Improving our understanding of the equal weighted portfolio

by Byran H. Taljaard

Submitted in partial fulfilment of the requirements for the degree
Philosophiae Doctor

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Natural and Agricultural Sciences
Department of Actuarial Science

Supervisor: Professor Eben Maré
April 24, 2022

Byran H. Taljaard
Improving our understanding of the equal weighted portfolio
Philosophiae Doctor, April 24, 2022
Supervisor: Professor Eben Maré

University of Pretoria
Faculty of Natural and Agricultural Sciences
Department of Actuarial Science
Corner of Lynnwood Road and Roper Street, Hatfield
Pretoria
0083

Abstract

This thesis analyses the performance of the equal weighted portfolio using an approach from stochastic portfolio theory. This framework allows for the decomposition of the relative performance of the equal weighted portfolio into four main parts; the change in the concentration of the cap weighted portfolio, the excess return generated by a diversification benefit, the difference in dividend rates, and a term called the leakage effect. In general equal weighted portfolios do outperform their cap weighted portfolio counterparts, although with varying degrees across different countries. In South Africa, for example, high levels of leakage over the past ten years and increasing concentration have led to poor relative performance of the equal weighted portfolio. In other countries such as the United Kingdom and Japan, equal weighted portfolios have done very well, with high levels of diversification benefits and low levels of leakage. Two models are presented in an attempt to reduce the relative drawdowns of the equal weighted portfolio and to blend the two weights (equal and cap) in an optimal manner. These models appear to do well in markets where the equal weighted portfolio has poor performance and large relative drawdowns.

This page has been intentionally left blank

Acknowledgements

I would like to express my sincere gratitude to Prof. Eben Maré for all his motivation, support, and assistance without which this thesis would have almost surely not come together. It goes without saying that I could not have asked for a better supervisor.

Jeanette, my loving wife, for putting up with the long hours and weekends I spent in front of my computer and motivating me to continue.

Reviewers and editorial staff who provided suggestions and comments on article submissions and which have added tremendous value to the finished product. I would also like to specifically mention the examiners, whose input contributed to an improved final product.

My family and friends for their words of encouragement and support over the past few years.

This page has been intentionally left blank

Declaration

I, the undersigned, declare that this thesis, which I hereby submit for the degree Philosophiae Doctor at the University of Pretoria, is my own independent work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Pretoria, April 24, 2022

Byran H. Taljaard

Contents

List of Figures

List of Tables

This page has been intentionally left blank

Definitions and terms

Cap weighted portfolio This is a short-hand term used for the market capitalisation weighted portfolio and refers to the portfolio with each stock's weight equal to the share of total market capitalisation in the universe of stocks. The universe of stocks is defined by the respective market index.

Cap weights This refers to the market capitalisation weights of each stock in the cap weighted portfolio.

Classification and regression tree (CART) A CART model is a supervised learning model resembling a decision tree, with each node (or split) determined by the maximisation of the homogeneity of the resulting sub-group.

Concentration Concentration in this thesis refers specifically to the concentration in the cap weights of the respective market. For example, in South Africa a handful of stocks account for a very large portion of the cap weighted index or portfolio by virtue of their significantly larger share of the overall market capitalisation.

Diversification Diversification in this thesis has a very specific form; the excess growth rate of the equal weighted portfolio (Equation 3.6). Intuitively, this measure is a function of the reduction in portfolio volatility due to the covariance structure of the underlying stocks.

Drift process The drift process is defined in Definition A.1.2 and is a contributing factor to the relative return of a portfolio. In the case of the equal weighted portfolio the drift process is equal to the excess growth rate.

Entropy Entropy is another measure for concentration used in information theory. Setting the parameters to the cap weights of an index allows one to calculate the entropy of an index or market.

Equal weighted portfolio This is a portfolio with equal weights assigned to each stock. In this thesis the universe of stocks is set to the constituents of the respective cap weighted portfolio or index.

Excess growth rate The excess growth rate is defined in Equation (3.6) and is a measure of diversification. Intuitively, this measure is a function of the reduction in portfolio volatility due to the covariance structure of the underlying stocks.

Functionally generated portfolios Functionally generated portfolios are portfolios whose weights can be generated systematically using a function with the cap weights as inputs (see Section 3.4). The equal weighted portfolio is an example of a functionally generated portfolio.

Geometric mean of cap weights The geometric mean of cap weights is an important measure of concentration as this function also represents the portfolio generating function of the equal weighted portfolio and, therefore, changes in the geometric mean of cap weights contributes directly to the relative return of the equal weighted portfolio (see Section 3.3).

Information Ratio The information ratio of a portfolio is a measure of the active returns (relative to the cap weighted portfolio in this thesis) scaled by the volatility of those active returns.

Leakage Leakage is a term from stochastic portfolio theory that describes the effect of stocks "leaking" out of the respective market index used to define the universe. This index is, however, only a subset of the overall market and, therefore, defining a portfolio's constituents using this subset leads to stock selection by rank. Given that the equal weighted portfolio has a larger weight in these stocks moving in and out of the index than the cap weighted portfolio, its effect on the equal weighted portfolio's relative return is generally negative.

Market capitalisation weighted portfolio This term refers to the portfolio with each stock's weight equal to the share of total market capitalisation in the universe of stocks. Cap weighted portfolio is used as a short-hand reference, however, throughout this thesis. Although the respective market capitalisation index is used to define the constituents in each equity market, cap weighted portfolios are constructed to ensure that rebalance periods match the equal weighted portfolio in question.

Portfolio generating function This is the function of the cap weights that generates a specific portfolio (see Section 3.4). In the case of the equal weighted portfolio this is the geometric mean of cap weights which is also a measure of concentration in the cap weighted portfolio. Changes in the portfolio generating function have a direct impact on the relative returns of a portfolio (see Section 3.3).

Random Forest model The Random Forest (RF) model is an extension of the CART model. Instead of only considering one decision tree, a RF model build numerous decision trees, with each tree being fit using a random sample of the available features and only a bootstrap sample of the data. The prediction from a RF model is then the average prediction across all the individual decision trees.

Relative return In this thesis the mention of relative return refers to a portfolio's return relative to the respective cap weighted portfolio.

Sharpe ratio This represents the ratio of excess return over the risk-free rate to the volatility of returns.

Sortino ratio This is a modification of the Sharpe ratio, whereby the volatility of returns is replaced by the volatility of negative returns (downside deviation).

Stochastic portfolio theory Stochastic portfolio theory is a framework for constructing and analysing portfolios, specifically against the market capitalisation weighted index or portfolio.

This page has been intentionally left blank

List of symbols

$X_i(t)$: Denotes the stock price process for stock i at time t.

$\gamma_i(t)$: The geometric growth rate of X_i at time t.

$\sigma_i(t)$: Volatility of stock price process X_i at time t.

$\sigma_{ij}(t)$: Covariance between X_i and X_j at time t.

$dW_i(t)$: Denotes a Brownian motion.

$\alpha_i(t)$: Arithmetic growth rate of X_i at time t.

$\pi_i(t)$: The weight of a stock i in a portfolio with weights denoted $\pi(t)$. Usually π represents the equal weighted portfolio in this thesis.

$\mu_i(t)$: Exclusively used in this thesis to denote the weight of a stock i in the market capitalisation weighted portfolio. See Equation (3.12).

$Z_\pi(t)$: $Z_\pi(t)$ denotes the portfolio price process for a portfolio with weights $\pi(t)$. $Z_\mu(t)$ exclusively denotes the portfolio price process of the market capitalisation weighted portfolio.

$\gamma_\pi(t)$: Denotes the geometric growth rate of the portfolio price process $Z_\pi(t)$ for a portfolio with weights $\pi(t)$. See Equation (3.5).

$\gamma_\pi^*(t)$: A part of $\gamma_\pi(t)$ which acts as a measure of diversification in the portfolio with weights $\pi(t)$. This term is referred to as the excess growth rate, see Equation (3.6), and is also the drift process for the equal weighted portfolio, see Equation (3.21).

$\delta_i(t)$: The dividend process for stock $X_i(t)$.

$\delta_\pi(t)$: The dividend process for a portfolio with weights π. See Equation (3.11).

$\hat{X}_i(t)$: The total return process for stock i, including dividends. See Equation (3.9).

$\hat{Z}_\pi(t)$: The total return process for a portfolio with weights π, including dividends. See Equation (3.10).

$\mathbf{S}(\mu(t))$: The portfolio generating function for a portfolio as per Definition 3.16. $\mathbf{S}(\mu(t))$ generates a portfolio as per Equation (A.23) in Theorem A.2.1. Changes in the portfolio generating function have a direct impact on relative returns, see Definition A.2.1.

$\Theta(t)$: The drift process, see Equation (A.24), of a portfolio with portfolio generating function $\mathbf{S}(\mu(t))$. The drift process has a direct impact on relative returns, see Definition A.2.1.

$\tau_i(t)$: The relative variance (or tracking error) of stock i, relative to a portfolio price process, such as the cap weighted portfolio.

$\tau_{ij}(t)$: The relative covariance of stock i and stock j, relative to a portfolio price process, such as the cap weighted portfolio.

$L_\pi(t)$: Denotes the net effect of leakage on a portfolio with weights π relative to the cap weighted portfolio with weights μ. See Equation (3.25).

$\Lambda_X(t)$: A semi-martingale local time process defined in Equation (3.26).

$D_p(\mu)$: A portfolio generating function that uses a parameter p, with $0 < p \leq 1$, to generate a portfolio with weights between equal weight (as $p \to 0$) and cap weight (as $p \to 1$). See Equations (5.3) and (5.4).

f_i: Denotes the i-th feature used in a decision tree. See Section 5.3.

f_i^S: Denotes the split criteria f_i^S selected for the feature f_i such that the split maximises the homogeneity of the resulting child nodes. See Section 5.3.

G_i: Denotes the Gini index, used as a determinant of homogeneity in CART and Random Forest models. See Equation (5.5). G_i represents the Gini index at node i in a decision tree.

List of equity indices

A key aim in this thesis is the analysis of market capitalisation weighted (cap weighted) and equal weighted portfolios among a selection of eight countries. A cap weighted index is selected to represent the equity universe in each country and, furthermore, the constituents of the respective indices are used to define the universe for the construction of the cap weighted and equal weighted portfolios within each country. The index name and related country are used interchangeably throughout this thesis. For example, a reference to US equities would, in this thesis, refer to the universe of US stocks defined by the S&P500 Index. The same would hold for references to SA equities and the FTSE/JSE Top40 index.

A description of these cap weighted equity indices are provided below.

CAC 40 The CAC 40 represents the 40 largest stocks, by market capitalisation, listed on the Euronext Paris stock exchange. Index weights are typically capped at 15% if necessary, although, in this thesis the cap weighted portfolio is re-constructed without such rules.

DAX The DAX is an index representing the top 40 (expanded from 30 in 2021) stocks, by market capitalisation, trading on the Frankfurt Stock Exchange.

FTSE 100 The FTSE 100 index represents the top 100 stocks, by market capitalisation, listed on the London Stock Exchange (UK).

FTSE/JSE Top 40 Index The FTSE/JSE Top 40 Index consists of the 40 largest companies, by market capitalisation, listed on the Johannesburg Stock Exchange (JSE) and is, therefore, representative of the largest stocks in South Africa (SA). In this thesis the FTSE/JSE Top 40 Index is referred to as the Top 40.

S&P/ASX 200 The S&P/ASX 200 Index represents the 200 largest companies, by market capitalisation, listed on the Australian Securities Exchange (ASX).

S&P/TSX 60 The S&P/TSX 60 Index represents large-cap stocks listed on the Toronto Stock Exchange (TSX). Although the index aims to represent the largest stocks on the TSX, it is also structured to reflect the sector weights of the larger S&P/TSX Composite Index which covers the broader listed equity market in Canada (CA).

S&P500 The S&P500 Index represents the top 500 listed companies in the United States (US) by market capitalisation.

TOPIX The Tokyo Stock Price Index (TOPIX) is a market capitalisation weighted index representing all the stocks listed on the Tokyo Stock Exchange (TSE) and categorised in the so-called First Section. The First Section contains the largest companies listed on the TSE. The TOPIX represents the largest index in this thesis by number of constituents with total membership of over 2000 stocks in this index.

1 Introduction

1.1 The equal weighted portfolio

The market capitalisation weighted index (cap weighted) index remains the most widely used equity benchmark both in practice and in academia. This is due, in part at least, to the Capital Asset Pricing Model (CAPM) developed by Sharpe (1964) which assumes that the cap weighted index will be mean-variance efficient. Secondly, from a practical perspective, cap weighted portfolios make intuitive sense, representing the total stock market by market capitalisation (or at least the largest firms within that market). Cap weighted indices, furthermore, require minimal effort to maintain beyond typical inclusions and exclusions in the index, implying little turnover in an underlying portfolio.

A number of studies have, however, challenged the efficiency of the cap weighted index. Haugen and Baker (1991) discuss the conditions necessary for cap weighted portfolios to be efficient. Among the conditions investigated are the assumptions that all investors agree about the risk and expected return for all securities, all investors can short-sell all securities, all investors have the same investable universe, and that no investor is subject to taxation. In the absence of these assumptions, the authors argue that the cap weighted portfolio is inefficient. Furthermore, they show that an ex-ante efficient portfolio consistently outperforms the Wilshire 5000 index with lower volatility. Hsu (2004) shows that a cap weighted portfolio suffers from a "return drag" which is directly proportional to the volatility of the individual stocks. Therefore, the higher the level of price inefficiency (volatility) in the market, the higher the return drag. This assumption is equivalent to an assumption of negative autocorrelation in returns. See also, DeMiguel et al. (2009), Martellini (2009) as well as Amenc et al. (2012) which question the efficiency of cap weighted indices.

Numerous studies have specifically shown equal weighted portfolios to be more efficient than cap weighted portfolios. Dash and Loggie (2008) provide empirical evidence of an equal weighted S&P500 index outperforming the cap weighted index. Dash and Zeng (2010) confirm similar results for the S&P International 700 index. Bolognesi et al. (2013) examined the outperformance of an equal weighted Dow Jones Eurostoxx 50 index compared with a cap weighted index. In terms of the efficiency of equal weighted portfolios, DeMiguel et al. (2009) evaluated the performance of optimal asset allocation against a naïve equal weighed portfolio. They found that none of the 14 models (extensions of the sample based mean variance optimisation) performed better out-of-sample than the equal weighted portfolio on a risk-adjusted basis. Plyakha et al. (2015) also show, empirically, that equal weighted portfolios, rebalanced monthly, outperform the cap weighted portfolio in terms of total return, alpha (relative to a four-factor model) and Sharpe ratio.

Table 1.1.: Annualised returns, volatility and risk-adjusted returns of the cap- and equal weighted portfolios for a selection of countries analysed in this thesis. The period starts 1 January 2003 and ends 31 May 2021. Portfolios are rebalanced monthly and transaction costs are excluded.

	CAGR (%)		Volatility (%)		Sharpe ratio[a]		Information ratio[a]
Country	**Cap**	**Equal**	**Cap**	**Equal**	**Cap**	**Equal**	**Equal**
Australia	12.1	11.5	16.4	17.6	0.57	0.50	-0.01
Canada	9.7	10.6	18.1	18.3	0.52	0.58	0.23
France	8.0	8.2	21.9	23.0	0.42	0.42	0.21
Germany	8.7	9.7	22.5	22.1	0.45	0.49	0.25
Japan	7.1	11.3	21.2	20.0	0.43	0.63	0.60
South Africa	14.7	14.6	20.2	19.4	0.42	0.42	0.00
United Kingdom	7.2	10.0	18.6	19.4	0.36	0.49	0.61
United States	11.3	13.1	19.4	21.6	0.58	0.62	0.41

[a]Respective risk-free rates used in Sharpe and Information ratio calculations.

Table 1.1 shows the returns (after transaction costs) for the cap- and equal weighted portfolios for selected equity markets based on analysis conducted in this thesis (Table 1.1 is a reproduction of Table 2.5 shown here for ease of reference). These results largely support the literature with the equal weighted portfolio outperforming the cap weighted portfolio in all but two equity markets (Australia and South Africa) and doing so with higher risk-adjusted returns as measured by the Sharpe and Information ratios.

1.1.1 Where does this outperformance come from?

In Malladi and Fabozzi (2017), the authors attempt to find the source of this outperformance in the US stock market by comparing the alpha of the equal weighted portfolio over the cap weighted portfolio against various factors. They find that the equal weighted portfolio has a bias towards small stocks (that is, the size factor) as expected but also has significant exposure to the reversal factor. This makes intuitive sense as the equal weighted portfolio continues to sell outperforming stocks and buy underperforming stocks through the rebalance process. This leads the authors to conclude that a significant portion of the excess return over the cap weighted portfolio is derived from the rebalancing process itself.

The act of rebalancing an equal weighted portfolio is also sometimes referred to as volatility harvesting or volatility return (see Bouchey, Nemtchinov, Paulsen, et al., 2012; Bouchey, Nemtchinov, and Wong, 2015; Hallerbach, 2014, for example), although, the mechanism is similar to the reversal factor studied in Malladi and Fabozzi (2017). A dynamic rebalancing frequency is, therefore, also an important decision to make given the analogy with volatility harvesting. That is, at times volatility may be so low that rebalancing is counterproductive after transaction costs, and vice versa, when volatility is so high that not rebalancing regularly creates a drag on performance.

The regular rebalancing to equal weights, while generating a return through reversals in individual stock returns, also has the effect of keeping the equal weighted portfolio more diversified than a cap weighted portfolio. The higher diversification generates additional return over the cap weighted portfolio and this leads Booth and Fama (1992), for example, to refer to the additional return from rebalancing as diversification return.

Under the assumption of a lognormal model of stock prices and a diversified market, Fernholz (2002) sets out a framework (stochastic portfolio theory) which allows for a comparison of expected returns of portfolios generated by some function against the expected returns of the cap weighted portfolio. Under this framework, it can be shown that an equal weighted portfolio outperforms a cap weighted portfolio in the long run under the assumption that the market does not become overly concentrated in one stock. Stochastic portfolio theory also allows for a decomposition or attribution of the excess return generated by the equal weighted portfolio over the cap weighted portfolio. Chapter 3 provides an introduction to this framework, however, it is interesting to note here that movements in the geometric mean of cap weights plays an important role in the relative return of the equal weighted portfolio. As the geometric mean decreases (i.e. as stock weights become less concentrated), it provides a positive contribution to the equal weighted portfolio's relative return. This is a very similar effect to the reversal factor identified in Malladi and Fabozzi (2017) and related to the "return drag" mentioned in Hsu (2004) for the cap weighted portfolio.

Stochastic portfolio theory is a key element of this thesis, and its empirical implementation in the following chapters attempts to provide a complete picture of the sources of the equal weighted's return (or lack thereof). Understanding these drivers of relative performance for the equal weighted portfolio is important. Many studies have shown the equal weighted portfolio to outperform the cap weighted portfolio on both a risk-adjusted and outright return basis on a longer-term basis. In the short-term, however, the equal weighted portfolio can suffer significant underperformance relative to the cap weighted portfolio. It may be difficult for investors to hold equal weighted portfolios during these periods that may affect overall long-term performance. Understanding where this underperformance is coming from can go a long way to improve decision making in portfolio construction especially since the equal weighted portfolio can be viewed as a proxy for contrarian strategies in general.

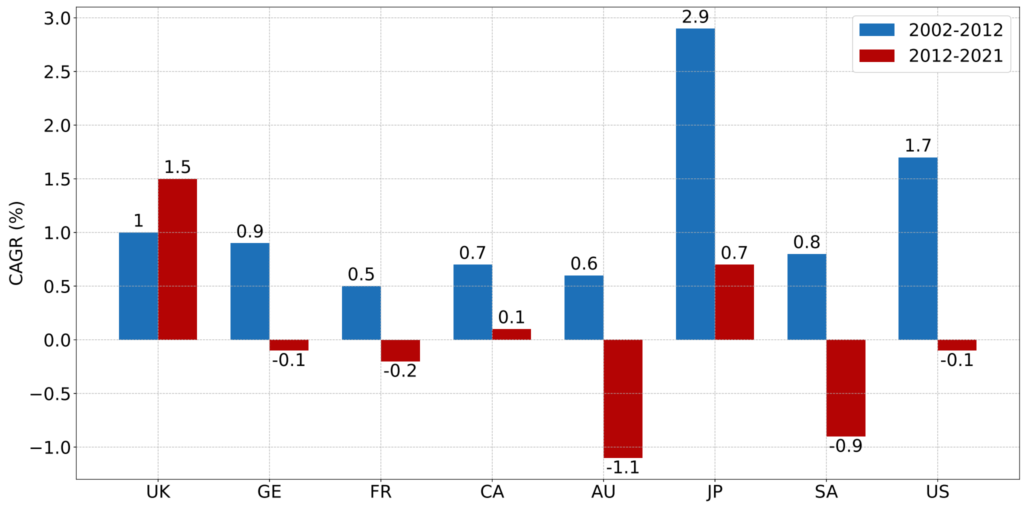

Figure 1.1.: Active returns (CAGR) of equal weighted portfolio relative to the cap weighted portfolio, pre-2012 and post-2012.

Figure 1.1 highlights one example of this by showing the relative returns of the equal weighted portfolio across various markets pre- and post-2012. Only in the UK equity market have relative returns for the equal weighted portfolio been higher after 2012. In the case of Japan, where the equal weighted portfolio does well overall, relative returns been significantly lower after 2012. Relative returns for markets such as South Africa and Australia have been significantly negative (likely due to the high resource sector exposure), however, relative returns are marginally negative even in developed markets such as Germany, France, and the US since 2012.

There can be various reasons for short-term underperformance. As mentioned before, the act of rebalancing an equal weighted portfolio can be viewed as exposure to some reversal effect, with the portfolio essentially selling stocks that have outperformed the average stock and buying those that have underperformed. In the short-term this would require some form of mean reversion (the reversal effect) in stocks to outperform the cap weighted portfolio. In momentum driven markets, however, this is unlikely to be the case and in such instances the equal weighted portfolio is bound to underperform the cap weighted portfolio. Another source of underperformance could be the diversification benefits of the equal weighted portfolio. While the act of rebalancing an equal weighted portfolio is typically thought to generate additional return through increased diversification, it is possible that the diversification return is too low, either because individual stock volatilities are low or correlation is high, and does not warrant rebalancing of the portfolio back to equal weights. This is analysed further in Taljaard and Maré (2019b) and Cuthbertson et al. (2016).

While the focus of this thesis is on the equal weighted portfolio, stochastic portfolio theory provides a general framework for analysing portfolios with weights which are systematically generated (see Theorem A.2.1). Therefore, it is possible to extend the discussion and analysis to so-called "smart beta" portfolios in general. The equal weighted portfolio is but one type of "smart-beta" portfolio. A value-based approach with stocks weighted by some fundamental price ratio, such as price-to-book ratios, is another example of a smart-beta portfolio where stochastic portfolio framework may be able to assist with empirical attribution. This is left as one area for further research.

1.1.2 The equal weighted portfolio as a benchmark?

Typically active equity portfolio managers do not stray too far from their given benchmark. Petajisto (2013), for example, find that 16% of all so-called active funds could be classified as closet indexers, and a further 48% of funds could only be classified as moderately active (with an average tracking error of less than 6%) for US all-equity funds from 1990 – 2009, using active share and tracking error as metrics. Taljaard and Maré (2019a) show in the South African equity unit trust universe that the median tracking error is about 5.5% and nearly half of these unit trusts have a tracking error of less than 5.5% w.r.t. the FTSE/JSE All Share Index. A quarter of the unit trusts have a tracking error of less than 4.5% and only 23% of funds have a tracking error of more than 8%.

Given the evidence that active portfolio managers are typically not benchmark agnostic, that is, they do not stray too far from their benchmark, and the literature showing that the cap weighted portfolio is less efficient and underperforms the equal weighted portfolio, it raises the question of whether an investor would be better served by providing active portfolio managers with an equal weighted portfolio or index as a benchmark. This is a question analysed by Taljaard and Maré (2019a). To analyse this question the authors in Taljaard and Maré (2019a) generate random active portfolios with active positions specified by a multiple of the stocks benchmark weight. This is not too far removed from happens in practice. The authors find that active portfolios display much higher Sharpe ratios than those using a cap weighted portfolio. This result is shown in Figure 1.2 as presented in Taljaard and Maré (2019a).

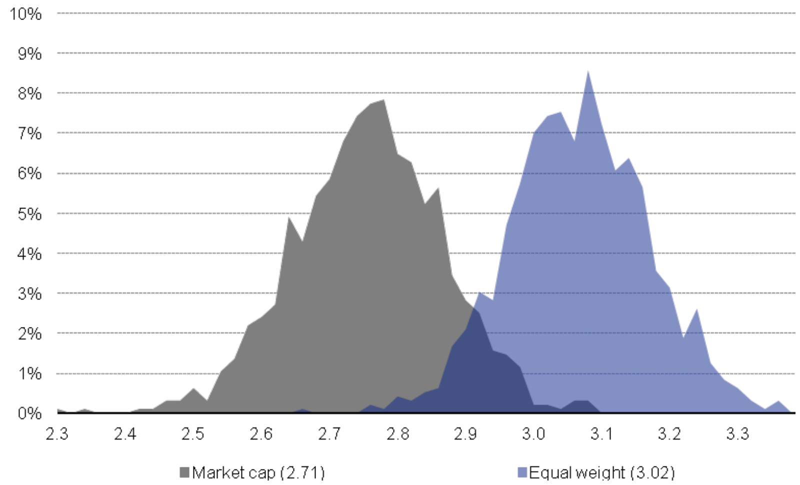

Figure 1.2.: Sharpe ratio (risk-free rate of 0 used) for random portfolios under a market cap and equal weight benchmark as presented in Taljaard and Maré (2019a).

The analysis presented in Taljaard and Maré (2019a) considered the FTSE/JSE Top 40 index until 2015. Much has, however, changed in the South African equity market and specifically with regards to the performance of the equal weighted portfolio in South Africa (see the results in Chapter 2). Understanding the factors driving this performance can provide additional information to equity investors about whether an equal weighted index would be appropriate as a benchmark for active portfolio managers.

1.2 Thesis Scope

While numerous papers have considered the long-term performance of the equal weighted portfolio and potential sources thereof, there has been a lack of discussion and analysis in relation to the empirical short-term relative performance of the equal weighted portfolio and why the equal weighted portfolio has underperformed significantly in certain periods despite its longer-term outperformance. Furthermore, while the equal weighted portfolio has outperformed the cap weighted portfolio in most equity markets, there is a large disparity between the performance of the equal weighted portfolio in different markets.

Stochastic portfolio theory, however, provides a framework that allows one to understand and analyse the relative performance of an equal weighted portfolio. Stochastic portfolio theory was first introduced by Robert Fernholz (Fernholz and Shay, 1982; Fernholz, 1999a; Fernholz and Square, 1998), which culminated in a book (Fernholz, 2002). This approach allows for an attribution of the equal weighted portfolio's return relative to the cap weighted portfolio, providing analysis of when and why the equal weighted portfolio outperforms or underperforms in certain periods and markets.

The scope of this thesis is, therefore, to analyse the performance of the equal weighted portfolio within this framework and to provide an empirical attribution of the relative performance. This is primarily done for the SA and US equity markets given that the SA equity market is the local equity market and is a market where the equal weighted portfolio has struggled (see Section 2.3). The US equity market, and the S&P500 Index, provides a good comparator to the SA equity market given the US market's role as one of the largest and most liquid equity markets in the world. This analysis is extended to consider a selection of other, primarily developed, equity markets which is particularly important when comparing the drivers of the equal weighted portfolio's relative return within different markets (see Section 4.6).

1.3 Thesis Aims

This thesis aims to analyse the empirical performance of the equal weighted portfolio in various equity markets, but with a focus on the SA and US markets. This analysis is performed within the stochastic portfolio theory framework, which allows for an empirical attribution of the relative performance of the equal weighted portfolio. Although the main focus is on the S&P500 and the Top 40, other countries are also included to provide broader context. Table 1.2 displays the list of countries and the respective indices used. Also included are start dates used for data sourced from Bloomberg.

A further aim of this thesis is to attempt to utilise this framework and the empirical analysis of the various drivers of relative performance to understand when holding the equal weighted portfolio may be sub-optimal, and to attempt to create an

Table 1.2.: List of countries and indices used for analysis.

Country	Index	Start Date
Australia (AU)	S&P/ASX 200	1 January 2002
Canada (CA)	S&P/TSX 60	1 January 2002
France (FR)	CAC 40	1 June 2001
Germany (GE)	DAX	1 January 2000
Japan (JP)	TOPIX	1 January 2000
South Africa (SA)	FTSE/JSE Top 40	1 November 2002
United Kingdom (UK)	FTSE 100	1 January 1998
United States (US)	S&P500	1 January 1996

improved portfolio which benefits from the longer-term outperformance by the equal weighted portfolio while reducing short-term drawdowns relative to the cap weighted portfolio.

The core of this thesis is based on four journal articles, three of which are peer reviewed, and have focused on the importance of the equal weighted portfolio and the components of its relative performance against the cap weighted portfolio. These articles are set out in Table 1.3.

Table 1.3.: Published doctoral research: Journal articles and online working papers.

Peer-reviewed publications	
1.	**Taljaard, B. H.**, & Maré, E. (2019). Considering the use of an equal-weighted index as a benchmark for South African equity investors. *South African Actuarial Journal*, 19(1), 53-70.
2.	**Taljaard, B. H.**, & Maré, E. (2021). Why has the equal weight portfolio underperformed and what can we do about it?. *Quantitative Finance*, 1-14.
3.	**Taljaard, B. H.**, & Maré, E. (2021). If the equal weighted portfolio is so great, why isn't it working in South Africa?. *Investment Analysts Journal*, 50(1), 32-49.
Online working paper (submitted to Journal of Investment Strategies)	
1.	**Taljaard, B.**, & Mare, E. (2019). Too much rebalancing is not a good thing. *Available at SSRN 3484337*.

1.4 Thesis structure

Chapter 2 sets the stage by discussing the historical performance of the equal weighted portfolio with a focus on both United States (S&P500) and South African (Top 40) equities over the past 18 and 25 years, respectively. An analysis and discussion of the performance of the equal weighted portfolio in the remaining countries in Table 1.2 is included in this chapter. Chapter 3 then introduces the main model, based on stochastic portfolio theory, that is used to analyse this performance and attempts to provide an intuitive explanation for each component. An empirical analysis and return attribution of the equal weighted portfolio is presented in Chapter 4 using the stochastic portfolio theory framework.

The final chapter aims to develop a forward-looking element to the concepts explored in the earlier chapters with direct impact on practical portfolio management. Chapter 5 introduces two models; a simplistic regression model to dynamically switch between the equal weighted and cap weighted portfolios, and a Random Forest model used to predict the probability of the equal weighted portfolio outperforming in the next month and setting weights accordingly.

1.5 Data and methodology

The methodology throughout this thesis is kept as consistent as possible with portfolios constructed using daily stock price and total return data from Bloomberg. Portfolios are rebalanced monthly, on the following business day's closing prices. Dividends and the effect of other capital adjustments are placed in cash until the next rebalance date. Cash balances between rebalance dates are assumed to earn no interest. Unless otherwise stated, transaction costs are excluded from the analysis. The cap weighted portfolio is constructed from market capitalisation weights using the universe of stocks in the respective index as at the end of the month prior to the rebalance date. These portfolios will, therefore, differ slightly from the official indices although the constituents mirror the official indices at all times.

All analysis is conducted on the period ending 30 May 2021, and starting on the dates provided in Table 1.2.

Important excerpts from the Python code used to perform the analysis in this thesis are provided in Appendix B.

1.6 Research questions

While the discussion has touched on various aims in this thesis, the main research question of this thesis is:

What are the components driving the performance of the equal weighted portfolio relative to the cap weighted portfolio and is it possible to utilise these to improve portfolio performance?

The attempt to address this research question is done by addressing the following specific research questions:

1. How has the equal weighted portfolio performed in various equity markets?

   This is discussed in Chapter 2, with performance for the SA and US equity markets presented in Sections 2.2 and 2.3, respectively. Historical performance is summarised for other markets in Section 2.5.

2. How can the framework in stochastic portfolio theory assist in analyses of the relative performance of the equal weighted portfolio?

   Chapter 3 provides a brief introduction to stochastic portfolio theory with an attempt to present an intuitive understanding of the various elements in the framework as it pertains to the equal weighted portfolio. Mathematical expressions are presented in Appendix A. Chapter 3 identifies the main drivers of the equal weighted portfolio's relative returns as per the stochastic portfolio theory framework in Section 3.5 and Equation 3.28.

3. What are the empirical drivers in performance across these various equity markets?

   Results presented in Chapter 3 are used to present an attribution of the relative return of the equal weighted portfolio. Again the focus is on the US and SA equity markets in Sections 4.2 and 4.3, respectively. Results for the other selected equity markets are presented in Section 4.6.

4. How much does concentration play a role in the relative returns of equal weighted portfolios?

   Section 2.4 touches on the role of concentration in the SA and US market briefly. The stochastic portfolio framework, however, allows for a more direct analysis of concentration and this is discussed in Sections 4.2 and 4.3 for the US and SA equity markets, respectively, and in Section 4.6 for other equity markets. Section 4.5 presents an estimate of the direct impact to growing concentration by specific stocks in the US and SA.

5. How have specific stocks, such as technology stocks, influenced the index concentration and what is the impact on the equal weighted portfolio?

   This is discussed in Section 4.5 where the impact of technology stocks in the US and Naspers (NPN) in SA are analysed.

6. Is increasing concentration a global phenomenon or is it more specific to each market?

   This is discussed throughout Chapter 4 and Section 4.6. Furthermore, Table 4.3, presents a summarised attribution of the main factors, including concentration, affecting the relative returns of the equal weighted portfolio to aid this discussion. Table 4.5 presents the correlations of one measure of concentration specific to the equal weighted portfolio (the portfolio generating function) across the different equity markets.

7. What role has diversification played and is this factor correlated across various markets?

   The role of diversification is discussed throughout Chapter 4 and in Section 4.6. Table 4.4 presents the correlation of one measure of diversification (the excess growth rate) across the various equity markets to aid this discussion and analysis.

