 0 && \text{otherwise},
\end{aligned}$$

for all $l \in \{1, 2, \ldots, m\}$ and all $i \in \{1, 2, \ldots, n\}$. This will result in m initial portfolio solutions, each containing a maximum of 2 unit trusts with $p_{li} > 0$. Let $\mathcal{P} = \{P_1, P_2, \ldots, P_m\}$ and let the efficient frontier set, $\mathcal{P}^*$, be initialised by including all P_l. Part of the initialisation is to define an initial temperature, t_l^p for every P_l, to set the non-acceptance counters, c_l^1, and the reheat counters, c_l^2 to 0 for every P_l, to define upper limits for the counters, C_f^u and C_s^u, respectively, to define a stopping time, T, and to set a cooling schedule coefficient, S^c.

4.3.3.3 Search phase

The search phase of the algorithm will select one of three operators with probabilities $P(1) = 0.45$, $P(2) = 0.4$, and $P(3) = 0.15$, to create a candidate solution, V_l, for each P_l. The three operators are

Algorithm 4.5: Pseudo-code of the Simulated Annealing algorithm in this paper.

Input : T, n, C_f^u, C_s^u, S^c.
Output: $\mathcal{P}^*$, the non-dominated set of portfolios.

```
initialise P and P* as per §4.3.3.2;
define initial temperatures, t_l^p for every P_l;
set C^1 = {c_1^1, c_2^1, ..., c_m^1};
set C^2 = {c_1^2, c_2^2, ..., c_m^2};
define T, C_f^u, C_s^u, S^c;
while t < T do
    determine candidate solution set, V;
    for l in V do
        determine acceptance using P^A(V_l) = min(1, exp((-f(V_l) - f(P_l)) / t_l^p)) ;
        if solution accepted then
            P_l <- V_l;
            c_l^1 <- 0;
        end
        else
            c_l^1 <- c_l^1 + 1;
        end
        if f(V_l) < f(P_l^*) then
            P_l^* <- V_l;
        end
        if c_l^1 >= C_f^u then
            t_l^p <- 100;
            c_l^2 <- c_l^2 + 1;
        end
        if c_l^2 >= C_s^u then
            t_l^p <- 100;
            P_l <- P_l^*;
            c_l^1 <- 0;
            c_l^2 <- 0;
        end
        t_l^p -> sc × t_l^p
    end
end
```

1. A redistribution operator, which identifies the two securities with $p_{li} > 0$ that contribute the most risk relative to their return and removes them before redistributing the weight to other securities.

2. A swap operator, which will swap the weights of two randomly selected unit trusts with $p_{li} > 0$.

3. An exploration operator, where a randomly selected unit trust with $p_{li} = 0$ is added to the solution with a random weight, or a randomly selected unit trust with $p_{li} > 0$ is removed. These two modifications are selected with equal probability, however, the algorithm will only choose the increase in unit trusts if the portfolio currently has less than three unit trusts in it.

The first function is designed to include both the ability to find the best combination of unit trusts in the portfolio and to reshape the portfolio by removing unit trusts. The second function tries to find the corner portfolio of the unit trusts currently in the portfolio. The third function varies the unit trusts in the portfolio to find new corner portfolios. These functions are designed for the unit trust selection problem, where the standard search function is to swap two weights [3].

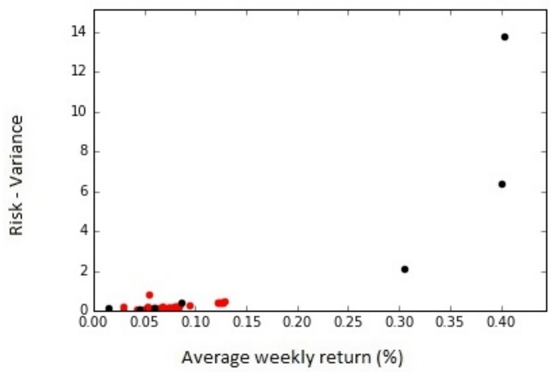
Each V_l is evaluated to see whether it will be accepted by the algorithm, using an acceptance function,

$$P^A(V_l) = \min(1, \exp(\frac{-(f(V_l) - f(P_l))}{t_l^p})),$$

where $f(\cdot)$ is the fitness value of the solution, calculated using (4.7) and the λ value for solution l. If the solution is accepted, $P_l = V_l$ and the fail counter is reset ($c_l^1 = 0$) and if V_l is also better than P_l^*, $P_l^* = V_l$. If the solution is not accepted, then the first counter is increased ($c_l^1 = c_l^1 + 1$) and if the resulting increase causes the counter to reach its limit ($c_l^1 = C_f^u$) then the second limit is increased ($c_l^2 = c_l^2 + 1$). If the second increase also causes the counter to reach its limit ($c_l^2 = C_s^u$) then the solution is reset to its best known solution, P_l^*, the temperature is reset to the initial value, and all counters are reset to 0.

4.3.3.4 Algorithm and parameter discussion

The modified initialisation and the search moves are contrary to literature, and are designed to emulate Sharpe's corner portfolio model in a similar fashion to the ABC. The modifications were included as it was found to improve the solution quality in terms of catering to multiple investor types, as shown by in Figure 4.3 (right). Three of the parameters have an effect on the



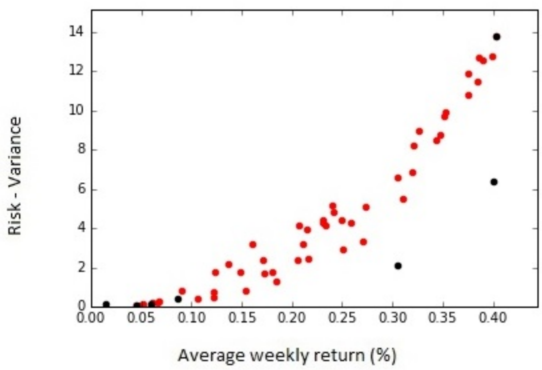


FIGURE 4.3: *Two scatter plots depicting the SA at 2 hours with a fully random initialisation (left) vs SA at 2 hours with the fixed start (right) in which the red points represent the initial best set of solutions for the problem and the black points represent select points on a frontier gained using optimisation software.*

outcome of the algorithm, namely, the cooling factor, the initial temperature, and the limits for non-acceptance. If the cooling factor is set too low, then the algorithm quickly becomes a hill climbing search as no worse solutions are accepted, and if set too high, it becomes a random search as the algorithm just accepts all answers. For this project, the cooling factor is set to 0.75 as any higher led to the solution moving away from good solutions too soon, and any lower meant that the local minima are explored to often. The initial temperature is set to 100 (and reheated to 100) as this gives the algorithm enough time to search through the solution space by accepting a larger number of worse solutions for longer, but still allowing the algorithm to enforce a hill climbing effect. The limits for resetting the temperature are kept low, at 20, to keep the algorithm from remaining at local minima for too long. The parameter values for the SA algorithm were determined by parameter calibration.

4.4 Results metrics

As discussed in Chapter 2, there are various result metrics that can be used to evaluate the efficient frontier determined by which ever solution method was chosen. Three different metrics are defined with which the meta-heuristics will be evaluated and compared in this project, namely the diversity metric, the spacing metric, and the frontier metric. A concise description of each metric and the calculation thereof is given below.

The Diversity metric, Δ, shows how well spaced the algorithm results are through the solution space [14, 22]. It is defined as the mean Euclidean distance between each point in the proposed frontier and the nearest points above and below it on the proposed frontier. The Euclidean distance between two solutions in the proposed frontier, l and q, is defined as

$$D_{lq} = \sqrt{\left(\sum_{i=1}^{n} p_{li}\mu_i - \sum_{i=1}^{n} p_{qi}\mu_i\right)^2 + \left(\sum_{i=1}^{n}\sum_{j=1}^{n} p_{li}p_{lj}\sigma_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{n} p_{qi}p_{qj}\sigma_{ij}\right)^2}.$$

The Spacing metric, or S-metric, shows how much of the solution space is captured by the algorithm and is calculated by determining the area of the smallest rectangle that would encompass the entire frontier [22]. The S-metric is defined as

$$S = \left(\max_l \left[\sum_{i=1}^{n} p_{li}\mu_i\right] - \min_l \left[\sum_{i=1}^{n} p_l i\mu_i\right]\right) \cdot \left(\max_l \left[\sum_{i=1}^{n}\sum_{j=1}^{n} p_{li}p_{lj}\sigma_{ij}\right]\right).$$

In order to determine how close the algorithms manage to get to the selected set of points from the frontier developed by optimisation software a frontier metric is defined [14]. This metric is calculated by finding the Euclidean distances from all points in the selected set to the nearest solution in the frontier and determining the average, where the distance from one point on the optimised frontier, w, to its nearest point on the proposed frontier, l, is defined as

$$\min_l \left[F_{wl}\right] = \sqrt{\left(\sum_{i=1}^{n} p_{wi}\mu_i - \sum_{i=1}^{n} p_{li}\mu_i\right)^2 + \left(\sum_{i=1}^{n}\sum_{j=1}^{n} p_{wi}p_{wj}\sigma_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{n} p_{li}p_{lj}\sigma_{ij}\right)^2}.$$

4.5 Chapter summary

Chapter 4 contains a description of all techniques used to achieve some of the objectives listed in Chapter 1. The chapter begins with the model formulation of the adjusted Markowitz model using an additional constraint to ensure the short term growth is also positive to keep investor confidence in the chosen portfolio. Following this is an explanation of the covariance estimate used in this study in place of the sample covariance matrix. The new estimate is used as it is distribution free and robust against the high dimensionality problem as described in Chapter 2.

Three algorithms chosen for use in this project, namely, the ABC, the PESA, and SA, are then discussed. For each algorithm, a brief introduction and overview is given before describing the important phases of the algorithm steps as they pertain to solving the portfolio optimisation problem. For each algorithm, an explanation of how the parameters of the algorithm affect the outcomes is included.

Finally, the performance metrics that will be used to determine efficacy of the algorithms are discussed and the formulae for determining the metrics are shown.

CHAPTER 5

Results

Contents

In this chapter the results obtained from the three algorithms described in Chapter 4 are given. In each section, the performance of the respective algorithms is reported on by means of graphs showing the solutions found by the algorithm as time progresses as well as the values of the performance metrics as stipulated in §4.6. In all graphs, the true efficient frontier is represented by a few choice points plotted in black and the algorithm frontier is represented by red points.

5.1 Computer specifications

The speed and resulting quality to time ratio of the algorithms will rely on the specifications of the computer and the software used to develop the algorithms. Table 5.1 provides a summary of the specifications and software used in this study. This computer was used to run each algorithm 10 times to gain an idea of the average performance of each algorithm.

Specification	Details
Operating System	Windows 10 Enterprise 64-bit
Processor	Intel Core i7-3770
RAM	8GB
Software	Anaconda Spyder 2.3 for Python 3.5

TABLE 5.1: *A summary of specification of the computer used to run the all three algorithms used in this study.*

5.2 The efficient frontier

To determine how well the output from the algorithms compares to the efficient frontier, a small set of points from this frontier was obtained by implementing the problem formulation given in §4.1 with the adjusted covariance as described in §4.2 in Lingo [45]. As the problem is quadratic in nature, Lingo uses a Generalized Reduced Gradient (GRG) algorithm while also incorporating Successive Linear Programming (SLP) to provide a quick feasible initial solution [46]. To gain the set of points, the value of λ was changed after each model rerun. The selected values of λ are $\lambda \in \{0, 0.25, 0.5, 0.75, 0.9, 0.95, 1\}$. Initially the 0.9 and 0.95 points were not included, however, it was found that the space between $\lambda = 1$ and $\lambda = 0.75$, is not equal to the space between $\lambda = 0.5$ and $\lambda = 0.25$, so points were added near the top of the frontier to try and find a more equal spread. The time taken to gain these 7 points were approximately a week, which makes using the exact model infeasible to gain portfolios in a usable time frame.

5.3 Artificial Bee Colony

The expected returns and variances associated with the non-dominated set of portfolios obtained by the ABC algorithm may be seen in Figures 5.1 to 5.6. In Figure 5.1, it may be seen that within a short time, the algorithm has already managed to find portfolios along the upper and bottom most section of the efficient frontier as well as being close at all other points. When observing the 30 and 60 minute graphs in Figures 5.2 and 5.3, it may be seen that while the overall position of the solutions does not move too much, it does become noticeably smoother. After 2 - 4 hours, the algorithm only finds new solutions to fill in the gaps between points in the proposed frontier, as may be seen in Figures 5.4 to 5.6. This indicates that while the ABC algorithm does have the ability to find better solutions as time progresses, resulting in a smooth proposed frontier, it struggles to converge to the optimal frontier and stays at a good solution to the problem. This correlates with the observation by Tuba and Bacanin [74], who also found that the ABC algorithms tend to have problems converging. A recommendation to improve the convergence is by replacing the employed bees search with their Firefly Algorithm [73].

5.4 Pareto Envelope-based Selection Algorithm

The expected returns and variances associated with the non-dominated set of portfolios obtained by the PESA may be seen in Figures 5.7 to 5.12. The PESA begins with a smooth curve at 10 minutes as seen in Figure 5.7, but quite far from the true efficient frontier. In Figures 5.8 and 5.9, it may be seen that at 30 minutes the algorithm begins to move towards the true frontier while by 60 minutes it has also managed to smooth itself out as the time progressed. The main

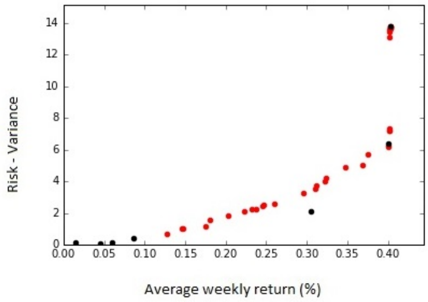


FIGURE 5.1: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the ABC at a 10 minute runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

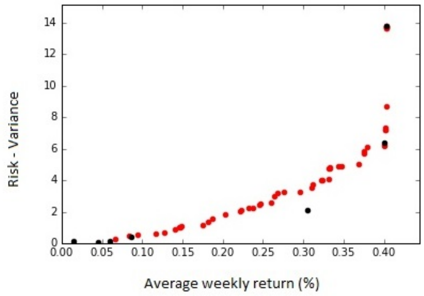


FIGURE 5.2: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the ABC at a 30 minute runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

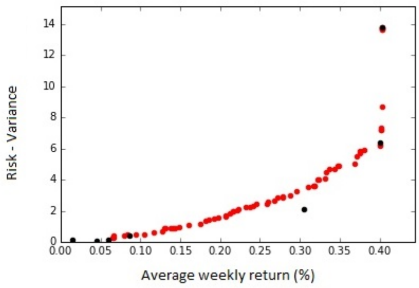


FIGURE 5.3: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the ABC at a 60 minute runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

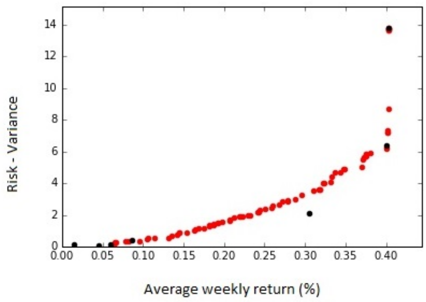


FIGURE 5.4: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the ABC at a 2 hour runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

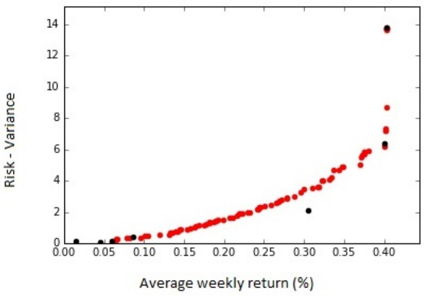


FIGURE 5.5: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the ABC at a 3 hour runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

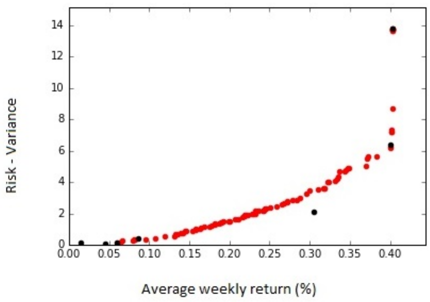


FIGURE 5.6: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the ABC at a 4 hour runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

features of the PESA is that it continues to move towards the true frontier as time progresses, as may be seen in Figures 5.10 to 5.12, and that it is best able to cover the entire efficient frontier.

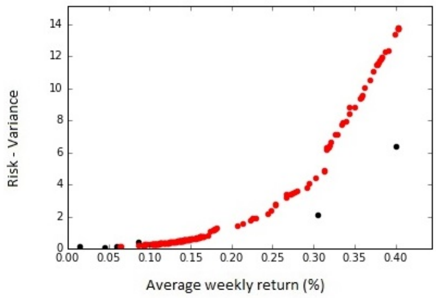


FIGURE 5.7: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the PESA at a 10 minute runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

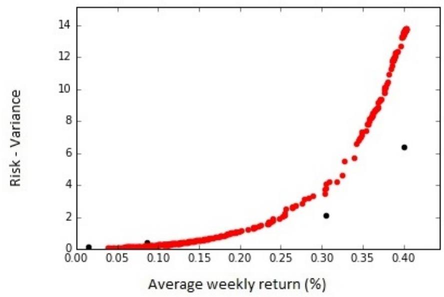


FIGURE 5.8: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the PESA at a 30 minute runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

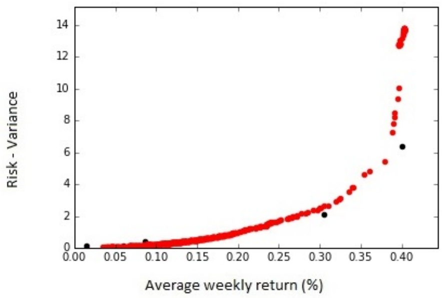


FIGURE 5.9: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the PESA at a 60 minute runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

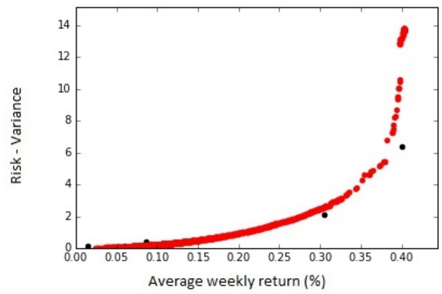


FIGURE 5.10: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the PESA at a 2 hour runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

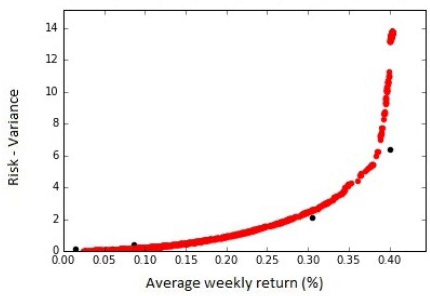


FIGURE 5.11: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the PESA at a 3 hour runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

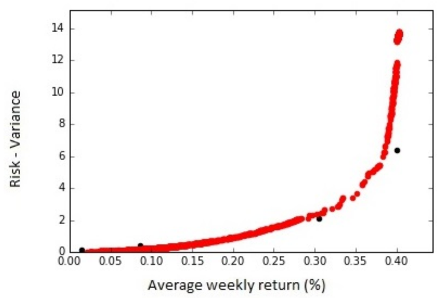


FIGURE 5.12: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the PESA at a 4 hour runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

5.5 Simulated Annealing

The expected returns and variances for different values of λ obtained by SA may be seen in Figures 5.13 to 5.18. At ten minutes, the simulated annealing algorithm resembles an elongated scatter formation stretched throughout the true frontier, as seen in Figure 5.13. In Figure 5.14, we see that by 30 minutes, very small, if no, changes have been made to the solution. This is better displayed by the result metric for diversity which only changes in the 5th decimal place between 10 and 30 minutes. After this, however, the algorithm shows no improvement at all and seems to be stuck at poor solutions as seen in Figures 5.16 to 5.18.

This can be attributed to SA being a single solution based algorithm where each point in the graph represents the return and risk values that contribute to the single objective (4.7) for each corresponding value of λ. As there is only one solution that will optimise the equation for each value of λ, the algorithm requires that the search associated with that λ value finds that solution. Therefore, if a solution is found that might be considered optimal for a value of λ, it may not be optimal for the λ value being evaluated and will therefore not be recorded. When combined with the large solution space, the probability of finding the solution that will optimise the solution for the value of λ is very low. While in a population based algorithm this would not be an issue, for a single solutions based algorithm it becomes a problem.

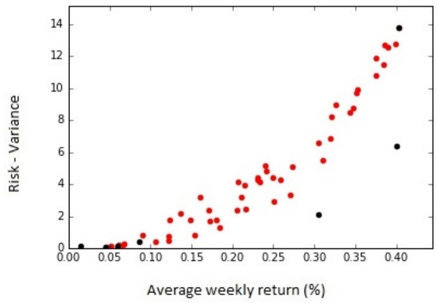


FIGURE 5.13: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the SA at a 10 minute runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

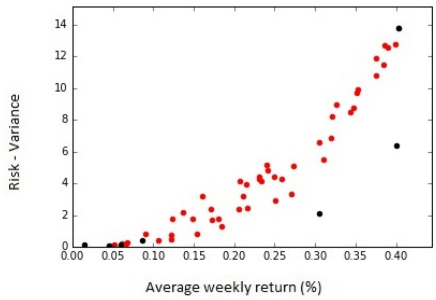


FIGURE 5.14: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the SA at a 30 minute runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

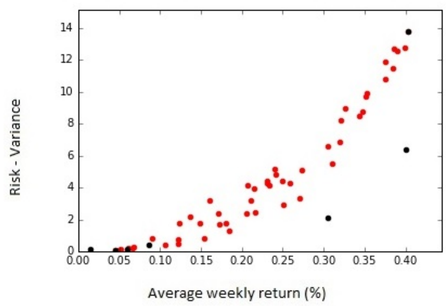


FIGURE 5.15: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the SA at a 60 minute runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

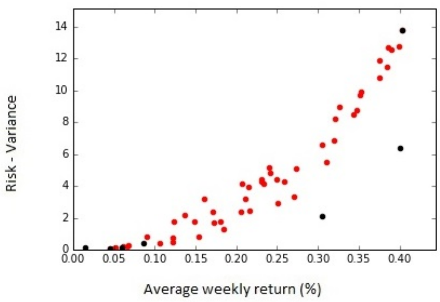


FIGURE 5.16: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the SA at a 2 hour runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

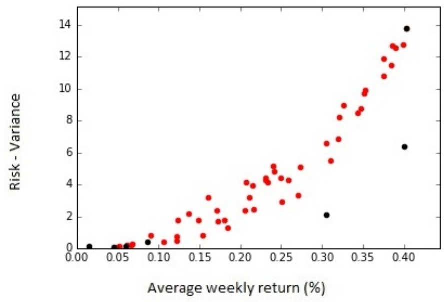


FIGURE 5.17: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the SA at a 3 hour runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

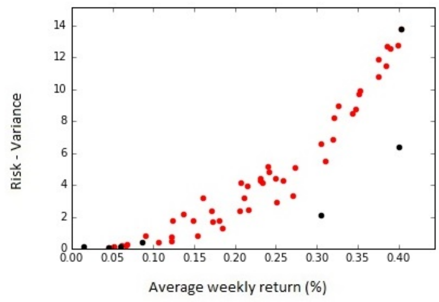


FIGURE 5.18: *A scatterplot depicting the risk and return levels for the proposed solutions (red) of the SA at a 4 hour runtime in one of 10 instances, versus the risk and return levels of the frontier determined using optimisation software (black).*

5.6 Performance metrics and comparison of the algorithms

An initial comparison of the three algorithms is given by a scatterplot showing a combined frontier of all three algorithms, where only solutions that are non-dominated when compared to other algorithms solutions are added to the frontier. This may be see in Figure 5.19.

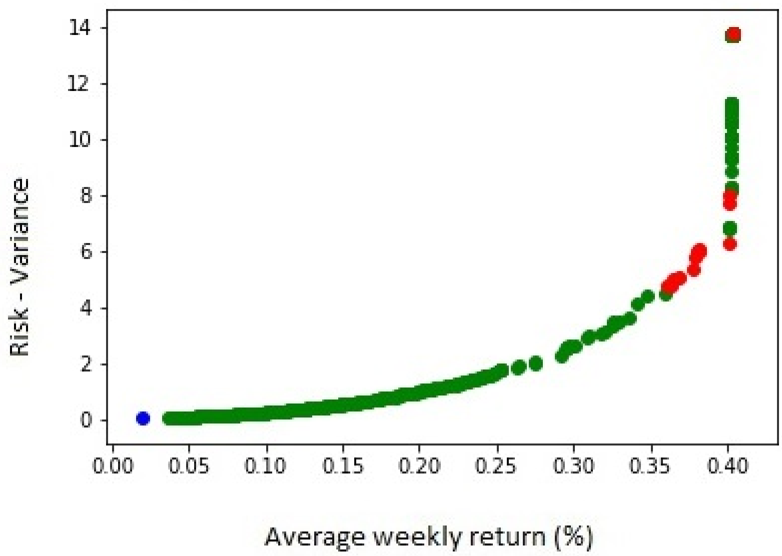


FIGURE 5.19: *A scatterplot depicting the risk and return levels for the combined frontier consisting of solutions of SA (blue), PESA (green), and the ABC (red).*

From Figure 5.19 it is clear that the PESA contributes the most to the combined frontier and will, most likely, be the algorithm with the better performance, as the ABC and SA only manges to contribute very few solutions to the frontier. What may also be seen is that the ABC tends to provide solutions around the middle of the frontier, while the SA provides only a solution at the lower of the frontier, which will have approximately 0% growth, which will defeat the purposes of investing. The ABC solutions in the middle of the frontier suggest that using the ABC and PESA in combination with each other may yield a better overall result, however, the added solutions from the ABC may not be worth the additional requirements of running the two simultaneously and further testing will be required.

The averages for the three metrics given in §4.4 for each algorithm over time are summarised in Tables 5.2, 5.3, and 5.4. From a cursory glance at the tables we may see that in Table 5.3, all three algorithms are able to cover the same area of the solution space, meaning that each algorithm does manage to gain portfolios at either end of the efficient frontier. If we observe Table 5.2, it may be seen that while all three algorithms cover the same space, the PESA may have a clear advantage in terms of diversity, thereby being able to represent more investor preferences in terms of the portfolios it generates. The PESA is followed by the ABC, which in turn is followed by SA. The frontier metric, in Table 5.4, shows us that SA is very far away from the frontier at all points in time and it may be concluded as the worst performing meta-heuristic. Furthermore, while the metric may indicate an eventual out-performance of the PESA by the ABC, results may be biased as a result of using only a small set of representative points on the best frontier found by optimisation software in the metric calculation instead of all points on the true frontier. This result may be improved by increasing the set of representative points.