8. Why has the equal weighted portfolio underperformed in the South African market? Is this only a matter of concentration or are there other factors?

   This is analysed and discussed in Section 4.3 where the empirical contributions to relative returns for the equal weighted portfolio are presented and analysed.

9. What are the differences in the equal weighted portfolio's performance across various equity markets and what has contributed to the equal weighted portfolio performing well in some countries and poorly in others?

   This is analysed and discussed in detail in Section 4.6 and Figures 4.20 to 4.27 show the main drivers per country.

10. Is there a way to utilise the main drivers of relative performance to create an optimal blend of cap and equal weights to generate outperformance but with lower relative drawdowns?

Chapter 5 attempts to use the empirical drivers of the equal weighted portfolio's relative return to achieve this using two methods; a rudimentary linear regression approach and a Random Forest model.

2 Performance of the equal weighted portfolio

2.1 Introduction

In this chapter the performance of monthly rebalanced equal and cap weighted portfolios in both the United States (US) and South African (SA) stock markets are analysed, making use of the Standard & Poor's 500 Index (S&P500) and the FTSE/JSE Top 40 Index (Top 40) as the universe of stocks in each country, respectively. Sections 2.2 and 2.3 look at the S&P500 and Top 40, respectively, before comparing the two and briefly touching on the levels of concentration and how this could contribute to underperformance. The majority of the analysis and results in Sections 2.2 and 2.3 are updated from Taljaard and Maré (2021b) and Taljaard and Maré (2021a), respectively. The chapter concludes with a look at the performance of the equal weighted portfolio within the equity markets of other countries in Table 1.2. While Taljaard and Maré (2021b) include a brief discussion on the equal weighted portfolio across these equity markets, Section 2.5 provides a deeper discussion of the differing performance of the equal weighted portfolio across the selected equity markets.

In practice, transaction costs are an important contributor to portfolio returns and are particularly significant when comparing the equal and cap weighted portfolios. The equal weighted portfolio has to constantly rebalance to bring the weights back to an equal weight. That is, selling outperforming stocks and buying underperforming stocks. The cap weighted portfolio, on the other hand, has very little need to rebalance (beyond corporate action activity and index constituent changes) as the change in market value has a natural rebalancing effect. This already places the equal weighted portfolio at a disadvantage and why it is important to include transaction costs. At the same time, however, it is necessary to get some understanding of the underlying strategies and their historical performance, and therefore, the analysis in this chapter assumes no transaction costs to get a better sense of the raw performance of the equal and cap weighted portfolios. The analysis in Chapter 4, however, includes costs of 15 bps, and the attribution analysis allows for the distinction

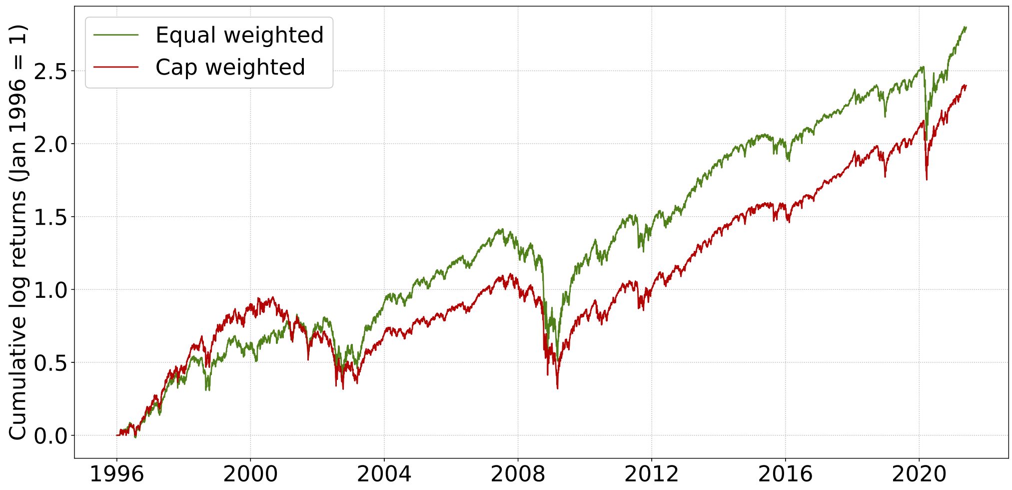

Figure 2.1.: Log cumulative returns for the S&P500 equal- and cap weighted portfolios with monthly rebalancing.

between the various factors affecting relative performance, with transaction costs being one such factor.

2.2 S&P500 performance

Table 2.1.: Risk adjusted performance of the S&P500 equal and cap weighted portfolios, monthly rebalanced. No transaction costs included.

Portfolio	CAGR[a]	Volatility	Sharpe ratio[b]	Sortino ratio
Equal weighted	12.1%	20.8%	0.55	0.70
Cap weighted	10.3%	19.7%	0.49	0.62

[a] Compound annual growth rate.

[b] One-month US Treasury bill used as risk free rate for both Sharpe and Sortino ratios.

Figure 2.1 shows the cumulative return (log scale) since 1996 of both the equal and cap weighted portfolios with monthly rebalancing. Table 2.1 shows the risk adjusted returns over the entire period. Notwithstanding the underperformance at the beginning of our sample period, the equal weighted portfolio does indeed outperform the cap weighted portfolio over the entire period. Although the equal weighted portfolio has a higher volatility, its higher annualised growth rate leads to higher Sharpe and Sortino ratios than the cap weighted portfolio.

Figure 2.2 shows the relative cumulative return over time of the equal weighted portfolio. The relative returns of the equal weighted portfolio on a rolling one-year basis are shown in Figure 2.3. Although the equal weighted portfolio does outperform the cap weighted portfolio over the entire period, the majority of this outperformance is generated during the period 2000 to 2008. The relative performance being largely sideways from 2008 onwards and somewhat negative since 2016. There has, however, been some sign of recovery more recently with relative returns improving following the low in March 2020.

Figure 2.2.: Relative cumulative return for the S&P500 equal weighted portfolio.

The lacklustre performance of the equal weighted portfolio relative to the cap weighted portfolio since 2016 is also evident from Table 2.2, which shows three-year total returns for the equal and cap weighted portfolios since 1996. Relative returns are the highest in the early 2000s and degrade slightly for later three-year periods. In particular, the period 2015 to 2020 has been especially tough for the equal weighted portfolio which has underperformed by approximately 7%.

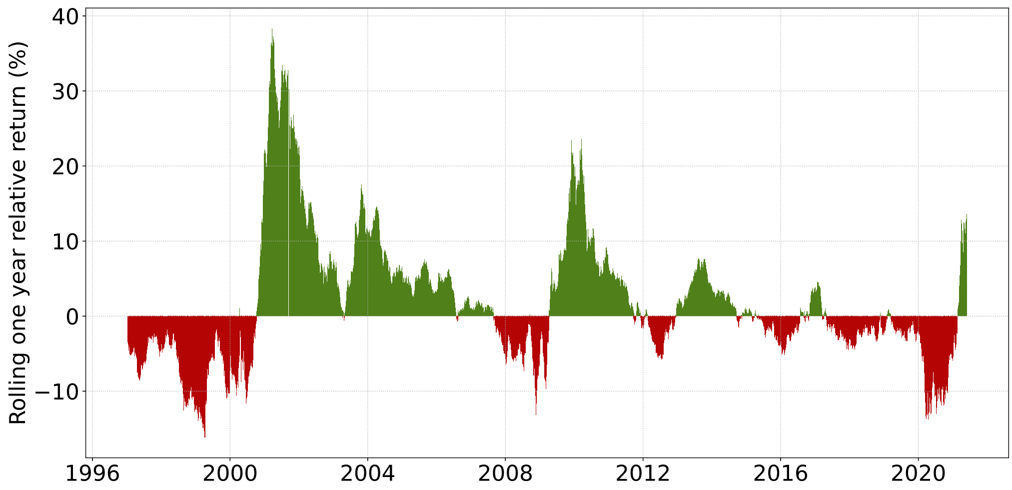

Figure 2.3.: Rolling one year relative return for the S&P500 equal weighted portfolio.

Table 2.2.: Total return over non-overlapping three-year periods for the S&P500 (%).

Period Start[a]	Period End	Equal weighted	Cap Weighted	Relative
1996	1998	69.1	105.5	-17.7
1999	2001	25.6	-2.8	29.1
2002	2004	37.6	11.2	23.7
2005	2007	27.5	28.9	-1.1
2008	2010	9.8	-7.4	18.6
2011	2013	61.6	56.6	3.2
2014	2016	28.6	28.9	-0.3
2017	2019	43.7	53.9	-6.6
2020	2021/05[b]	34.4	33.5	0.7

[a] Period Starts 1 January and ends 31 December of Period End year

[b] Period ends 30 May 2021

2.3 Top 40 performance

While the equal weighted portfolio in the US may have disappointed over the last five to ten years, it has at least outperformed the cap weighted portfolio over a longer period. The case in South Africa is a little more dire, with the equal weighted portfolio barely keeping up with the cap weighted portfolio since 2002. Figure 2.4 shows the cumulative return for the Top 40 equal- and cap weighted portfolios. Much like the S&P500, the performance of the equal weighted portfolio was ahead of the cap weighted portfolio until 2016. The proceeding period was, however, much more severe for the Top 40 as illustrated in the relative return in Figure 2.5 and the rolling one-year relative returns in Figure 2.6.

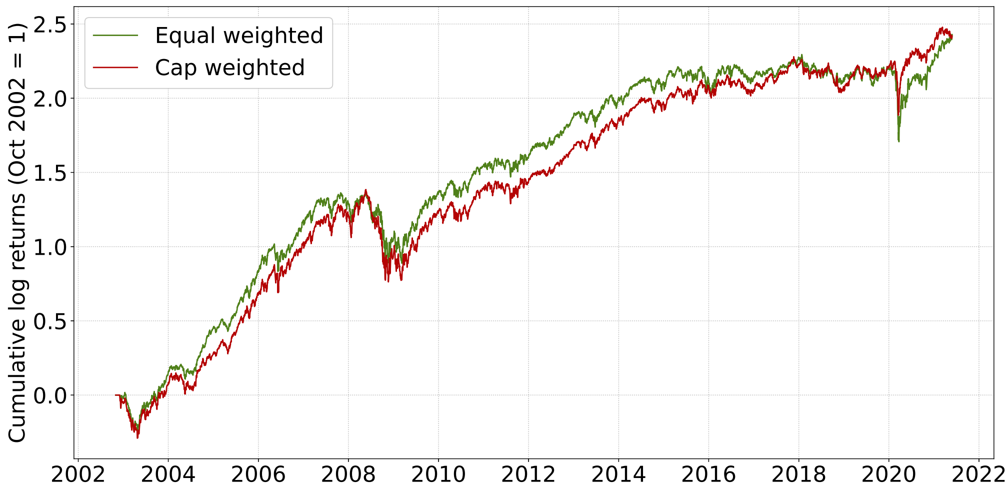

Figure 2.4.: Log cumulative returns for the Top 40 equal- and cap weighted portfolios with monthly rebalancing.

Table 2.3.: Risk adjusted performance of the Top 40 equal and cap weighted portfolios, monthly rebalanced.

Portfolio	CAGR[a]	Volatility	Sharpe ratio[b]	Sortino ratio
Equal weighted	14.6%	19.4%	0.42	0.57
Cap weighted	14.7%	20.2%	0.42	0.58

[a]Compound annual growth rate.

[b]One-month JIBAR used as risk free rate for both Sharpe and Sortino ratios.

Table 2.3 highlights the risk-adjusted performance of the two portfolios, indicating once again that the equal weighted portfolio has only just kept up with the cap weighted portfolio in South Africa (and bearing in mind that this excludes transaction

Figure 2.5.: Relative cumulative return for the Top 40 equal weighted portfolio.

costs). The Sharpe and Sortino ratios are very similar with the equal weighted portfolio showing moderately lower volatility than the cap weighted portfolio.

In comparison to the S&P500 rolling one-year relative returns (Figure 2.3), the drawdowns in the past five to ten years have been particularly deep for the Top 40 and have regularly breached -10% for the Top 40 equal weighted portfolio. Table 2.4 confirms that the equal weighted portfolio in South Africa has underperformed significantly over the past five to ten years with relative returns of -8.6% and -13.2% in the period 2015 - 2017 and 2018 - 2020, respectively. As in the S&P500, the Top 40 equal weighted portfolio has staged somewhat of a recovery with a 10.6% relative return in the first five months of 2021.

Table 2.4.: Total return over non-overlapping three-year periods for the Top 40(%).

Period Start[a]	Period End	Equal weighted	Cap Weighted	Relative
2003	2005	134.3	107.6	12.9
2006	2008	25.6	30.7	-3.9
2009	2011	71.7	62.8	5.4
2012	2014	66.0	70.7	-2.7
2015	2017	19.0	30.2	-8.6
2018	2020	-2.6	12.2	-13.2
2021	2021/05[b]	19.5	8.0	10.6

[a]Period Starts 1 January and ends 31 December of Period End year

[b]Period ends 30 May 2021

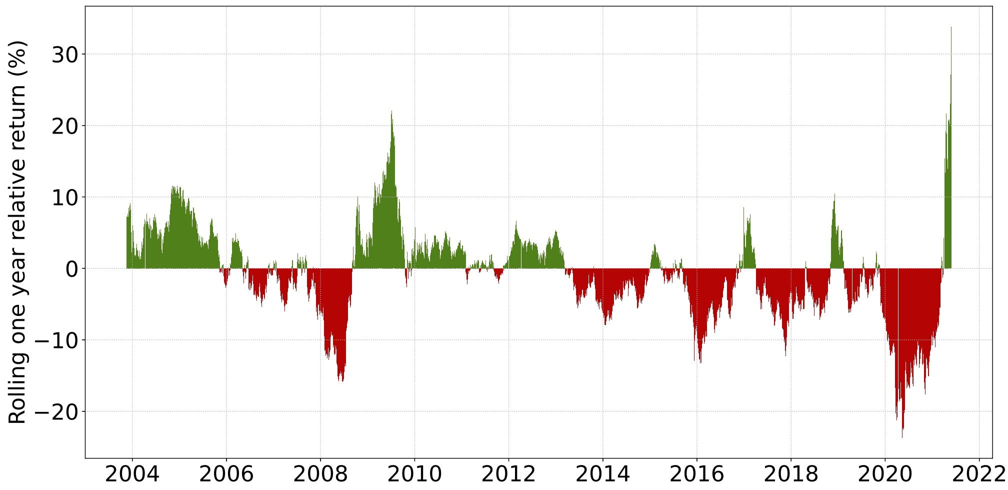

Figure 2.6.: Rolling one year relative return for the Top 40 equal weighted portfolio.

2.4 A matter of concentration?

Despite the more severe drawdowns since 2016 in the SA equal weighted portfolio, a cursory glance at Figures 2.3 and 2.6 seem to indicate some similarities between the relative performance in the US and SA equal weighted portfolios. Figure 2.7, which shows the rolling three year correlation of the relative returns for the equal weighted portfolios in both the US and SA, confirms a mostly positive relationship between the two. Generally the relationship is fairly strong, averaging 0.41 over the entire period, except for the period 2013 to 2017.

In this period, the Top 40 equal weighted portfolio seems to have underperformed while that of the S&P500 was largely flat (see Tables 2.2 and 2.4). If the cap weighted portfolio was outperforming the equal weighted portfolio in SA, it would be likely that stocks with higher cap weights would outperform those with lower cap weights. One example of this is Naspers (NPN) and the cap weight of NPN over the period 2012 to 2018 is shown in Figure 2.8. Over this period the cap weight of NPN increased from about 3% to over 15% in almost a straight line.

While it may be tempting to place all the blame for the equal weighted portfolio's performance with NPN, the SA equity market has a long history of battling with high levels of concentration. This is well documented, for example, in Kruger and Van Rensburg (2008) and Raubenheimer (2010), and led to the introduction of the FTSE/JSE Shareholder Weighted Top 40 Index (SWIX40) and the FTSE/JSE Capped Top 40 Index (CAPI40) as an attempt to reduce the levels of concentration.

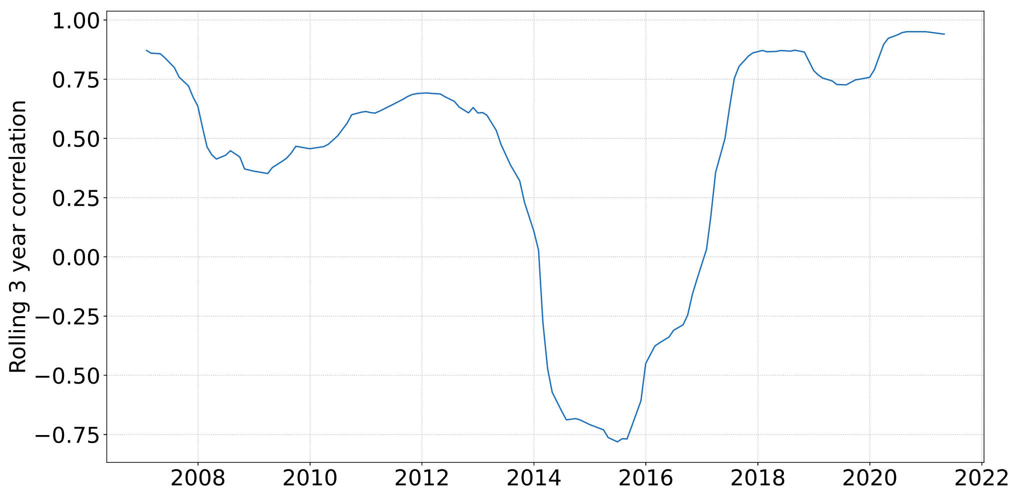

Figure 2.7.: Rolling three year correlation of monthly relative returns for the S&P500 and Top 40 equal weighted portfolios.

Figure 2.8.: Cap weight of Naspers (NPN) from 2012 to 2018.

Entropy is one way of visualising the changes in concentration. Entropy is a concept used in information theory and introduced by Shannon (1948). Entropy has also been used as a measure of diversity in stock markets. See Fernholz (1999a), for example.

Mathematically, entropy is defined as,

$$S(\mu(t)) = -\sum_{i=1}^{N} \mu_i(t)\log\mu_i(t), \qquad t \in [0, T], \tag{2.1}$$

where $\mu_i(t)$ represents the cap weight of stock i (given N stocks in the universe) at time t.

Figure 2.9 shows the level of entropy for both the S&P500 and the Top 40 since their respective start dates. Since 2012, entropy has declined in both the S&P500 and Top 40, although the decline has been much more significant in the case of the Top 40.

Figure 2.9.: Level of entropy normalised to starting entropy for relevant index.

These periods of increasing concentration (decreasing entropy) seem to coincide with periods of underperformance by the equal weighted portfolio in both the S&P500 and Top 40. In the period 1999 to 2004, for example, the S&P500 equal weighted portfolio outperformed the cap weighted portfolio by approximately 50% (Table 2.2) while over the same period the entropy for the S&P500 cap weights increased substantially from its lows in 1999. This points to some relationship between the level of concentration and the relative return of the equal weighted portfolio as one would expect.

At this stage, however, this analysis is somewhat anecdotal and lacks any mathematical rigor. As a result one can only guess about the direct impact these changes in concentration have on the equal weighted portfolio and ignoring other factors that may impact the relative returns of the equal weighted portfolio. In the next section, however, stochastic portfolio theory is introduced. This is a framework that will allow for the analysis of the various components driving the relative returns of the equal weighted portfolio. Our use of stochastic portfolio theory is on the basis that it requires relatively few assumptions, is based on a widely used model of stocks, and provides a direct approach to modelling and understanding the drivers of returns of the equal weighted portfolio relative to the cap weighted portfolio.

2.5 Markets other than the US and SA equity markets

This section takes a brief look at the performance of the equal weighted portfolio in countries other than the US and SA equity markets, relative to the respective cap weighted indices or portfolios. The respective cap weighted indices used are provided in Table 1.2, although as is the case in the US and SA analysis, these indices are used for membership purposes only and the cap weighted portfolios are constructed on a monthly basis. This analysis begins in 2003 to align all the countries to the same start date (this is the SA equal weighted portfolio start date).

Table 2.5 is a reproduction of Table 1.1 showing the annualised returns and risk-adjusted returns since 2003 for each equity market and shown here for ease of reference. Generally the equal weighted portfolio performs well relative to the cap weighted portfolio. Only two equity markets show an underperformance for the equal weighted portfolio (Australia and South Africa). Risk-adjusted returns are, furthermore, higher for the equal weighted portfolio except for Australia, and South Africa and are equal for France and South Africa. Information ratios are also positive for the equal weighted portfolios, except for Australia which is moderately negative.

Figure 2.10 shows the log cumulative relative return of the various equal weighted portfolios. That is, the cumulative return relative to the respective cap weighted portfolio on a log-scale. Japan, appears to be the best performing country for the equal weighted portfolio with returns of approximately 2.7% per annum since 2003. The TOPIX index used here has over 2000 members and is perhaps indicative of improved performance of the equal weighted portfolio against cap weighted indices with a large membership.

Conversely, the worst performing country is Australia with relative returns of approximately -1.3% per annum. The majority of this underperformance is, however, due to the 2012 to 2016 period when the resources sector underperformed the overall equity market. Relative returns have been largely flat since then. South Africa's

Table 2.5.: Annualised returns, volatility and risk-adjusted returns of the cap- and equal weighted portfolios for a selection of countries analysed in this thesis. All portfolios start 1 January 2003. No transaction costs included.

	CAGR (%)		Volatility (%)		Sharpe ratio[a]		Information ratio[a]
Country	**Cap**	**Equal**	**Cap**	**Equal**	**Cap**	**Equal**	**Equal**
Australia	12.1	11.5	16.4	17.6	0.57	0.50	-0.01
Canada	9.7	10.6	18.1	18.3	0.52	0.58	0.23
France	8.0	8.2	21.9	23.0	0.42	0.42	0.21
Germany	8.7	9.7	22.5	22.1	0.45	0.49	0.25
Japan	7.1	11.3	21.2	20.0	0.43	0.63	0.60
South Africa	14.7	14.6	20.2	19.4	0.42	0.42	0.00
United Kingdom	7.2	10.0	18.6	19.4	0.36	0.49	0.61
United States[b]	11.3	13.1	19.4	21.6	0.58	0.62	0.41

[a]Respective risk-free rates used in Sharpe and Information ratio calculations.

[b]These results for the US are slightly different from those in Table 2.1 due to the change in start dates to align to 1 January 2003.

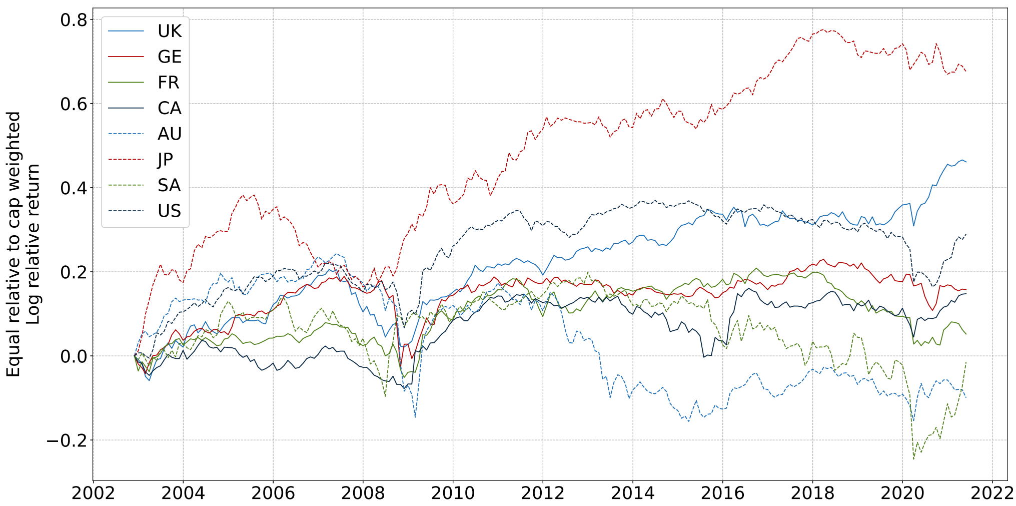

Figure 2.10.: Equal weighted portfolio log cumulative return relative to cap weighted portfolio.

performance on the other hand has been in a steady decline since 2012, somewhat in line with France's relative performance. As noted above, the SA equal weighted portfolio had a poor start to 2020 but recovered well thereafter. Still, the equal weighted portfolio in South Africa has a poor relative performance overall.

Overall performance is somewhat mixed among the different countries, although the equal weighted portfolio does outperform the cap weighted portfolio in all but three equity markets. Furthermore, the equal weighted portfolio has performed exceptionally well in some countries (Japan and the UK) and particularly poorly in others (Australia, France and South Africa). The size of the index membership was mentioned in relation to the Japanese experience, although it is less clear when comparing the performance of South Africa (which has 40 members) and Australia (which has 200 members) and so this explanation is not sufficient to describe all the disparities in relative performance.

Figure 2.11 highlights the calendar relative returns of the various equal weighted portfolios. This data is also reflected in Table 2.6. Most countries show a similar structure in relative returns up until 2012. Positive relative returns are recorded until 2007 for the most part and thereafter almost all equal weighted portfolio experience negative returns in 2007 and 2008 before recovering in the following two or three years. Thereafter, some countries (such as France and Germany) experience very small positive relative returns before a run of underperformance from 2018 onwards. Canada and Australia, for example, also experience similar relative returns with negative relative returns since 2012 to 2016 and a brief period of outperformance before returns become somewhat mixed.

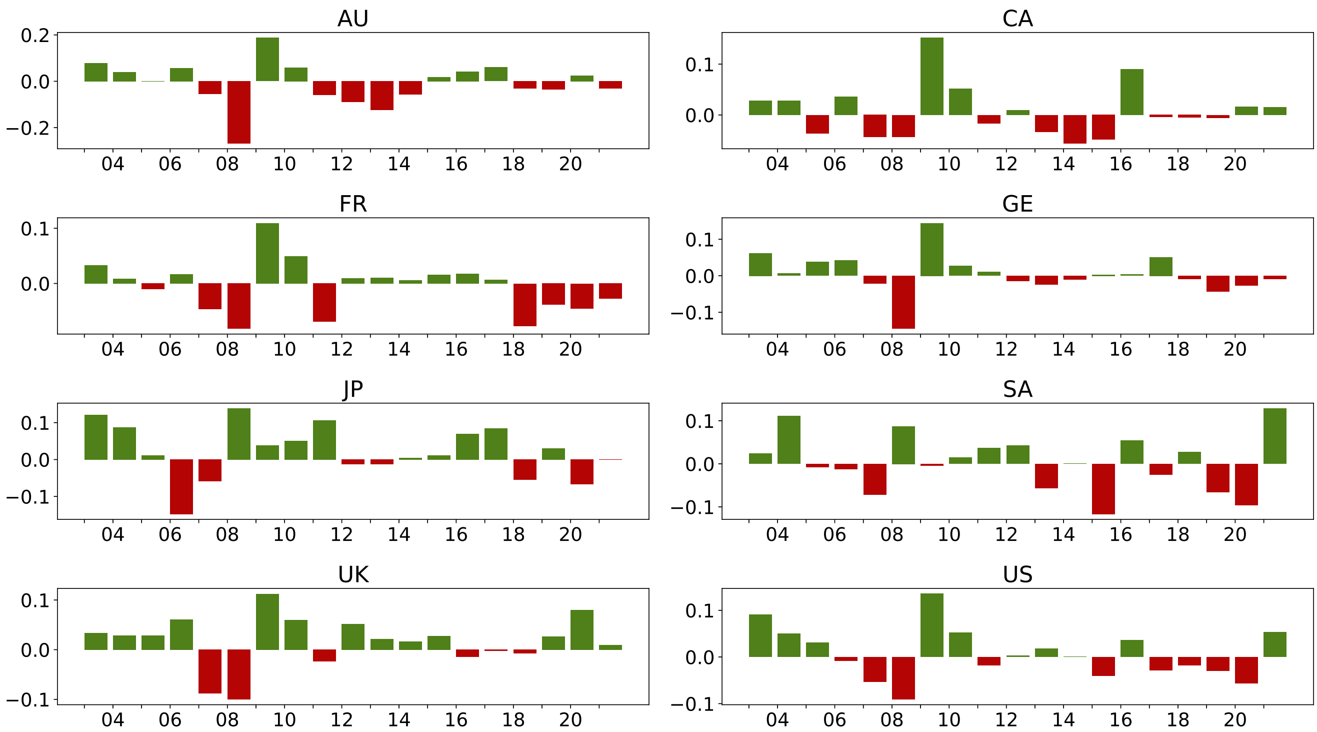

Figure 2.11.: Annual relative performance of equal weighted portfolio by country.

Table 2.6.: Annual relative performance of equal weighted portfolio by country.

Year	AU	CA	FR	GE	JP	SA	UK	US
2003	7.8	2.8	3.3	6.2	12.2	2.4	3.4	9.1
2004	4.0	2.8	0.8	0.6	8.7	11.0	2.9	5.0
2005	0.2	-3.6	-1.1	3.8	1.1	-0.7	2.8	3.1
2006	5.8	3.6	1.7	4.2	-14.8	-1.3	6.1	-0.8
2007	-5.5	-4.4	-4.6	-2.1	-5.9	-7.1	-8.8	-5.3
2008	-26.8	-4.3	-8.2	-14.5	13.8	8.7	-10.0	-9.1
2009	18.8	15.2	10.9	14.4	3.9	-0.4	11.3	13.5
2010	5.9	5.2	4.9	2.7	5.0	1.4	6.0	5.2
2011	-5.9	-1.6	-6.9	1.0	10.6	3.7	-2.3	-1.7
2012	-9.0	0.9	1.0	-1.5	-1.3	4.3	5.2	0.3
2013	-12.4	-3.3	1.0	-2.4	-1.3	-5.6	2.2	1.8
2014	-5.8	-5.6	0.6	-1.1	0.5	0.1	1.7	0.1
2015	1.9	-4.9	1.6	0.3	1.2	-11.7	2.8	-4.0
2016	4.3	9.0	1.8	0.4	6.9	5.4	-1.5	3.6
2017	6.0	-0.4	0.7	5.1	8.5	-2.5	-0.2	-2.9
2018	-3.1	-0.5	-7.7	-0.9	-5.5	2.8	-0.8	-1.8
2019	-3.5	-0.5	-3.9	-4.4	3.0	-6.6	2.6	-3.0
2020	2.4	1.6	-4.5	-2.6	-6.7	-9.6	8.0	-5.7
2021[a]	-3.1	1.5	-2.8	-0.9	0.0	12.9	1.0	5.3

[a]Period Starts 1 January and ends 31 May 2021

On the whole, however, most countries, experience a similar trend; the period before 2012 shows high levels of outperformance with some underperforming years, which is followed by the period from 2012 onwards where positive relative returns are muted but there are periods of deep underperformance. Although this general trend can be seen in Figure 2.11, the overall correlation between the different countries is relatively small. In Figure 2.7, the rolling three-year correlation between the US and SA equal weighted portfolio is quite high. Figure 2.12, however, shows a much lower average correlation between the different equal weighted portfolios on a rolling three-year basis. This has averaged between 15% and 20% since 2006 in contrast to Figure 2.7 which shows an average of 60% ot 70% between the US and SA equal weighted portfolio.

At this stage it is difficult to discern where these differences come from. Chapter 3, however, presents one framework that can help in attributing the relative return of the equal weighted performance.

Figure 2.12.: Rolling 12 month average correlation between equal weighted relative performances.

2.6 Conclusion

This chapter presented the historical performance of the monthly rebalanced equal weighted portfolio against the cap weighted portfolio. This was primarily done for the US and SA equity markets, with the final section discussing the performance in a handful of other countries.

In terms of the US and SA equity markets, there appears to be a period of strong positive correlation between the relative returns of the US and SA equal weighted portfolios with both portfolios performing well in the early 2000s, while underperforming over the past five to ten years. In the case of the SA equal weighted portfolio, relative returns have been particularly poor since 2012. This has coincided with a large increase in concentration in the Top 40. Although the concentration in the S&P500 had also increased over this period, the increase in the Top 40's concentration was much more pronounced. Relative returns have, however, improved substantially since March 2020 for both the US and SA equal weighted portfolios.

The relative performance of the other countries analysed is rather mixed. The standout equal weighted portfolio is that of the TOPIX in Japan. The UK equal weighted portfolio has also performed well against its cap weighted portfolio. While the SA equal weighted portfolio performs poorly, its Australian counterpart performs even worse. The majority of this underperformance comes in the period just after 2012, however, with relative performance flat thereafter. This is in contrast with the equal weighted portfolio in SA which has experienced a steady decline in relative performance since 2012.

This page has been intentionally left blank

3 Introduction to stochastic portfolio theory

3.1 Introduction

This chapter provides a brief overview of the necessary expressions from stochastic portfolio theory, as developed by Fernholz (2002). A more formal, mathematical treatment is presented in Appendix A. Karatzas (2006) and Fernholz and Karatzas (2009) are, furthermore, suggested for a complete overview of stochastic portfolio theory.

Two main parts of stochastic portfolio theory are covered in the following subsections, namely the portfolio price process and portfolio generating functions. Other topics of stochastic portfolio theory, such as the capital distribution curve, see Fernholz (2001b), and stock selection by rank, see Fernholz (2001a), are not considered here.

The chapter combines an intuitive explanation of the portfolio price process using the familiar geometric Brownian motion model for the stock price process with the necessary mathematical expressions. This is extended to consider a portfolio's return relative to the cap weighted portfolio and the concept of leakage. This chapter is structured as follows:

- The stock price process and extending this to a portfolio of stocks and the resulting portfolio price process.
- Returns relative to the cap weighted portfolio.
- Portfolio generating functions.
- Portfolios on subsets of the market and the concept of leakage.

The intuitive discussion of stochastic portfolio theory and the equal weighted portfolio's relative return found in Taljaard and Maré (2021a) form the basis of this chapter and the more technical expressions found in Appendix A can also be found in Taljaard and Maré (2021b).