Diversity Metric								
Algorithm	**10 Min**	**30 Min**	**60 Min**	**2 Hrs**	**3 Hrs**	**4 Hrs**	**Runs**	**Variance**
ABC	0.162	0.1230	0.0935	0.0793	0.0726	0.0676	10	$6.69 \cdot 10^{-4}$
PESA	0.0356	0.0142	0.0060	0.0024	0.0015	0.0010	10	$1.7 \cdot 10^{-6}$
SA	0.1339	0.1266	0.1234	0.1220	0.1228	0.1228	10	$8.5 \cdot 10^{-4}$

Table 5.2: *A summary of the Diversity metric results over time.*

Spacing Metric								
Algorithm	**10 Min**	**30 Min**	**60 Min**	**2 Hrs**	**3 Hrs**	**4 Hrs**	**Runs**	**Variance**
ABC	3.9864	4.4489	4.8277	4.7398	4.9826	5.0288	10	0.057
PESA	4.3889	4.9182	5.0914	5.0784	5.1012	5.0938	10	0.031
SA	5.001	5.0261	5.0261	5.0261	5.0261	5.0261	10	1.204

Table 5.3: *A summary of the Spacing metric results over time.*

Distance to Frontier								
Algorithm	**10 Min**	**30 Min**	**60 Min**	**2 Hrs**	**3 Hrs**	**4 Hrs**	**Runs**	**Variance**
ABC	0.2545	0.1504	0.0986	0.0481	0.0398	0.0359	10	$3.5 \cdot 10^{-4}$
PESA	0.2146	0.1145	0.0756	0.0468	0.0411	0.0522	10	0.001
SA	3,1338	3,1354	3,1354	3,1354	3,1354	3,1354	10	2.192

Table 5.4: *A summary of the Frontier metric results over time.*

In order to assess whether any of the algorithms is significantly better than the rest, a Student t-test was performed on the metrics. Due to the results coming from different algorithms, independent variance is assumed and a pooled t-test was performed. The Student t-test determines if the average values of the metrics for each algorithm significantly differ, where the null hypothesis is, $H_0 : \mu_1 = \mu_2$. The Student t-tests at $\alpha = 0.05$ have a critical value of ± 2.1. The summary of these results may be seen in Table 5.5.

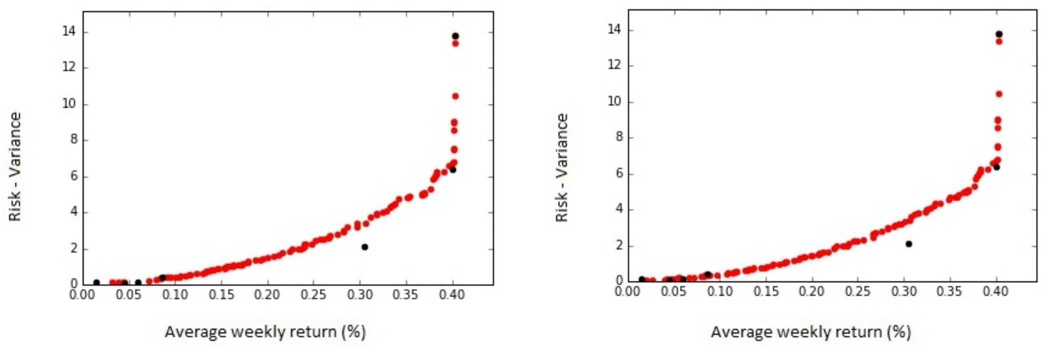
A true result means that the null hypothesis holds for every metric in the 4 hour period, or that

Diversity Metric		Spacing Metric		Frontier Metric	
Pairing	$H_0 : \mu_1 = \mu_2$	Pairing	$H_0 : \mu_1 = \mu_2$	Pairing	$H_0 : \mu_1 = \mu_2$
ABC - PESA	False	ABC - PESA	True	ABC - PESA	True
ABC - SA	False	ABC - SA	True	ABC - SA	False
PESA - SA	False	PESA - SA	True	PESA - SA	False

TABLE 5.5: *A summary of the hypotheses and results of the Student t-tests for the three different performance metrics.*

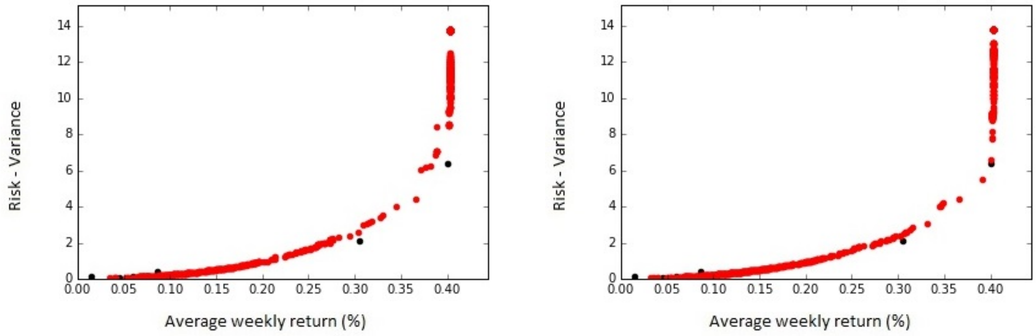
it holds for the metrics that are associated with the later time periods. This will indicate that neither algorithm outperforms the other, or that as time passes the algorithm's metric results become very similar despite one of the algorithms having better initial values. A false result in turn would be if the null hypothesis does not hold for every metric in the 4 hour period, or that it only holds for the metrics that are associated with the earlier time periods. This indicates that one of the algorithms always outperforms the other, or that as time progresses one of the algorithms starts to outperform the other despite similar initial values. Based on the results of the t-tests, the algorithms can be said to represent the same total area of the solution space, however, the individual preference for risk and return is not equally covered. In this regard, the PESA can be said to outperform both the ABC and SA, and that the ABC outperforms SA. In terms of distance to the optimised frontier, both the ABC and PESA outperform SA, however, do not manage to significantly outperform the other.

Since the ABC and PESA outperform the SA in all aspects, other than spacing, SA should be excluded for use in a decision support tool. The PESA is recommended for use in a decision support tool as it is equal to the ABC in two categories, and the best in the other, allowing the decision support tool to gain better portfolios for any investor risk profile that wishes to use it. This is corroborated by running the algorithms for 20 hours to see how the extra time affects the results, shown in Figures 5.20, 5.21, and 5.22. These graphs show that SA does not change from 5 to 20 hours, showing that any extra time does not help the single solution search algorithm; the ABC continues to find solutions within its current range, but the extra time does not help the non convergence issue; and that the PESA continues to move closer to the pareto front over time.



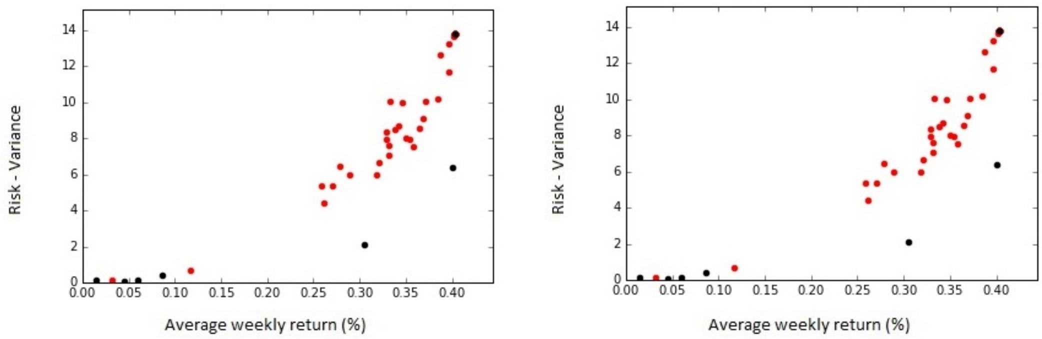
(a) *5 hours* (b) *20 hours*

FIGURE 5.20: *Two scatterplots depicting the risk and return levels for the proposed solutions (red) of the ABC at a 5 hour runtime (left) and a 20 hour runtime (right), versus the risk and return levels of the frontier determined using optimisation software (black).*



(a) *5 hours* (b) *20 hours*

FIGURE 5.21: *Two scatterplots depicting the risk and return levels for the proposed solutions (red) of the PESA at a 5 hour runtime (left) and a 20 hour runtime (right), versus the risk and return levels of the frontier determined using optimisation software (black).*



(a) *5 hours* (b) *20 hours*

FIGURE 5.22: *Two scatterplots depicting the risk and return levels for the proposed solutions (red) of the SA at a 5 hour runtime (left) and a 20 hour runtime (right), versus the risk and return levels of the frontier determined using optimisation software (black).*

In Figure 5.23, 5.24, and 5.25, the values of the respective performance metrics are given as a function of simulation run time. From these results, suitable stopping criteria can be defined for each algorithm for future use. In Figure 5.25, the SA results have been excluded due to having a significantly higher value than either the ABC or PESA, which makes comparison of the ABC and the PESA difficult.

It may be seen that the ABC algorithm has large decreases in the diversity metric until it reaches the second hour, at which point the decreases begin to slow, with the cumulative decreases from hour 2 to 4 being approximately equal to the decrease from hour 1 to 2. The frontier metric improves until the 2 hour mark and then plateaus with little improvement from that point on. The spacing metric, however, does seem to continue to change as time progresses. As the frontier and diversity metric could reasonably be considered as more relevant than the spacing metric, a choice between 1 and 2 hours will be required by the user of the algorithm. The PESA has relatively small improvements in the diversity metric, which can be attributed to it being effective from the start. The spacing and frontier metrics seem to improve rapidly until the 1 and 2 hour marks, respectively. Termination at 2 hour may therefore be a reasonable choice. Finally, the results of the SA algorithm show relatively no improvement in any of the runs,

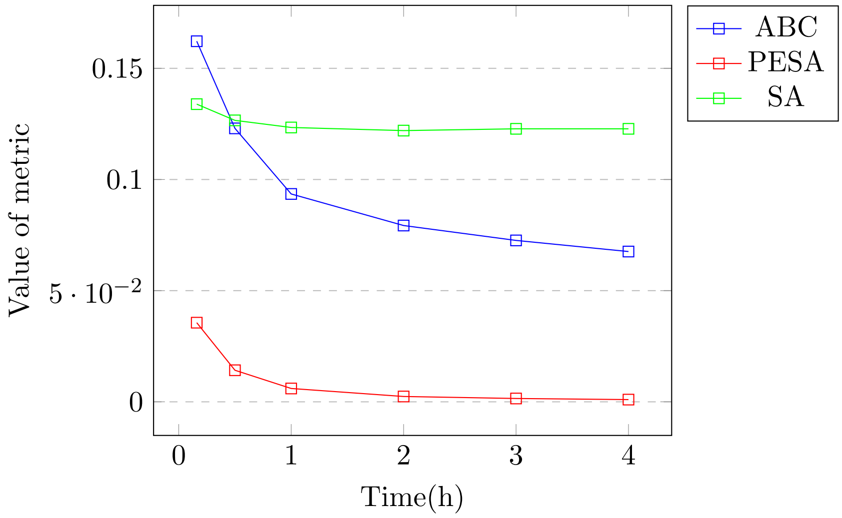


FIGURE 5.23: *A line graph showing the averages of the Diversity metric over time for the three algorithms used in this study.*

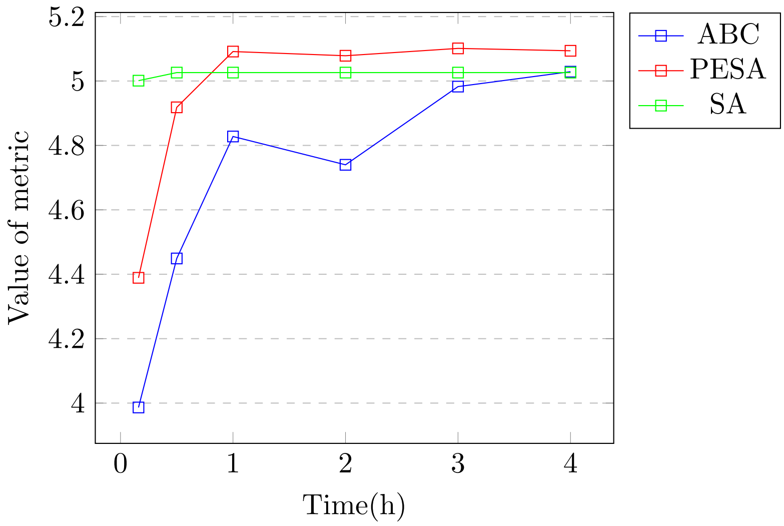


FIGURE 5.24: *A line graph showing the averages of the Spacing metric over time for the three algorithms used in this study.*

including that of the 20 hour run. However, in the cases where any improvement is found, it is found within the first two hours, leading to the same conclusion of the PESA, that the algorithm may be terminated after two hours.

5.7 Parameter calibration and sensitivity analysis

The parameter calibration section provides a comparison of the results of other configurations to the parameters discussed in §4.3.5, §4.4.5, and §4.5.5 as compared to the results gained from configuration used in §4.3.1 to 4.3.2.

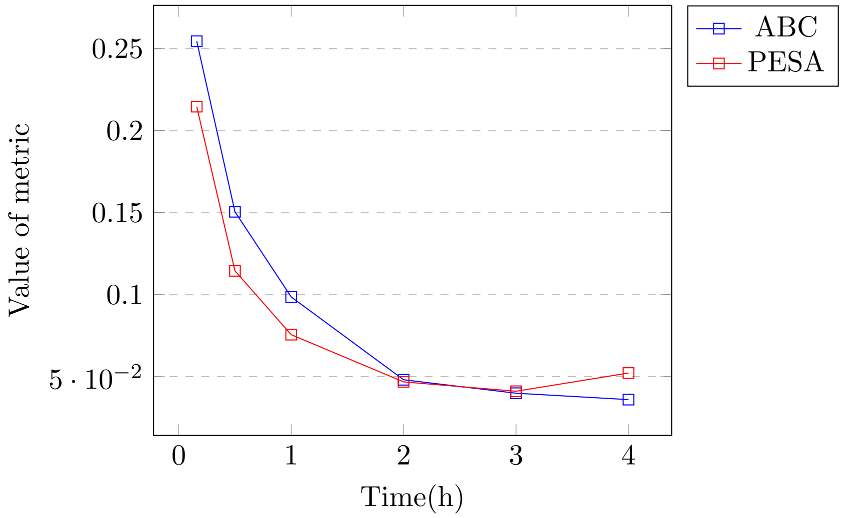


FIGURE 5.25: *A line graph showing the averages of the Frontier metric over time for the three algorithms used in this study.*

5.7.1 Parameter calibration for the ABC algorithm

In Figure 5.26, the effects of changing the hive size, and therefore the number of employed and onlooker bees, has on the solution quality at both the two hour mark and the four hour mark. These two marks are chosen as they were the chosen stopping time for a proposed decision support tool and the preliminary stopping time chosen for this study, respectively. What may be seen from the two graphs in Figure 5.26, is that for the frontier metric, the metric results are relatively low at both the two and four hour mark for lower values of the hive, however, the mid range values, approximately 5000 bees, are better at the four hour mark, but worse at the two hour mark. The metric results then tend to get worse as the hive size increase from this mid range set of values. The reason for the poorer solution quality as time progresses is due to the fact that too many bees are searching around the same solutions per iteration, slowing down the exploration of the algorithm.

This shows that if one uses the two hours as the stopping criteria as per a decision support tool, it may be more useful to use a small hive size, but if running the algorithm for longer periods of time, to stay at mid ranged hive sizes. The decision to reduce the hive size is corroborated by the diversity and spacing metrics as they show improved results at smaller hives. In fact, the diversity and spacing metrics contradict the frontier metric, as they show that the mid range hive size perform the worst, while the frontier metric shows that the mid ranges perform the best.

In Figure 5.27, the effect that size of the hive has on the time per iteration is seen. This is an important result as it affects the control the user will have in a decision support tool. What may be seen is that the relationship between hive size and iteration time is approximately linear, meaning that the amount of time needed per iteration for a given hive size should be relatively easy to determine. It was found that the larger hive sizes reach average iteration times of approximately one hour, which means that the algorithm will attempt to terminate in multiples of an hour. While this does not conflict with the proposed time in this study, if the results had shown that a fraction of an hour was a better termination time, the iteration time would cause an issue.

In Figure 5.28, one may see the effects of the limit given to the algorithm for how long it may

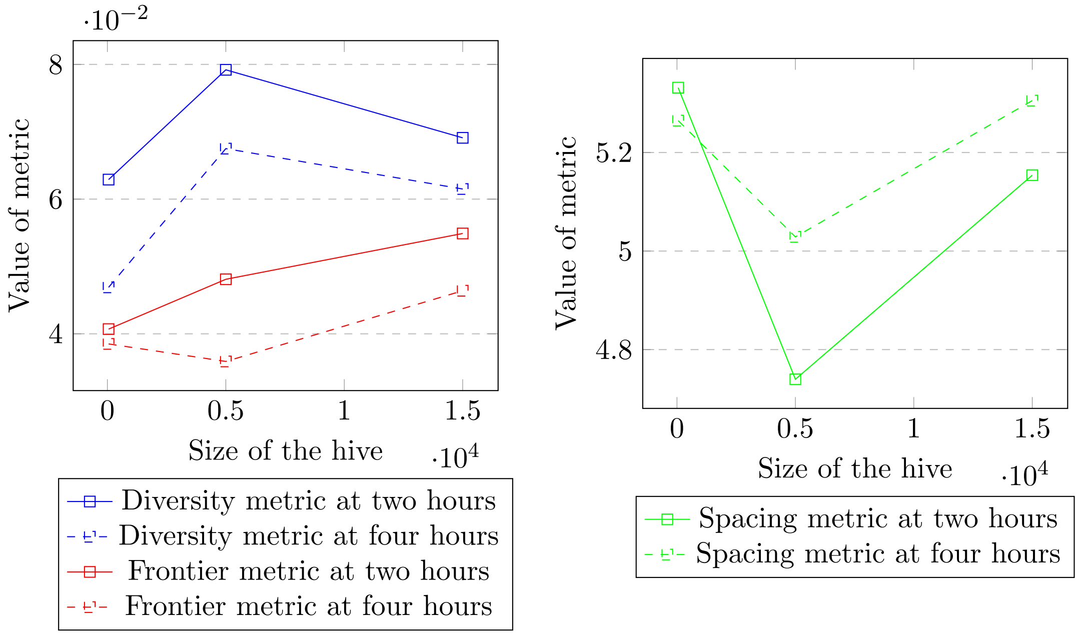


FIGURE 5.26: *A line graph showing the results of the parameter calibration of the size of the ABC hive with respect to its effect on the three solution metrics.*

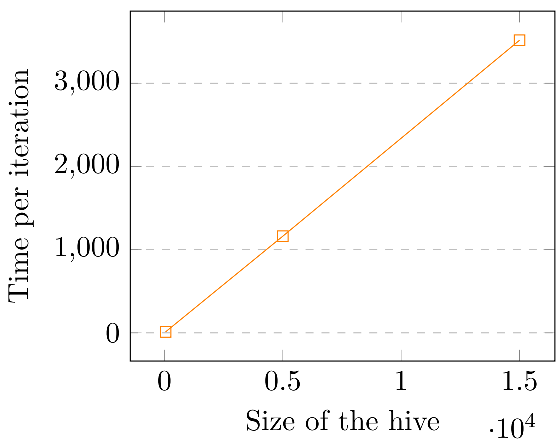


FIGURE 5.27: *A line graph shwoing the analysis of the time per iteration for different hive sizes for the ABC algorithm.*

continue to explore a given solution before needing to abandon it in favour of a new random solution. This limit controls the exploration and exploitation of the algorithm. The left hand side graphs shows that mid range values are better for all stages of the algorithm, while the spacing metric shows a small limit may be able to determine a better spread throughout the solution space after 4 hours, however, at two hours it shows that larger sizes will provide better results. This would be due to the fact that smaller limit values favour exploration and larger values favour exploitation, therefore the larger sizes may not have changed position significantly at the shorter time period. Due to the fact that mid range values seem to give a better combination of both exploration and exploitation, if one assumes that the frontier metric is more important than the spacing metric, using mid range values will be better for the algorithm configuration.

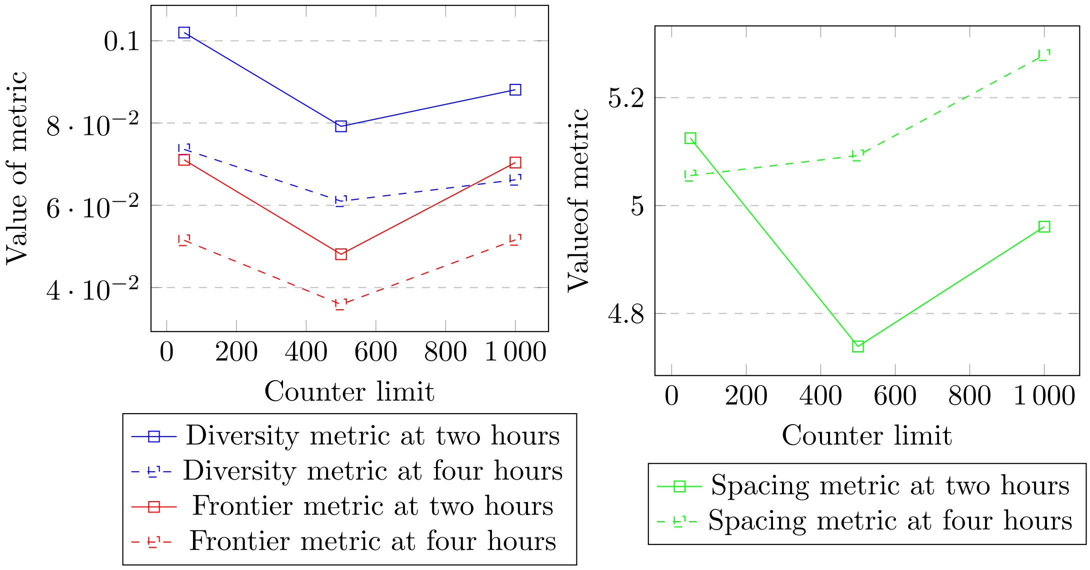


FIGURE 5.28: *A line graph showing the results of the parameter calibration of the non-improvement counter limit of the ABC with respect to its effect on the three solution metrics.*

5.7.2 Parameter calibration for the PESA

In Figure 5.29, the effect of different internal population sizes on the solution metrics for the PESA are shown. In Figure 5.29, it may be seen that the PESA is relatively insensitive to changes in the population size between the smaller and mid ranges, however the solution quality does start to decline when the population becomes significantly higher. This would be due to the fact that in the early stages the archive set of non-dominated solutions may be small and the internal population may spend time crossing the same parents over. This would make the large internal population redundant and reduce solution quality over time.

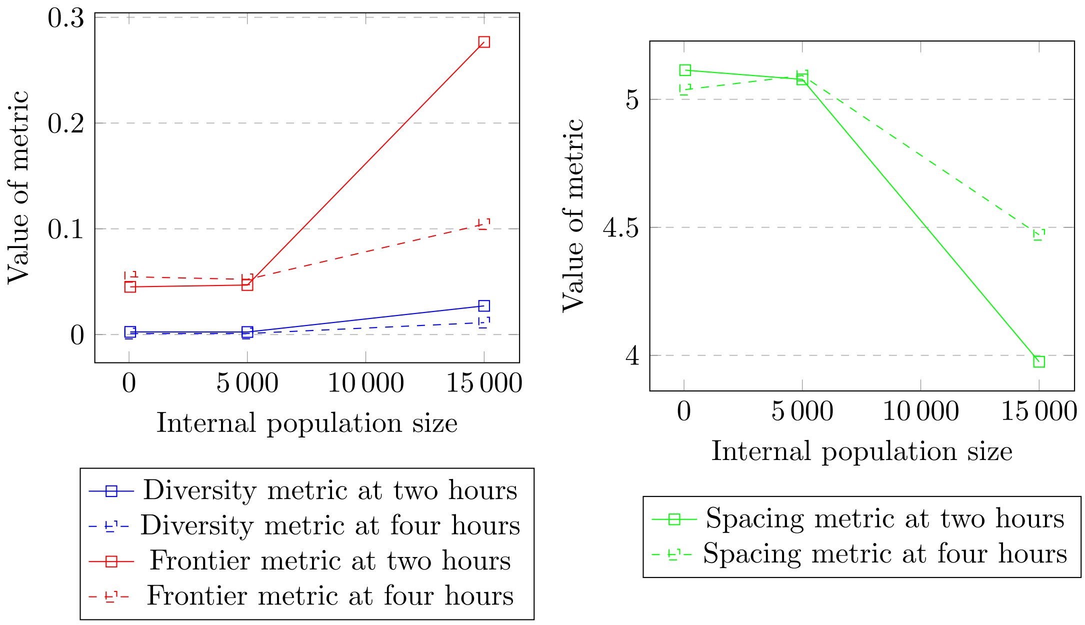


FIGURE 5.29: *A line graph showing the results of the parameter calibration of the PESA internal population size with respect to its effect on the three solution metrics.*

In Figure 5.30, we see the effect that the size of the internal population has on the time per iteration. Once again, this is an important result as it affects the control the user may have in a decision support tool. The relationship is approximately linear, similar to the results for the ABC in Figure 5.27. The larger hive sizes reach average iteration times of approximately 25 minutes, which means the algorithm will terminate in multiples of 25 minutes. This will affect a decision support tool as it will give very fixed times for the termination criteria which is not a positive trait for any possible software that could be designed to allow an investor to use the algorithms.