3.2 The portfolio price process

Consider two stocks, A and B, that are negatively correlated. In each period, a stock has a 50% chance of a 25% gain, and a 50% chance of a 20% loss. Stock A, for example, may have a 25% gain in one period, which would mean stock B has a 20% loss in the same period (and vice versa). The return in each period is independent of any other period. Consider an equal weighted portfolio between the two which is rebalanced each period. The geometric growth rate of each stock is 0% and yet the expected growth rate of the equal weighted portfolio would be 2.5% in each period. This return is a function of the volatility of the two stocks and the correlation structure between them. In Bouchey, Nemtchinov, Paulsen, et al. (2012), Bouchey, Nemtchinov, and Wong (2015) and Hallerbach (2014), this is referred to as volatility harvesting or volatility return. In stochastic portfolio theory, this return is referred to as the portfolio excess growth rate and represents the return attributable to diversification in the portfolio. The idea of diversification directly impacting the return of a portfolio is echoed by Booth and Fama (1992) and Cuthbertson et al. (2016).

Mathematically, consider a stock i with stock price represented by X_i and assume that the price process follows the popular logarithmic model for the continuous-time stock process:

$$d\log X_i(t) = \gamma_i(t)dt + \sigma_i(t)dW_i(t) \qquad t \in [0, \infty), \tag{3.1}$$

where $\sigma_i(t)$ is the volatility of stock $X_i(t)$ and $dW_i(t)$ represent a Brownian motion. $\gamma_i(t)$ is the geometric growth rate of stock $X_i(t)$ related to the arithmetic growth rate, $\alpha_i(t)$, by

$$\gamma_i(t) = \alpha_i(t) - \frac{1}{2}\sigma_i^2(t).$$

Consider a set of stocks $X_1, \ldots, X_n$ that each follow the price process as in Equation (3.1) and construct a long-only portfolio with weights given by $\pi(t) = (\pi_1(t), \ldots, \pi_n(t))$, such that

$$\sum_{i=1}^{n} \pi_i(t) = 1,$$

and

$$\pi_i(t) \geq 0, \quad \forall i = 1, \ldots, n.$$

The portfolio price process, denoted $Z_\pi(t)$, follows:

$$\frac{dZ_\pi(t)}{Z_\pi(t)} = \sum_{i=1}^{n} \pi_i(t) \frac{dX_\pi(t)}{X_\pi(t)}. \tag{3.2}$$

Equation (3.1) can be rewritten as,

$$dX_i(t) = \alpha_i(t)dt + \sigma_i(t)dW_i(t), \tag{3.3}$$

and substitute it into Equation (3.2) and, after some simplification, obtain

$$d\text{log}Z_\pi(t) = \gamma_\pi(t) + \sum_{i=1}^{n} \pi_i(t)\sigma_i dW_i(t), \tag{3.4}$$

where

$$\gamma_\pi(t) = \sum_{i=1}^{n} \pi_i(t)\gamma_i(t) + \frac{1}{2}\left(\sum_{i=1}^{n} \pi_i(t)\sigma_{ii}(t) - \sum_{i,j=1}^{n} \pi_i(t)\pi_j(t)\sigma_{ij}(t)\right), \tag{3.5}$$

where $\sigma_{ij}(t)$ represents the covariance process for stocks X_i and X_j, and $\sigma_{ii}(t)$ representing the variance process of X_i. See Proposition A.1.1 for a formal proof.

The portfolio's growth rate, $\gamma_\pi(t)$, consists of two distinct parts: the weighted growth rates of the individual stocks and a second term involving the stocks' weighted volatilities and covariances. This second term is referred to as the excess growth rate and is given by

$$\gamma_\pi^*(t) = \frac{1}{2}\left(\sum_{i=1}^{n} \pi_i(t)\sigma_{ii}(t) - \sum_{i,j=1}^{n} \pi_i(t)\pi_j(t)\sigma_{ij}(t)\right). \tag{3.6}$$

Intuitively, Equation (3.6) can be thought of as a measure of diversification since it is the difference between the average stock volatility and the resulting portfolio's volatility. Although $\gamma_\pi^*(t) \geq 0$, its absolute level will vary depending on the individual stock volatilities and their correlations. If the correlation between stocks is small, then γ_π^* would be larger leading to a much higher contribution to the overall portfolio growth rate. Conversely, if correlations are high then there is less diversification and γ_π^* is lower. Similarly, should individual stock volatilities be low, then γ_π^* would also be low as there would be less scope for a reduction in portfolio volatility.

This implies, therefore, that there is a direct relationship between a portfolio's return and its level of diversification. Returning to the example above, since the individual stock growth rates are zero, $(1.25)(0.8) - 1 = 0$, the portfolio growth rate is equal to the excess growth rate.

$$\gamma_\pi^* = \frac{1}{2}\left(\pi_1(1-\pi_1)\sigma_1^2 + \pi_2(1-\pi_2)\sigma_2^2 - 2\sigma_{1,2}\pi_1\pi_2\right),$$

where

$$\pi_1 = \pi_2 = 0.5,$$

and

$$\sigma_1^2 = \sigma_2^2 = -\sigma_{1,2} = 0.5.$$

This yields a portfolio growth rate of

$$\gamma_\pi = \gamma_\pi^* = 0.025,$$

or 2.5%, as expected.

Note that, by inserting Equation (3.5) into Equation (3.4) and rearranging the terms, one can rewrite the portfolio price process as,

$$d\log Z_\pi(t) = \sum_{i=1}^{n} \pi_i(t) d\log X_i(t) + \gamma_\pi^*(t)dt. \tag{3.7}$$

Again the direct impact of diversification on the portfolio price process is highlighted with $d\log Z_\pi(t)$ a function of two parts: the weighted price processes of the individual stocks and the excess growth rate, representing the diversification benefit within the portfolio.

The above formulae are easily extended to include dividends. Given a dividend process for stock $X_i(t)$, represented as $\delta_i(t)$, then the total return for stock $X_i(t)$ is

$$\hat{X}_i(t) = X_i(t)\exp\left(\int_0^t \delta_i(s)ds\right), \qquad t \in [0, \infty). \tag{3.8}$$

Therefore, Equation (3.1) can be extended for the total return process,

$$d\log\hat{X}_i(t) = d\log X_i(t) + \delta_i(t)dt, \tag{3.9}$$

and the total return process for the portfolio given by weights $\pi(t)$ becomes,

$$d\log\hat{Z}_\pi(t) = d\log Z_\pi(t) + \delta_\pi(t)dt, \tag{3.10}$$

where

$$\delta_\pi(t) = \sum_{i=1}^{n} \pi_i(t)\delta_i(t). \tag{3.11}$$

3.3 Relative portfolio returns

Although stand-alone portfolio performance is important to consider, in practice most portfolio managers must consider portfolio performance relative to some benchmark. The most often used benchmark for portfolio managers of equities is the cap weighted portfolio or index.

Consider the cap weighted portfolio price process given by $Z_\mu(t)$, where the portfolio weight of stock i at time t, $\mu_i(t)$, is given by

$$\mu_i(t) = \frac{X_i(t)}{\sum_{i=1}^{n} X_i(t)}, \tag{3.12}$$

where $X_i(t)$ is stock i's market capitalization at time t. $Z_\mu(t)$ is just like any other portfolio price process and, therefore, the expressions in Section 3.2 apply to $Z_\mu(t)$, with the only difference being the weights themselves.

Equation (3.12) can also be written as,

$$\mu_i(t) = \frac{X_i(t)}{Z_\mu(t)}. \tag{3.13}$$

Although, this portfolio process follows the same structure in Equation (3.4), the difference is that its weights change given each stock's relative performance over a period. Consider the example used in Section 3.2. Of course, each period in the example is random and independent from the previous period; however, consider an example where the returns for each of the stocks alternates from one period to the next. If in period 1, for example, stock A has a return of -20% and stock B returns 25%, then in period 2 stock A would return 25% and stock B would return -20%. The equal weighted portfolio would be unaffected, as at each period the portfolio would be rebalanced. The cap weighted portfolio, however, would become concentrated in the stock which had outperformed in the prior period only for that stock to underperform in the next.

Stock A, in this example, would have a cap weight of 39% after period 1, which would revert to 50% after period 2, in which it outperforms stock B. In relative terms, the equal weight outperforms the cap weighted portfolio (by about 5% in this example), and it does so for two reasons:

- The cap weighted portfolio becomes more concentrated in a single stock (stock B in this example) after the first period only for that stock to then underperform the other. In other words, the cap weights become concentrated in period 1 and then less concentrated in period 2 and this directly impacts the cap weighted portfolio's return relative to the equal weighted portfolio.

- The cap weighted portfolio has a lower excess growth rate than that of the equal weighted portfolio since it is more concentrated in a single stock.

The simple example above demonstrates how movements in the cap weights affect the relative return of the equal weighted portfolio. In the example, the cap weight of each stock mean reverts, and this benefits the equal weighted portfolio's relative return. The reverse can of course occur as well; the cap weights could exhibit momentum and the stocks with larger cap weights could continue to outperform the average stock leading to further concentration in the cap weighted portfolio. This would have the reverse effect on the equal weighted portfolio and create a drag on relative performance.

Consider Equation (3.7) and subtract $d\log Z_\mu(t)$ on each side:

$$d\log Z_\pi(t) - d\log Z_\mu(t) = \sum_{i=1}^{n} \pi_i(t) d\log X_i(t) - d\log Z_\mu(t) + \gamma_\pi^*(t)dt.$$

Noting that the weights $\pi_i(t)$ sum to one, this can be rewritten as:

$$d\log\left(\frac{Z_\pi(t)}{Z_\mu(t)}\right) = \sum_{i=1}^{n} \pi_i(t) d\log\left(\frac{X_i(t)}{Z_\mu(t)}\right) + \gamma_\pi^*(t)dt. \tag{3.14}$$

Equation (3.14) is true for any two portfolios, however, in the case of the cap weighted portfolio, and using Equation (3.13), Equation (3.14) can be simplified into:

$$d\log\left(\frac{Z_\pi(t)}{Z_\mu(t)}\right) = \sum_{i=1}^{n} \pi_i(t) d\log\mu_i(t) + \gamma_\pi^*(t)dt. \tag{3.15}$$

That is, the equal weighted portfolio's return relative to the cap weighted portfolio is equal to a term involving the change in the log cap weights and the portfolio's excess growth rate. Equation (3.15) holds for any portfolio with weights $\pi(t)$, although the focus here is on the equal weighted portfolio.

When considering the equal weighted portfolio, in particular, the first term in Equation (3.15) can be rewritten as:

$$\begin{aligned}\sum_{i=1}^{n} \pi_i(t) d\log\mu_i(t) &= \frac{1}{n}\sum_{i=1}^{n} d\log\mu_i(t) \\ &= \frac{1}{n} d\log\left(\prod_{i=1}^{n}\mu_i(t)\right) \\ &= d\log\left(\prod_{i=1}^{n}\mu_i(t)\right)^{\frac{1}{n}}.\end{aligned}$$

That is when considering the equal weighted portfolio relative to the cap weighted portfolio, the first term becomes the change in the log of the geometric mean of the cap weights. The log of the geometric mean will be negative, given that the cap weights are less than one. Therefore, as the geometric mean increases, the log of the geometric mean decreases. That is, as the cap weights become more concentrated, the log of the geometric mean will decline (become more negative) and cause a drag on the equal weighted portfolio's relative performance, through Equation (3.15). In summary then, the relative return of the equal weighted portfolio is a function of the change in the geometric mean of cap weights and the excess growth rates.

An important assumption in stochastic portfolio theory is that no one stock can dominate the market or, alternatively, the maximum cap weight for a stock is bounded. Under this assumption the geometric mean of cap weights, and the term containing it in Equation (3.15), is bounded and over the long-term its contribution to the relative performance of the equal weighted portfolio should be negligible. In the context of Equation (3.15), this implies that over the long-term the equal weighted portfolio is likely to outperform the cap weighted portfolio since the excess growth rate is greater than, or equal to, zero and the geometric mean term is bounded.

This confirms, theoretically at least, the findings of empirical studies into equal weighted portfolio performance such as Plyakha et al. (2012), Bolognesi et al. (2013), and Malladi and Fabozzi (2017). Although the geometric mean term is bounded over the long-term, in the short-term, this geometric mean term can be a significant driver of relative performance, dominating the excess growth rate term in Equation (3.15).

3.4 Portfolio generating functions

In stochastic portfolio theory the geometric mean of cap weights is called the portfolio generating function of the equal weighted portfolio. Portfolio generating functions are functions that generate various portfolio weights using, as an input, the market capitalization weights, $\mu(t)$. They are defined as follows (see Fernholz (2002), Definition 3.1.1, reproduced below for convenience):

Definition 3.4.1. Let **S** be a positive continuous function defined on Δ^n (the unit n-simplex) and let π be a portfolio. Then **S** generates π if there exists a measurable process of bounded variation Θ such that:

$$\log\left(Z_\pi(t)/Z_\mu(t)\right) = \log\mathbf{S}(\mu(t)) + \Theta(t), \qquad t \in [0,T], \quad \text{a.s.} \tag{3.16}$$

The process Θ is called the drift process corresponding to **S**.

Equation (3.16) can also be expressed in differential form,

$$d\log\left(Z_\pi(t)/Z_\mu(t)\right) = d\log\mathbf{S}(\mu(t)) + d\Theta(t), \qquad t \in [0,T], \quad \text{a.s.} \tag{3.17}$$

If dividends are included then, using Equation (3.10), the total relative return process becomes:

$$d\log\left(\hat{Z}_\pi(t)/\hat{Z}_\mu(t)\right) = d\log\mathbf{S}(\mu(t)) + \int_0^t (\delta_\pi(s) - \delta_\mu(s))ds + d\Theta(t), \qquad t \in [0,T], \quad \text{a.s.} \tag{3.18}$$

In other words, the relative performance of a portfolio with weights $\pi(t)$ is a function of the portfolio generating function **S**, the difference in dividend rates and the drift process.

As mentioned above, the geometric mean of cap weights is the portfolio generating function of the equal weighted portfolio and one can, therefore, rewrite Equation (3.15) as follows:

$$d\log\left(\frac{Z_\pi(t)}{Z_\mu(t)}\right) = d\log\mathbf{S}(\mu(t)) + \gamma_\pi^*(t)dt, \tag{3.19}$$

where,

$$\mathbf{S}(\mu(t)) = \left(\prod_{i=1}^{n}\mu_i\right)^{\frac{1}{n}}. \tag{3.20}$$

Comparing this with Equation (3.16), it is clear that the drift term, $d\Theta(t)$, for the equal weighted portfolio should be equal to the excess growth rate of the equal weighted portfolio. More formally, the drift process of the equal weighted portfolio, generated by the function in Equation (3.20), is given by

$$d\Theta(t) = \gamma_\pi^*(t)dt, \tag{3.21}$$

where $\gamma_\pi^*(t)$ is defined as the excess growth rate in Equation (3.6).

Considering this in the context of Equation (3.17), the relative performance of the equal weighted portfolio over some time period is given by the addition of the change in $\frac{1}{n}\log\left(\mu_1(t)\cdots\mu_n(t)\right)$ and the rate of diversification, given by $\gamma_\pi^*(t)dt$.

Clearly, $\gamma_\pi^*(t)$ is always positive, however, the change in $\log\mathbf{S}(\mu)$ is dependent on the change of the distribution of weights, $\mu_i(t)$ in the cap weighted portfolio. If the concentration increases, the change in $\log\mathbf{S}(\mu)$ is negative and detracts from the equal weight portfolio's performance relative to the cap weighted portfolio and vice versa.

In a scenario where the cap weighted portfolio becomes increasingly concentrated, the change in $\log\mathbf{S}(\mu)$ becomes a short-term, but consistent, drag on relative performance through Equation (3.17). As a result, the excess growth rate will have to offset this term in order for the equal weighted portfolio to outperform the cap weighted portfolio. Although always positive, it is not a given that in every period the excess growth rate will be high enough to offset the growing concentration of market capitalization weights and this can, therefore, lead to short-term underperformance.

The expressions for the weights, $\pi_i(t)$, as well as the drift process, $\Theta(t)$, can be derived using Theorem 3.1.5 in Fernholz (2002). This theorem and its proof are reproduced in Appendix A, Theorem A.2.1. The relevant expressions for the weights and drift process are reproduced below:

$$\pi_i(t) = \left(D_i \log \mathbf{S}(\mu(t)) + 1 - \sum_{j=1}^{n} \mu_j(t) D_j \log \mathbf{S}(\mu(t)) \right) \mu_i(t), \tag{3.22}$$

for $t \in [0, T]$ and $i = 1, \ldots, n$ and with a drift process Θ such that a.s., for $t \in [0, T]$,

$$d\Theta(t) = \frac{-1}{2\mathbf{S}(\mu(t))} \sum_{i,j=1}^{n} D_{ij}\mathbf{S}(\mu(t))\mu_i(t)\mu_j(t)\tau_{ij}(t)dt. \tag{3.23}$$

The notation D_i represents the partial derivative with respect to the ith variable and $\tau_{ij}(t)$ represents the relative covariance of stock i and j. That is,

$$\tau_{ij}(t) = \langle \log\left(X_i/Z_\mu\right), \log\left(X_j/Z_\mu\right) \rangle .$$

See Appendix A and Equations (A.17) and (A.19) for a formal definition of $\tau_{ij}(t)$.

Using the expressions above one can formally derive the expressions for the portfolio generating function and drift process of the equal weighted portfolio. This is shown in Appendix A.3.

3.5 Portfolios on subsets of the whole market

In the previous sections, portfolios formed on the entire equity market were considered. However, in practice, an index formed on the top stocks is used as a benchmark and its constituents are used to form active portfolios. So far the S&P500 and Top 40 have been considered, both of which would be a subset of the entire US and SA equity markets, respectively. As a result, these indices are implicitly selecting stocks by rank in both the cap weighted and equal weighted portfolios.

This requires a modification of the drift process, $d\Theta(t)$, in Equation (3.23) by introducing a new term $dL_\pi(t)$, namely

$$d\Theta(t) = \frac{-1}{2\mathbf{S}(\mu(t))} \sum_{i,j=1}^{n} D_{ij}\mathbf{S}(\mu(t))\mu_i(t)\mu_j(t)\tau_{ij}(t)dt + dL_\pi(t), \tag{3.24}$$

where

$$dL_\pi(t) = \frac{1}{2}\sum_{k=1}^{n-1} \left(\pi_{(k+1)}(t) - \pi_{(k)}(t)\right) d\Lambda_{\log \mu_{(k)} - \log \mu_{(k+1)}}(t), \qquad t \in [0,T]. \tag{3.25}$$

The terms $\pi_{(k)}(t)$ and $\mu_{(k)}(t)$ represent the portfolio weight and the market cap weight of the k-th largest stock, respectively, where the stocks are ranked by market cap weights from largest to smallest over the entire market.

Λ is the semi-martingale local time process defined as,

$$\Lambda_X(t) = \frac{1}{2}\left(|X(t)| - |X(0)| - \int_0^t sgn(X(s))dX(s)\right), \qquad t \in [0,T]. \tag{3.26}$$

This new term dL_π can be thought of as the shift in weights within the market, and specifically how that affects portfolio selection both in terms of the equal weighted portfolio and the benchmark, which is a subset of the whole market. In our case here, with the S&P500 or any other index, the changes that would affect portfolio selection occurs at the edges. That is, the 500th stock, for example, becoming the 501st stock and falling out of the index and therefore out of the individual portfolios. This term is, therefore, defined as leakage, given that its size is determined by stocks "leaking" out of the portfolio.

As a result, and in the specific case of the equal weighted portfolio, this leakage term, over small time periods, usually only involves the last position in the index. That is, in the case of the S&P500, only changes in the 500th position over a small time period are of interest. This change (the last stock being replaced) is likely to affect the equal weighted portfolio much more than a cap weighted portfolio, given that the final few weights are likely to already be small in the case of the cap weight. The equal weight on the other hand, typically, has larger weights in those last few stocks and is, therefore, more affected by exclusions.

The leakage term defined in Equation (3.25) is of a portfolio relative to the entire market. In the case of the equal weighted portfolio with weights $\pi_{(k)}(t)$ relative to the cap weighted portfolio with weights $\xi_{(k)}(t)$, where both portfolios are a subset of the entire market, the leakage term corresponding to the relative return of the equal weighted against this cap weighted portfolio can be expressed as:

$$dL_{\pi/\xi}(t) \quad = \quad \frac{1}{2}\sum_{k=1}^{n-1}\Big([\xi_{(k+1)}(t)-\xi_{(k)}(t)]-[\pi_{(k+1)}(t)-\pi_{(k)}(t)]\Big)\,d\Lambda_{\log\mu_{(k)}-\log\mu_{(k+1)}}(t). \tag{3.27}$$

This is the difference in impact of the leakage term on the cap weighted subset portfolio and the equal weighted portfolio in our case. As explained above, the equal weighted portfolio is likely to hold more weight in the stocks exiting the index and, therefore, Equation (3.27) is a negative drag on the equal weighted portfolio's performance relative to the cap weighted portfolio in most cases.

Therefore, rewriting Equation (3.18), the relative return of the equal weighted portfolio (with weights $\pi(t)$) to a cap weighted portfolio (with weights $\xi(t)$) formed on a subset of the market can be given by:

$$\begin{aligned} d\log\left(Z_\pi(t)/Z_\xi(t)\right) \quad &= \quad d\log\mathbf{S}(\xi(t)) + d\Theta(t) \\ &+ \quad \int_0^t(\delta_\pi(s)-\delta_\mu(s))ds + dL_{\pi/\xi}(t), \qquad t\in[0,T], \quad \text{a.s.} \end{aligned} \tag{3.28}$$

3.6 Conclusion

This chapter introduced the basic concepts of stochastic portfolio theory as it relates to the return of functionally generated portfolios relative to the cap weighted portfolio (a common portfolio benchmark). Stochastic portfolio theory provides a robust theoretical framework for understanding this relative performance by considering the drivers of the relative performance, namely, the level of concentration in cap weights (portfolio generating function), the level of diversification benefits (the excess growth rate), and the impact of leakage (as stocks move in and out of the index). This thesis is specifically interested in the equal weighted portfolio and discussed how the generating function of this portfolio is the geometric mean of cap weights, how this should be bounded, and what this implies for the long-term relative return. The concepts discussed in this chapter, and specifically Equation (3.28) are crucial for the empirical analysis performed in Chapter 4.

4 Attribution of the equal weighted portfolio's performance

4.1 Introduction

Chapter 3 provided an introduction to stochastic portfolio theory and set out the necessary theoretical concepts to analyse the performance of the equal weighted portfolio against a cap weighted portfolio. This chapter makes use of these expressions, and more specifically Equation (3.28), to provide an attribution of the performance of the equal weighted portfolio. This chapter extends Chapter 2, where the empirical performance of the S&P500 and Top 40 equal weighted portfolios were discussed.

Sections 4.2 and 4.3 look at the S&P500 and Top 40, respectively, on their own before discussing some similarities between the two markets. Taljaard and Maré (2021b) and Taljaard and Maré (2021a) form the basis of the results for the S&P500 and Top 40, respectively, in these two sections. The final section of this chapter, Section 4.6, extends this analysis and presents the attribution of the equal weighted portfolios in the other selected equity markets noted in Table 1.2.

The four main components to calculate in order to provide an attribution of relative performance of the equal weighted portfolio are:

A Portfolio generating function, given by Equation (3.20) for the equal weighted portfolio,

B Portfolio drift process, which in the case of the equal weighted portfolio is given by the equal weighted portfolio's excess growth rate, calculated using Equation (3.6),

C The difference in dividend rates, calculated as the difference in dividends received for each month, and

D The net impact of leakage, estimated using Equation (3.27).

The drift process is an instantaneous rate, however, it requires the estimation of the covariance of the underlying stocks which requires the use of historical rates of return. The estimate of the covariance used in this analysis is made using the Exponentially Weighted Moving Average (EWMA) method using a decay factor (λ) of 0.97 and fit over the preceding three-year period. This is similar to the use of the EWMA covariance matrix within the JP Morgan RiskMetrics methodology (see RiskMetrics, 1996). The standard EWMA model was selected given its simplicity and literature suggesting that it does not perform much worse than a multivariate generalized autoregressive conditional heteroscedasticity (GARCH) model. See, for example, Zakamulin (2015), Pantaleo et al. (2011), and Kuang (2021).

The analysis in this chapter, furthermore, includes trading costs of 15 basis points, applied to both the equal weighted and cap weighted portfolio. This is roughly in line with analysis done by Frazzini et al. (2018) for stocks over the period 1998 to 2016.

4.2 Attribution of the S&P500

Figure 4.1 highlights the attribution of the relative return for the S&P500 equal weighted portfolio on a rolling one-year basis. Table 4.1 shows the decomposition of relative returns for consecutive three-year periods.

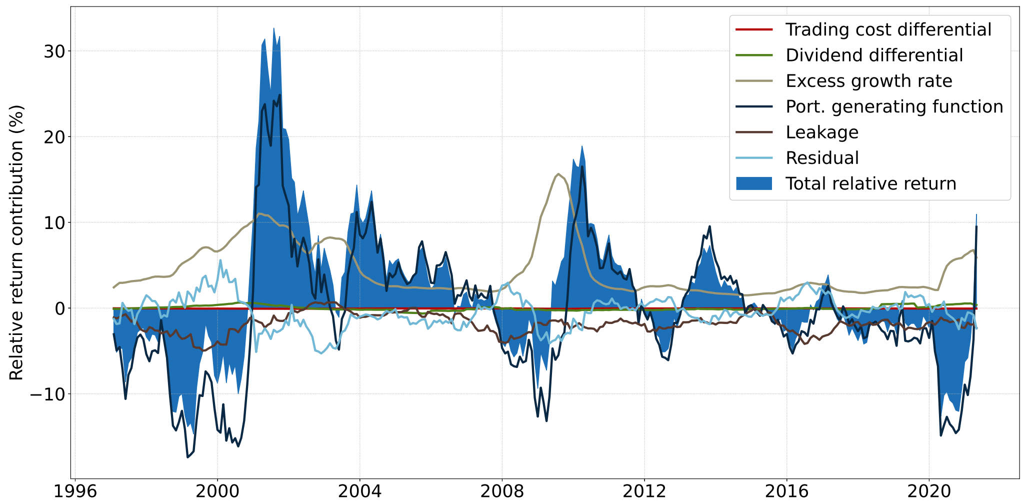

Figure 4.1.: Annual decomposition of S&P500 equal weighted portfolio relative return as per Equation (3.28) and including transaction cost differentials.

As expected, the portfolio generating function displays the highest volatility, with contribution to relative returns oscillating between 10% and −10% on a regular basis. In the late 1990s, the portfolio generating function reduced the relative return of the equal weighted portfolio by over 20% before contributing positively for the period 2000 to 2005. These changes are related to changes in concentration in the cap weighted portfolio weights and coincide with periods where concentration increased in the late 1990s, before concentration declined in the period 2000 to 2005. This is also evident from Figure 2.9, where entropy declines show increasing levels of concentration until 2000, and vice versa for the following five years.

Table 4.1.: Decomposition of the S&P500 equal weighted portfolio relative return over consecutive three-year periods.

Period Start[a]	Period End	Excess growth rate	Port. gen. function	Leakage	Dividends	Costs	Residual	Total
1996	1998	11.4	-22.5	-6.4	0.1	-0.1	0.0	-19.0
1999	2001	28.4	3.3	-6.2	1.1	-0.2	0.6	26.8
2002	2004	15.2	15.9	-0.8	-0.8	-0.2	-6.5	22.7
2005	2007	7.1	-1.2	-6.1	-0.5	-0.1	-0.4	-1.5
2008	2010	23.9	5.3	-5.4	-0.7	-0.2	-4.4	17.2
2011	2013	6.5	4.3	-6.0	-0.5	-0.1	-1.0	-2.8
2014	2016	6.6	-2.9	-5.9	-0.3	-0.1	2.3	-0.6
2017	2019	6.5	-7.8	-5.5	0.5	-0.1	-0.2	-7.0
2020	2021/05[b]	7.7	-5.4	-2.6	0.6	-0.1	-1.4	-1.4

[a]Period Starts 1 January and ends 31 December of Period End year

[b]Period ends 30 May 2021

Figure 4.2 shows the portfolio generating function for the equal weighted portfolio and also highlights the significant increase in concentration until 2000, after which concentration decreased. There appear to be two further periods of increasing concentration in cap weights. Firstly, in 2007-2008 during the financial crisis and then a slow increase in concentration since 2016 until the first half of 2020. Unlike the first two periods of increased concentration this slow increase in concentration led to a consistent negative impact on the relative return, with contributions of -2.9%, -7.8%, and -5.4% for the consecutive three-year periods starting in 2015 and ending in 2021 (Table 4.1).

The excess growth rate also appears to have declined since 2011, evident in Table 4.1 and evident in the contribution from the excess growth rate of 5.4% per annum from 1996 to 2010 and only 2.6% per annum from 2011 to 2021. Figure 4.3 shows the drift process (the excess growth rate for the equal weighted portfolio). There appears to be three periods of heightened excess growth rates; 2000 to 2003, 2008 to 2011, and 2020 more recently. These periods lead to strong outperformance by the equal weighted performance and usually coincide with a positive contribution from the portfolio generating function. Therefore, while the equal weighted portfolio should outperform the cap weighted portfolio over the long-term, much of the

Figure 4.2.: Portfolio generating function for the S&P500 equal weighted portfolio.

excess growth rate contribution to this long-term outperformance appears to come in clusters following a large market dislocation.

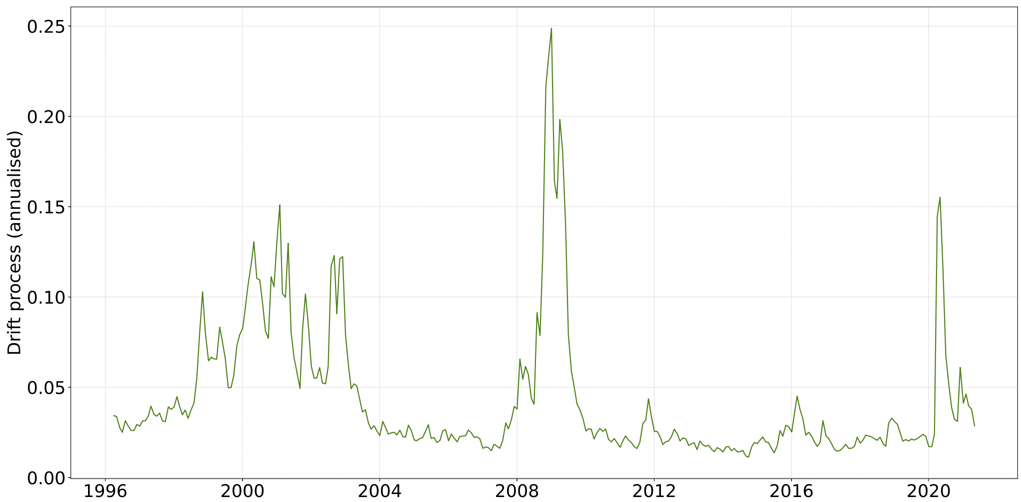

Figure 4.3.: Drift process (excess growth rate) for the S&P500 equal weighted portfolio.

Another large contributor (negatively) and an often unmentioned factor is that of leakage. As a reminder, leakage was presented in Equation (3.27) and is intuitively the result of having to replace stocks when they move in and out of the respective index. Typically, the impact of leakage is higher for the equal weighted portfolio than for the cap weighted portfolio and is, therefore, a net drag on the equal weighted portfolio's relative performance. This is because the equal weighted portfolio holds a larger weight in the lower ranked stocks (by cap weight) than the cap weighted portfolio and, therefore, having to replace stocks as they leave the index has a larger impact.

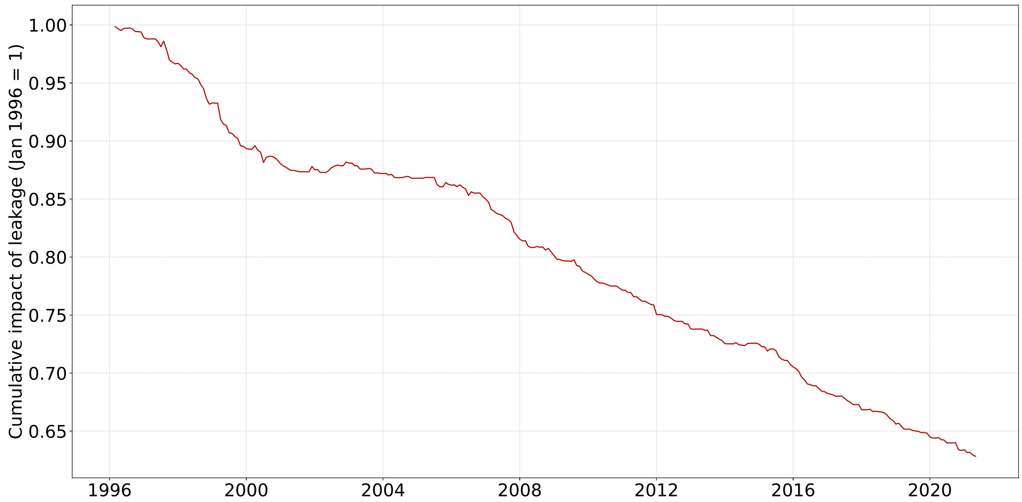

Figure 4.4.: Impact of leakage for the S&P500 equal weighted portfolio.

Table 4.1 seems to suggest that this contribution is a consistent -5% to -6% over three year periods. While the overall contribution appears to be consistent, Figure 4.1 seems to suggest that the impact of leakage can oscillate on a monthly basis. Figure 4.4 displays the cumulative impact of leakage on the equal weighted portfolio since 1996 which highlights the consistent negative contribution leakage makes on the relative return of the equal weighted performance. Interestingly, there appears to be a period between 2000 and 2005 when the impact of leakage is small. Figure 4.5 highlights the rolling 12-month change in index membership for the S&P500 which sheds light on this period. During this period the change in index membership declined from almost 60 stocks in 2001 to a low of 10 in 2004. This has a direct impact on leakage. Furthermore, during the period 2012 to 2015, the change in index membership was relatively small at around 25 changes per year. This same period coincides with a small period of slower than normal declines from leakage in Figure 4.5.

Dividends and transaction costs appear to be negligible contributors with costs being a fairly consistent -0.1% impact, as expected, and the impact of dividends contributing between -1% and 1% over consecutive three-year periods. Finally, Figure 4.6 shows the monthly residuals from the theoretical attribution of relative returns against the actual relative returns. The residuals are generally well behaved with an average of 0% and a standard deviation of 1%. Residuals are generally

Figure 4.5.: Rolling 12 month change in S&P500 index membership.

higher when a shift in the excess growth rate occurs, which could indicate some error in the covariance matrix estimation during these periods.

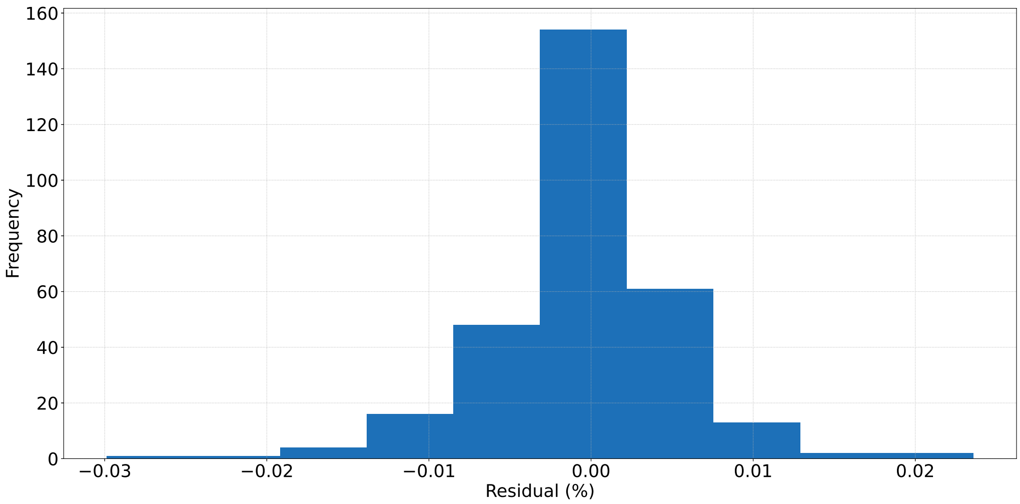

Figure 4.6.: Histogram of monthly residuals from theoretical attribution of relative returns and actual relative returns for the S&P500. Average monthly residual is 0.00% with a standard deviation of 1%.

4.3 Attribution of the Top 40

Figure 4.7 shows the attribution of the relative return for the Top 40 equal weighted portfolio on a rolling one-year basis. Table 4.2 shows the decomposition of relative returns for consecutive three-year periods, mirroring the periods in Table 2.4.

Table 4.2.: Decomposition of the Top 40 equal weighted portfolio relative return over consecutive three-year periods.

Period Start[a]	Period End	Excess growth rate	Port. gen. function	Leakage	Dividends	Costs	Residual	Total
2003	2005	8.6	5.4	-5.5	0.8	-0.1	0.6	10.1
2006	2008	14.9	-4.5	-21.0	1.5	-0.1	0.3	-5.3
2009	2011	9.5	-9.0	-2.8	1.2	-0.1	6.1	4.9
2012	2014	5.9	-0.9	1.4	-0.1	-0.1	-9.0	-3.2
2015	2017	8.8	-15.0	-9.5	0.5	-0.2	5.4	-9.7
2018	2020	16.5	-16.1	-17.3	1.0	-0.2	-2.8	-15.7
2021	2021/05[b]	1.4	3.6	-0.1	0.2	0.0	-1.0	4.2

[a] Period Starts 1 January and ends 31 December of Period End year

[b] Period ends 30 May 2021

The main drivers of relative performance are the excess growth rate, portfolio generating function, and leakage. Dividends and costs add very little volatility, although the impact of dividends are moderately positive in most years. As expected, the portfolio generating function adds the majority of volatility in relative returns with three-year periods oscillating between 5.4% and -16.1%. Another point to note is that the impact of the portfolio generating function within the Top 40, is almost always negative to some degree over consecutive three-year periods. This is evident

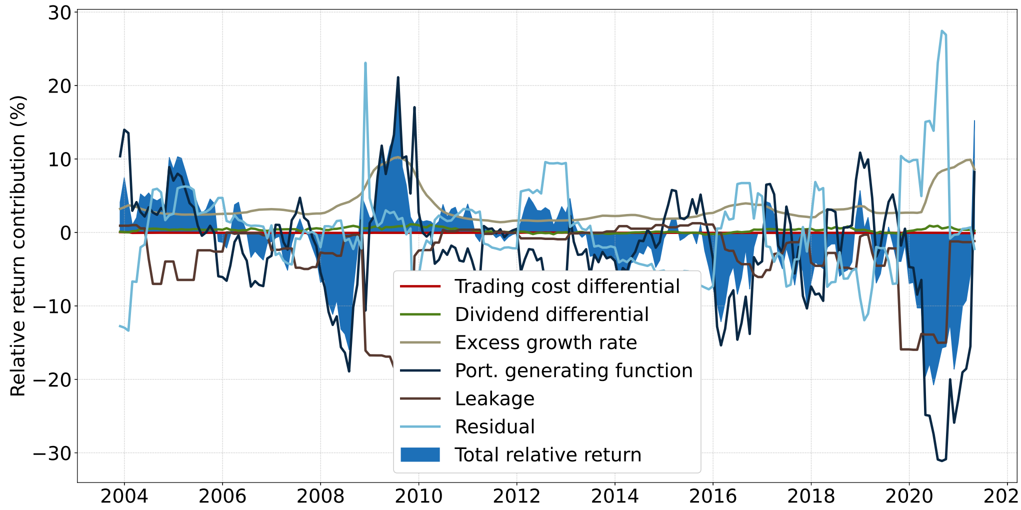

Figure 4.7.: Annual decomposition of Top 40 equal weighted portfolio relative return as per Equation (3.28) and including transaction cost differentials.

from Table 4.2, although Figure 4.7 highlights that there are annual periods where the impact of the portfolio generating function is positive. This may be indicative of constant concentration issues within the Top 40.

Figure 4.8 confirms the consistent negative contribution from the portfolio generating function. Although there are periods where the impact is positive (an increase in the portfolio generating function), the trend is a steady decline from 2010 onwards. It appears, at least from Table 4.2, that positive relative performance in the Top 40 equal weighted portfolio is more a function of the excess growth rate overcoming the negative contribution from the portfolio generating function.

Figure 4.9 shows the excess growth rate of the Top 40 equal weighted portfolio. Typically the excess growth rate is relatively low for the Top 40. This could be due to the high concentration in the cap weighted index leading to higher correlations as an "indexing" effect takes hold. In this scenario, since relatively few stocks determine the direction of the cap weighted index, portfolio manager decisions (related to smaller stocks) are highly influenced by how the larger stocks behave (as that drives index performance). For example, a large stock declines significantly, leading to the index as a whole declining and dragging smaller stocks lower as well. In this case, stocks have less of a beta exposure to the whole market but rather a beta to a very volatile and small set of stocks.

Figure 4.10 shows the rolling 12-month average correlations and volatilities among the constituents in the Top 40. Average correlation is fairly high, consistently above

Figure 4.8.: Portfolio generating function for the Top 40 equal weighted portfolio.

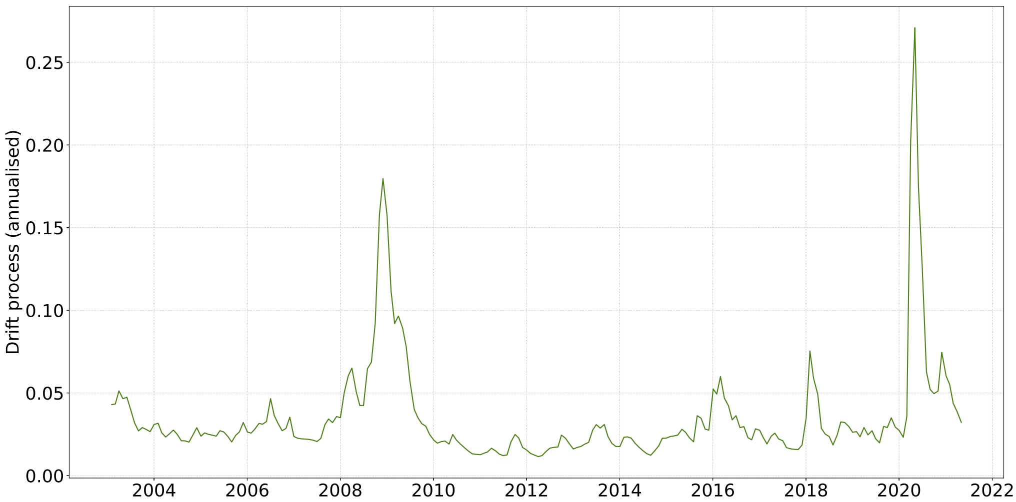

Figure 4.9.: Drift process (excess growth rate) for the Top 40 equal weighted portfolio.

30% and routinely as high as 50%. This is higher than the average correlations present in the S&P500 (see Figure 4.11), where average correlations are usually between 20% and 30%. Although the excess growth rate does show two periods of excess contribution to returns (following the 2008 financial crisis and in early 2020), these periods are offset by the negative impact of leakage, and thereby offsetting any positive gains from the excess growth rate. As one would expect leakage has a much bigger impact on the Top 40 index (a smaller index that suffers from high concentration) than the S&P500.

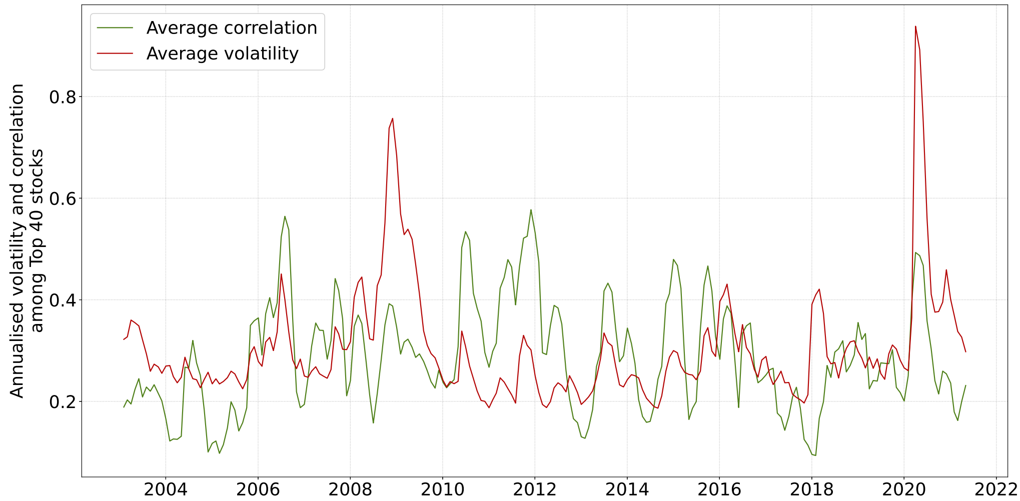

Figure 4.10.: Average correlations and volatilities among Top 40 stocks.

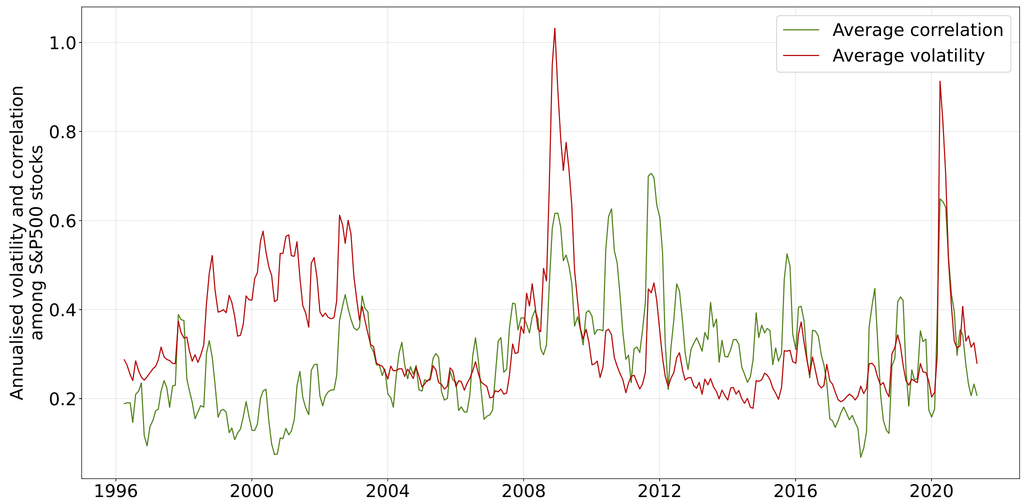

Figure 4.11.: Average correlations and volatilities among S&P500 stocks.

Figures 4.12 and 4.13 show the cumulative impact of leakage on the equal weighted portfolio and the rolling 12-month index change, respectively. The impact of leakage shows a high level of discontinuity (unlike in the case of the S&P500), with periods of only small changes punctuated by very dramatic declines. These declines coincide with large shifts in index membership (Figure 4.13). Furthermore, Figure 4.13 implies that index changes have increased since 2015. Although concentration has increased and contributed to the underperformance of the equal weighted portfolio, leakage has also played an important role over the past five years. The net impact of leakage between 2003 and 2011 was -2.1% per annum compared to -8.6% per annum for the period 2011 to 2021, highlighting the sharp increase in its impact on the equal weighted portfolio.

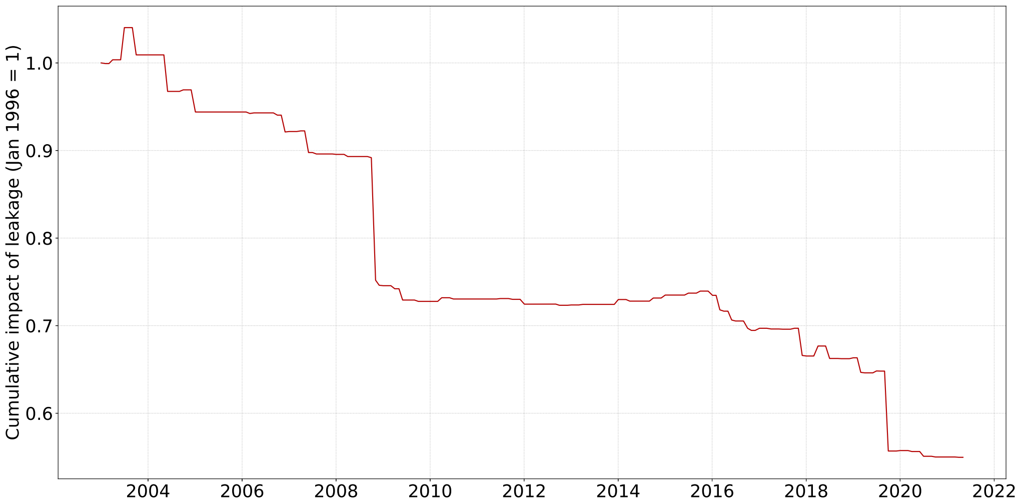

Figure 4.12.: Impact of leakage for the Top 40 equal weighted portfolio.

Figure 4.14 shows the histogram of residuals from this model. Although the residuals shown in Table 4.2 seem to indicate that overall residuals are a small (further supported by the average monthly residual of 0%), there are periods where residuals are significant. This is also evident in Figure 4.7, where in 2008 and 2020 residuals were large. In 2008, and similar to the observations in the S&P500, this was most likely due to a sudden shift in the covariance matrix. In 2020 another sudden shift occurs in the covariance matrix, however, a large decline in the portfolio generating function also occurs at this time. This is related to capital changes within the index (Prosus and Naspers) that had a large impact on the index and overall market caps. This may not be properly captured in the model leading to the model assigning too much of a negative impact to the portfolio generating function for changes that are

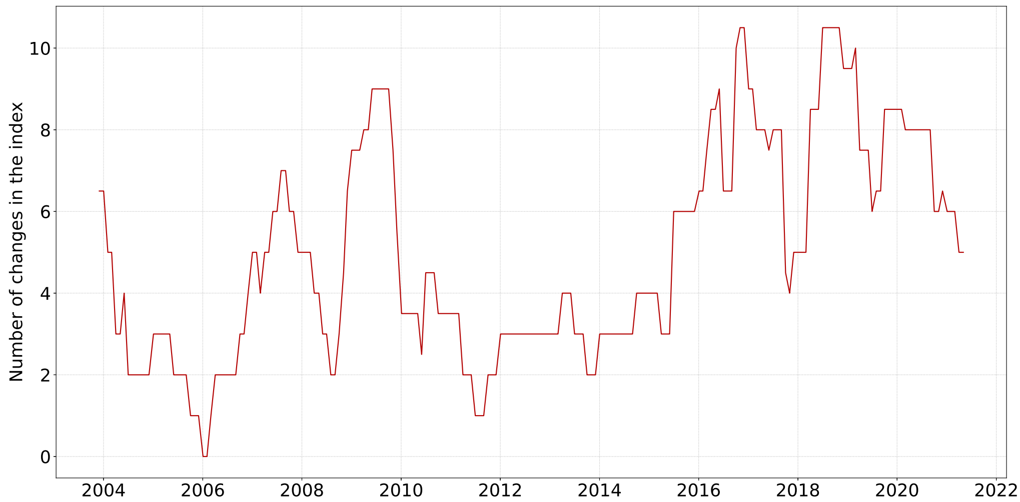

Figure 4.13.: Rolling 12 month change in Top 40 index membership.

actually structural in nature and whose impact is somewhat smaller on the equal weighted portfolio.

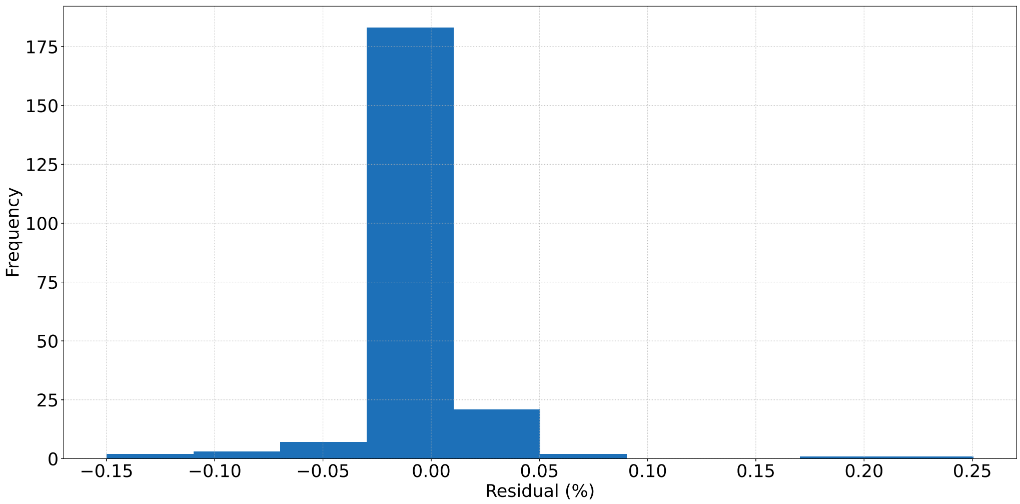

Figure 4.14.: Histogram of monthly residuals from theoretical attribution of relative returns and actual relative returns for the Top 40. Average monthly residual is 0.00% with a standard deviation of 3%.

4.4 Comparison of the main factors driving relative returns

A comparison of the Top 40 and S&P500 equal weighted portfolio performance provides interesting insight into the performance of the equal weighted portfolio in large and, relatively, smaller equity markets. While the S&P500 does at times display levels of increasing concentration, these periods pale in comparison to the effects of concentration in the Top 40. This is evident from the comparison of the portfolio generating functions for the S&P500 and the Top 40, Figures 4.2 and 4.8, respectively. The equal weighted Top 40 portfolio's portfolio generating function shows a steady decline from about 2005, whereas that of the S&P500 shows a much more bounded function with periods of both increasing and decreasing concentration.

The excess growth rate, however, appears similar for both markets and, since 2003, the correlation in monthly excess growth rates is quite high at 86%. Figure 4.15 shows the rolling one-year correlation of the monthly excess growth rates between the two equal weighted portfolios. Generally, the correlation on a one-year basis is quite high with correlations reaching almost 100% after the crises in this sample period (2008 and 2020). There are also three periods where correlations turn negative. Furthermore, Figure 4.16 shows the scatter plot of the S&P500 which

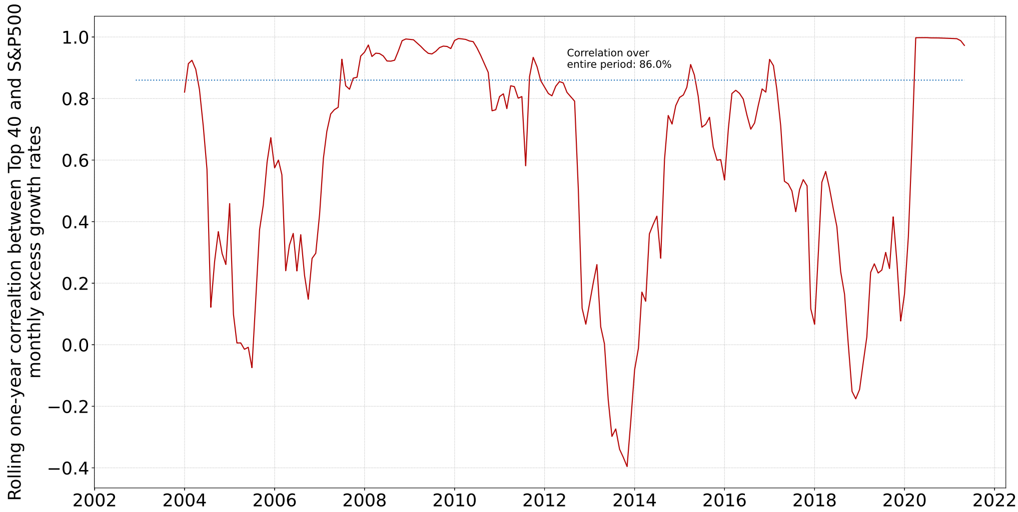

Figure 4.15.: Rolling one-year correlations between Top40 and S&P500 excess growth rates on a monthly basis.

highlights the generally positive relationship. That is, when the excess growth rate in one market is higher, the other is generally higher as well.

Furthermore, while leakage does create a drag on relative performance in both equal weighted portfolios, it has a far greater impact on the Top 40. This is largely expected given that the last stock in the Top 40 has a much larger weight in the equal weighted portfolio than the last stock in the S&P500 (broadly 1/40 vs. 1/500). The average annual impact of leakage on the Top 40 is -3.2% since 2003, with the last ten years being particularly poor at -8.6% per annum. In comparison, leakage has contributed a steady -1.8% since 2003. This amount of leakage, while not ideal, could be made up for by the contribution from the excess growth rate in the case of the S&P500 (3.4% per annum contribution since 2003). In contrast, an impact of leakage of -3.2% per annum is barely counteracted by the excess growth rate (also 3.4% per annum since 2003 for the Top 40).

4.5 Estimating the impact of specific stocks

The characterisation of the relative performance of the equal weighted portfolio in to the main components using stochastic portfolio theory allows for a better indication of the contribution to market concentration from one or more specific stocks. In the US the focus has been on technology stocks and in particular the set of stocks referred to as the FAANG stocks or more recently expanded version FAAANM:

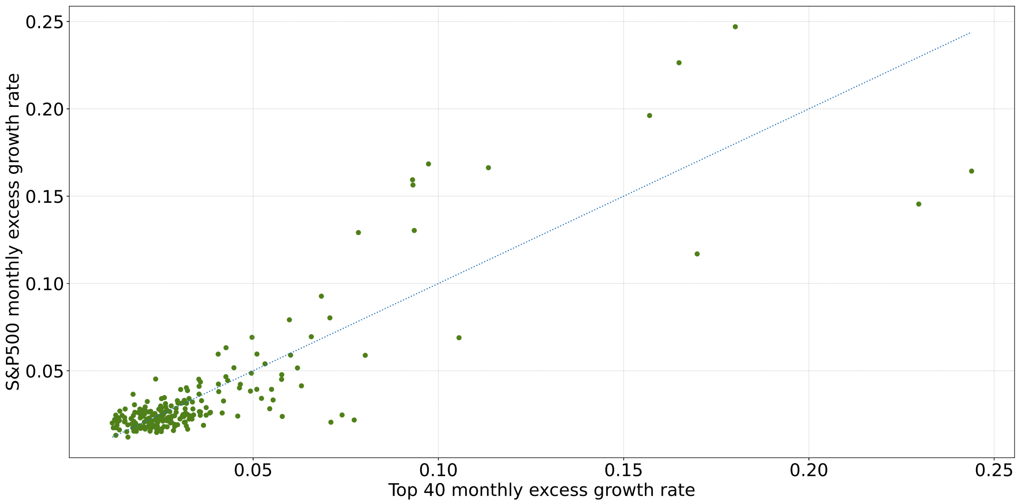

Figure 4.16.: Top40 (x-axis) and S&P500 (y-axis) excess growth rates a monthly basis.

Facebook, Alphabet (Google),Amazon, Apple, Netflix, and Microsoft. In South Africa, the focus has been on Naspers (NPN) and Prosus (PRX).

To gauge the impact on the equal weighted portfolio's relative return, these stocks are removed and their cumulative weight is redistributed proportionate to the remaining market capitalisation weights. The portfolio generating function as per Equation (3.20) is recalculated and compared to the actual portfolio generating function. Any difference in the two would correspond to a difference in the relative performance of the equal weighted portfolio.

Figures 4.17 and 4.18 show the impact on a rolling five-year basis for the FAAANM stocks and the Naspers related stocks, respectively. In the case of the FAAANM stocks, there was only a moderate impact of between 2% and 4% until 2017. Since then these stocks have had a large impact on the underperformance of the equal weighted portfolio through the increased concentration. More specifically, over the last five years these stocks have had a negative contribution of between 10% and 12% to the equal weight portfolio's relative performance.

Naspers and Prosus have, however, had a much larger negative drag on the equal weighted portfolio's relative performance in the Top 40. Since 2013, Naspers has had a steady and increasing drag on the relative performance of the equal weight portfolio. Over the last five-years Naspers and Prosus have led to a drag of between 20% and 25% on the equal weighted portfolio. As an interesting, and opposite view Figure 4.19 shows the same impact on the portfolio generating function when removing the resource stocks Anglo American and BHP Billiton.

Figure 4.17.: Change in the portfolio generating function of the equal weighted portfolio (over a rolling five year basis) when removing the stocks: Facebook, Alphabet (Google), Amazon, Apple, Netflix, and Microsoft and re-distributing their weight proportionately among the remaining stocks.

Removing these stocks during the period 2012 to 2018 would have had the impact of increasing concentration by up to 15%. The fact that this impact (on a five-year rolling basis) becomes increasingly negative over this period is indicative of their continued underperformance of the Top 40 over this period. That is, these stocks increasingly reduce concentration as their weights get smaller in the index due to underperformance. This would have led to some relative underperformance in the equal weighted portfolio because the portfolio would be overweight these stocks relative the cap weights.

4.6 Attribution of other countries

Attribution of returns for each country by source is provided in Table 4.3 while the three main components (portfolio generating function, excess growth rate and leakage) are shown per country in Figures 4.20 to 4.27. Most of the residual returns are in line with those found in the US and SA markets, however, residuals for Australia and Japan are quite significant (3.1% and -5.5% per annum, respectively). In both cases, the excess growth rates are significantly higher than peers which may indicate this as a source of the model error (higher residual).

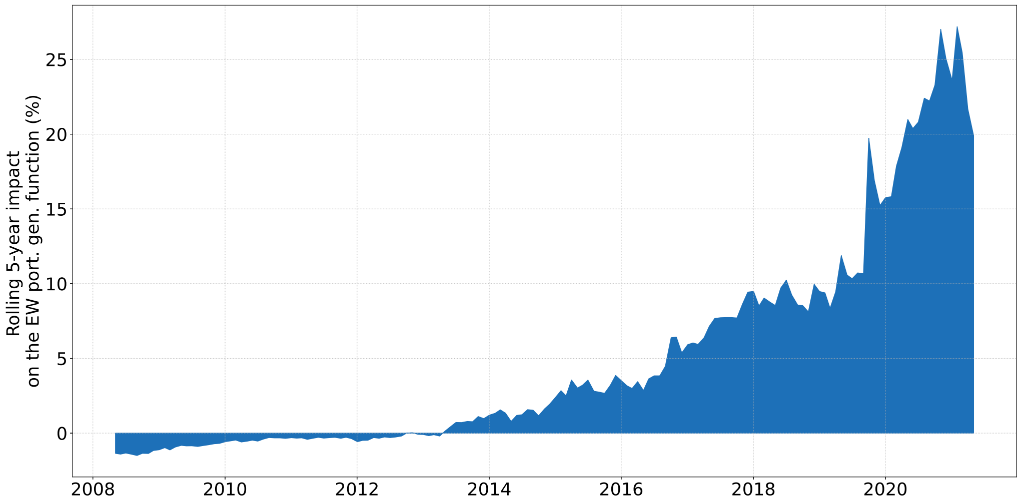

Figure 4.18.: Change in the portfolio generating function of the equal weighted portfolio (over a rolling five year basis) when removing Naspers and Prosus, and re-distributing their weight proportionately among the remaining stocks.

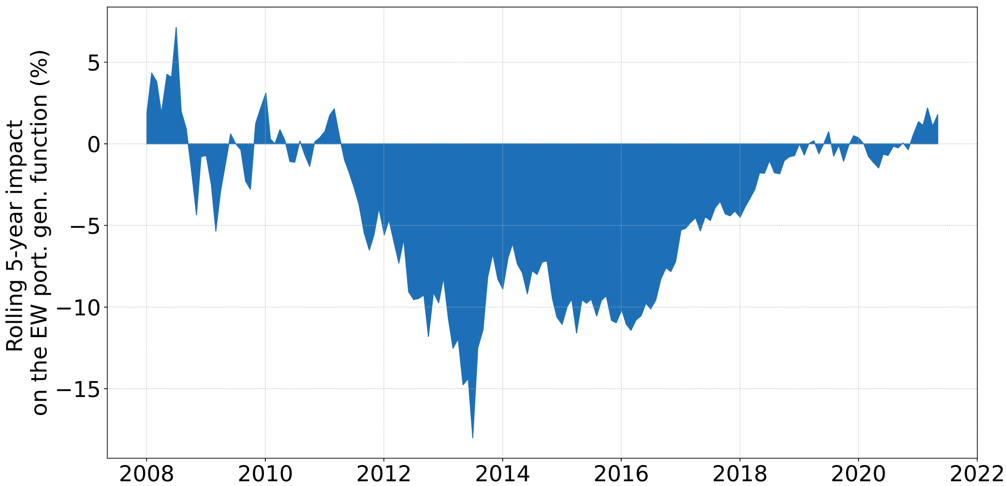

Figure 4.19.: Change in the portfolio generating function of the equal weighted portfolio (over a rolling five year basis) when removing Anglo American (AGL) and BHP Billiton (BHP), and re-distributing their weight proportionately among the remaining stocks.

In the case of Japan, the equal weighted portfolio has performed the best with outperformance of 3.5% per annum over the respective cap weighted portfolio. This is despite the worst contribution from the portfolio generating function of -3.4% per annum, double the next worst level of -1.7% in South Africa. The excess growth rate in the Japanese equal weighted portfolio is, however, one of the highest at 4.7% per annum and the impact of leakage is the lowest at -0.7% (likely due to the index's size). Interestingly the portfolio generating function for Japan has consistently trended downwards, indicating a continued increase in concentration. The starting level, however, of the portfolio generating function is much lower (between -8.6 to -9.4) compared to, for example, the US which is in a range of -6.8 to -7.3 (see Figure 4.27). That is, although the concentration has been consistently increasing, the base level of concentration is still much lower than other markets (at least by the geometric mean of cap weights).

Another notable equal weighted portfolio is that of the United Kingdom (UK) which had the highest relative return after Japan. Furthermore, the residual return is among the smallest in our sample, indicating a decent fit in this market. The majority of this outperformance has come from the portfolio generating function which contributed a positive 1.5% per annum. The top, left panel in Figure 4.21 shows the portfolio generating function of the equal weighted portfolio in the UK in which, although there is some increase in concentration in the early 2000s and 2008, has for the most part been a trend of lower concentration. This is especially true since 2011 in contrast to many other markets, such as SA, which actually showed the complete opposite. That is, a significant increase in concentration in contrast to the UK equity market.

Another market where the general trend in the portfolio generating function has been for lower concentration has been that of Germany (Figure 4.22). This has also led to a positive contribution in relative returns from the portfolio generating function in the German equal weighted portfolio. The excess growth rate in this market is, however, among the lowest (2.5% per annum), with France the only country showing a lower contribution to relative return. Interestingly the portfolio generating functions of these two countries are quite different with that of France remaining very range bound over the sample period, contributing -0.6% per annum to relative returns (Figure 4.23).

Another point to notice is how similar the excess growth rates across all the equity markets appear, at least visually. In general, excess growth rates rise substantially following a crises such as in 2000, 2008, and 2020. This similarity is confirmed by the high correlation between these excess growth rates, shown in Table 4.4.

Table 4.3.: Attribution of relative equal weighted portfolio returns per annum by country (CAGR).