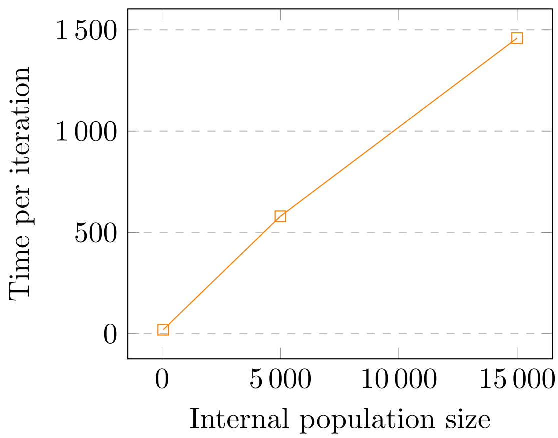


FIGURE 5.30: *A line graph showing the analysis of the time per iteration for different internal population sizes for the PESA.*

In Figure 5.31, the effects of the mutation rate may be seen. The clearest result is that the spacing metric becomes steadily worse as the mutation rate increases. This may be attributed to the fact that the mutations are considered simple local searches, and as seen in §5.5, the local search procedures alone are inadequate for finding solutions to the portfolio optimisation problem. In the frontier and diversity metrics, the mid range mutation rate is the best choice at both two and four hours, allowing the conclusion that this configuration should be used. It is worth noting that in the frontier metric, some mutations were necessary to gain a better solution, by diversifying the solution pool a little, but as the mutation value increased the local search problem of §5.5 once again worsened the solution quality.

5.7.3 Parameter calibration for the SA algorithm

The effect of the temperature parameter is shown in Figure 5.32. It may be seen that the frontier metric is relatively insensitive to changes in the initial temperature. This is especially true between the mid range and higher temperatures. This could be due to the resetting of the solution to its best known solution after becoming stuck at a local minimum [48]. As a result, we rely on the other two metrics to make a decision on the more suitable configuration for the SA algorithm. Viewing the spacing metric, it may be seen that the mid range and higher temperatures outperform the lower temperatures, with a preference to choosing the mid range temperatures. The diversity metric has the opposite result, where the lower temperatures give better results, with the solution quality worsening as the temperature rises. This is due to the higher temperatures allowing the solutions to explore more of the solution space, therefore being more spread out and not as close together. Given the better frontier and spacing metrics, the mid range temperatures seem to be the better fit.

In Figure 5.33, we see the effects of the limit given to the algorithm which will control how

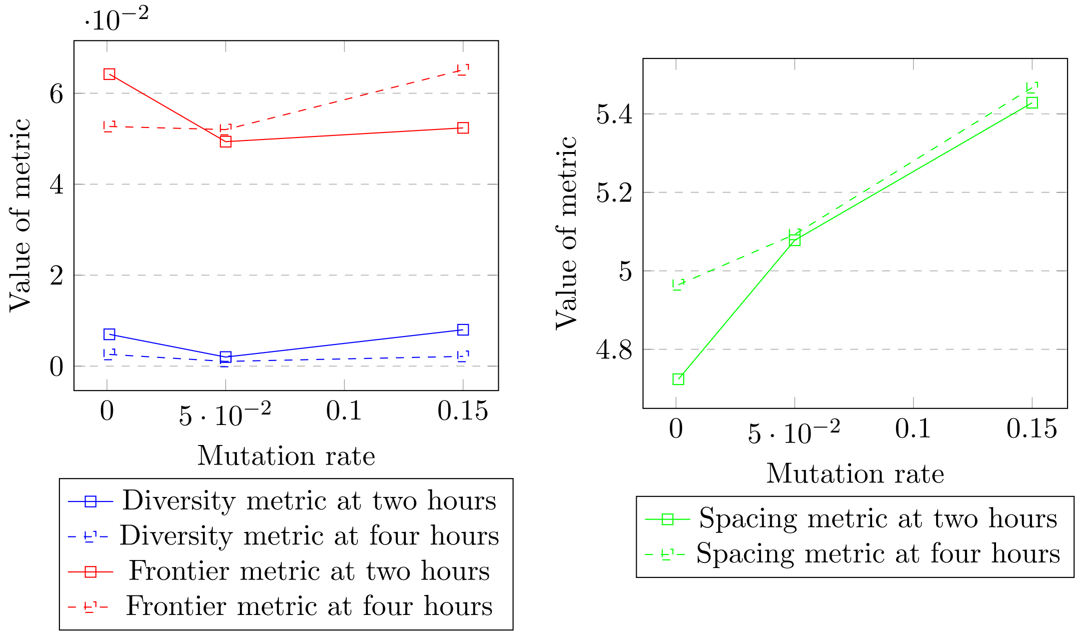


FIGURE 5.31: *A line graph showing the results of the parameter calibration of the PESA mutation rate with respect to its effect on the three solution metrics.*

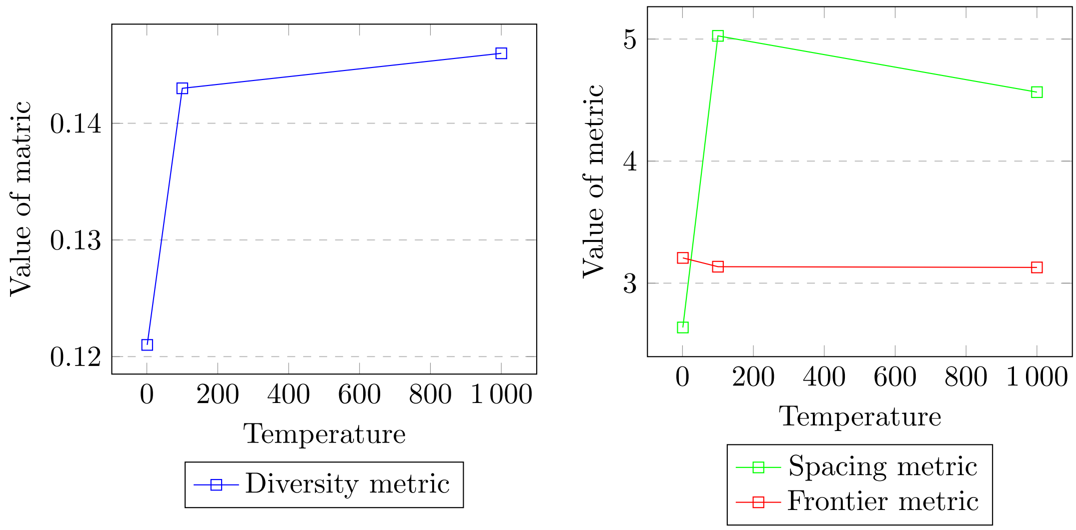


FIGURE 5.32: *A line graph showing the results of the parameter calibration of the SA temperature control with respect to its effect on the three solution metrics.*

long the algorithm stay at a local minimum until the temperature is reheated to the initial temperature and the solution reset to the best solution. Once again, the frontier metric is relative insensitive to any changes, therefore the other metrics are used to decide which configuration to use. By viewing the graphs, it may be seen that the best spacing value belongs to the smaller limit and the best diversity belongs to the higher limits. The better spacing metric can be attributed to the smaller limits limiting the time spent at local minima so allowing the algorithm to explore more solutions. This could indicate that the solution space contains many local minima. The same rationale results in the opposite effect for the diversity metric, as when the algorithm spends more time at the local minima and not exploring the solutions will be located closer to each other. While using smaller limits seems to be the better configuration, it may not be prudent to make the assumption of a local minimum too quickly, so a better choice may be to use a value between small and mid range values.

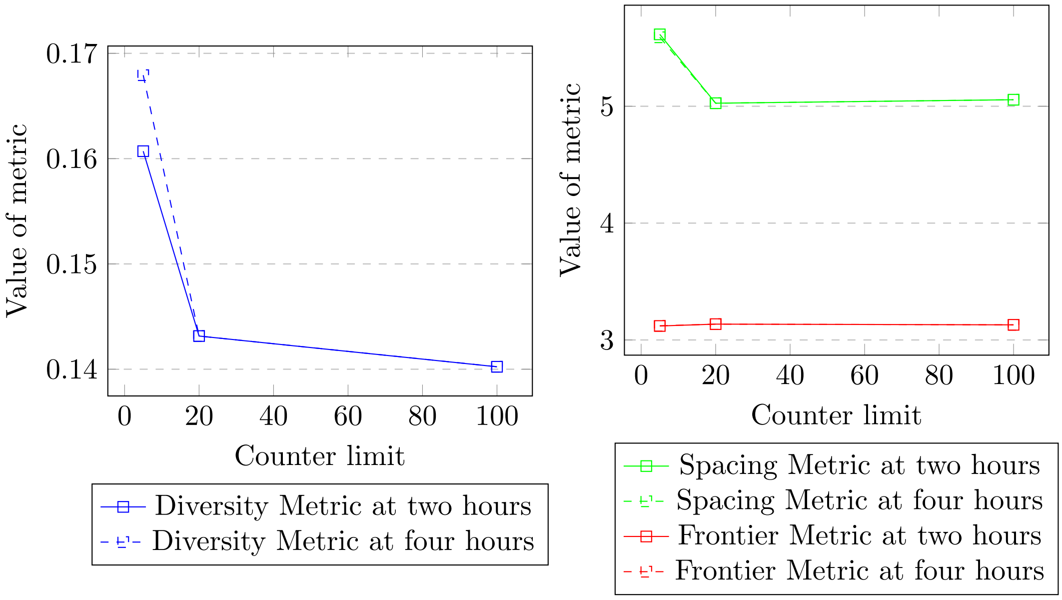


FIGURE 5.33: *A line graph showing the results of the parameter calibration of the SA counter towards non-acceptance with respect to its effect on the three solution metrics.*

In Figure 5.34, the effects of the cooling schedule parameter may be seen. The frontier metric is once again relatively insensitive to changes in the cooling parameter, however, the metric does seem to marginally worsen as the parameter moves towards one. The other two metrics, however, point towards using a mid range cooling parameter to balance the exploration and exploitation of the algorithm. This is attributed to the cooling parameter determining how quickly SA will become a hill climbing search or conversely how long it will spend exploring worse solutions before becoming stricter on the acceptance function.

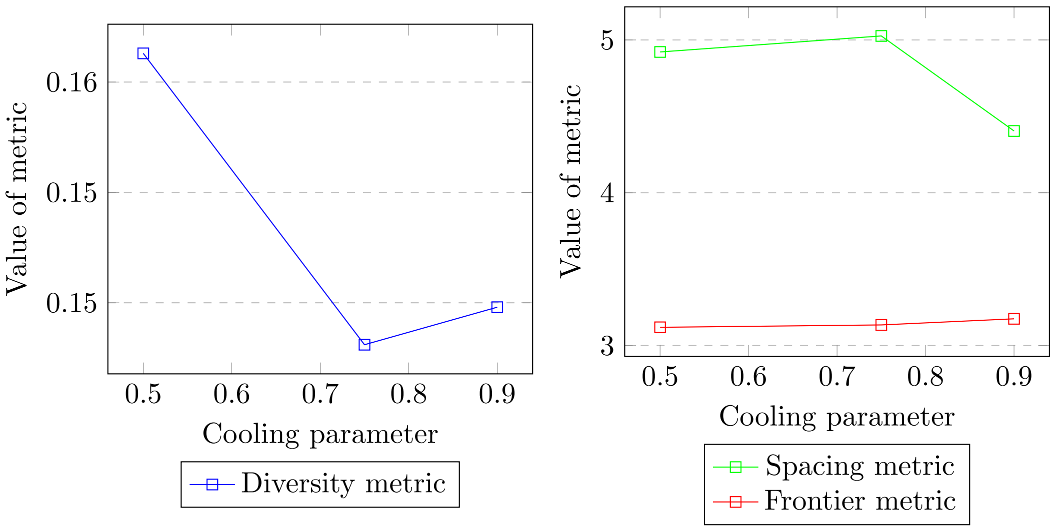


FIGURE 5.34: *A line graph showing the results of the parameter calibration of the SA cooling schedule with respect to its effect on the three solution metrics.*

5.7.4 Sensitivity analysis on the dataset size

In Figure 5.35 the effect of different data sizes on the convergence of the PESA is given as an example of how the size of the data set may affect algorithm performance. What may be seen is that when the frontier and spacing metrics are observed, the assumption of a relatively linear relationship between the size of the dataset and the time to convergence may be made. The diversity metric, however, does not seem to follow any relationship between the two other than stationary, with occasional outliers. This shows that with a larger dataset than the current one, more time may be needed before termination.

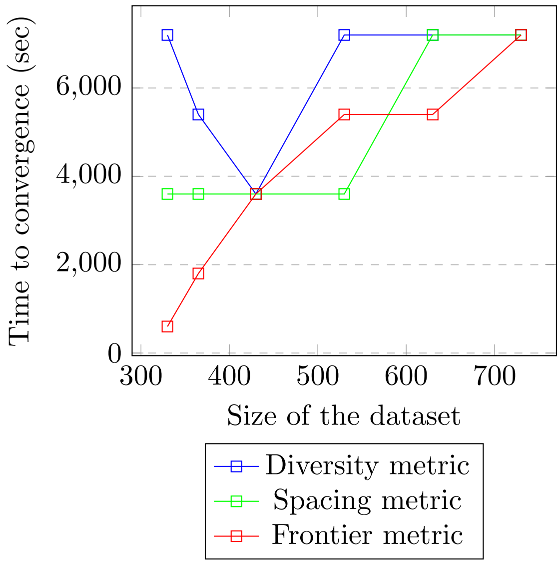


Figure 5.35: *A line graph showing the results of the sensitivity result on the size of the dataset with respect to the convergence times for the solution metrics.*

5.8 Typical portfolio composition

An interesting result to explore is the composition of the funds that the algorithms determine. This will determine whether funds that have a high rating in one objective are included in the efficient portfolios, or whether or not the idea of optimising multiple objectives has a significant effect on the inclusion of certain funds thereby reinforcing the adage of "a jack of all trades is a master of none, but often times better than a master of one". Tables 5.7 and 5.6 show the funds that represent possible portfolios for low, medium, and high risk investors from the frontiers developed by the ABC and PESA. The different risk profiles are shown by the different values of λ. In Table 5.6, the PESA results show that the PESA over diversifies the portfolio, with many small investments, $< 1\%$, being made. Therefore only the significant investments are shown in the table.