Country	Excess growth rate	Generating Func.	Leakage	Dividends	Costs	Residual	Total
Japan	4.7	-3.4	-0.7	0.0	-0.1	3.1	3.5
United Kingdom	3.0	1.5	-1.6	-0.2	-0.1	-0.4	2.3
United States	3.4	0.3	-1.8	-0.1	0.0	-0.4	1.3
Germany	2.5	0.8	-1.3	-0.2	0.0	-1.5	0.7
Canada	3.7	0.0	-1.6	-0.2	-0.1	-1.6	0.6
France	2.4	-0.6	-0.9	-0.3	0.0	-0.6	0.2
South Africa	3.4	-1.7	-3.2	0.3	-0.1	-0.4	-0.8
Australia	4.9	1.6	-1.3	-0.5	-0.1	-5.5	-0.8

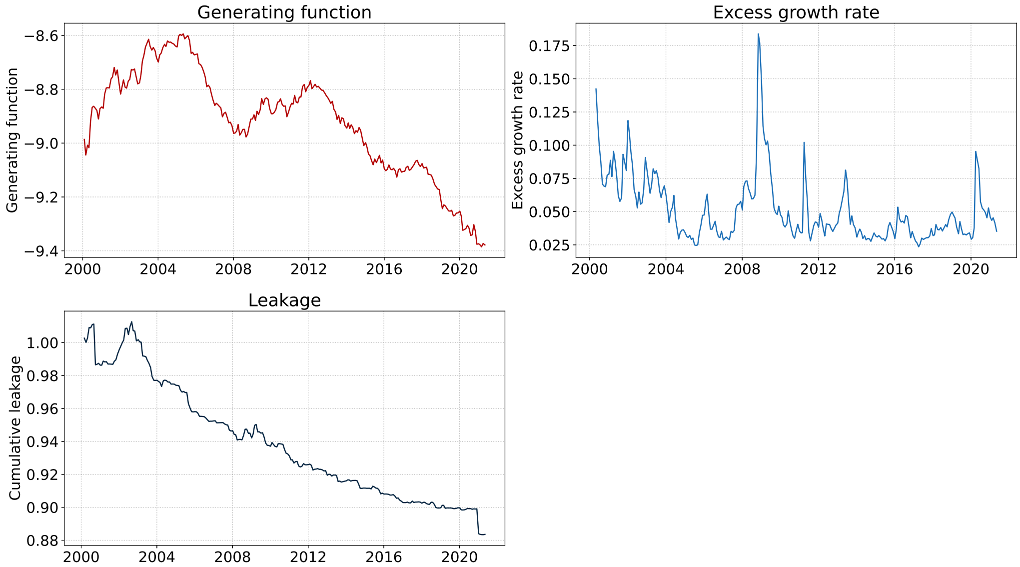

Figure 4.20.: Japan (JP): Portfolio generating function, excess growth rate and cumulative leakage for the equal weighted portfolio.

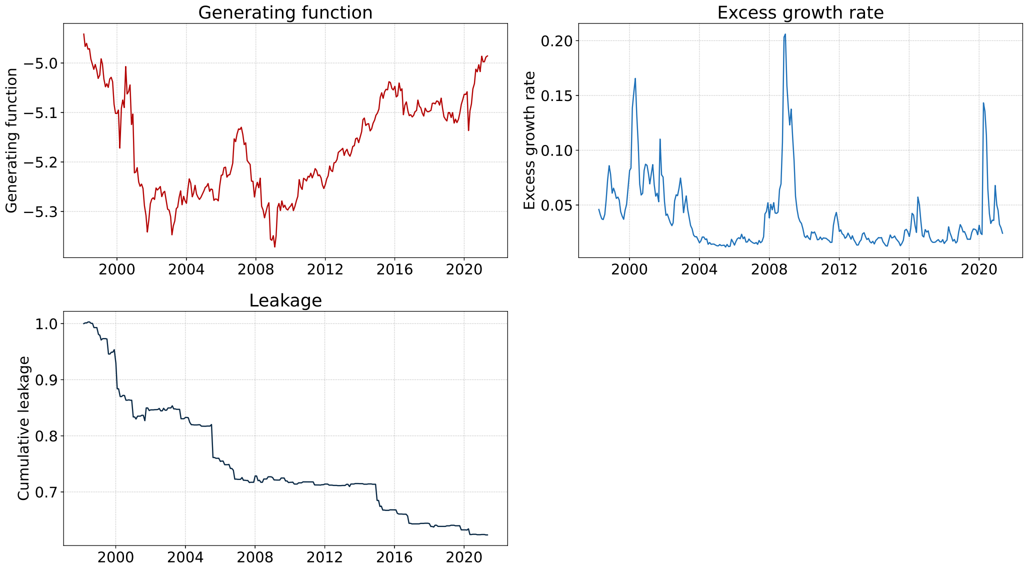

Figure 4.21.: United Kingdom (UK): Portfolio generating function, excess growth rate and cumulative leakage for the equal weighted portfolio.

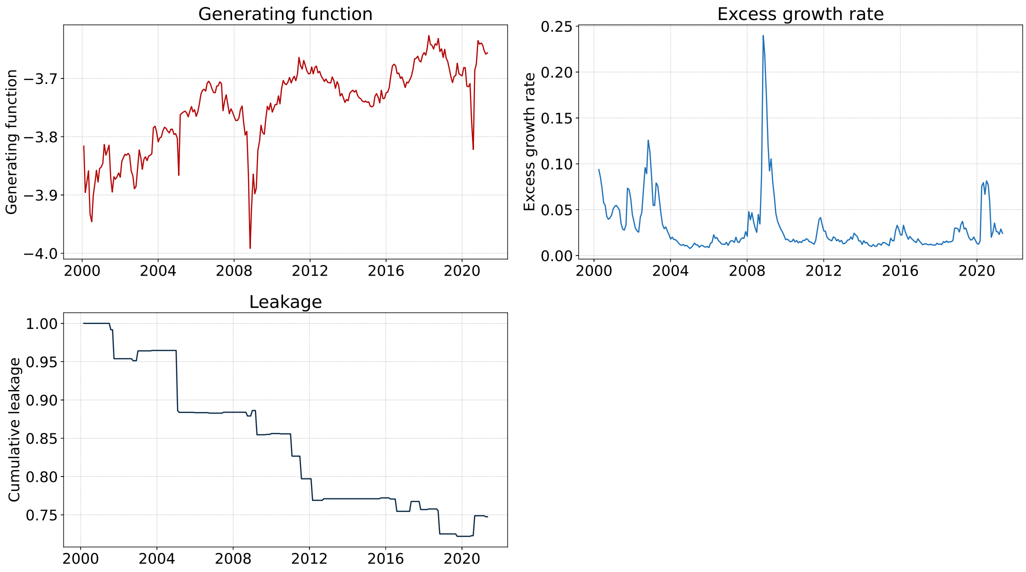

Figure 4.22.: Germany (GE): Portfolio generating function, excess growth rate and cumulative leakage for the equal weighted portfolio.

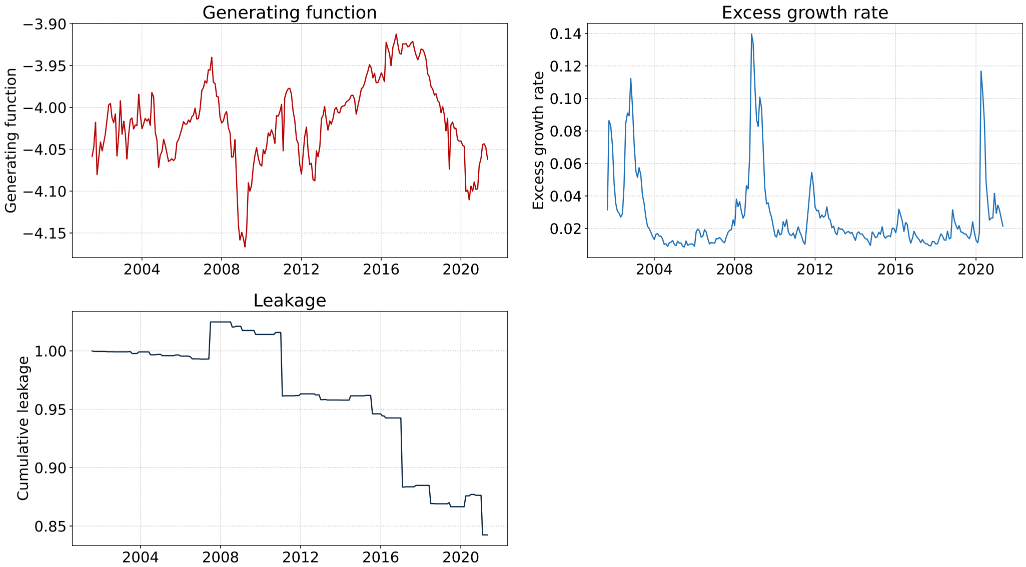

Figure 4.23.: France (FR): Portfolio generating function, excess growth rate and cumulative leakage for the equal weighted portfolio.

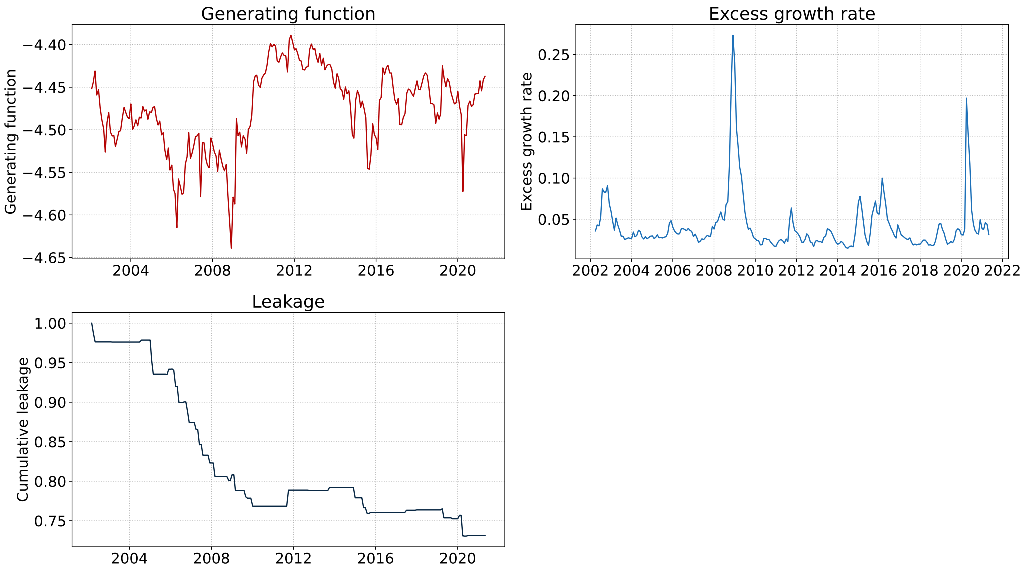

Figure 4.24.: Canada (CA): Portfolio generating function, excess growth rate and cumulative leakage for the equal weighted portfolio.

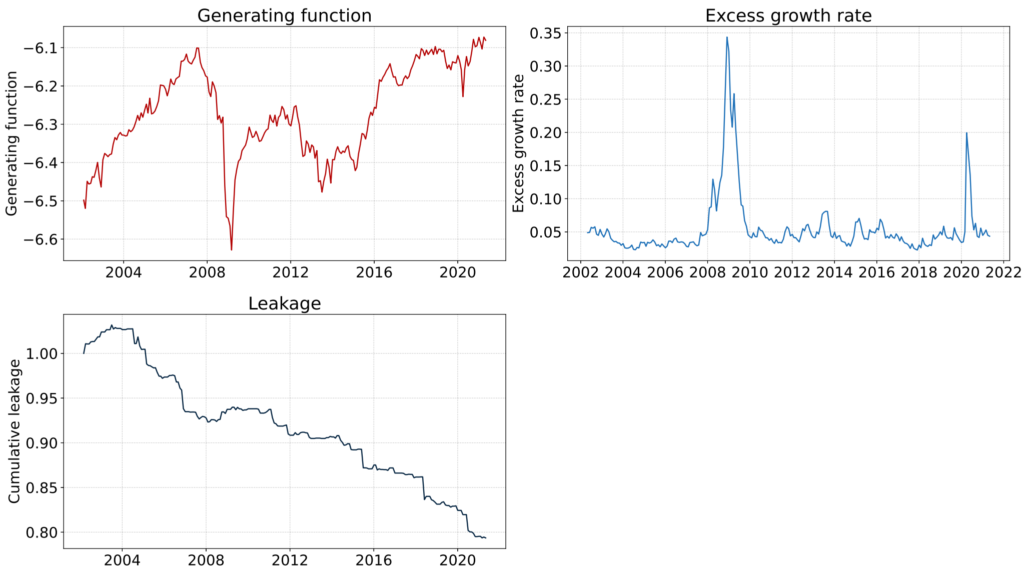

Figure 4.25.: Australia (AU): Portfolio generating function, excess growth rate and cumulative leakage for the equal weighted portfolio.

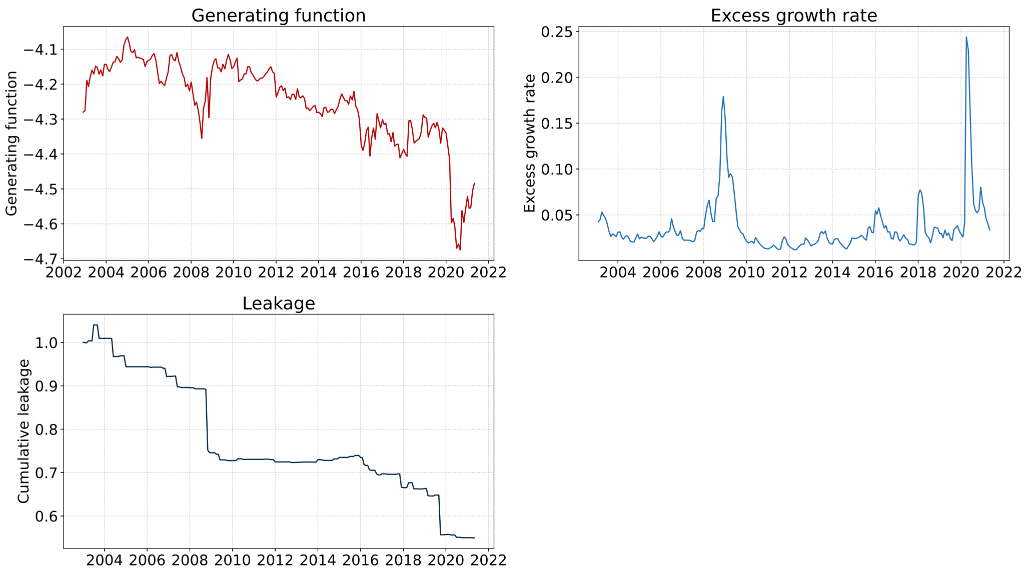

Figure 4.26.: South Africa (SA): Portfolio generating function, excess growth rate and cumulative leakage for the equal weighted portfolio.

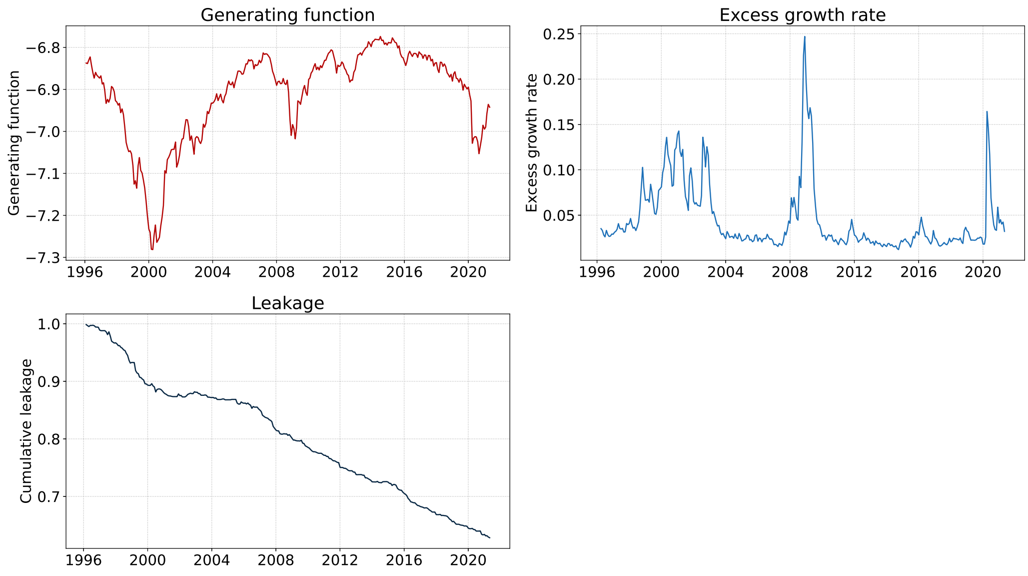

Figure 4.27.: United States (US): Portfolio generating function, excess growth rate and cumulative leakage for the equal weighted portfolio.

The average correlation is 85% among the excess growth rates, with the lowest correlation between Japan and South Africa (70%). This compares to the average correlation of portfolio generating functions of only 8.5%. The correlations between the portfolio generating functions are shown in Table 4.5.

Table 4.4.: Correlations between the excess growth rate of equal weighted portfolios in the different countries.

	UK	GE	FR	CA	AU	JP	SA	US
UK	1.00	0.88	0.92	0.92	0.91	0.85	0.87	0.94
GE		1.00	0.92	0.85	0.80	0.82	0.76	0.89
FR			1.00	0.84	0.78	0.78	0.84	0.95
CA				1.00	0.89	0.77	0.85	0.90
AU					1.00	0.80	0.78	0.88
JP						1.00	0.70	0.84
SA							1.00	0.85
US								1.00

Table 4.5.: Correlations between the portfolio generating functions of equal weighted portfolios in the different countries.

	UK	GE	FR	CA	AU	JP	SA	US
UK	1.00	0.53	0.48	0.26	0.57	-0.77	-0.74	-0.17
GE		1.00	0.43	0.49	0.65	-0.51	-0.46	0.69
FR			1.00	0.19	0.34	-0.17	-0.07	0.61
CA				1.00	-0.04	-0.15	-0.18	0.26
AU					1.00	-0.52	-0.45	0.16
JP						1.00	0.86	-0.08
SA							1.00	0.21
US								1.00

4.7 Conclusion

This chapter represents an empirical implementation of the model presented in Chapter 3 and explains the relative performance of the equal weighted portfolio through the components set out in Equation (3.28). That is, primarily as a function of the portfolio generating function (a measure of concentration in the case of the equal weighted portfolio), the drift process (the excess growth rate), and the impact of leakage (the impact of stocks moving in and out of the index).

Section 4.2 looked at the attribution of US equities. The portfolio generating function of the S&P500 equal weighted portfolio, while displaying periods of increasing concentration, appears to be well bounded and this leads to periods of both positive and negative contributions which should offset themselves for an overall net zero contribution as stochastic portfolio theory would predict. This is in contrast to the Top40 in South Africa (Section 4.3) where the portfolio generating function has contributed negatively to relative performance on a very consistent basis. Furthermore, a visual inspection of the portfolio generating function (Figure 4.8) shows a function that does not appear at all to be bounded and consequently a stock market that shows continuously higher levels of concentration.

The excess growth rate for the S&P500 and Top 40 equal weighted portfolios appears very similar, although there are some periods where they display low levels of correlation (Figure 4.15). Tables 4.1 and 4.2 show the contribution from each factor on a rolling three-year basis. In the US, there is a distinct difference in the contribution from the excess growth rate before and after 2010. Excess growth rates in the S&P500 equal weighted portfolio appear much higher pre-2010, contributing an average of 5.4% per annum to the equal weighted portfolio's relative performance.

This declines to an average of 2.6% per annum after 2010. The contribution from the excess growth rate in the SA equity market does not show this decline with the highest contribution of 16.5% coming in the period 1 January 2018 to 31 December 2020.

Leakage was identified as an important, and in some years a major, negative contributor to the equal weighted portfolio's relative performance. In the S&P500 equal weighted portfolio this contributed -1.8% per annum to the relative performance of the equal weighted portfolio compared to a much higher -3.2% for the Top 40. Furthermore, leakage has become an increasingly negative drag on the relative performance of the equal weighted portfolio in South Africa with an average contribution of -8.6% per annum since 2011 as the rotation of stocks in and out of the index in the last ten years has increased.

This characterisation of the relative performance of the equal weighted portfolio, allows for an estimate of the impact of specific stocks on the equal weighted portfolios performance due to their contribution to the cap weighted portfolio's concentration (Section 4.5). For example, the increased concentration from technology stocks in the S&P500 (Facebook, Alphabet (Google), Amazon, Apple, Netflix, and Microsoft) is estimated to have contributed negatively to the equal weighted portfolio's relative return by 10% to 12% on a rolling five year basis. In a similar manner, Naspers and Prosus appear to have had a negative impact of between 20% and 25% on the relative performance of the Top 40 equal weighted portfolio through a direct increase in concentration over the last five years (Figure 4.18).

This chapter concluded by comparing the attribution between different countries in Section 4.6. Most countries experienced high contributions to relative returns for the equal weighted portfolio from the excess growth rate, with rates of 3.0% to 4.9% per annum. France and Germany experienced the lowest contribution from the excess growth rate with rates of 2.4% and 2.5% per annum, respectively. While in South Africa the portfolio generating function contributed negatively to relative returns, for most countries the contribution was relatively small. In some countries, however, the contribution was actually positive (such as in the UK and Germany) as the cap weighted portfolio became less concentrated over time. Finally, it was noted how correlated the excess growth rates were across most markets, with an average correlation of 85%. The portfolio generating functions, on the other hand, showed very little correlation among the different equity markets.

This page has been intentionally left blank

5 Optimal choice between equal and cap weighted portfolios

5.1 Introduction

As mentioned in Chapter 3, Equation (3.20), the portfolio generating function of the equal weighted portfolio, implies that $\log\mathbf{S}(\mu(t))$ is bounded, assuming that the stock market does not become concentrated in a single stock. Furthermore, the drift process of the equal weight portfolio, represented by the excess growth rate in Equation (3.21), is always greater than, or equal to, zero. As a result, Equation (3.17), the relative return process of the equal weighted portfolio, implies that the equal weighted portfolio will almost surely outperform the cap weighted portfolio in the long-run.

A complicating factor, however, is the effect of leakage. That is, the effect of forming a portfolio or index only on a subset of the market (the top 40 stocks, for example) and having to completely divest or invest as stocks move in and out of the index. In a small market, such as South Africa and the Top 40 for example, this effect can have a large impact as noted in Chapter 4.

In the case of the Top 40, the equal weighted portfolio has underwhelmed, primarily between 2016 and 2021, largely due to large increases in the concentration (the negative impacts from the portfolio generating function), and the large impact of leakage. Although the equal weighted portfolio performs much better in the S&P500, given a more bounded portfolio generating function and much smaller impact of leakage, there still appears to be long periods of relative drawdowns.

This chapter presents two models aimed at combining the cap weighted and equal weighted portfolios in an optimal way to reduce the short-term relative drawdowns while still benefiting from the longer-term outperformance of the equal weighted portfolio. These models use the main components in Equation (3.28), in an attempt to negate some of these periods of steep drawdowns relative to the cap weighted portfolio. In the first, section the aim is to create as simple a model as possible and in turn highlight that monitoring the portfolio generating function, as given by Equation (3.20), the excess growth rate, and the impact of leakage, Equation

(3.27), are important when improving the short-term performance of the equal weighted portfolio relative to a cap weighted index. The second model aims at improving performance by creating an optimal blend of the cap and equal weighted portfolios.

The first section in this chapter introduces the rudimentary model (see Taljaard and Maré, 2021b; Taljaard and Maré, 2021a) and provides an analysis of the results for the S&P500 and Top 40. The results for the other countries are also presented, before results are compared across different switch thresholds. The second section presents an alternative approach, using a Random Forest classifier, which seeks to blend the cap weighted and equal weighted portfolios using the forecasted probability of either portfolio outperforming the other. Features, used as inputs in the model, are constructed from the main components in Equation (3.18) and using the Python library called *tsfresh* (Christ et al., 2018).

5.2 A rudimentary linear regression approach

The rudimentary approach used here is to make use of the portfolio generating function, the excess growth rate and leakage to forecast what the relative performance of the equal weight will be in the following month and switching between the cap weight and equal weighted portfolios. This is done given that in the previous sections it was shown how, theoretically and empirically, the relative return of an equal weighted portfolio is a combination of the changes in concentration of the cap weighted portfolio weights (the portfolio generating function), the excess growth rate (heuristically the benefits of diversification of the equal weighted portfolio) and the rate of leakage as stocks move out of the cap weighted portfolio. Therefore, this chapter focuses on optimising the equal weighted portfolio by making use of these three components.

A rudimentary linear regression model is implemented that looks to forecast the next month's relative performance using the average monthly change in the log portfolio generating function, Equation (3.20), over the past three months, the most recent estimate of the drift process, and the average of the last three month's leakage as defined in Equation (3.27). In this case the drift process is the excess growth rate of the equal weighted portfolio, Equation (3.6), where volatilities and correlations are estimated using the EWMA method with a decay factor of 0.97 as in Chapter 4. The model is fit on the prior three years' of data and attempt to forecast the next month's relative return.

Mathematically this would be expressed as

$$\hat{y}_{t+1} = \beta_0 + \beta_1 X_t, \qquad t = 3, \ldots, n \tag{5.1}$$

where $\hat{y}_{t+1}$ is the next month's relative return of the equal weighted portfolio and X_t is given by

$$X_t = \frac{1}{12}\gamma^*_\pi(t) + \frac{1}{3}\sum_{i=t-2}^{t} d\log\mathbf{S}(\mu(i)) + \frac{1}{3}\sum_{i=t-2}^{t} dL_{\pi/\mu}(i), \tag{5.2}$$

where $d\log\mathbf{S}(\mu(i))$ is the one-month change in the log of the portfolio generating function and $dL_{\pi/\mu}(i)$ represents the net leakage effect, Equation (3.27), for the equal weighted portfolio at time $t = i$ with μ representing the cap weights of the S&P500 or Top 40 constituents. The intercept β_0 and coefficient β_1 are fit using an ordinary least squares approach. The model is fit over the preceding three years' monthly data at the beginning of each month prior to any monthly rebalancing in the portfolios.

Since the dynamic portfolio will be shifting between equal and cap weighted, turnover is likely to increase and, therefore, trading costs of 15 basis points are included. This is roughly in line with analysis done by Frazzini et al. (2018) for stocks over the period 1998 to 2016. As the analysis in this chapter includes transaction costs of 15 basis points for all portfolios, the optimal portfolio only switches from equal weights to cap weights (or vice versa) if the relative return is predicted to, at least, offset these transaction costs.

Practically, to generate the equal and cap weights for the dynamic portfolio, the model makes use of the diversity weighted portfolio (see Fernholz, 2002). This portfolio has the following generating function

$$D_p(\mu) = \left(\sum_{i=1}^{n} \mu_i^p\right)^{\frac{1}{p}}. \tag{5.3}$$

This generates weights given by

$$\pi_i(t) = \frac{\mu_i^p(t)}{\sum_{i=1}^{n} \mu_i^p(t)}. \tag{5.4}$$

As $p \to 0$ the portfolio's weights tend to the equal weight portfolio and as $p \to 1$, the portfolio's weights tend to the cap weighted portfolio. There have been attempts to

Table 5.1.: Risk adjusted performance of S&P500 optimised, equal, and cap weighted portfolios, monthly rebalanced with 15bps of costs.

Portfolio	CAGR[a]	Volatility	Sharpe ratio[b]	Information ratio[c]	Sortino ratio
Optimised	11.9%	20.3%	0.547 (0.224)	0.434 (0.520)	0.700
Equal weighted	12.0%	20.9%	0.545 (0.765)	0.321 (–)	0.682
Cap weighted	10.3%	19.7%	0.486 (–)	–	0.618

[a]Compound annual growth rate.

[b]One-month US Treasury bill used as risk free rate for both Sharpe and Sortino ratios. P-value for two sided hypothesis test as per Ledoit and Wolf (2008) shown in brackets. The null hypothesis is that the Sharpe ratio is equal to that of the cap weighted portfolio.

[c]Information ratio relative to the cap weighted portfolio. P-value for two sided hypothesis test as per Ledoit and Wolf (2008) shown in brackets. The null hypothesis is that the IR is equal to that of the equal weighted portfolio.

directly optimise this value p (see, for example, Samo and Vervuurt, 2016); although, this tends to result in portfolios with $p < 0$. This leads to an inverse cap weighted portfolio where smaller stocks have much larger weights than the largest stocks in the market.

In the implementation of the linear model in this chapter p is restricted to either very close to 0, or equal to 1, to select directly between the equal weighted or cap weighted portfolios. In the next chapter this analysis is extended to consider a floating value for p between 0 and 1.

5.2.1 Results for the S&P500

The main results for the S&P500 are highlighted in Table 5.1. A statistical test for the significance is performed on the difference in the Sharpe ratios of the equal weighted and optimal portfolios relative to the cap weighted portfolio. This test is also performed for the Information ratios of the two portfolios. The method used is that of Ledoit and Wolf (2008), which makes use of a studentised circular bootstrap approach to construct a confidence interval at a given significance level. This test can be altered to provide a p-value for the null hypothesis. The hypothesis test is a two-sided test with the null hypothesis as H_0 : the difference in Sharpe ratios is zero.

The optimised portfolio performs in line with the equal weighted portfolio, with slightly lower volatility and higher Information and Sortino ratios. Note that in both cases the p-values are large. It is, however, worth bearing in mind that the optimal portfolio is a combination of the equal weight and cap weighted portfolio. Therefore, the bootstrapping approach would cover many areas where both the optimal portfolio and cap weighted portfolio have the same return series. The proportion of time spent in each specific portfolio weighting scheme is shown in Figure 5.1. The optimal portfolio spends approximately two-thirds of the time within an equal weighting and reverts to cap weighting the remainder of the time. In this context one could argue the p-value associated with the difference in the Sharpe ratios (0.224) is not too extreme considering the proportion of time the optimal portfolio spends in each weighting methodology.

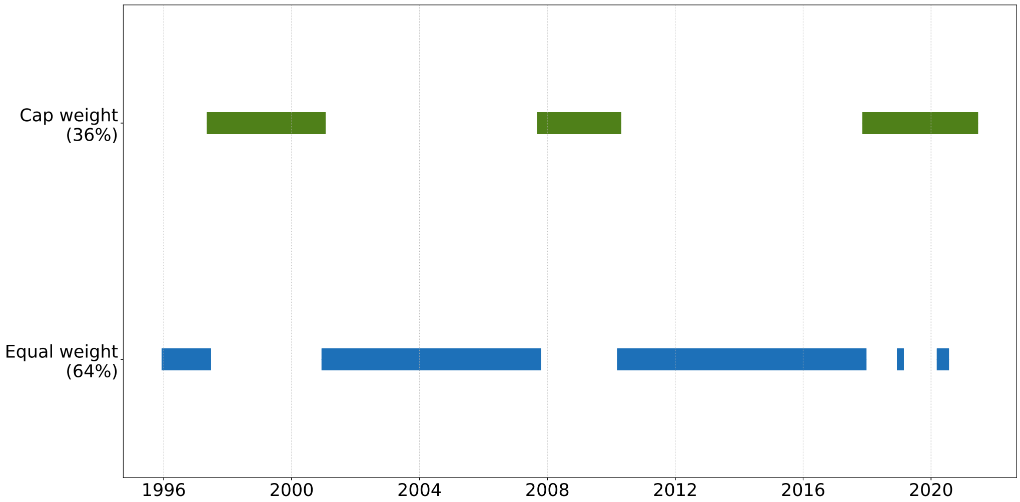

Figure 5.1.: Allocation of optimal portfolio between S&P500 cap weighted and equal weighted portfolios. Proportion of overall time allocated to a specific portfolio is shown in brackets.

Although outright performance is similar, the optimised portfolio does avoid the steep drawdowns relative to the cap weighted portfolio. In Table 5.2 the maximum relative return drawdowns of the optimised and equal weighted portfolios (relative to the cap weighted portfolio) are shown. This is reduced from a maximum underperformance of 31.8% for the equal weighted portfolio to 10.9% for the optimised portfolio. Furthermore, the alpha (over the cap weighted portfolio) per unit of average relative drawdown shows a marked improvement from 6.44 times to 9.02 for the optimised portfolio.

This improvement is most evident in a chart of the cumulative returns of the optimised and equal weighted portfolios (Figure 5.2). The optimised portfolio does well to switch into the cap weighted portfolio during the three main drawdown periods: around the 2000s, post-2008, and more recently post-2016. Where the model does underperform, however, is during each period's recovery following a large relative drawdown. For example, the equal weighted portfolio outperforms

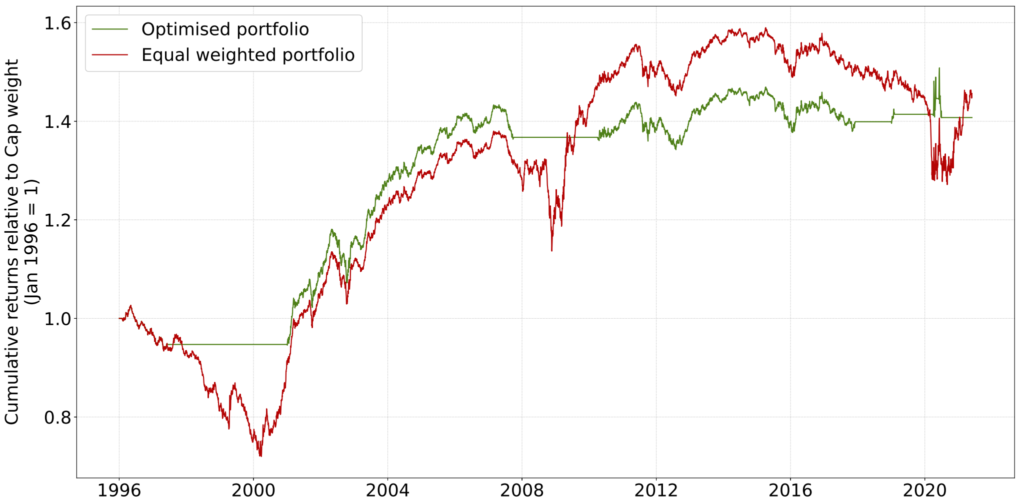

Figure 5.2.: S&P500 optimised and equal weighted portfolio cumulative returns relative to cap weighted portfolio with monthly rebalancing.

Table 5.2.: Relative return drawdowns of S&P500 optimised and equal weighted portfolios.

Portfolio	Maximum relative drawdown[a]	Alpha / Avg DD
Optimised	10.9%	9.02
Equal weighted	31.8%	6.44

[a]Maximum relative return drawdown.

[b]Alpha (relative to cap weighted portfolio) per unit of average relative return drawdown.

the cap weighted portfolio significantly starting in 2000, but the optimised portfolio only switches to equal weight much later, thereby missing out on at least a year's outperformance. That said, given the simplicity of the model, the optimised portfolio appears to do well to generate a higher return than the equal weighted portfolio while avoiding large relative drawdowns.