Of the 12 unique funds in the tables, 2 of the listed funds are located in the top 10 tables from Chapter 3. This may indicate that while investing in funds with high single objective values is of worth, there is still value in using an optimisation tool to find combinations of funds that may not seem obvious at first. This is more evident when observing the investment values of the funds from the top 10 tables, the Allan Grey Equity Fund and the PSG Wealth Enhanced Interest Fund. The values actually invested in these funds are relatively low, not exceeding 25%, showing that even when included, funds with a high single objective may not be the bulk of the

Pareto Envelop-based Selection Algorithm			
Unit trust name	$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 0.75$
Nedgroup ABIL Retention Fund	4%	4%	11.37%
Tantalum BCI Strategic Income Fund	9.4%	9.4%	17.39%
PSG Wealth Enhanced Interest Fund	13.48%	13.48%	24.83
Seed Stable Fund	4.3%	4.3%	5.9%
Satrix Bond Index Fund	1%	1%	–
Momentum Enhanced Yield Fund	50%	50%	–
Rezco Value Trend Fund	–	–	1.49%
ABSA - Protected Accumulator	–	–	10.08%
Other $w_i < 1\%$	17.82%	17.82%	28.94%

TABLE 5.6: *A view of the typical composition of a portfolio determined by the PESA for different investor goals.*

Artificial Bee Colony			
Unit trust name	$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 0.75$
Allan Grey Equity Fund	2.9%	2.9%	1.1%
Emperor IP Momentum Equity Fund	4.5%	4.5%	20.9%
Sanlam Namibia Floating Rate Fund	92.5%	92.5%	–
Evolve BCI Conservative Fund	–	–	77.8%

TABLE 5.7: *A view of the typical composition of a portfolio determined by the ABC for different investor goals.*

investment.

If one then compares the example composition of the ABC versus that of the PESA, three observations can be made. The first is that the the PESA contains a larger number of funds in the portfolio when compared to the ABC, which would mean the PESA develops more diversified portfolios. The second is that none of the funds are shared between the two portfolios, i.e. that there is no fund that is in both algorithm's high-, medium-, and low-risk portfolios. A possible explanation of the difference in fund composition is the fact that the two algorithms are occupying different areas of the solution space, as seen in Figures 5.1 to 5.12. Finally, there is a large proportion of the funds in the PESA composition that have a weight of below 1% of the total. This creates an issue as it is difficult for an investor to invest less than 1% of their investment in a fund, but the sum of these insignificant weights is a relatively large proportion of the total. This issue comes as a result of the increased diversification of the PESA and can be fixed by creating a cardinality constraint in the model.

5.9 Shrinkage theory

To demonstrate the effects of the shrinkage theory, two sets of results are given. The first, shown in Table 5.12, is an excerpt from the covariance matrix gained through the steps presented in §4.1, and the second is that of an excerpt of the sample covariance for the dataset, shown in Table 5.13. If one observes the corresponding values in Tables 5.12 and 5.13, it may be seen that, on average, the larger covariances in the sample covariance matrix has been made smaller and visa versa for the smaller ones, which is in line with what the literature, presented in §4.2, suggests should happen.

The second result is the real growth gained from May 2012 to April 2017 for a medium risk, or balanced, portfolio and a high risk portfolio using the shrinkage theory estimate tested against the same types of portfolios gained using the sample covariance matrix (SCM). To determine the real growth, the PESA is run using the data from May 1990 until April 2012 and the resulting portfolios are tracked to see what the growth would have been at the end of the original dataset time period, April 2017. The portfolios are described in Tables 5.8 to 5.10. Due to the over diversification of the PESA, as mentioned in §5.8, the insignificant funds, i.e. with $< 1\%$ investment amounts are excluded and the weight redistributed equally to the significant funds.

Shrinkage Estimate Portfolio $\lambda = 0.5$	
Fund name	**Percentage**
Old Mutual Global Emerging Market Fund	35.3%
Old Mutual Premium Equity Fund	45.8%
BCI Bst Blend Worldwide Flexible Fund	18.8%

Table 5.8: *A view of the composition of a portfolio gained using the shrinkage theory estimate and $\lambda = 0.5$.*

Shrinkage Estimate Portfolio $\lambda = 0.75$	
Fund name	**Percentage**
Prescient RECM Global Fund	1.6%
BCI Bst Blend Worldwide Flexible Fund	92.4%
Eastspring Investments Balanced Fund	5.9%

Table 5.9: *A view of the composition of a portfolio gained using the shrinkage theory estimate and $\lambda = 0.75$.*

SCM Portfolio $\lambda = 0.5$ and $\lambda = 0.75$	
Fund name	**Percentage**
Evolve BCI Managed Fund	62.2%
Investec Equity	37.7%

Table 5.10: *A view of the composition of a portfolio gained using the sample covariance matrix for both λ values.*

The results of the portfolios are summarised in Table 5.11. From the results it may be seen that the best performing portfolio was the one gained from using the shrinkage estimate with $\lambda = 0.5$, far outperforming the same λ when using the sample covariance. This may present some evidence that using the shrinkage theory estimate may provide a better combination of unit trusts. This is despite the poor performing shrinkage estimate with $\lambda = 0.75$, as this is a possibility that comes with trying to invest in a portfolio that is designed to be quite risky. This result is not conclusive as the algorithms do provide many alternative portfolios that can be tested to provide a bigger sample size for testing to see if the shrinkage theory estimate does continue to provide the better portfolios.

Portfolio	Total growth value	Ranking
Shrinkage estimate $\lambda = 0.75$	-8.79%	3
Shrinkage estimate $\lambda = 0.5$	57.91%	1
Sample covariance matrix	33.67%	2

TABLE 5.11: *A summary of the real growth of the portfolios determined by using the different covariances matrices.*

5.10 Chapter summary

The results chapter begins by presenting a description of the computer and specifications used to gain the results in this chapter. A brief explanation of the methodology used to determine the representative points of the true efficient frontier is then given, before presenting the individual results of each algorithm. The results presented for each algorithm are the graphs of a typical run of the algorithm along with a discussion of the graphs and why the algorithm may present the results shown. Following this, the comparison of the algorithms results are given as tables containing the values of the solution metric from §4.4, tables showing the results of the student t-tests performed on the metrics to determine if the algorithm's performances differ, and a comparison of the metrics over time. This is followed by a sensitivity analysis on the parameters of the algorithms. The typical portfolios for different investor types determined by the algorithms are then presented, before giving the results of the shrinkage theory compared to that of using the sample covariance matrix.

	LIOOFGE	ELEBALA	CORGLOB	PRPCFA1	OMGEMKA	ABXPRA1	RMBAFRA	SIMINDR	PRPCMA1	CORCGRO	NGCIARF	ANBGEQA	COROPTG
LIOOFGE	2.997	1.026	1.969	1.913	1.423	0.649	-0.847	3.061	0.987	3.017	0.271	0.159	1.785
ELEBALA	1.026	0.984	1.244	1.643	1.444	0.792	-0.325	1.902	1.078	1.823	0.290	0.295	1.365
CORGLOB	1.969	1.244	4.060	2.991	3.001	1.506	-0.243	3.988	1.878	3.771	0.531	0.725	3.743
PRPCFA1	1.913	1.643	2.991	4.442	3.422	2.037	-0.952	4.717	2.739	4.490	0.732	0.780	3.324
OMGEMKA	1.423	1.444	3.001	3.422	3.621	1.826	-0.582	4.012	2.315	3.790	0.651	0.784	3.149
ABXPRA1	0.649	0.792	1.506	2.037	1.826	1.429	-0.428	2.504	1.438	2.336	0.485	0.596	1.850
RMBAFRA	-0.847	-0.325	-0.243	-0.952	-0.582	-0.428	9.048	-1.331	-0.542	-1.527	-0.235	0.178	-0.136
SIMINDR	3.061	1.902	3.988	4.717	4.012	2.504	-1.331	7.784	3.014	7.182	0.949	0.892	4.738
PRPCMA1	0.987	1.078	1.878	2.739	2.315	1.438	-0.542	3.014	1.993	2.861	0.510	0.591	2.200
CORCGRO	3.017	1.823	3.771	4.490	3.790	2.336	-1.527	7.182	2.861	8.828	0.889	0.820	5.239
NGCIARF	0.271	0.290	0.531	0.732	0.651	0.485	-0.235	0.949	0.510	0.889	0.352	0.166	0.650
ANBGEQA	0.159	0.295	0.725	0.780	0.784	0.596	0.178	0.892	0.591	0.820	0.166	0.541	0.881
COROPTG	1.785	1.365	3.743	3.324	3.149	1.850	-0.136	4.738	2.200	5.239	0.650	0.881	6.368

TABLE 5.12: *An excerpt from the shrinkage theory covariance matrix.*

	LIOOFGE	ELEBALA	CORGLOB	PRPCFA1	OMGEMKA	ABXPRA1	RMBAFRA	SIMINDR	PRPCMA1	CORCGRO	NGCIARF	ANBGEQA	COROPTG
LIOOFGE	3.926	0.952	1.748	1.105	0.688	-0.044	-0.784	2.424	0.367	2.498	0.037	-0.127	1.193
ELEBALA	0.952	0.804	0.567	0.600	0.555	0.066	0.127	0.503	0.362	0.531	0.016	0.069	0.441
CORGLOB	1.748	0.567	3.811	1.254	1.763	0.366	0.709	1.907	0.676	1.843	0.084	0.508	2.843
PRPCFA1	1.105	0.600	1.254	1.981	1.132	0.308	0.151	1.214	1.020	1.226	0.064	0.263	1.019
OMGEMKA	0.688	0.555	1.763	1.132	1.926	0.336	0.562	0.973	0.760	0.950	0.073	0.408	1.325
ABXPRA1	-0.044	0.066	0.366	0.308	0.336	0.550	0.333	0.283	0.276	0.236	0.123	0.383	0.441
RMBAFRA	-0.784	0.127	0.709	0.151	0.562	0.333	15.730	-0.019	0.276	-0.534	-0.011	0.691	1.117
SIMINDR	2.424	0.503	1.907	1.214	0.973	0.283	-0.019	3.755	0.575	3.401	0.135	0.170	1.879
PRPCMA1	0.367	0.362	0.676	1.020	0.760	0.276	0.276	0.575	0.883	0.589	0.056	0.278	0.630
CORCGRO	2.498	0.531	1.843	1.226	0.950	0.236	-0.534	3.401	0.589	4.680	0.121	0.128	2.256
NGCIARF	0.037	0.016	0.084	0.064	0.073	0.123	-0.011	0.135	0.056	0.121	0.360	0.043	0.095
ANBGEQA	-0.127	0.069	0.508	0.263	0.408	0.383	0.691	0.170	0.278	0.128	0.043	0.729	0.602
COROPTG	1.193	0.441	2.843	1.019	1.325	0.441	1.117	1.879	0.630	2.256	0.095	0.602	3.921

TABLE 5.13: *An excerpt from the sample covariance matrix.*

CHAPTER 6

Conclusion

Contents

A summary of the work contained in this thesis is presented in §6.1 as well as indications on how each of the objectives for this study as given in Chapter 1 was fulfilled. A summary and a discussion of the results obtained in Chapter 5 is presented in §6.2. Contributions made by the research are then presented in §6.3. The chapter concludes with recommendations for possible future work and research in §6.4.

6.1 Thesis summary

In Chapter 1, the portfolio optimisation problem is introduced and the importance of diversifying the investments made by an investor is highlighted. Two challenges arise from a diversification strategy, namely that of struggling to purchase a fully diversified portfolio from the beginning of the investment period and that choosing a selection of financial assets from the multitude of options is not a trivial task. Unit trusts are given as a possible solution to the first problem. The history of unit trusts are explored with the major event points of the financial sector being highlighted and the different types of unit trusts are also discussed. The history of the portfolio optimisation problem is then discussed; highlighting the contribution of the major mathematical models developed to quantify financial risk before discussing the computational problems that are associated with the portfolio optimisation problem. The scope and objectives are listed at the end of Chapter 1.

Chapter 2 contains a review (in completion of Objective I(i)) on the portfolio optimisation problem along with a discussion on the different quantifications of risk and the different solution approaches followed to solve the portfolio optimisation models formulated. Furthermore, a review on meta-heuristics is presented in fulfilment of Objective I(ii), beginning with a description and classification of different types of meta-heuristics, and then moving onto a discussion of the papers that use meta-heuristics to solve the portfolio optimisation problem (in fulfilment of Objective I(iii)).

A discussion and exploration of the data gathered and used in this project is provided in Chapter 3 (in fulfilment of Objective II(i)). It begins with an outline of the data, including but not limited to, where the data has been gained from and the classification of the data. Some descriptive statistics are then provided along with a few top 10 tables of the funds with the best expected growth, the lowest associated risk, and the best Sharpe or risk-to-return ratio. The chapter concludes with the outcomes of the normality tests performed on the dataset.