5.2.2 Results for the Top 40

Similar results are presented for the Top 40 with Table 2.3 reflecting the risk adjusted performance of the three portfolios. The outright performance of the optimised portfolio is moderately ahead of the equal weighted portfolio (0.3% CAGR outperformance per annum) but with differences in Sharpe, Information and Sortino ratios statistically insignificant. This is slightly different from the results in Taljaard and Maré (2021a) due to the period in 2020, where the optimised portfolio switched into equal weights as the equal weighted portfolio suffered a significant relative drawdown. This is most notable in Figure 5.3 which shows the relative performance

Table 5.3.: Risk adjusted performance of Top 40 optimised, equal, and cap weighted portfolios, monthly rebalanced with 15bps of costs.

Portfolio	CAGR[a]	Volatility	Sharpe ratio[b]	Information ratio[c]	Sortino ratio
Optimised	14.7%	20.2%	0.417 (0.76)	0.045 (0.78)	0.556
Equal weighted	14.4%	19.3%	0.411 (0.88)	-0.016 (–)	0.555
Cap weighted	14.6%	20.1%	0.412 (–)	–	0.573

[a]Compound annual growth rate.

[b]One-month JIBAR used as risk free rate for both Sharpe and Sortino ratios. P-value for two sided hypothesis test as per Ledoit and Wolf (2008) shown in brackets. The null hypothesis is that the Sharpe ratio is equal to that of the cap weighted portfolio.

[c]Information ratio relative to the cap weighted portfolio. P-value for two sided hypothesis test as per Ledoit and Wolf (2008) shown in brackets. The null hypothesis is that the IR is equal to that of the equal weighted portfolio.

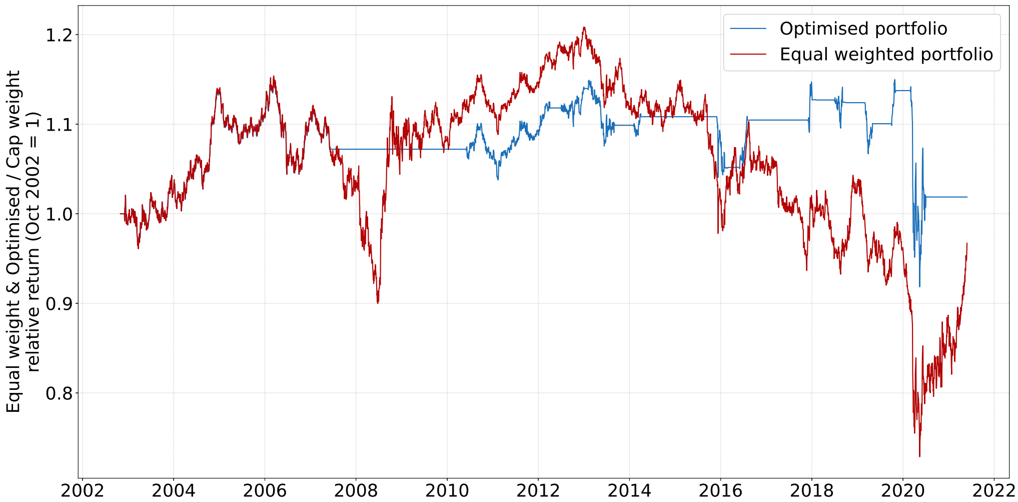

Figure 5.3.: Top 40 optimised and equal weighted portfolio cumulative returns relative to cap weighted portfolio with monthly rebalancing.

of both the optimised and equal weighted portfolio. Another possible issue with the model are the irregular changes around this period in the generating function that are due to corporate actions as noted in Section 4.3.

Figure 5.4 highlights the optimal allocation over time with the model spending less than 50% of the time in the equal weighted portfolio. This contrasts with the S&P500 portfolio where almost two-thirds of the time was spent in the equal weighted portfolio. The majority of time spent in cap weights highlights again the significant underperformance of the equal weighted portfolio in South Africa over the last decade.

Table 5.4.: Relative return drawdowns of Top 40 optimised and equal weighted portfolios.

Portfolio	Maximum relative drawdown[a]	Alpha / Avg DD
Optimised	23.4%	0.33
Equal weighted	48.0%	-0.30

[a]Maximum relative return drawdown.

[b]Alpha (relative to cap weighted portfolio) per unit of average relative return drawdown.

While risk adjusted performance is somewhat disappointing in the Top 40 case, the model does improve the relative drawdown experience of the equal weighted portfolio. The maximum relative drawdown for the optimal portfolio is lower at 23% compared to the equal weighted portfolio's 48%, and alpha per unit drawdown is higher at 0.3 while that of the equal weighted portfolio is negative (-0.3).

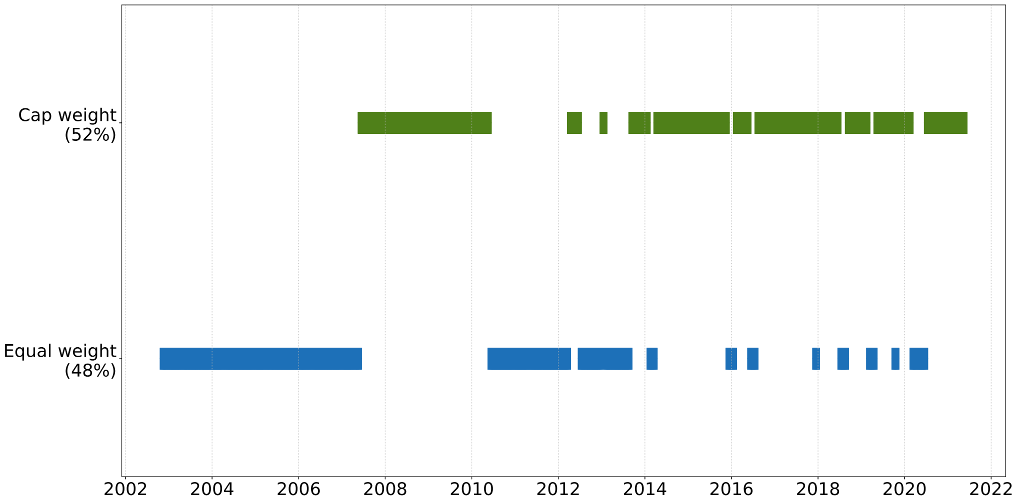

Figure 5.4.: Allocation of optimal portfolio between Top 40 cap weighted and equal weighted portfolios. Proportion of overall time allocated to a specific portfolio is shown in brackets.

5.2.3 Results for other countries

The results for other countries are presented in Tables 5.5 and 5.6. In general, the results follow a similar pattern as that of the US and South Africa. That is, the optimal portfolio gives up some return over the equal weighted portfolio but does reduce relative drawdowns. The optimal portfolio does reduce drawdowns relative to the equal weighted portfolio in all cases except for Japan and South Africa.

Table 5.5.: CAGR and volatility of the optimised, equal, and cap weighted portfolios for a selection of countries[a].

	CAGR (%)			Volatility (%)		
Country	**Cap**	**Equal**	**Optimised**	**Cap**	**Equal**	**Optimised**
Australia	11.0	10.0	9.2	16.2	17.3	16.4
Canada	8.6	9.7	8.6	18.1	18.4	18.4
France	4.8	5.5	5.5	23.0	24.0	23.3
Germany	3.5	5.9	4.9	24.0	22.9	22.9
Japan	2.8	8.5	8.4	21.4	19.9	20.5
South Africa	14.6	14.4	14.7	20.1	19.3	20.2
United Kingdom	4.7	7.2	5.1	19.4	19.7	19.5
United States	10.3	12.0	11.9	19.7	20.9	20.3

[a]Portfolios are rebalanced monthly with 15bps of costs.

Table 5.6.: Risk adjusted performance of the optimised, equal, and cap weighted portfolios for a selection of countries[a].

	Sharpe ratio[b]			Information Ratio[c]	
Country	**Cap**	**Equal**	**Optimised**	**Equal**	**Optimised**
Australia	0.50	0.43	0.40	-0.09	-0.34 (0.164)
Canada	0.46	0.51	0.45	0.15	-0.10 (0.184)
France	0.28	0.30	0.30	0.05	0.05 (0.928)
Germany	0.21	0.31	0.27	0.16	-0.07 (0.004)
Japan	0.23	0.51	0.49	0.58	0.65 (0.556)
South Africa	0.41	0.41	0.42	-0.02	0.05 (0.912)
United Kingdom	0.20	0.32	0.22	0.52	0.16 (0.012)
United States	0.49	0.54	0.56	0.32	0.43 (0.520)

[a]Portfolios are rebalanced monthly with 15bps of costs.

[b]Respective risk-free rates used in Sharpe ratio calculations.

[c]Information ratio relative to the cap weighted portfolio. P-value for two sided hypothesis test as per Ledoit and Wolf (2008) shown in brackets. The null hypothesis is that the IR is equal to that of the equal weighted portfolio.

Table 5.6 contains the Sharpe and Information Ratios for the portfolios, including the same statistical test for the Information Ratios as used in Table 5.3, for example. The results for the optimal portfolio are largely mixed, with an improvement in Information Ratios across France, Japan, South Africa and United States. Information ratios in the remaining countries are lower than the equal weighted portfolios, largely due to lower CAGRs.

Table 5.7 shows the relative return drawdowns and average alpha per unit of average drawdown for the equal weighted and optimal portfolios. This largely mirrors the results in Table 5.6 with only some countries showing a large improvement in alpha per unit average drawdown. This is largely a result of lower CAGRs for the optimal

Table 5.7.: Relative return drawdowns of optimised and equal weighted portfolios.

	Maximum relative drawdown (%)		Alpha / Avg DD	
Country	Equal	Optimised	Equal	Optimised
Australia	43.4	44.6	-0.80	-1.12
Canada	19.0	16.0	4.17	0.04
France	24.5	10.7	2.83	3.40
Germany	56.4	55.5	8.88	1.86
Japan	52.1	39.8	17.40	20.18
South Africa	48.0	23.2	-0.30	0.38
United Kingdom	27.3	21.6	15.29	1.06
United States	31.8	10.9	6.44	12.10

portfolios as all countries, except for Australia, show an improvement in drawdowns relative to the cap weighted portfolio.

Figure 5.5 shows the cumulative relative returns for the equal weighted and optimal portfolios per country (on a log-scale). The results are largely in line with that of the US and South Africa. Namely, the model does well generally to avoid the large drawdown periods but is slow to switch back to the equal weighted portfolio as it begins to outperform.

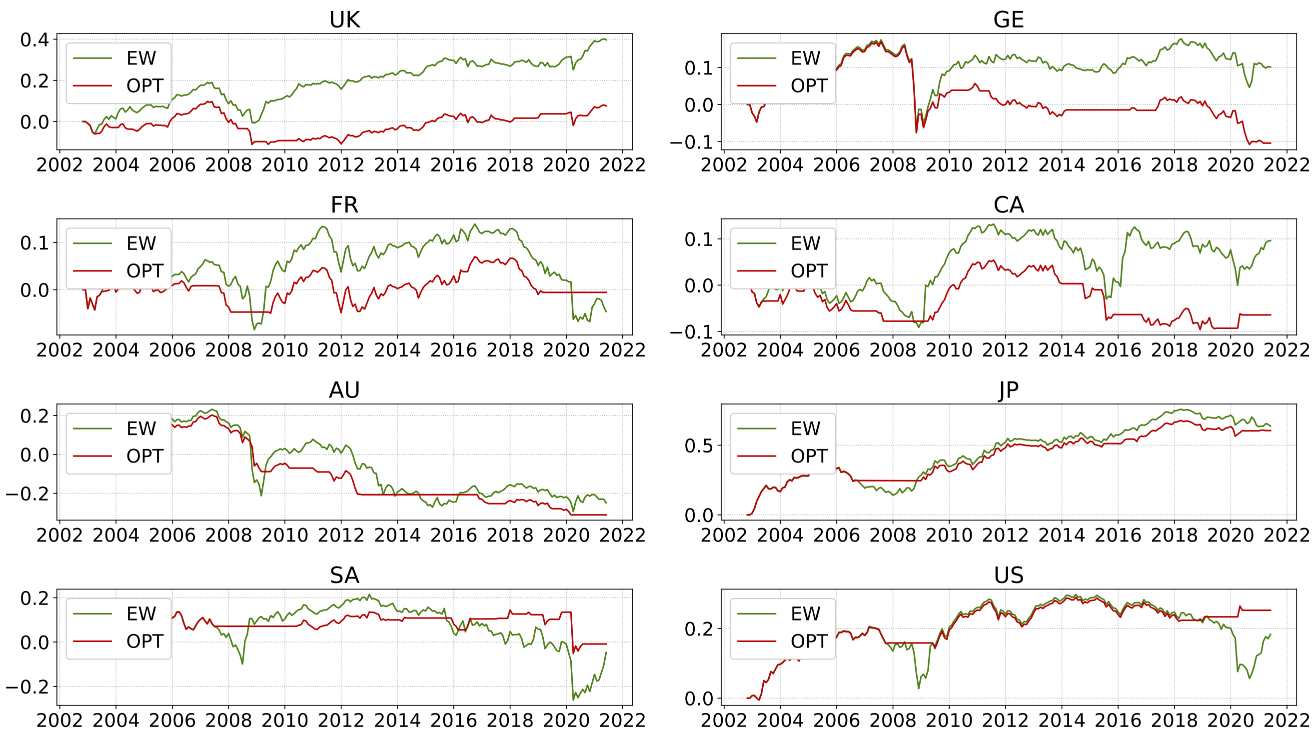

Figure 5.5.: Annual cumulative relative performance (log-scale) of equal weighted and optimal portfolios by country.

5.2.4 Optimising the model

One of the issues with the rudimentary linear model is that it is slow to switch from cap weight to equal weight as the equal weighted portfolio begins to outperform. This could be a result of the threshold imposed in the model to determine whether a switch is made. Recall that the model only switches if the predicted difference in returns is greater than the trading costs (15 bps). This is somewhat arbitrary in this context as it is unlikely that more than half of the portfolio would be churned as a switch is made from equal to cap weight, or vice versa. The trading activity would be confined only to the difference in the weights and not the entire portfolio. Furthermore, a switch to another weighting scheme would likely have a cumulative impact and so checking one month's expected relative return against the full transaction cost could be misleading.

In this subsection both the threshold for switching to, or from, cap weights (and vice versa) and the arbitrary choice of three years of data for the linear regression fit are optimised for each country.

The results for SA and US equities are shown in Figures 5.6 and 5.7, respectively. These figures show CAGR and alpha per unit of average relative drawdown across various thresholds for switching (in terms of expected alpha) and the length of the regression lookback window used. In general results for CAGR and alpha per unit of average relative drawdown are similar with only minor differences. In the case of SA, it appears that a very short lookback window with a high threshold results in the best performance. This could, however, be indicative of poor performance in the model in general as the performance of the model drops off significantly thereafter.

The best performing combination of parameters for the US also appears to be for a short regression window, although here the results are fairly similar across threshold levels. The best combination, however, is very similar to the SA case with a short regression window and higher threshold levels. There also appears to be a peak around the 90 month lookback window and a lower threshold level, where results are also quite strong.

Results for the remaining countries are provided in Figures 5.8 to 5.13 with the best performing combination summarised in Table 5.8. The optimal regression windows are largely between 60 and 90 months, with only Japan, South Africa and United States showing smaller regression windows of 24 to 36 months. Interestingly for these countries with shorter regression windows, higher thresholds are the most optimal. Germany and Japan also show relatively higher threshold levels. Australia, Canada, and France appear to have much lower threshold levels than the other countries with the optimal threshold in Australia at zero.

Table 5.8.: Optimised threshold and regression window parameters by country.

Country	Regression window	Threshold (bps)	Alpha/Avg DD	CAGR (%)	Sharpe ratio
Australia	90	0	-0.7	10.1	0.45
Canada	90	3	1.1	8.9	0.47
France	90	3	5.4	5.8	0.32
Germany	90	12	8.3	5.8	0.31
Japan	24	9	23.6	8.6	0.50
South Africa	24	12	0.6	15.0	0.43
United Kingdom	60	12	10.8	6.4	0.28
United States	36	15	12.1	12.2	0.56

The largest improvements in performance after optimisation are for Australia, Japan, and United Kingdom although, as expected, there is some improvement across all the countries. Improvement in terms of CAGR is best veiwed using a comparison between Figures 5.14 and 5.5 which show the cumulative relative performance against the cap weighted portfolio. In the UK, where the equal weighted portfolio has performed very well, the optimal model improves to now roughly match this performance. This is similar to Japan, although the model does protect its relative gains during the 2006 to 2008 period in this case.

Other notable improvements are in Australia and South Africa where the model outperforms the equal weighted portfolio overall in markets where the equal weighted portfolio generally underperforms the cap weighted portfolio. In South Africa the optimised model actually generates some outperformance against the cap weighted portfolio in a market where the equal weighted portfolio has struggled to keep up with the benchmark.

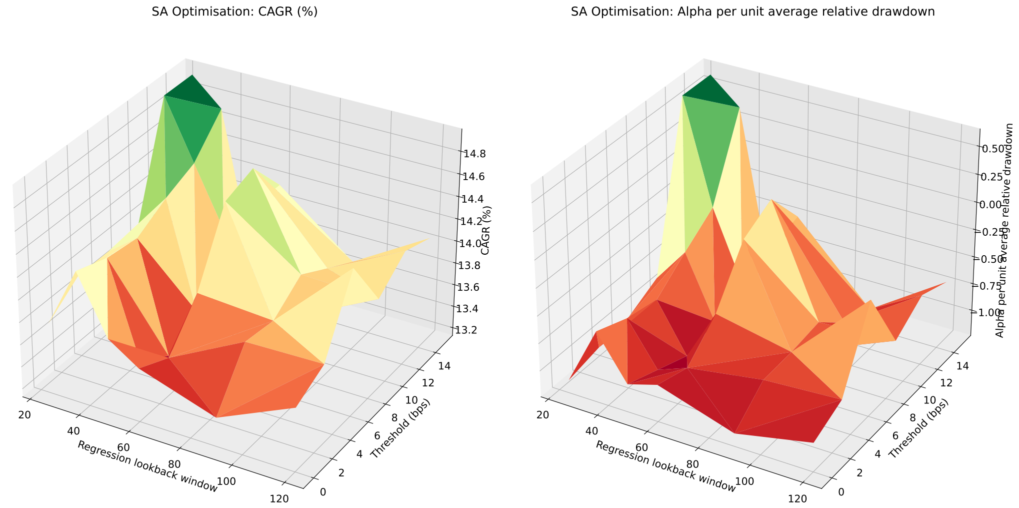

Figure 5.6.: CAGR and alpha per unit average relative drawdown by regression lookback window and threshold for switching for South African equities.

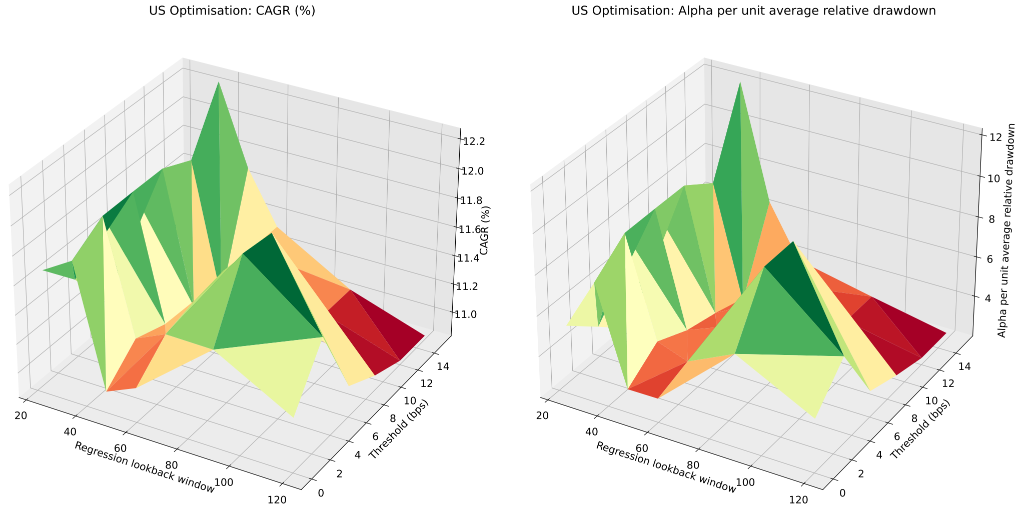

Figure 5.7.: CAGR and alpha per unit average relative drawdown by regression lookback window and threshold for switching for United States equities.

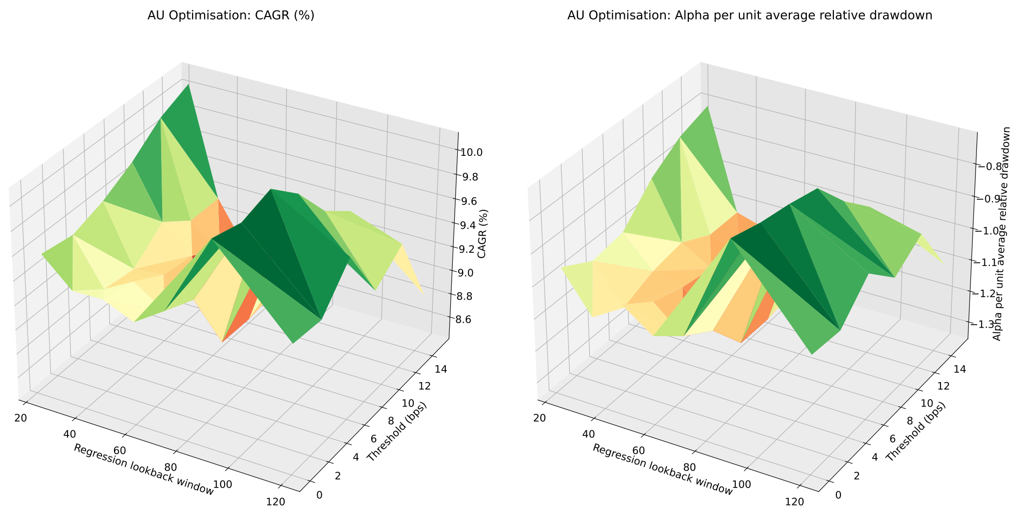

Figure 5.8.: CAGR and alpha per unit average relative drawdown by regression lookback window and threshold for switching for Australian equities.

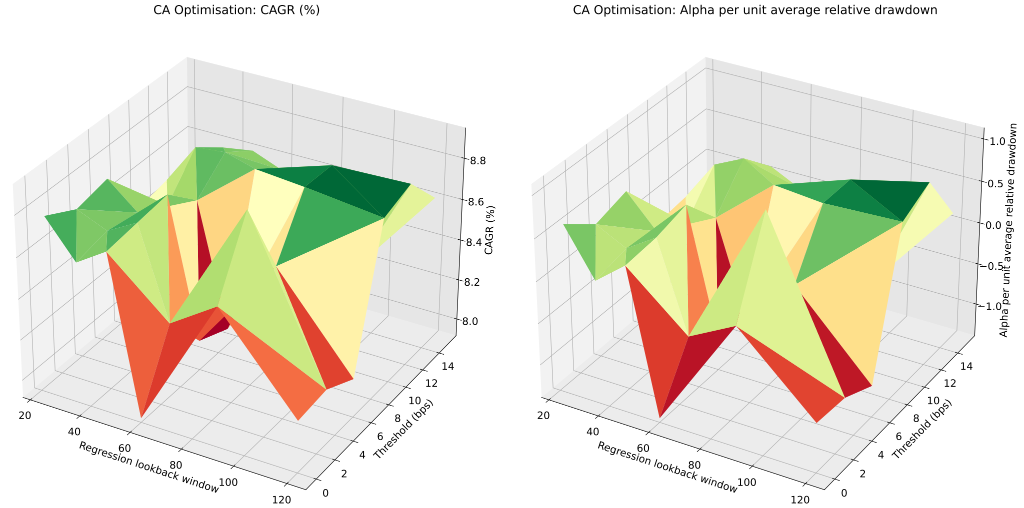

Figure 5.9.: CAGR and alpha per unit average relative drawdown by regression lookback window and threshold for switching for Canadian equities.

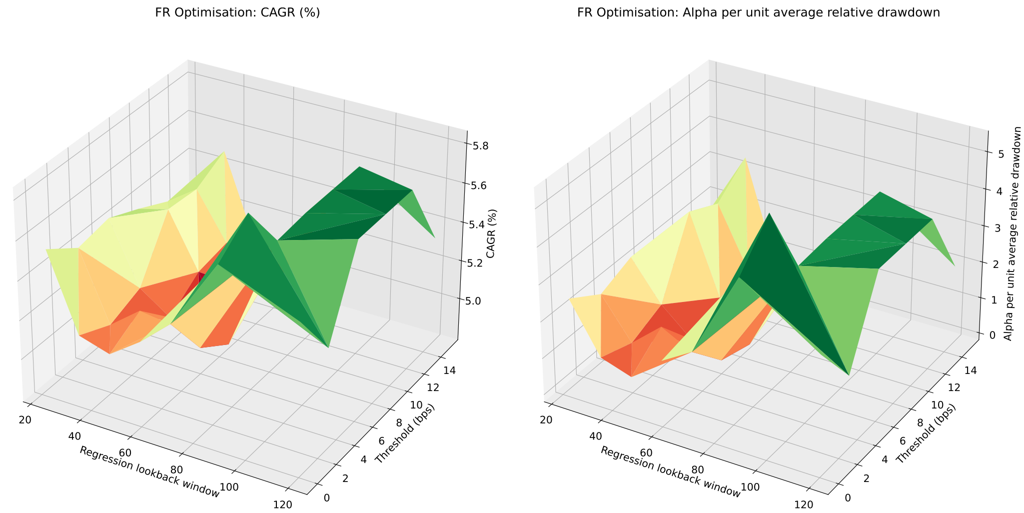

Figure 5.10.: CAGR and alpha per unit average relative drawdown by regression lookback window and threshold for switching for French equities.

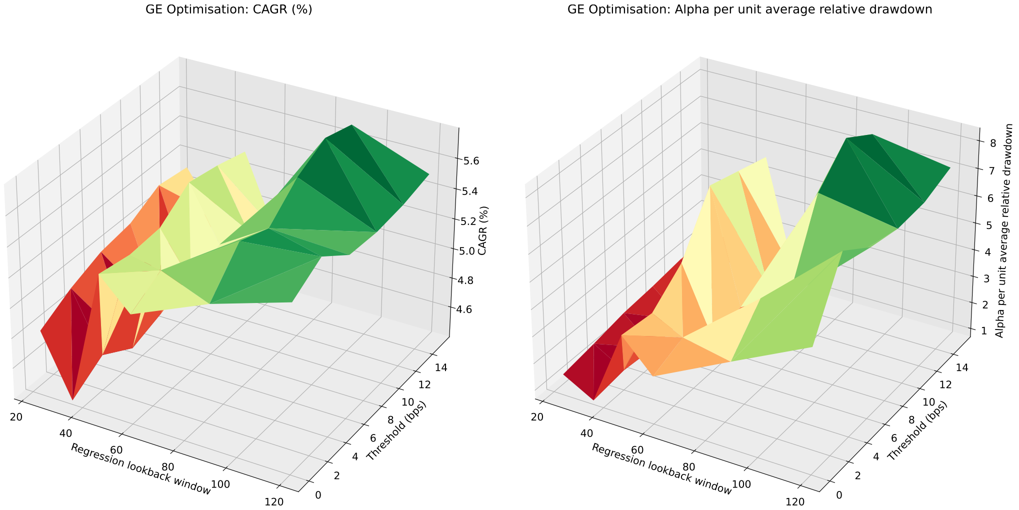

Figure 5.11.: CAGR and alpha per unit average relative drawdown by regression lookback window and threshold for switching for German equities.

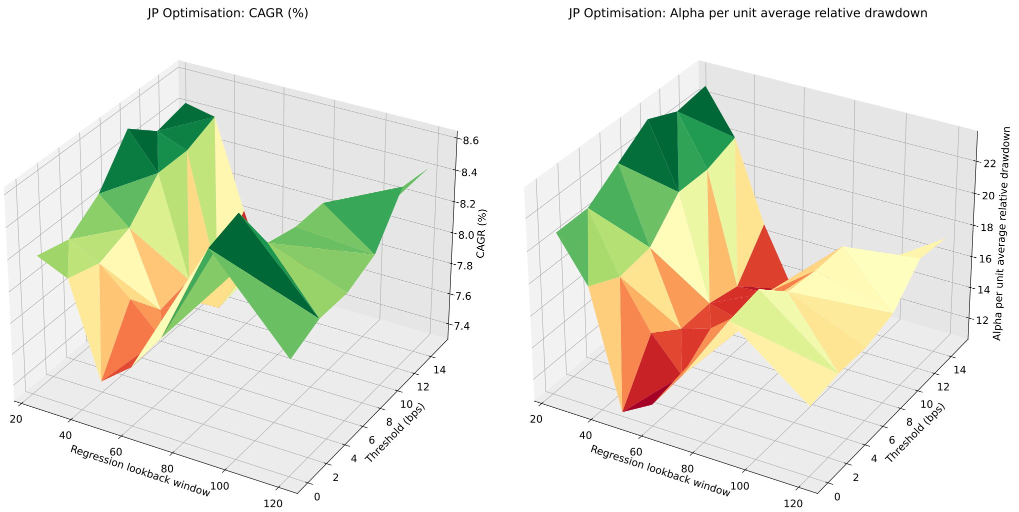

Figure 5.12.: CAGR and alpha per unit average relative drawdown by regression lookback window and threshold for switching for Japanese equities.

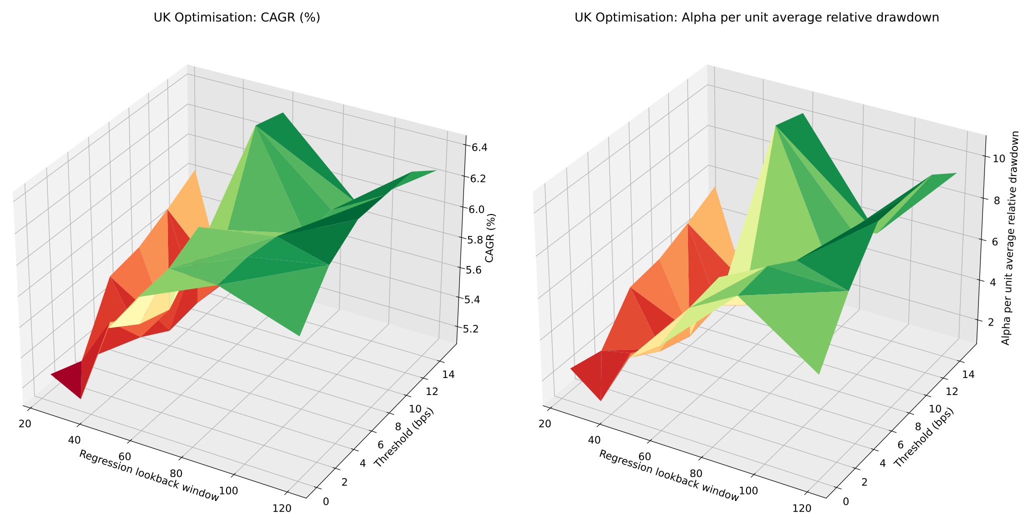

Figure 5.13.: CAGR and alpha per unit average relative drawdown by regression lookback window and threshold for switching for British equities.

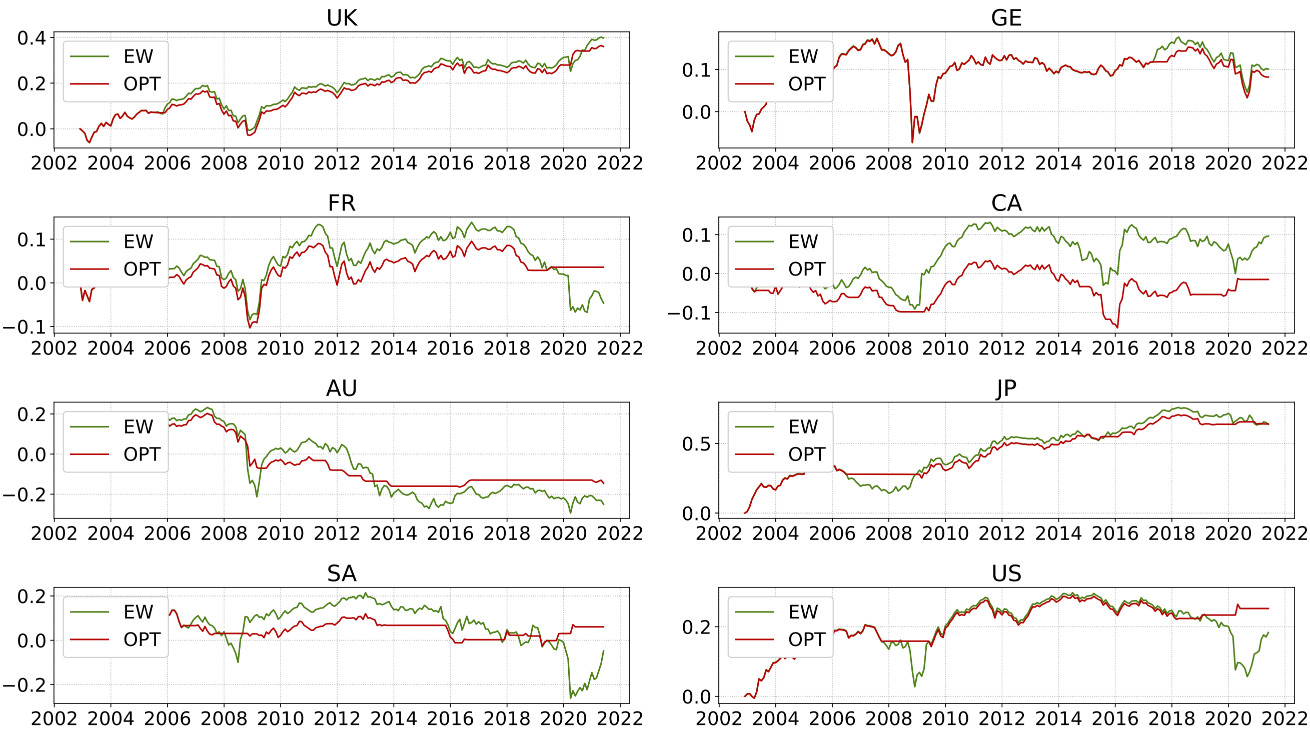

Figure 5.14.: Annual cumulative relative performance (log-scale) of equal weighted and optimal portfolios by country using optimised threshold for switching and linear regression lookback window.