The methodologies adapted in this thesis are presented in Chapter 4. Firstly an adapted Markowitz model formulation is presented (in fulfilment of Objective III(ii)). The model uses an improved covariance estimate, as well as a constraint for short term growth. The explanation of the shrinkage theory estimate used to determine a usable covariance matrix is then provided (in fulfilment of Objective III(i)). This estimate overcomes the issue of the normality assumption that comes from using the Markowitz model and the high dimensionality problem that comes with financial data. Following this the three algorithms in this study; the ABC, the PESA, and SA; are described (in fulfilment Objective III(iii)). For each algorithm, an overview of the main steps to the algorithm are explained along with the algorithm's origins. This is followed by a description of the different phases of the algorithm as they have been used in this paper with a focus on how they pertain to the portfolio optimisation problem. Psuedo-code is also given for each algorithm.

The results of this project are provided in Chapter 5. The results from each algorithm are presented individually with graphs illustrating the efficient frontiers obtained at different simulation stopping criteria, along with explanations of why the results may present the way they do. The results of the three algorithms are then compared in fulfilment of Objective III(iv) using three metrics; a diversity metric to determine how evenly spaced the results are throughout the solution space, a spacing metric to determine how much of the solution space has been explored, and a frontier metric to determine how close the algorithm manges to get to a select set of points on the true frontier set. Finally, the composition of the portfolios gained are viewed in order to see how the end combinations compare to the data exploration findings in Chapter 3 and the results of applying shrinkage theory to the data are explored.

6.2 Summary and discussion of results

A summary of the main results of this study as well as their corresponding implications using the results in portfolio optimisation decision support are listed below:

1. Three algorithms; the ABC, the PESA, and SA; were applied and used to find efficient frontiers for the dataset. The modifications to the algorithms have been described in Chapter 4. Only the ABC and the PESA were able to find efficient solutions to the problem, thereby indicating the possibility of using them in a decision support tool.

2. T-tests were performed on the solution metrics to determine which of the algorithms are more effective than the others. The results showed that the ABC algorithm only outperforms the SA, and the PESA in turn outperforms the ABC and the SA making it the most effective for this study.

3. Analysis was done on the metric results over time in two regards. The first being to see what would happen if the algorithms were left to run for an extended time. These results showed that the SA and ABC algorithms start to stall and no longer move towards the efficient frontier and would require adapted search phases to help convergence. The second was to determine when the algorithm results start to slow and therefore can be considered

a stopping time for a user as the extra time would begin to result in improvements that are not worth the time required to find them. This result showed that at two hours the algorithms had reached a point where the improvements could be argued to no longer increase with the increase in time and may be concluded as a suitable stopping criteria for future use in a decision support tool.

4. An analysis on the parameters of each algorithm is presented. The analysis includes graphs showing the effects that the change of the parameter has on the overall solution quality, as well as a discussion on why the parameter has the effect it does and a conclusion on which configuration to be used. The configurations used in this study are summarised below.

Parameter	**Configuration**
Hive size	5000
First upper limit on non-improvement	500
Second upper limit on non-improvement	10

TABLE 6.1: *A summary of the parameters of the ABC algorithm used in this study and the configurations thereof.*

Parameter	**Configuration**
Internal population size	5000
Mutation rate	0.05
Crossover rate	0.95

TABLE 6.2: *A summary of the parameters of the PESA used in this study and the configurations thereof.*

Parameter	**Configuration**
Initial temperature	100
Cooling coefficient	0.75
Upper limit on non-acceptance of solutions	20

TABLE 6.3: *A summary of the parameters of the SA algorithm used in this study and the configurations thereof.*

5. The portfolio compositions were also explored. It was found that some of the funds with better results in one characteristic are included in the portfolios, however they were not always the majority of the portfolio investment, showing that there is value in the use of a optimising tool. In terms of the find composition of the individual algorithms, the only significant results was that the PESA over-diversifies the portfolio and should be modified with a cardinality constrained model.

6. Shrinkage theory was successfully applied to the dataset to calculate an improved covariance estimate to be used in the Markowitz model, thereby overcoming the issues of non-normality and high dimensionality in the data. Most articles ignore the incorrect normality assumption, or use a different measure of risk to improve the incorrect covariance estimate. What was found was that the medium risk portfolio obtained with the new estimate outperformed the portfolios determined using the traditional sample covariance matrix when using a 5 year hold out period from the original data set.

6.3 Contributions of this study to the field

The following list summarises the possible contributions made by this study, to the best of the author's knowledge:

1. *Solving an adjusted Markowitz model with a covariance estimate gained from shrinkage theory using a selection of meta-heuristics.* The articles found by the author either dealt with replacing the sample covariance matrix with the shrinkage theory estimate, or solving the portfolio optimisation with meta-heuristics, but not both.

2. *A comparison of the efficacy of the PESA, ABC, and SA algorithms in solving the portfolio optimisation problem.* No direct comparison was found by the author.

3. *The initialisation processes of all algorithms were modified, with a focus on the unit trust with the highest expected return, and found to improve the final solution quality. The searching phases of the ABC and SA were modified to resemble Sharpe's corner portfolio model, and this was found to improve solution quality as well.* The changes made were not found in any other literature by the author.

6.4 Future work

Possible future work may include the following proposals (in fulfilment in Objective IV):

1. Expand the research by considering other heuristics and meta-heuristics that may be applied to the portfolio optimisation problem and tested against the results of this thesis. The Particle Swarm Optimisation algorithm is also an attractive option for testing within the swarm intelligence set of algorithms as the articles that deal with it have positive results in terms of its use in solving multi-objective optimisation and the portfolio optimisation problem in specific [64, 59, 19, 34]. Also, as mentioned in §2.1.3, the Successive Regression Algorithm for solving stochastic programming problems should be explored as a possible method for creating efficient portfolios.

2. Expand the research by modifying all three algorithms. The SA algorithm can be adjusted from a single search based local optimisation to a population based algorithm with evolutionary strategies as presented by Mariner and Kellerer [49]. In their paper, multiple solutions are annealed simultaneously. In every iteration, elitism and mutation concepts from evolutionary algorithms are borrowed and performed on the solutions by allowing better solutions to have a chance of replacing the poorer solutions. This is either by replacing them entirely or by transforming the poorer solutions by an averaged solution of the weights of the best solutions overall as well as the best of the current solutions. The effects of the evolutionary concepts were found to improve the results of SA significantly. Another possible modification to the meta-heuristics are to attempt the modifications to the Artificial Bee Colony mentioned in §5.1 in order to see how the ABC algorithm will compare to the PESA once the convergence issue has been improved.

3. As this study only dealt with the Markowitz model, there is room for testing the algorithms in conjunction with one of the downside risk models. This will allow for testing to see if the algorithms find more effective portfolios outright when using the downside risk models. It will also allow to see if using the downside risk affects the time needed to find efficient portfolios.

4. The results from the shrinkage theory portfolios versus those of the sample covariance matrix (SCM) are not fully conclusive. This leaves room for more extensive testing on whether the shrinkage theory portfolios significantly outperform those of the SCM.

5. To bring a more real world result to the project, the algorithms can be tested against a selection of portfolios from investment companies that were being offered at a certain date in the past. This date will be used as the end point for the dataset and we construct portfolios from the algorithm at this date. The portfolios from both the investment companies and the algorithms can then be evaluated over time to see how the portfolios compare against each other in real growth and not hypothetical growth.

6. To further improve the real-world application, the model formulated can be expanded in two manners:

   (a) The objectives can be expanded by the addition of other concerns for investors such as:

   - Maximising liquidity. A concern if the funds invested in has too low of a liquidity characteristic is when one wishes to gain the return on the initial investment, the fund may struggle to find a buyer for the investment and struggle to reimburse the investor, adding to the inherent risk of investing.
   - Maximise the total investment rating. A score given to the fund by an investment rating board that will take into account various characteristics of the fund, such as; fund size, history of dividend payments from companies in the fund, the fund's style of investing, and whether the fund is well priced compared to its return [60], which may help to quantify qualitative aspects for inclusion in an optimisation model.
   - Minimise the expected costs associated with investment. The return of the funds can be affected by bringing into account the fees that are linked to each fund. Some funds with high returns may also come with higher than average cost structures which would result in the model picking lower expected returns due to higher costs. These fees include initial investment fees, monthly management fees, and in some rare cases, a fee for leaving the investment before a certain amount of time has passed.

   (b) The constraints of the model can be expanded by including cardinality constraints to limit the over diversification of funds that arises from some of the algorithms' operations.

7. One end product for the algorithms and recommendations included in this paper is to create a decision support tool to help the ordinary investor or a financial advisor gain recommendations as to what they should invest in or recommend to clients. For this tool to be effective, recommendations 3 and 4 would be helpful to allow the tool to compete with fund managers. In addition to these recommendations, the tool will require an evaluation of the user's risk profile through the use of something similar to an input questionnaire that gathers information like age, earning potential, long term goals, current investment assets, etc. This evaluation will allow the algorithm to select portfolio(s) that are more suited towards the user's goals, which would also mean that a sorting and selection technique should be developed to aid in selecting the correct portfolio(s) for the user. See Figure 6.1 for a flow chart of the proposed system.

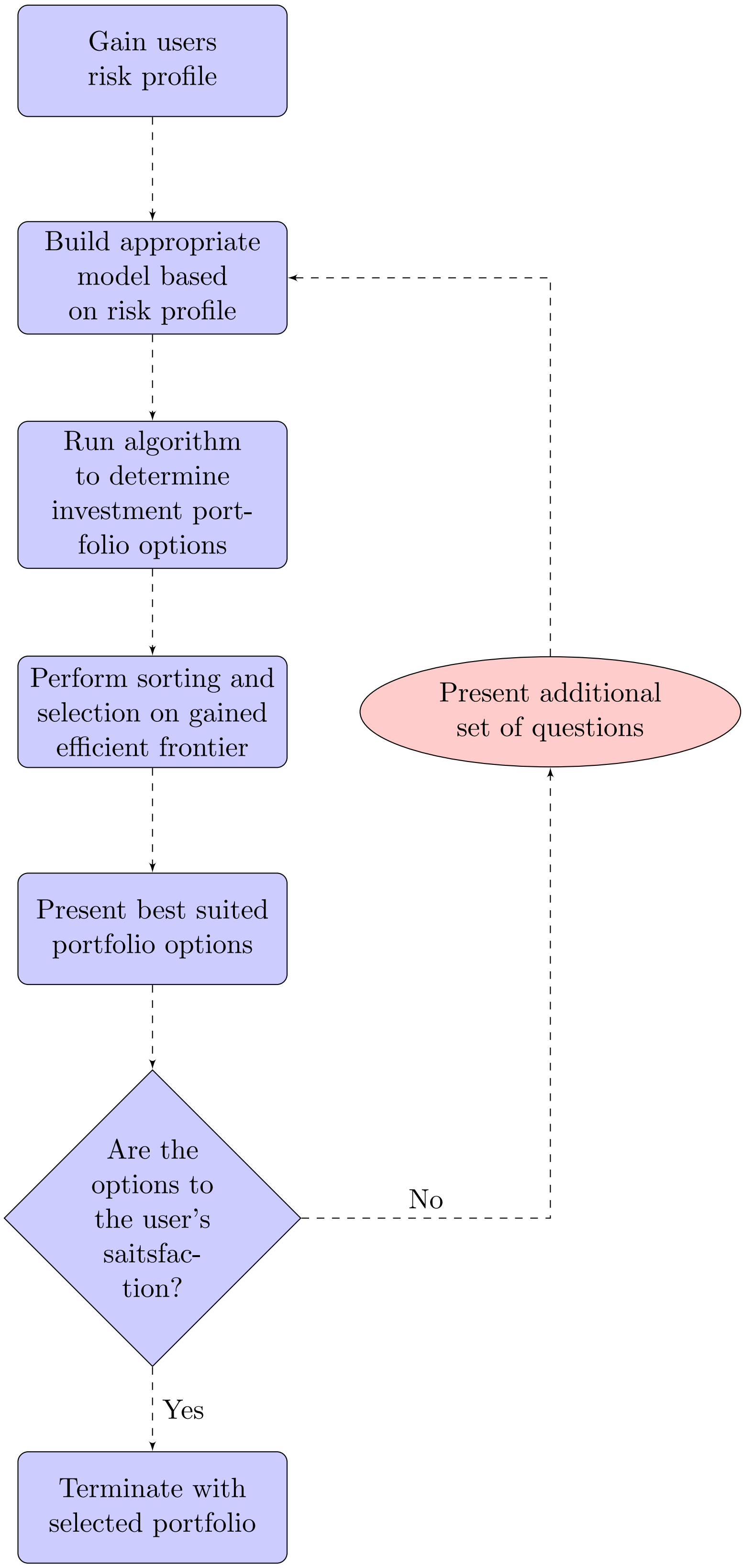


FIGURE 6.1: *A flowchart of the proposed decision support tool for ordinary investors outlining the steps to gain a final recommendation.*

References

[1] Ágoston KC, 2012, *CVaR minimization by the SRA algorithm*, Central European Journal of Operations Research, **20(4)**, pp. 623–632.