5.3 Blending equal and cap weights using a Random Forest approach

One of the disadvantages of the rudimentary model in Section 5.2 is the abrupt and large changes in the portfolio weights as the portfolio switching between the cap weight and equal weight. Secondly, predicting both the direction and size of the relative performance of the equal weighted portfolio is a very tough task and one that creates significant risk in a real-world implementation should the regression model prediction be wrong. An approach which may be much more amenable to real-world implementation is rather skewing a portfolio towards either the equal weighted portfolio and thereby creating a blended portfolio based on whether the equal weighted portfolio is expected to outperform the cap weighted portfolio. In this case, the aim is to improve the performance of the cap weighted portfolio by accessing some of the alpha in the equal weighted portfolio, if the expectation is for the equal weighted portfolio to outperform.

This transforms the problem from one of regression to classification. That is, the problem to solve is whether or not the equal weighted portfolio is expected to outperform the cap weighted portfolio in the next period. This approach can be extended to not only consider whether or not the equal weighted portfolio is expected to outperform, but what probability can be assigned to this outcome. Let p_e be the probability of the equal weighted portfolio outperforming in the next period. Equation (5.3) provides a straight-forward way to translate this probability to the portfolio weight space. Recall that setting the parameter p in Equation (5.3) equal to 1 generates weights equal to the cap weighted portfolio as per Equation (5.4). As this parameter p tends towards 0, the weights tend towards an equal weight.

Therefore, setting p equal to $1-p_e$ will generate a portfolio closer to the cap weighted portfolio when the probability of the equal weighted portfolio outperforming is low, and therefore p_e is low, and vice versa. The next question is how to estimate p_e and how can the portfolio generating function, the drift process and the impact of leakage for the equal weighted portfolio be used to generate p_e? It is fair to assume that the features and output are likely to be related in a non-linear fashion and that this lends itself to an implementation of some machine learning (ML) approach. Secondly, an algorithm that is easy to understand, implement, optimise and interpret is important. For these reasons, a random forest classifier was selected.

Random forest models tend to perform well against other ML algorithms, are robust against over-fitting, and are easy to interpret with feature importance readily available. The starting point for random forests are classification and regression trees (CART) (Breiman et al., 1984). A CART model begins with the entire population and the resulting classes (or states) are already defined (supervised learning essentially). The algorithm begins splitting the population with the aim of increasing or maximising the homogeneity of the resulting sub-group at each node. An illustrative example is provided in Figure 5.15.

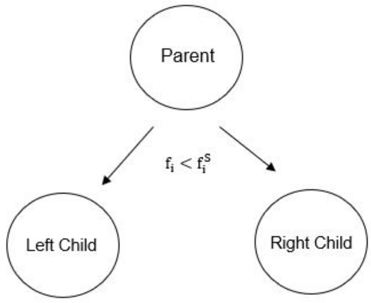

Figure 5.15.: Example of a node-split in a Classification and Regression Tree. The split criteria f_i^S is selected for the feature f_i such that the split maximises the homogeneity of the resulting child nodes.

The determinant of homogeneity can be any of a number of methods. One popular method, and the one utilised here, is the Gini index. In this setting, it is represented as,

$$G_i = \sum_{\forall k \neq j} p(k|i)p(j|i), \tag{5.5}$$

where k and j represent classes and i represents the node in question. In the model presented in this section there are only two states: equal weighted portfolio outperforms or underperforms the cap weighted portfolio. $p(k|i)$ is, therefore, the probability of being in class k at node i. In this instance, the goal would be to minimize G_i at each node and, as this is done for each consecutive node, build a tree such that the ends provide a final classification. As a result, it is possible to trace the exact determinants of a particular classification. It is also possible to determine feature importance using the reduction in G_i (or any other measure used) per feature. The higher the total reduction, the more important that particular feature is.

CART models represent one decision tree. Breiman (1996) extends this idea by building numerous trees (using the procedure above) independently with a specific classification determined by taking a "vote" among all the constructed classification trees. Furthermore, these trees are built using different bootstrap samples from the main sample of data, adding a layer of randomness to them. This is a concept referred to as "bagging". This was further extended in Breiman (2001) to not only use random bootstrap samples but also use a random selection of the available features for each tree. In other words, before building a specific tree, a random sample of data is bootstrapped and a random selection of the available features are made. That specific tree is built using only those features and that bootstrap sample of data. A collection of trees are then used to make predictions through a "voting" procedure. This results in the Random Forest model.

5.3.1 Methodology

The same data per country is used in these simulations and backtests as in the prior section and chapters. At the end of each month the portfolio generating function, drift process (excess growth rate for the equal weighted portfolio) and net impact of leakage are used as base features. These base features are transformed further by making use of the *tsfresh* Python package on the prior six months rolling data. The *tsfresh* package uses over 60 time series techniques to transform the base inputs into almost 800 features. A list of these techniques are provided in Appendix C.

After extracting the features, a basic Random Forest model is built using 100 estimators (trees) and building these to a maximum depth. This model is then fit on the features extracted, with the objective to classify the next month equal weighted portfolio performance as outperforming or underperforming the cap weighted portfolio. This is done using the *sklearn* package in Python. The purpose of this step is to filter out some of the features and the bottom 10% of features, ranked by feature importance (using the Gini index as explained above), are removed.

The final step is to optimize the Random forest model along the parameters number of estimators (trees) and the maximum depth of the tree (layers of nodes). This is done using k-fold cross validation by splitting the sample period into three, non-overlapping, subsamples or folds. Then, for each combination of parameters being optimized, fit the model on two of the three folds, and hold out the remaining fold as a pseudo-test set. Each fold is given the opportunity to act as a test set and an average score or accuracy is taken across all three for each combination of parameters.

The probability of the equal weighted portfolio underperforming in the next month is used from this model to set the parameter p in Equation 5.4 to derive weights

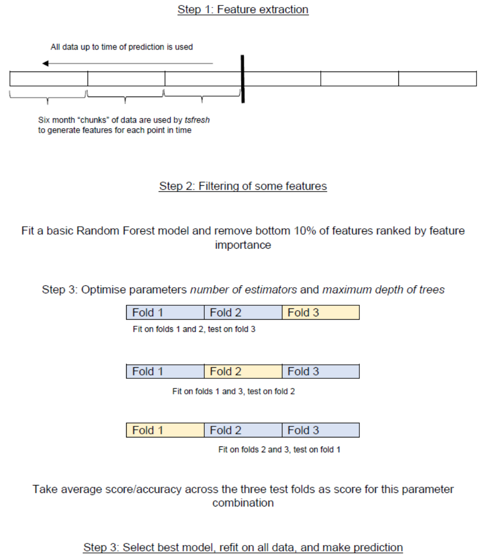

Figure 5.16.: Process for selecting features, optimising model and making prediction at the end of each month. Target variable is 0: Equal weighted portfolio underperforms or 1: Equal weighted portfolio outperforms cap weighted portfolio in the next month. Probability of target variable being 0 in the next month is used to derive weights.

for the next month. If this probability is low, then the equal weighted portfolio is expected to outperform and so p is closer to 0. If the probability is high, then p will be closer to 1 as the equal weighted portfolio is expected to underperform.

5.3.2 Results

The results for each country are presented in Table 5.9 for the Random Forest (RF), equal weighted and cap weighted portfolios. These results include 15bps of transaction costs as before.

Across all the countries, the RF model outperforms the cap weighted portfolio, especially if the equal weighted portfolio in the respective countries outperform the cap weighted portfolio. The RF model does, however, do so with lower volatility and for the countries where the equal weighted portfolio outperforms the cap weighted portfolio, the RF model portfolio does well to match the cap weighted portfolio. One example is Australia, where the RF model matches the cap weighted portfolio and thereby outperforms the equal weighted portfolio.

Table 5.9.: CAGR and volatility of the Random Forest (RF), equal, and cap weighted portfolios for a selection of countries[a].

	CAGR (%)			Volatility (%)		
Country	Cap	Equal	RF	Cap	Equal	RF
Australia	11.0	10.0	11.0	16.2	17.3	16.2
Canada	8.6	9.7	9.2	18.1	18.4	18.1
France	4.8	5.5	5.5	23.0	24.0	23.4
Germany	3.5	5.9	5.7	24.0	22.9	22.8
Japan	2.8	8.5	6.3	21.4	19.9	20.3
South Africa	14.6	14.4	14.8	20.1	19.3	19.3
United Kingdom	4.7	7.2	6.3	19.4	19.7	19.4
United States	10.3	12.0	11.2	19.7	20.9	20.1

[a]Portfolios are rebalanced monthly with 15bps of costs.

Table 5.10 shows the Sharpe ratio for all three portfolios, together with the statistical test for differences in Sharpe ratios. The test is the same as performed in Section 5.2 and tests the equal weighted and RF model's Sharpe ratio against that of the cap weighted portfolio. Across all the countries the RF model portfolio generates a Sharpe ratio between that of the equal and cap weighted portfolios. That is, better than the worst of the cap and equal weighted portfolios, but slightly worse than the best. P-values for the statistical test in differences between Sharpe ratios are generally still quite high. There are, however, some countries where the p-value is lower than the average among countries. These countries include Japan and United Kingdom where blending the equal weighted portfolio weights into the RF model creates additional alpha over the cap weighted portfolio. Australia also shows a lower p-value for the Sharpe ratio, where blending equal and cap weights does seemingly create a somewhat different return profile for the RF model portfolio even though the equal weighted portfolio does not do well in Australia.

Figure 5.17 presents the performance of the RF model, together with the equal weighted portfolio, relative to the cap weighted portfolio. In general, the RF model outperforms the cap weighted portfolio but does lag the equal weighted portfolio in markets where the equal weighted portfolio does well. In countries such as Australia and South Africa, where the equal weighted portfolio does poorly relative to the cap weighted portfolio, the RF model portfolio moderately outperforms the cap weighted portfolio.

Figure 5.18 shows the optimal p used in Equation 5.4 to generate the weights for the RF model portfolio in each country. These are not similar across the different countries but there are some patterns within each country. In the UK for example, the optimal p is generally below 0.5, with the same pattern occurring in Japan.

Table 5.10.: Risk adjusted performance of the optimised, equal, and cap weighted portfolios for a selection of countries[a].

	Sharpe ratio[b,c]		
Country	**Cap**	**Equal**	**RF**
Australia	0.502	0.429	0.505 (0.172)
Canada	0.458	0.508	0.488 (0.96)
France	0.278	0.305	0.305 (0.436)
Germany	0.212	0.311	0.304 (0.516)
Japan	0.234	0.506	0.401 (0.176)
South Africa	0.412	0.411	0.431 (0.636)
United Kingdom	0.195	0.315	0.275 (0.192)
United States	0.486	0.545	0.521 (0.652)

[a]Portfolios are rebalanced monthly with 15bps of costs.

[b]Respective risk-free rates used in Sharpe ratio calculations.

[c]P-value for two sided hypothesis test as per Ledoit and Wolf (2008) shown in brackets. The null hypothesis is that the Sharpe Ratio is equal to that of the cap weighted portfolio.

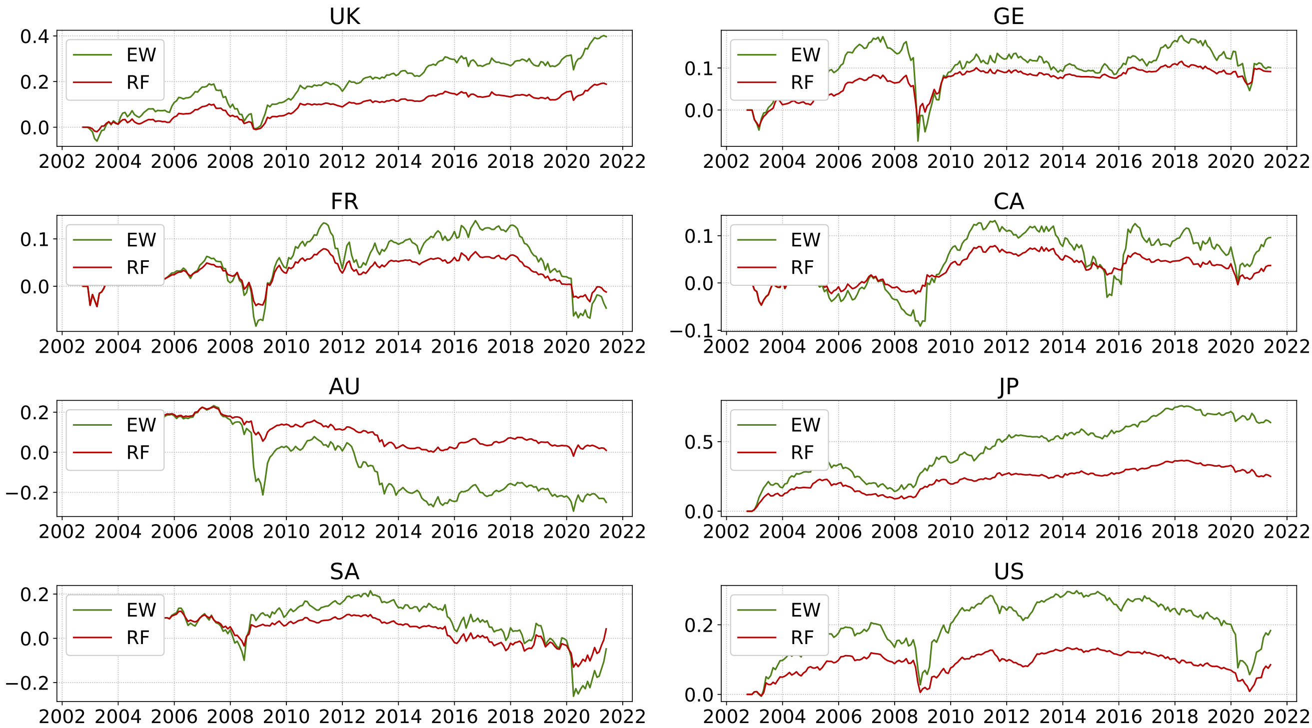

Figure 5.17.: Annual cumulative relative performance (log-scale) of equal weighted and Random Forest (RF) portfolios by country.

Within these countries the equal weighted portfolio has done well and so the model skewing towards equal weights makes sense.

In Australia the optimal p is mostly below 0.5 until 2012 where the model switches to values mostly above 0.5 and towards the cap weights. This coincides with the period when the Australian equal weighted portfolio underperformed. The US is also a fairly interesting case, with two distinct periods; pre-2014 when the optimal p was largely below 0.5, and post-2014 where the optimal p is mostly above 0.5 indicative of the cap weighted portfolio performing well in this latter period. South Africa shows a stranger pattern, where the optimal p is largely mixed but rising moderately overall until 2016, that is, skewing towards cap weights, before actually averaging optimal values below 0.5. Over this former period the RF model, therefore, underperforms the cap weighted portfolio as the equal weighted portfolio has poor performance in this period. Performance of the RF model portfolio is, however, still ahead of the equal weighted portfolio largely due to the increase in the optimal value in early parts of 2020 and thereby missing most of this underperformance. The RF model portfolio in SA does, however, skew back towards equal weights in time to capture most of the recovery post-March 2020.

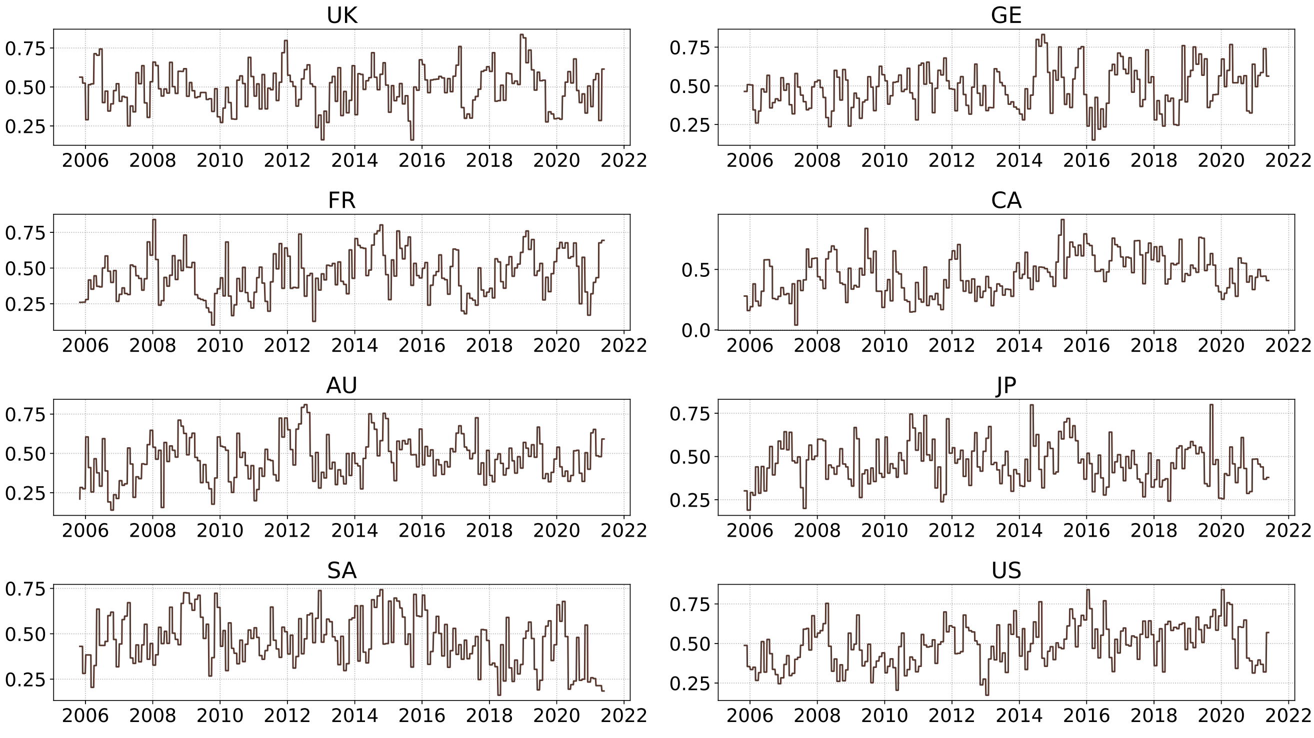

Figure 5.18.: Optimal p used to determine weights as per the Random Forest (RF) model by country. A value closer to 1 results in weights closer to cap weights and a value closer to 0 results in weights closer to equal weights.

As mentioned, across countries the optimal values are not very similar and, in fact, the average rolling 12 month correlation is very small among these optimal values at 3%. There are, however, periods where this correlation spikes such as around 2008 and more recently in early 2020. This could be indicative of shifts in the market regime and one extension of this work could be to investigate the use of these to identify turning points in the general market regime.

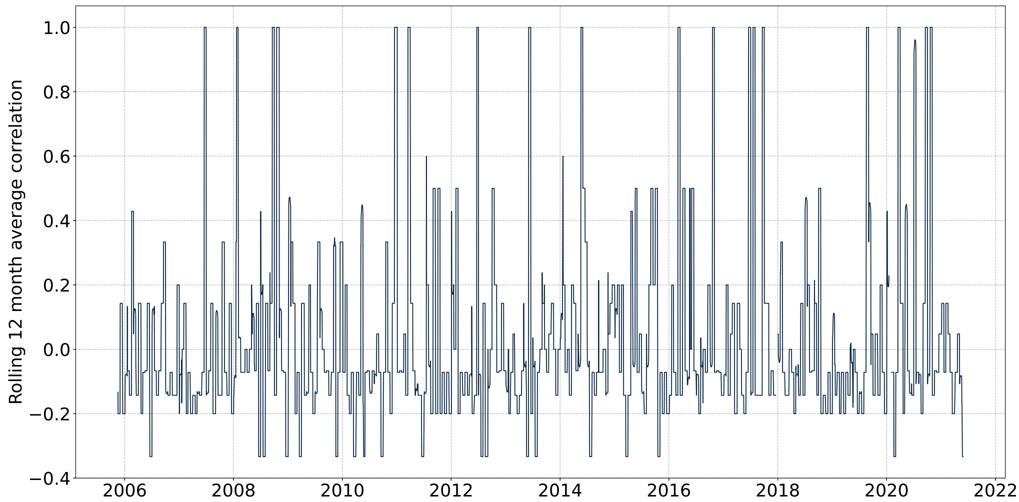

Figure 5.19.: Correlations over a rolling 12 month period for the optimal p used to determine the weights in the Random Forest (RF) model for each country.

5.4 Conclusion

This chapter attempted to use the main components in Equation (3.28), to generate portfolios that either reduce the relative drawdown of the equal weighted portfolio or attempt to blend the cap and equal weighted portfolios to achieve improved risk-adjusted returns. The first section attempted to create as simple a model as possible by using a rudimentary linear regression and in turn highlight that monitoring the portfolio generating function, as given by Equation (3.20), the excess growth rate, and the impact of leakage, Equation (3.27), are important when improving the short-term performance of the equal weighted portfolio relative to a cap weighted index. This model largely achieved its aims of reducing relative drawdowns in the US and SA equal weighted portfolios, but performance was mixed across other countries. After some optimisation, the model performed reasonably well in countries such as the US, SA, UK, France, Australia, and Japan. The model generally underperformed in countries where the equal weighted portfolio performed very well relative to the cap weighted portfolio.

The second model aimed at improving performance by creating an optimal blend of the cap and equal weighted portfolios. In general this model combined the best of both worlds, generating portfolios which performed in between the cap and equal weighted portfolios in each market. Therefore, in markets where the equal weighted portfolio performs very well, this model underperformed the equal weighted portfolio but outperformed the cap weighted portfolio. The UK and Japanese equity markets are examples of this. Importantly, in markets where the equal weighted portfolio struggled, this model outperformed the equal weighted portfolio with performance not too far from that of the cap weighted portfolio. Examples include Australian and South African equity markets.

Optimal values in the RF model are generally not correlated across different countries with a very low average correlation, however there are periods where correlations spike which might be useful in identifying changes in market regimes.

6 Conclusion

The equal weighted portfolio has an intuitively appealing portfolio construction methodology in that it requires no prediction or estimation of any parameters. This makes it a somewhat unique alternative to the cap weighted portfolio and even other "risk-aware" portfolios such as the minimum variance portfolio. Risk in the portfolio tends to be allocated more uniformly (ignoring correlations of course) across all the constituent stocks as opposed to the cap weighted portfolio which can sometimes become concentrated in a handful of stocks. This has the intuitive result of increasing the diversification of the resulting portfolio relative to the cap weighted portfolio. Although the equal weighted portfolio tends to outperform in many equity markets over the long-term (see Table 1.1 and Chapter 2) on a risk-adjusted basis, its performance over the short-term can lead to significant underperformance of the respective cap weighted index. Furthermore, although most equity markets show the equal weighted portfolio outperforming, the level of outperformance is very different across the equity markets.

In order, therefore, to be comfortable in using equal weighted portfolios as alternatives to cap weighted approaches in practice, it is important to understand the drivers of its relative performance over time and across different markets. The main aim of the thesis, therefore, was to contribute to the existing literature by broadening the understanding of the equal weighted portfolio relative to the cap weighted portfolio. This was primarily done for South African and US equities but also with some insight from a handful of other countries, and, in the process, to address the research questions set out in Section 1.6.

Chapter 2 focused on the historical performance of the equal weighted portfolio in the US and SA equity markets, as well as in a selection of other developed equity markets. In the US markets, the equal weighted portfolio has performed relatively well against the cap weighted portfolio, however, the case is very different in SA where the equal weighted portfolio has only just kept up with the cap weighted portfolio and this has largely been due to the performance since March 2020. Prior to that, the equal weighted portfolio had been on a long underperformance streak starting in 2012. Despite this, the performance of the equal weighted portfolio in SA is highly correlated with that of the US. As one would expect, however, the key difference does appear to be in the level of concentration in the cap weights where SA equities show much higher levels of concentration.

Although the SA and US equal weighted portfolios appear highly correlated, the same is not true across the selection of countries highlighted in this thesis. Correlations have largely been low, but have increased substantially since the start of 2020, although this is still small compared to the correlation between US and SA equal weighted portfolios. Performance is fairly mixed across the different countries with seemingly good performance in the Japanese and UK equity markets and particularly poor performance in the Australian equity market. In fairness, the underperformance in Australia is largely due to the period 2012 to 2014, and relative performance was muted thereafter. This is in contrast to the SA market, where underperformance continued well after 2012.

Chapter 3, provides the basic theoretical understanding for stochastic portfolio theory with the objective of showing how this framework can be used to analyse the relative performance of the equal weighted portfolio. Stochastic portfolio theory looks to understand the various parts of a portfolio's relative performance against a cap weighted benchmark or portfolio with a specific focus on functionally generated portfolios. That is, portfolios that have weights generated by some function of the cap weights. Stochastic portfolio theory provides a robust theoretical framework for understanding this relative performance by considering the main drivers, namely, the level of concentration in cap weights (portfolio generating function), the level of diversification benefits (the excess growth rate), and the impact of leakage (as stocks move in and out of the index).

Stochastic portfolio theory and the resulting explanations of relative performance of functionally generated portfolios allows for an attribution and understanding of empirical performance, in this case for the equal weighted portfolio. One such function for generating a portfolio is the geometric mean of the cap weights which is shown to generate equal weights and using the equations in Chapter 3 allows for a decomposition of the drivers of relative performance. This framework also confirms previous research showing that the equal weighted portfolio should outperform the cap weighted portfolio over the long term. The results in stochastic portfolio theory show this should be expected as long as the market's level of concentration is bounded. Long-term outperformance is generated, in this case, by the positive effects of diversification.

This empirical analysis and attribution is detailed in Chapter 4. The portfolio generating function of the S&P500 equal weighted portfolio, while displaying periods of increasing concentration, appears to be well bounded and this leads to periods of both positive and negative contributions which should offset themselves for an overall net zero contribution as stochastic portfolio theory would predict. This is in contrast to SA where the portfolio generating function has contributed negatively to relative performance on a very consistent basis, especially since 2012. In terms of the benefits of diversification (termed excess growth rate here), in the US, there is a distinct difference in the contribution from the excess growth rate before and after 2010. Excess growth rates in the S&P500 equal weighted portfolio appear much higher pre-2010. Excess growth rates do, however, appear to spike during market turmoil as one might have expected given that volatilities increase and this allows for a higher impact from diversification.

These two factors, the level of concentration (portfolio generating function) and the benefits of diversification (excess growth rate) are usually discussed when analysing the relative performance of the equal weighted portfolio. In the stochastic portfolio theory framework another factor can also be identified that is usually hard to quantify or rarely discussed. This is a factor termed leakage in stochastic portfolio theory and is the result of stocks moving in and out of the benchmark. This usually affects the equal weighted portfolio negatively as it holds a larger weight in those lower ranked stocks than the cap weighted portfolio. Therfore, as these stocks move in and out of the index, the equal weighted portfolio has to churn a larger portion of the portfolio. Leakage was identified as an important, and in some years a major, negative contributor to the equal weighted portfolio's relative performance. In the S&P500 equal weighted portfolio this contributed -1.8% per annum to the relative performance of the equal weighted portfolio compared to a much higher -3.2% for the Top 40. In SA, specifically, leakage has been a major drag on relative performance with an average contribution of -8.6% per annum since 2011 as the rotation of stocks in and out of the index in the last ten years has increased.

This approach also allows for an estimate of the increased concentration from various stocks and their impact on the relative performance of the equal weighted portfolio. This is done for the technology stocks in the S&P500, with an estimate of 10% to 12% on a rolling five year basis, and for Naspers and Prosus in SA, which appear to have had a negative impact of between 20% and 25% on the relative performance of the Top 40 equal weighted portfolio.

This analysis was also applied to a selection of other countries. Most of these countries experienced high contributions to relative returns for the equal weighted portfolio from the excess growth rate, with the notable exception of France and Germany. While in SA the portfolio generating function contributed negatively to relative returns, for most countries the contribution was relatively small and in some countries, the contribution was actually positive (such as in the UK and Germany). It was noted in Chapter 2 that the Japanese equal weighted portfolio showed the largest outperformance relative to the cap weighted portfolio. Interestingly, this outperformance appears to come from a very high excess growth rate and a very low impact of leakage as the generating function contribution in Japanese equities is actually the most negative across all the other countries.

An attempt was made in Chapter 5 to use these drivers (excess growth rate, generating function, and leakage) to build a model to try an improve the relative performance. Specifically to attempt to reduce the relative drawdowns in the equal weighted portfolio which can be quite significant. A rudimentary linear regression model was built that switches between the cap and equal weighted portfolios. This model largely achieved its aims of reducing relative drawdowns in the US and SA equal weighted portfolios, but performance was mixed across other countries. After some optimisation, the model performed reasonably well in countries such as the US, SA, UK, France, Australia, and Japan.

A second, more complicated, model was also analysed where the aim was to outperform the cap weighted portfolio by blending weights across the cap and equal weight spectrum using the predicted probability of the equal weighted portfolio outperforming over the following month using a Random Forest model. This model resulted in combining the best of both worlds, generating portfolios which performed in between the cap and equal weighted portfolios in each market. Therefore, in markets where the equal weighted portfolio performs very well, this model underperformed the equal weighted portfolio but outperformed the cap weighted portfolio. The UK and Japanese equity markets are examples of this. In markets where the equal weighted portfolio struggled (such as Australia and South Africa), this model outperformed the equal weighted portfolio with performance not too far form that of the cap weighted portfolio.

There are a few areas of future research that could extend the analysis presented in this thesis:

- **Use of the excess growth rate as a market regime indicator:** The excess growth rates were presented in Chapter 4 for each country. Furthermore, it appeared as though spikes in the excess growth rates were correlated with market regime changes, specifically spiking when the euqiyt markets experienced large declines. Could this be used to identify and pre-empt market regime changes?

- **Use of models other than Geometric Brownian Motion:** The basic stock price process in Equation (3.1) uses the common Geometric Brownian Motion model for stocks. It may be of interest to consider other models, such as the Ornstein-Uhlenbeck model. This could also be of interest in any application to bond portfolios, or portfolios of other assets.

- **Attempting to reduce leakage:** Leakage was shown to be a significant component of the equal weighted portfolio's relative return in Chapter 4. Is it possible to reduce the impact of leakage on the equal weighted portfolio by systematically reducing the weight of the lowest weighted two or three stocks in the index?

- **Understanding the link between index size (by number of index members) and performance:** In Chapter 4 it was found that Japan had the lowest impact from leakage and South Africa had the largest. These impacts had a large effect on the relative performance of the respective equal weighted portfolios. Is it possible to show this link directly using both different indices in each market of varying sizes and within a simulated market (using a Monte-Carlo approach for example)?

- **Conditions for which the equal weighted portfolio should outperform:** Is it possible to define a general set of real-world conditions for which to evaluate whether an equal weighted portfolio in a specific equity market should perform well (or poorly) relative to the respective cap weighted index? Given that leakage plays an important role in reducing the expected long-term outperformance of the equal weighted portfolio, it would likely be beneficial to thoroughly understand leakage in different market environments to address this question.

- **Optimal capped weights:** Some index providers have used "capped" weights as a way of limiting the concentration of weights in the cap weighted portfolio/index. These "caps" are usually arbitrarily selected at levels such as 10%, for example, and any weights above this level are redistributed to the remaining stocks. This is repeated until all stocks have a weight lower than, or equal to, the "cap" weight. Is it possible to use the stochastic portfolio theory framework to set an optimal level for these caps?

- **Optimal rebalancing for the equal weighted portfolio:** In this thesis rebalancing was performed each month. The rebalancing frequency itself can, however, have a large impact on portfolio performance. Is it possible, therefore, to determine optimal times to rebalance based on levels of diversification (the excess growth rate) and movements in the portfolio generating function (increasing or decreasing concentration)?

- **Analysing smart beta portfolios:** Could the same empirical analysis be performed on some smart beta portfolios? This would be of particular interest in the case of "value"-based smart beta portfolios and their underperformance over recent years. As long as the portfolio can be generated using the expressions in Theorem A.2.1, stochastic portfolio theory should be able to provide a framework for this analysis.

- **Using inputs from other countries in models**: In Chapter 5, the models were fit using only data from within the specific country in question. In Chapter 4, however, there was some indication of correlations between the excess growth rates in each country. Could this be used to improve the models in Chapter 5?

- **Modelling the portfolio generating function**: Is it possible to model and predict the equal weighted portfolio's portfolio generating function directly? This would likely involve modelling the ranked stock weights directly. That is, modelling the weight of position i in an index (for example, the weight of position 10 in the index) rather than on a specific stock basis. Achieving this would likely go a long way to improving the models in Chapter 5.

Bibliography

Amenc, Noël, Felix Goltz, Ashish Lodh, and Lionel Martellini (2012). "Diversifying the diversifiers and tracking the tracking error: Outperforming cap-weighted indices with limited risk of underperformance". In: *The Journal of Portfolio Management* 38.3, pp. 72–88.

Bartlett, Peter, Yoav Freund, Wee Sun Lee, and Robert E Schapire (1998). "Boosting the margin: A new explanation for the effectiveness of voting methods". In: *The annals of statistics* 26.5, pp. 1651–1686.

Bernard, Simon, Sébastien Adam, and Laurent Heutte (2007). "Using random forests for handwritten digit recognition". In: *Ninth international conference on document analysis and recognition (ICDAR 2007)*. Vol. 2. IEEE, pp. 1043–1047.

Bhattacharya, Utpal and Neal Galpin (2011). "The global rise of the value-weighted portfolio". In: *Journal of Financial and Quantitative Analysis*, pp. 737–756.

Bolognesi, Enrica, Giuseppe Torluccio, and Andrea Zuccheri (2013). "A comparison between capitalization-weighted and equally weighted indexes in the European equity market". In: *Journal of Asset Management* 14.1, pp. 14–26.

Booth, David G and Eugene F Fama (1992). "Diversification returns and asset contributions". In: *Financial Analysts Journal* 48.3, pp. 26–32.

Bouchey, Paul, Vassilii Nemtchinov, Alex Paulsen, and David M Stein (2012). "Volatility harvesting: Why does diversifying and rebalancing create portfolio growth?" In: *The Journal of Wealth Management* 15.2, pp. 26–35.