[2] Anderson TW & Darling DA, 1952, *Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes*, The annals of mathematical statistics, **23(2)**, pp. 193–212.

[3] Armananzas R & Lozano JA, 2005, *A multiobjective approach to the portfolio optimization problem*, Proceedings of the The 2005 IEEE Congress on Evolutionary Computation, Edinburgh, UK, pp. 1388–1395.

[4] Arntz A, 2013, *Fund manager at Prudential*, [Interview], Interview available in Profile's Unit Trusts & Collective Investments by Coronation Fund Managers, pg. 152-154.

[5] Bawa VS, 1975, *Optimal rules for ordering uncertain prospects*, Journal of Financial Economics, **2(1)**, pp. 95–121.

[6] Bernstein W, 2002, *The Four Pillars of Investing*, McGraw Hill Companies, Inc., New York, USA.

[7] Bertolis D & Hayes M, 2014, *An investigation into South African general equity unit trust performance during different economic periods*, South African Actuarial Journal, **14(1)**, pp. 73–99.

[8] Biger N & Page M, 1993, *Unit trusts' performance: does the yardstick matter*, Studies in Economics and Econometrics, **17(1)**, pp. 1–15.

[9] Bloomberg L.P , 2017, Historical price data for all South African Unit trusts, retrieved from Bloomberg Terminal March 2017.

[10] Blum C & Roli A, 2003, *Metaheuristics in combinatorial optimization: Overview and conceptual comparison*, ACM Computing Surveys (CSUR), **35(3)**, pp. 268–308.

[11] Bonabeau E, Dorigo M & Theraulaz G, 1999, *Swarm intelligence: from natural to artificial systems*, Oxford university press, New York, USA.

[12] Cable D & Groenewald H, 2013, *Fund managers at Coronation*, [Interview], Interview available in Profile's Unit Trusts & Collective Investments by Coronation Fund Managers, pg. 140-141.

[13] Campbell R, Huisman R & Koedijk K, 2001, *Optimal portfolio selection in a Value-at-Risk framework*, Journal of Banking & Finance, **25(9)**, pp. 1789–1804.

[14] Chen AH, Liang Y.-C & Liu C.-C, 2012, *An artificial bee colony algorithm for the cardinality-constrained portfolio optimization problems*, Proceedings of the 2012 IEEE Congress on Evolutionary Computation (CEC), Brisbane, Australia, pp. 1–8.

[15] Chen Y, Wiesel A & Hero AO, 2011, *Robust shrinkage estimation of high-dimensional covariance matrices*, IEEE Transactions on Signal Processing, **59(9)**, pp. 4097–4107.

[16] Corne DW, Knowles JD & Oates MJ, 2000, *The Pareto envelope-based selection algorithm for multiobjective optimization*, Proceedings of the Conference on Parallel Problem Solving from Nature, PPSN VI, Paris, France, pp. 839–848.

[17] Coronation (Ed), 2013, *Profile's Unit Trusts and Collective Investments*, 1st Edition.

[18] Cramér H, 1928, *On the composition of elementary errors*, Almqvist & Wiksells, Uppsala, Sweden.

[19] Cura T, 2009, *Particle swarm optimization approach to portfolio optimization*, Nonlinear Analysis: Real World Applications, **10(4)**, pp. 2396–2406.

[20] Daniel WW, 1978, *Applied nonparametric statistics*, Houghton Mifflin, Boston, USA.

[21] Deb K, Pratap A, Agarwal S & Meyarivan T, 2002, *A fast and elitist multiobjective genetic algorithm: NSGA-II*, IEEE Transactions on Evolutionary Computation, **6(2)**, pp. 182–197.

[22] Dioşan L, 2005, *A multi-objective evolutionary approach to the portfolio optimization problem*, Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, Vienna, Austria, pp. 183–187.

[23] Dixon WJ & Massey Frank J, 1950, *Introduction To Statistical Analsis*, McGraw-Hill Book Company, Inc; New York.

[24] Dorigo M, Birattari M & Stutzle T, 2006, *Ant colony optimization*, IEEE computational intelligence magazine, **1(4)**, pp. 28–39.

[25] Eberhart RC & Kennedy J, 1995, *A new optimizer using particle swarm theory*, Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, pp. 39–43.

[26] Feo TA & Resende MG, 1995, *Greedy randomized adaptive search procedures*, Journal of global optimization, **6(2)**, pp. 109–133.

[27] Foo T & Eng S, 2000, *Asset allocation in a downside risk framework*, Journal of Real Estate Portfolio Management, **6(3)**, pp. 213–223.

[28] freeInvestmentAdvice.org , 2017, *Investment Platforms in SA*, [Online], [Cited November 6th, 2017], Available from `http://www.freeinvestmentadvice.org/South-Africa/platforms/investment-platforms.php`.

[29] Ginsberg , 2014, *What is an Independent Financial Advisor?*, [Online], [Cited August 1st, 2017], Available from `http://www.gfb.co.za/what-is-an-independent-financial-advisor/`.

[30] Glover FW & Kochenberger GA, 2006, *Handbook of metaheuristics*, Springer Science & Business Media, Berlin, Germany.

[31] Glover F & Laguna M, 2013, *Tabu Search*, pp. 3261–3362 in , *Handbook of Combinatorial Optimization*, Springer, Berlin, Germany.

[32] Harlow WV, 1991, *Asset allocation in a downside-risk framework*, Financial analysts journal, **47(5)**, pp. 28–40.

[33] Holton GA, 2002, *History of Value-at-Risk*, Available from `https://pdfs.semanticscholar.org/edb7/883debc63b2789e3d1c40c95579775f8e9e1.pdf`.

[34] Hu X & Eberhart R, 2002, *Multiobjective optimization using dynamic neighborhood particle swarm optimization*, Proceedings of the Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on, Honolulu, USA, pp. 1677–1681.

[35] Huang X, 2008, *Portfolio selection with a new definition of risk*, European Journal of Operational Research, **186(1)**, pp. 351–357.

[36] Karaboga D, 2005, *An idea based on honey bee swarm for numerical optimization*, (Unpublished) Technical Report, Technical report-tr06, Erciyes university, engineering faculty, computer engineering department.

[37] Karaboga D & Basturk B, 2007, *A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm*, Journal of global optimization, **39(3)**, pp. 459–471.

[38] Kendall G & Su Y, 2005, *A Particle Swarm Optimisation Approach in the Construction of Optimal Risky Portfolios.*, Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, Innsbruck, Austria, pp. 140–145.

[39] Kirkpatrick S, Gelatt CD & Vecchi MP, "Optimization by simulated annealing", in: *Readings in Computer Vision*, Amsterdam, Netherlands: Elsevier, 1987, pp. 606–615.

[40] Knight E & Firer C, 1989, *The performance of South African unit trusts 1977–1986*, South African Journal of Economics, **57(1)**, pp. 37–47.

[41] Kruger C, 2016, *Financial advisor at Pentagon Solutions*, [Personal Communication], Contactable at calvink@pentag.com.

[42] Künzi-Bay A & Mayer J, 2006, *Computational aspects of minimizing conditional value-at-risk*, Computational Management Science, **3(1)**, pp. 3–27.

[43] Ledoit O & Wolf M, 2004, *A well-conditioned estimator for large-dimensional covariance matrices*, Journal of multivariate analysis, **88(2)**, pp. 365–411.

[44] Ledoit O & Wolf M, 2004, *Honey, I shrunk the sample covariance matrix*, The Journal of Portfolio Management, **30(4)**, pp. 110–119.

[45] Lindo Systems , 2008, *Lingo 11*, Linear programming solver.

[46] Lindo Systems , 2008, *Powerful LINGO Solvers*, 2017], Available from `http://www.lindo.com/index.php/products/lingo-and-optimization-modeling?catid=89&id=88:powerful-lingo-solvers`.

[47] Mansini R, Ogryczak W & Grazia Speranza M, 2003, *LP solvable models for portfolio optimization: A classification and computational comparison*, IMA Journal of Management Mathematics, **14(3)**, pp. 187–220.

[48] Maringer DG & Winker P, 2003, *Portfolio optimization under different risk constraints with modified memetic algorithms*, University Staatswiss.

[49] Maringer D & Kellerer H, 2003, *Optimization of cardinality constrained portfolios with a hybrid local search algorithm*, Or Spectrum, **25(4)**, pp. 481–495.

[50] Markowitz H, 1952, *Portfolio selection*, The journal of finance, **7(1)**, pp. 77–91.

[51] Markowitz HM, 1991, *Foundations of portfolio theory*, The journal of finance, **46(2)**, pp. 469–477.

[52] Meyer-Pretorius MC & Wolmarans HP, 2006, *The unit trust industry in South Africa from 1965 to June 2005: are investors better off?*, Meditari: Research Journal of the School of Accounting Sciences, **14(1)**, pp. 49–67.

[53] Michalowski W & Ogryczak W, 2001, *Extending the MAD portfolio optimization model to incorporate downside risk aversion*, Naval Research Logistics (NRL), **48(3)**, pp. 185–200.

[54] Middleton R, 2013, *Fund manager at Investec*, [Interview], Interview available in Profile's Unit Trusts & Collective Investments by Coronation Fund Managers, pg. 145-146.

[55] Mitchell M, 1998, *An introduction to genetic algorithms*, MIT press, Massachusetts, USA.

[56] MomentumGroup , 2013, *Fund manager at Momentum*, [Interview], Interview available in Profile's Unit Trusts & Collective Investments by Coronation Fund Managers, pg. 147-148.

[57] Moneyweb , 2013, *Does the size of your unit trust matter?*, [Online], [Cited May 29th, 2017], Available from `https://www.moneyweb.co.za/archive/does-the-size-of-your-unit-trust-matter/`.

[58] Moneyweb , 2016, *Independent financial advisors under pressure*, [Online], [Cited July 7th, 2017], Available from `https://www.moneyweb.co.za/mymoney/moneyweb-financial-planning/independent-financial-advisors-under-pressure/`.

[59] Moore J & Chapman R, 1999, *Application of particle swarm to multiobjective optimization*, Department of Computer Science and Software Engineering, Auburn University, **32**.

[60] Morningstar , 2011, *Morningstar Analyst Rating*, [Online], [Cited January 1st, 2017], Available from `http://media.morningstar.com/uk/dotcodotuk/Analyst_Research/analystratingforfundsfactsheet.pdf`.

[61] Moyle M, 2013, *Fund manager at Prudential*, [Interview], Interview available in Profile's Unit Trusts & Collective Investments by Coronation Fund Managers, pg. 154-155.

[62] Nofsinger JR & Varma A, 2007, *How analytical is your financial advisor?*, Financial Services Review, **16(4)**, pp. 245.

[63] Pollak I, 2011, *Weight shrinkage for portfolio optimization*, Proceedings of the 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), San Juan, PR, USA, pp. 37–40.

[64] Rabbani M, Bajestani MA & Khoshkhou GB, 2010, *A multi-objective particle swarm optimization for project selection problem*, Expert Systems with Applications, **37(1)**, pp. 315–321.

[65] Rom BM & Ferguson KW, 1994, *Post-modern portfolio theory comes of age*, The Journal of Investing, **3(3)**, pp. 11–17.

[66] SAS , 2016, *SAS Studio 3.7*, Software.

[67] Shaphiro S & Wilk M, 1965, *An analysis of variance test for normality*, Biometrika, **52(3)**, pp. 591–611.

[68] Sharpe WF, 1967, *A linear programming algorithm for mutual fund portfolio selection*, Management Science, **13(7)**, pp. 499–510.

[69] Sharpe WF, 1963, *A simplified model for portfolio analysis*, Management science, **9(2)**, pp. 277–293.

[70] Shi Y, 2001, *Particle swarm optimization: developments, applications and resources*, Proceedings of the Congress on Evolutionary Computation, Seoul, South Korea, pp. 81–86.

[71] Sokanu , 2017, *What is a Financial Advisor?*, [Online], [Cited November 6th, 2017], Available from `https://www.sokanu.com/careers/financial-advisor/`.

[72] Toplis GJ, 2016, *Portfolio selection from the JSE Top 40 using modern portfolio theory*, Honours, Stellenbosch University.

[73] Tuba M & Bacanin N, 2014, *Artificial bee colony algorithm hybridized with firefly algorithm for cardinality constrained mean-variance portfolio selection problem*, Appl. Math. Inf. Sci, **8(6)**, pp. 2831–2844.

[74] Tuba M & Bacanin N, 2014, *Artificial bee colony algorithm hybridized with firefly algorithm for cardinality constrained mean-variance portfolio selection problem*, Applied Mathematics & Information Sciences, **8(6)**, pp. 2831.

[75] van Taak N, 2017, *Analyst at Analytics*, [Personal Communication], Contactable at nadine@analytics.co.za.

[76] Voss S, 2000, *Meta-heuristics: The state of the art*, Proceedings of the Workshop on Local Search for Planning and Scheduling, Berlin, Germany, pp. 1–23.

[77] Wittig S, 2015, *Shrinkage Theory for Portfolio Optimization with Correlated Geometric Brownian Motion.*

[78] Yang L, Couillet R & McKay MR, 2014, *Minimum variance portfolio optimization with robust shrinkage covariance estimation*, Proceedings of the 48th Asilomar Conference on Signals, Systems and Computers, California, USA, pp. 1326–1330.

[79] Young MR, 1998, *A minimax portfolio selection rule with linear programming solution*, Management science, **44(5)**, pp. 673–683.