Bouchey, Paul, Vassilii Nemtchinov, and Ting-Kam Leonard Wong (2015). "Volatility harvesting in theory and practice". In: *The Journal of Wealth Management* 18.3, pp. 89–100.

Breiman, Leo (1996). "Bagging predictors". In: *Machine learning* 24.2, pp. 123–140.

– (2004). "Consistency for a simple model of random forests". In.

– (2001). "Random forests". In: *Machine learning* 45.1, pp. 5–32.

Breiman, Leo, Jerome H. Friedman, richard A. Olshen, and Charles J. Stone (1984). *Classification and regression trees*.

Buetow Jr, Gerald W, Ronald Sellers, Donald Trotter, Elaine Hunt, and Willie A Whipple Jr (2002). "The benefits of rebalancing". In: *Journal of Portfolio Management* 28.2, p. 23.

Chambers, Donald R and John S Zdanowicz (2014). "The limitations of diversification return". In: *The Journal of Portfolio Management* 40.4, pp. 65–76.

Christ, Maximilian, Nils Braun, Julius Neuffer, and Andreas W Kempa-Liehr (2018). "Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package)". In: *Neurocomputing* 307, pp. 72–77.

Cremers, KJ Martijn and Antti Petajisto (2009). "How active is your fund manager? A new measure that predicts performance". In: *The review of financial studies* 22.9, pp. 3329–3365.

Cuthbertson, Keith, Simon Hayley, Nick Motson, and Dirk Nitzsche (2016). "What Does Rebalancing Really Achieve?" In: *International Journal of Finance & Economics* 21.3, pp. 224–240.

Dash, Srikant and Keith Loggie (2008). "Equal weight indexing-five years later". In: *Available at SSRN 1321726*.

Dash, Srikant and Liyu Zeng (2010). "Equal Weight Indexing-Seven Years Later". In: *Available at SSRN 1691213*.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal (2009). "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" In: *The review of Financial studies* 22.5, pp. 1915–1953.

Dichtl, Hubert, Wolfgang Drobetz, and Martin Wambach (2016). "Testing rebalancing strategies for stock-bond portfolios across different asset allocations". In: *Applied Economics* 48.9, pp. 772–788.

– (2014). "Where is the value added of rebalancing? A systematic comparison of alternative rebalancing strategies". In: *Financial Markets and Portfolio Management* 28.3, pp. 209–231.

Ekren, Ibrahim, Ren Liu, and Johannes Muhle-Karbe (2018). "Optimal rebalancing frequencies for multidimensional portfolios". In: *Mathematics and Financial Economics* 12.2, pp. 165–191.

Fernholz, Robert (1998). "Crossovers, dividends, and the size effect". In: *Financial Analysts Journal* 54.3, pp. 73–78.

– (2001a). "Equity portfolios generated by functions of ranked market weights". In: *Finance and Stochastics* 5.4, pp. 469–486.

– (1999a). "On the diversity of equity markets". In: *Journal of Mathematical Economics* 31.3, pp. 393–417.

– (1999b). "Portfolio generating functions". In: *Quantitative Analysis in Financial Markets: Collected Papers of the New York University Mathematical Finance Seminar*. World Scientific, pp. 344–367.

– (2001b). "Stable models for the distribution of equity capital". In.

– (2002). "Stochastic portfolio theory". In: *Stochastic portfolio theory*. Springer, pp. 1–24.

Fernholz, Robert and Ioannis Karatzas (2009). "Stochastic portfolio theory: an overview". In: *Handbook of numerical analysis* 15, pp. 89–167.

Fernholz, Robert, Ioannis Karatzas, and Constantinos Kardaras (2005). "Diversity and relative arbitrage in equity markets". In: *Finance and Stochastics* 9.1, pp. 1–27.

Fernholz, Robert and Brian Shay (1982). "Stochastic portfolio theory and stock market equilibrium". In: *The Journal of Finance* 37.2, pp. 615–624.

Fernholz, Robert and One Palmer Square (1998). *Arbitrage in equity markets*. Tech. rep. Working Paper, Intech, Princeton.

Frazzini, Andrea, Ronen Israel, and Tobias J Moskowitz (2018). "Trading costs". In: *Available at SSRN 3229719*.

Ghojogh, Benyamin and Mark Crowley (2019). "The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial". In: *arXiv preprint arXiv:1905.12787*.

Hallerbach, Winfried G (2014). "Disentangling rebalancing return". In: *Journal of Asset Management* 15.5, pp. 301–316.

Haugen, Robert A and Nardin L Baker (1991). "The efficient market inefficiency of capitalization–weighted stock portfolios". In: *The Journal of Portfolio Management* 17.3, pp. 35–40.

Hilliard, Jimmy E and Jitka Hilliard (2018). "Rebalancing versus buy and hold: theory, simulation and empirical analysis". In: *Review of Quantitative Finance and Accounting* 50.1, pp. 1–32.

Hsu, Jason C (2004). "Cap-weighted portfolios are sub-optimal portfolios". In: *Journal of investment Management* 4.3.

Karatzas, Ioannis (2006). "A survey of stochastic portfolio theory". In: *The Eugene Lukacs Lectures, Bowling Green University*.

Kruger, R and P Van Rensburg (2008). "Evaluating and constructing equity benchmarks in the South African portfolio management context". In: *Investment Analysts Journal* 37.67, pp. 5–17.

Kuang, Wei (2021). "Conditional covariance matrix forecast using the hybrid exponentially weighted moving average approach". In: *Journal of Forecasting*.

Kumar, Manish and M Thenmozhi (2006). "Forecasting stock index movement: A comparison of support vector machines and random forest". In: *Indian institute of capital markets 9th capital markets conference paper*.

Ledoit, Oliver and Michael Wolf (2008). "Robust performance hypothesis testing with the Sharpe ratio". In: *Journal of Empirical Finance* 15.5, pp. 850–859.

Liaw, Andy, Matthew Wiener, et al. (2002). "Classification and regression by randomForest". In: *R news* 2.3, pp. 18–22.

Liu, Ding (2019). "Analytical solutions of optimal portfolio rebalancing". In: *Quantitative Finance* 19.4, pp. 683–697.

Malladi, Rama and Frank J Fabozzi (2017). "Equal-weighted strategy: Why it outperforms value-weighted strategies? Theory and evidence". In: *Journal of Asset Management* 18.3, pp. 188–208.

Martellini, Lionel (2009). *Towards the design of better equity benchmarks*. Tech. rep. EDHEC Working Paper.

Masters, Seth J (2003). “Rebalancing”. In: *Journal of Portfolio Management* 29.3, p. 52.

Mattei, Michael D and Nicholas Mattei (2016). “Analysis of fixed and biased asset allocation rebalancing strategies”. In: *Managerial Finance* 42.1, pp. 42–50.

Pantaleo, Ester, Michele Tumminello, Fabrizio Lillo, and Rosario N Mantegna (2011). “When do improved covariance matrix estimators enhance portfolio optimization? An empirical comparative study of nine estimators”. In: *Quantitative Finance* 11.7, pp. 1067–1080.

Petajisto, Antti (2013). “Active share and mutual fund performance”. In: *Financial Analysts Journal* 69.4, pp. 73–93.

Platen, Eckhard and Renata Rendek (2012). “Approximating the numéraire portfolio by naive diversification”. In: *Journal of Asset Management* 13.1, pp. 34–50.

Plyakha, Yuliya, Raman Uppal, and Grigory Vilkov (2014). “Equal or value weighting? Implications for asset-pricing tests”. In: *Implications for Asset-Pricing Tests (January 15, 2014)*.

– (2015). “Why Do Equal-Weighted Portfolios Outperform Value-Weighted Portfolios?” In: *Working paper, SSRN 2724535*.

– (2012). “Why does an equal-weighted portfolio outperform value-and price-weighted portfolios?” In: *Available at SSRN 2724535*.

Raubenheimer, Heidi (2010). “Concentration in the South African equity market and its implied restrictions on the long only equity fund manager’s opportunity set”. In: *South African Journal of Business Management* 41.4, pp. 1–10.

RiskMetrics (1996). *Risk Metrics TM, Technical Document. Risk Metrics Group*.

Rundo, Francesco, Francesca Trenta, Agatino Luigi di Stallo, and Sebastiano Battiato (2019). “Machine learning for quantitative finance applications: A survey”. In: *Applied Sciences* 9.24, p. 5574.

Samo, Yves-Laurent Kom and Alexander Vervuurt (2016). “Stochastic portfolio theory: a machine learning perspective”. In: *arXiv preprint arXiv:1605.02654*.

Shannon, Claude Elwood (1948). “A mathematical theory of communication”. In: *The Bell system technical journal* 27.3, pp. 379–423.

Sharpe, William F (1964). “Capital asset prices: A theory of market equilibrium under conditions of risk”. In: *The journal of finance* 19.3, pp. 425–442.

Sun, Walter, Ayres Fan, Li-Wei Chen, Tom Schouwenaars, and Marius A Albota (2006). “Optimal rebalancing for institutional portfolios”. In: *The Journal of Portfolio Management* 32.2, pp. 33–43.

Taljaard, Byran and Eben Maré (2019a). “Considering the use of an equal-weighted index as a benchmark for South African equity investors”. In: *South African Actuarial Journal* 19.1, pp. 53–70.

– (2021a). “If the equal weighted portfolio is so great, why isn’t it working in South Africa?” In: *Investment Analysts Journal* 50.1, pp. 32–49.

– (2019b). “Too Much Rebalancing Is Not a Good Thing”. In: *Available at SSRN 3484337*.

– (2021b). “Why has the equal weight portfolio underperformed and what can we do about it?” In: *Quantitative Finance*, pp. 1–14.

Timofeev, Roman (2004). “Classification and regression trees (CART) theory and applications”. In: *Humboldt University, Berlin*, pp. 1–40.

Willenbrock, Scott (2011). “Diversification return, portfolio rebalancing, and the commodity return puzzle”. In: *Financial Analysts Journal* 67.4, pp. 42–49.

Zakamulin, Valeriy (2015). “A test of covariance-matrix forecasting methods”. In: *The Journal of Portfolio Management* 41.3, pp. 97–108.

This page has been intentionally left blank

Stochastic portfolio theory

A

In this appendix a more formal introduction to stochastic portfolio theory than Chapter 3 is provided. The first section describes the stock price process and the portfolio price process. The second section describes portfolio generating functions before deriving the portfolio generating function and drift process for the equal weighted portfolio.

A.1 Portfolio price process

Definition A.1.1. The price process of a stock X is given by

$$\frac{dX(t)}{X(t)} = \alpha(t)dt + \sigma(t)dW(t), \quad t \in [0, \infty), \tag{A.1}$$

where $\alpha(t)$ and $\sigma(t)$ are the stock's arithmetic rate of return and volatility, respectively. $dW(t)$ is a Brownian motion.

Using Itô's lemma on the function $F(t) = \log X(t)$ results in the popular logarithmic model for the continuous-time stock price processes:

$$d\log X(t) = \gamma(t)dt + \sigma(t)dW(t), \tag{A.2}$$

where $\sigma(t)$ and $dW(t)$ are defined as before and $\gamma(t)$ is defined as:

$$\gamma(t) = \alpha(t) - \frac{1}{2}\sigma^2(t). \tag{A.3}$$

For completeness we present the following definition of the stock price process $X(t)$ along the lines of the same definition in Fernholz (2002).

Definition A.1.2. A stock price process X is a process that satisfies the stochastic differential equation

$$d\log X(t) = \gamma(t)dt + \sigma(t)dW(t), \qquad t \in [0, \infty),$$

where $W(t)$ is a Brownian motion, γ is measurable, adapted and satisfies $\int_0^T |\gamma(t)|\, dt < \infty$, for all $T \in [o, \infty)$, a.s. Furthermore, the $\sigma(t)$ is measurable, adapted and satisfies:

1. $\int_0^T \sigma^2(t)dt < \infty, \qquad T \in [0, \infty), \quad a.s.,$
2. $\lim_{t\to\infty} t^{-1}\sigma^2(t)\log\log t = 0, \quad a.s,$
3. $\sigma^2(t) > 0, \quad t \in [0, \infty), \quad a.s..$

Consider a market M of a family of stocks $X_1, ..., X_n$, each defined by Equation (A.2). A portfolio in the market M is a measurable, adapted vector-valued process π, with $\pi(t) = (\pi_1(t), ..., \pi_n(t))$, for $t \in [0, \infty)$ and

$$\sum_{i=1}^{n} \pi_i(t) = 1, \qquad t \in [0, \infty).$$

We say a market is non-degenerate if there exists a number $\epsilon_1 > 0$ such that,

$$x\sigma(t)x^T \geq \epsilon_1 \|x\|^2, \quad x \in \Re^n, \quad t \in [0, \infty). \tag{A.4}$$

That is, a market is non-degenerate if the variance of the market portfolio is bounded away from zero.

Furthermore, if there exists a number $\epsilon_2 > 0$ such that,

$$x\sigma(t)x^T \leq \epsilon_2 \|x\|^2, \quad x \in \Re^n, \quad t \in [0, \infty), \tag{A.5}$$

we say the market M has bounded variance.

Simply put, we have a market M, which has n stocks. A valid portfolio within this market is any combination of the stocks in M such that the weights, $\pi_i(t)$, sum to one.

The process $\pi_i(t)$ therefore represents the proportion of capital invested in the i-th stock. Now, let $Z_\pi(t)$ represent the value of some portfolio with weights π at time t. Obviously, the amount invested in stock X_i is given by,

$$\pi_i(t)Z_\pi(t).$$

Therefore, if the price X_i changes by $dX_i(t)$ then the change in the portfolio value $Z_\pi(t)$ is given by

$$\pi_i(t)Z_\pi(t)\frac{dX_i(t)}{X_i(t)}.$$

and the total change in the portfolio can therefore be expressed by,

$$\frac{dZ_\pi(t)}{Z_\pi(t)} = \sum_{i=1}^{n} \pi_i(t)\frac{dX_i(t)}{X_i(t)}. \tag{A.6}$$

The instantaneous rate of change for the portfolio, $Z_\pi(t)$, is therefore the weighted sum of instantaneous changes of the stocks in the portfolio.

Let the covariance processes for logX_i and logX_j be given by $\sigma_{ij}(t)$ such that,

$$\sigma_{ij}(t)dt = d\left\langle \log X_i, \log X_j \right\rangle_t, \quad t \in [0,\infty). \tag{A.7}$$

For $i = 1,...,n$, the process $\sigma_{ii}(t) = \langle \log X_i \rangle$ is called the covariance process of X_i.

We can now use Equation (A.6) to derive a price process for the portfolio $Z_\pi(t)$ in differential form. To do this we set out the following proposition as in Fernholz (1999a).

Proposition A.1.1. *Let π be a portfolio and let,*

$$d\log Z_\pi(t) = \gamma_\pi(t)dt + \sum_{i=1}^{n} \pi_i(t)\sigma_i(t)dW_i(t), \tag{A.8}$$

where

$$\gamma_\pi(t) = \sum_{i=1}^{n} \pi_i(t)\gamma_i(t) + \frac{1}{2}\left(\sum_{i=1}^{n} \pi_i(t)\sigma_{ii}(t) - \sum_{i,j=1}^{n} \pi_i(t)\pi_j(t)\sigma_{ij}(t)\right). \tag{A.9}$$

Then, for any initial value $Z_\pi(0) > 0$, (A.8) can be integrated directly to obtain

$$Z_\pi(t) = Z_\pi(0)\exp\left(\int_0^t \gamma_\pi(s)ds + \int_0^t \sum_{i=1}^{n} \pi_i(s)\sigma_i(t)dW_i(s)\right), \tag{A.10}$$

as a solution of (A.6), for $t \in [0, \infty)$.

Proof. To prove this proposition we will begin with Equation (A.10) and prove it, and therefore (A.8), are equivalent to Equation (A.6).

It follows from (A.10) that

$$d\log Z_\pi(t) = \gamma_\pi(t)dt + \sum_{i=1}^{n} \pi_i(t)\sigma_i(t)dW_i(t).$$

We apply Itô's Lemma to $Z_\pi = \exp(\log Z_\pi(t))$ to obtain

$$dZ_\pi(t) = Z_\pi(t)d\log Z_\pi(t) + \frac{1}{2}d\left\langle \log Z_\pi \right\rangle_t,$$

and therefore, by substituting in $d\log Z_\pi(t)$,

$$\frac{dZ_\pi(t)}{Z_\pi(t)} = \gamma_\pi(t)dt + + \sum_{i=1}^{n} \pi_i(t)\sigma_i(t)dW_i(t) + \frac{1}{2}Z_\pi(t)d\left\langle \log Z_\pi \right\rangle_t,$$

where $d\left\langle \log Z_\pi \right\rangle_t$ is the covariance process of $d\log Z_\pi$. Since $Z_\pi(t)$ is a portfolio of stocks and given the definition of the covariance process before, we have

$$d\left\langle \log Z_\pi \right\rangle_t = \sum_{i,j=1}^{n} \pi_i(t)\pi_j(t)d\left\langle \log X_i, \log X_j \right\rangle_t = \sum_{i,j=1}^{n} \pi_i(t)\pi_j(t)\sigma_{ij}(t)dt.$$

Now by definition we have

$$\gamma_\pi(t) = \sum_{i=1}^{n} \pi_i(t)\gamma_i(t) + \frac{1}{2}\left(\sum_{i=1}^{n} \pi_i(t)\sigma_{ii}(t) - \sum_{i,j=1}^{n} \pi_i(t)\pi_j(t)\sigma_{ij}(t) \right),$$

and therefore,

$$\frac{dZ_\pi(t)}{Z_\pi} = \sum_{i=1}^{n} \pi_i(t)\gamma_i(t)dt + \frac{1}{2}\sum_{i=1}^{n} \pi_i(t)\sigma_{ii}(t)dt + \sum_{i=1}^{n} \pi_i(t)\sigma_i(t)dW_i(t).$$

Furthermore, by Equation (A.2) we have,

$$dX_i(t) = \left(\gamma_i(t) + \frac{1}{2}\sigma_{ii}(t) \right) X_i(t)dt + X_i(t)\sigma_i(t)dW_i(t), \qquad \text{for t} \in [0, \infty).$$

This implies that,

$$\frac{dZ_\pi(t)}{Z_\pi(t)} = \sum_{i=1}^{n} \pi_i(t) \frac{dX_i(t)}{X_i(t)}.$$

□

A.2 Portfolio generating functions

In this section we now focus on portfolio generating functions, that is, functions that systematically generate portfolio weights. Furthermore, within the context of stochastic portfolio theory we are able to analyse the behaviour of these functions and the portfolio's they generate. In particular we will focus on the implications of portfolio generating functions which rely on some measure of stock market diversity, given the importance of diversity as presented in the previous sections.

We begin with a general definition of a portfolio generating function **S(x)** following along the lines of Fernholz (2002).

Definition A.2.1. Let **S** be a continuous function defined on Δ^n and let π be a portfolio. Then **S** generates π if there exists a measurable process of bounded variation Θ such that,

$$\log\left(Z_\pi/Z_\mu\right) = \log\mathbf{S}(\mu(t)) + \Theta(t), \qquad t \in [0,T]. \tag{A.11}$$

Where Δ^n is the set given by,

$$\left\{x \in \Re^n : x_1 + ... + x_n = 1, \quad 0 < x_i < 1, \quad i = 1, ..., n\right\}. \tag{A.12}$$

The function **S** in Definition A.11 is called the generating function of π, while Θ is called the drift process of **S**. We can also express Equation A.11 in differential form,

$$d\log\left(Z_\pi/Z_\mu\right) = d\log\mathbf{S}(\mu(t)) + d\Theta(t), \qquad t \in [0,T]. \tag{A.13}$$

It is important to notice from (A.13) that **S** operates in a relative space, in this case relative to the market portfolio μ. Furthermore, it is not unreasonable to assume that the portfolio generating function **S** is bounded on Δ^n. This implies that the long-term relative performance of the portfolio π generated by **S** is determined by the behaviour of $\Theta(t)$. In particular, if $\Theta(t)$ is increasing the portfolio π will outperform the market portfolio in the long-run.

Equation (A.13) excludes the impact of dividends. However, if we consider the definition of the portfolio process including a continuous dividend rate δ_π in Equation (**??**), (A.13) becomes

$$d\text{log}\left(\widehat{Z}_\pi/\widehat{Z}_\mu\right) = d\text{log}\mathbf{S}(\mu(t)) + \int_0^t (\delta_\pi(s) - \delta_\mu(s))\, ds + d\Theta(t), \qquad t \in [0,T]. \quad \text{(A.14)}$$

Therefore, including dividends, a functionally generated portfolio's performance relative to the market portfolio is dependent on the generating function, drift process and the difference in dividend rates.

Before continuing, we first consider the relative performance of an individual stock, X_i versus a portfolio Z_η. We define the relative return process of X_i versus Z_η as,

$$\text{log}(X_i(t)/Z_\eta(t)), \qquad t \in [0,\infty). \quad \text{(A.15)}$$

Then the cross-variation (covariance) process for the relative return process for stocks X_i and X_j is given by,

$$\begin{aligned} \langle \text{log}(X_i/Z_\eta), \text{log}(X_j/Z_\eta)\rangle_t &= \langle \text{log}(X_i), \text{log}(X_j)\rangle_t - \langle \text{log}(X_i), \text{log}(Z_\eta)\rangle_t \\ &\quad - \langle \text{log}(X_i), \text{log}(Z_\eta)\rangle_t + \langle \text{log} Z_\eta\rangle_t\,. \end{aligned} \quad \text{(A.16)}$$

We define $\tau_{ij}^\eta(t)$ as the relative covariance process for stocks X_i and X_j as in (A.16) with

$$\tau^\eta(t) = \left(\tau_{ij}^\eta(t)\right)_{1 \le i,j \ge n},$$

the relative covariance process $\tau^\eta(t)$ in matrix form.

Furthermore, we define the process $\sigma_{i\eta}(t)$ as,

$$\sigma_{i\eta}(t) = \sum_{j=1}^n \eta_j(t)\sigma_{ij}(t), \qquad t \in [0,\infty),$$

and we therefore have that,

$$d \langle log(X_i), log(Z_\eta) \rangle_t = \sigma_{i\eta}(t)dt.$$

Now we can write Equation (A.16) as,

$$\tau^{\eta}_{ij}(t) = \sigma_{ij}(t) - \sigma_{i\eta}(t) - \sigma_{j\eta}(t) + \sigma_{\eta\eta}(t), \qquad t \in [0, \infty), \tag{A.17}$$

for $i, j = 1, ..., n$ with $\sigma_{\eta\eta}(t)$ as the variance process of the portfolio η and therefore,

$$\sigma_{\eta\eta}(t) = \eta(t)\sigma(t)\eta^T(t), \qquad t \in [0, \infty). \tag{A.18}$$

Then for all i and j we have,

$$d \langle \log(X_i/Z_\eta), \log(X_j/Z_\eta) \rangle_t = \tau^{\eta}_{ij}(t)dt, \qquad t \in [0, \infty). \tag{A.19}$$

Equation (A.17) is the relative variance process of two stocks, X_i and X_j, both relative to the portfolio with weights $\eta(t)$.

We now extend this to consider the relative covariance process of a portfolio $\pi(t)$ versus a portfolio $\eta(t)$. Consider first that we have the variance process of a portfolio $\pi(t)$, which is given by,

$$\sigma_{\pi\pi}(t) = \pi(t)\sigma(t)\pi^T(t), \qquad t \in [0, \infty). \tag{A.20}$$

Therefore, to get the relative variance of portfolio $\pi(t)$ versus portfolio $\eta(t)$ we can substitute in the relative covariance of each stock i relative to the portfolio $\eta(t)$. That is instead of using $\sigma(t)$, we use $\tau^{\eta}(t)$ and we therefore have,

$$\tau^{\eta}_{\pi\pi}(t) = \pi(t)\tau^{\eta}(t)\pi^T(t). \tag{A.21}$$

Furthermore,

$$\pi(t)\tau^{\eta}(t)\pi^T(t) = (\pi(t) - \eta(t))\,\sigma(t)\,(\pi(t) - \eta(t))^T = \eta(t)\tau^{\pi}(t)\eta^T(t). \tag{A.22}$$

This implies that,

$$\tau^{\eta}_{\pi\pi} = \tau^{\pi}_{\eta\eta}.$$

Therefore if $\sigma(t)$ is singular then, the relative variance process of two portfolios is zero if and only if the two portfolios are equal.

Given the portfolio generating function **S**, we require firstly the resulting weights and secondly the drift process $\Theta(t)$ to determine the relative performance of the generated performance. The following theorem characterises these components and is reproduced from Fernholz (1999b) and Fernholz (2002).

Theorem A.2.1. *Let* **S** *be a positive* C^2 *function defined on a neighbourhood* U *of* Δ^n *such that for* $i = 1, ..., n$, $x_i D_i \log \mathbf{S}(x)$ *is bounded on* Δ^n. *Then* **S** *generates the portfolio* π *with weights*

$$\pi_i(t) = \left(D_i \log \mathbf{S}(\mu(t)) + 1 - \sum_{j=1}^{n} \mu_j(t) D_j \log \mathbf{S}(\mu(t)) \right) \mu_i(t), \quad t \in [0, \infty), \qquad \text{(A.23)}$$

for $i = 1, ..., n$, *and drift process*

$$d\Theta(t) = \frac{-1}{2\mathbf{S}(\mu(t))} \sum_{i,j=1}^{n} D_{ij}\mathbf{S}(\mu(t))\mu_i(t)\mu_j(t)\tau_{ij}(t)dt, \quad t \in [0, \infty). \qquad \text{(A.24)}$$

Proof. Recall that the weight process $\mu_i(t)$ is given by $X_i(t)/Z_\mu(t)$, and therefore we can represent the relative covariance process $\tau_{ij}(t)$ as given in Equation (A.19) as,

$$d \langle \log\mu_i, \log\mu_j \rangle_t = \tau_{ij}(t)dt, \quad t \in [0, \infty).$$

Now, applying Itô's Lemma to $\mu_i(t) = \exp(\log\mu_i(t))$ we have,

$$d\mu_i(t) = \mu_i(t) d\log\mu_i(t) + \frac{1}{2}\mu_i(t)\tau_{ii}(t)dt. \qquad \text{(A.25)}$$

and

$$d \langle \mu_i, \mu_j \rangle_t = \mu_i(t)\mu_j(t)\tau_{ij}dt, \quad t \in [0, \infty). \qquad \text{(A.26)}$$

Furthermore, if we apply Itô's Lemma to **S** and using (A.26), we get,

$$d\log\mathbf{S}(\mu(t)) = \sum_{i=1}^{n} D_i\log\mathbf{S}(\mu(t))d\mu_i(t) + \frac{1}{2}\sum_{i,j=1}^{n} D_{ij}\log\mathbf{S}(\mu(t))\mu_i(t)\mu_j(t)\tau_{ij}(t)dt.$$

Now,

$$D_{ij}\log\mathbf{S}(\mu(t)) = \frac{D_{ij}\mathbf{S}(\mu(t))}{\mathbf{S}(\mu(t))} - D_i\log\mathbf{S}(\mu(t))D_j\log\mathbf{S}(\mu(t)),$$

so,

$$\begin{aligned} d\log\mathbf{S}(\mu(t)) &= \sum_{i=1}^{n} D_i\log\mathbf{S}(\mu(t))d\mu_i(t) \\ &\quad + \frac{1}{2\mathbf{S}(\mu(t))}\sum_{i,j=1}^{n} D_{ij}\mathbf{S}(\mu(t))\mu_i(t)\mu_j(t)\tau_{ij}(t)dt \\ &\quad - \frac{1}{2}\sum_{i,j=1}^{n} D_i\mathbf{S}(\mu(t))D_j\mathbf{S}(\mu(t))\mu_i(t)\mu_j(t)\tau_{ij}(t)dt. \end{aligned} \quad \text{(A.27)}$$

For (A.13) to hold, the martingale components of $\log\mathbf{S}(\mu(t))$ and $\log(Z_\pi/Z_\mu)$ must be equal. Equation (3.7) implies that,

$$\begin{aligned} d\log(Z_\pi(t)/Z_\mu(t)) &= \sum_{i=1}^{n} \pi_i(t)d\log(X_i(t)/Z_\mu(t)) + \gamma_\pi^*(t)dt \\ &= \sum_{i=1}^{n} \pi_i(t)d\log\mu_i(t) + \gamma_\pi^*(t)dt \\ &= \sum_{i=1}^{n} \frac{\pi_i(t)}{\mu_i(t)}d\mu_i(t) - \frac{1}{2}\sum_{i,j=1}^{n} \pi_i(t)\pi_j(t)\tau_{ij}(t)dt, \end{aligned} \quad \text{(A.28)}$$

by Equation (3.6), Equation (A.17) and the fact that $\tau_{\pi\pi}^{\mu(t)} = \pi(t)\tau^{\mu}(t)\pi^T(t)$ as shown in Equation (A.21).

Now suppose that

$$\pi_i(t) = \left(D_i \log \mathbf{S}(\mu(t)) + \varphi(t)\right) \mu_i(t), \tag{A.29}$$

where $\varphi(t)$ is chosen such that $\sum_{i=1}^{n} \pi_i(t) = 1$. Then,

$$\begin{aligned} \sum_{i=1}^{n} \frac{\pi_i(t)}{\mu_i(t)} d\mu_i(t) &= \textstyle\sum_{i=1}^{n} D_i \log \mathbf{S}(\mu(t)) d\mu_i(t) + \varphi(t) \sum_{i=1}^{n} d\mu_i(t) \\ &= \textstyle\sum_{i=1}^{n} D_i \log \mathbf{S}(\mu(t)) d\mu_i(t), \end{aligned} \tag{A.30}$$

since $\sum_{i=1}^{n} d\mu_i(t) = 0$. Hence the martingale components $\log \mathbf{S}(\mu(t))$ and $\log\left(Z_\pi / Z_\mu\right)$, that is in Equations (A.2) and (A.28), respectively are equal.

Furthermore, if we choose $\varphi(t)$ such that,

$$\varphi(t) = 1 - \sum_{i=1}^{n} \mu_j(t) D_j \log \mathbf{S}(\mu(t)),$$

$\sum_{i=1}^{n} \pi_i(t) = 1$ is satisfied. Therefore (A.23) is proved and

$$\pi_i(t) = \left(D_i \log \mathbf{S}(\mu(t)) + 1 - \sum_{j=1}^{n} \mu_j(t) D_j \log \mathbf{S}(\mu(t)) \right) \mu_i(t), \quad t \in [0, \infty).$$

If $\pi_i(t)$ satisfies (A.29) then,

$$\begin{aligned} \sum_{i,j=1}^{n} \pi_i(t)\pi_j(t)\tau_{ij}(t) &= \sum_{i,j=1}^{n} D_i \log \mathbf{S}(\mu(t)) D_j \log \mathbf{S}(\mu(t)) \mu_i(t)\mu_j(t)\tau_{ij}(t) \\ &\quad + 2\varphi(t) \sum_{i,j=1}^{n} D_i \log \mathbf{S}(\mu(t)) \mu_i(t)\mu_j(t)\tau_{ij}(t) \\ &\quad + \varphi^2(t) \sum_{i,j=1}^{n} \mu_i(t)\mu_j(t)\tau_{ij}(t) \\ &= \sum_{i,j=1}^{n} D_i \log \mathbf{S}(\mu(t)) D_j \log \mathbf{S}(\mu(t)) \mu_i(t)\mu_j(t)\tau_{ij}(t), \end{aligned}$$

since $\mu(t)$ is in the null space of $\tau(t)$ by Equation (A.22). Hence,

$$d\log\left(Z_\pi / Z_\mu\right) = \sum_{i=1}^{n} D_i \log \mathbf{S}(\mu(t)) d\mu_i(t) - \frac{1}{2} \sum_{i,j=1}^{n} D_i \log \mathbf{S}(\mu(t)) D_j \log \mathbf{S}(\mu(t)) \mu_i(t)\mu_j(t)\tau_{ij}(t) dt.$$

This equation and (A.2) imply that,

$$d\log\left(Z_\pi/Z_\mu\right) = d\log\mathbf{S}(\mu(t)) - \frac{1}{2\mathbf{S}(\mu(t))}\sum_{i,j=1}^{n} D_{ij}\mathbf{S}(\mu(t))\mu_i(t)\mu_j(t)\tau_{ij}(t).$$

Comparing this to Equation (A.13) we see that,

$$d\Theta(t) = \frac{-1}{2\mathbf{S}(\mu(t))}\sum_{i,j=1}^{n} D_{ij}\mathbf{S}(\mu(t))\mu_i(t)\mu_j(t)\tau_{ij}(t), \quad t \in [0,\infty),$$

and therefore, (A.24) is proved. □

Theorem A.2.1 allows us to characterise the drift process which, for a given generating function, allows us to determine long-term behaviour for a corresponding portfolio relative to the market portfolio.

A.3 Portfolio generating function for the equal weighted portfolio

Section 3.4 provided an intuitive explanation of the portfolio generating function and drift process for the equal weighted portfolio. Here, however, we use the expressions in Theorem A.2.1 to confirm that the geometric mean of cap weights generates the equal weighted portfolio and its drift process is the excess growth rate.

Consider first the expression $D_i\log\mathbf{S}(\mu(t))$ which appears in Equation (A.23).

$$\begin{aligned} D_i\log\mathbf{S}(\mu(t)) &= D_i\left(\frac{1}{n}\log\prod_{k=1}^{n}\mu_k(t)\right) \\ &= \frac{1}{n}D_i\left(\sum_{k=1}^{n}\log\mu_k(t)\right) \\ &= \frac{1}{n\mu_i(t)}. \end{aligned} \quad \text{(A.31)}$$

If we substitute the expression in (A.31) into Equation (A.23) then,

$$\begin{aligned}
\pi_i(t) &= \left(\frac{1}{n\mu_i(t)} + 1 - \sum_{j=1}^{n} \mu_j(t) \frac{1}{n\mu_j(t)}\right) \mu_i(t) \\
&= \left(\frac{1}{n\mu_i(t)} + 1 - \frac{n}{n}\right) \mu_i(t) \\
&= \left(\frac{1}{n\mu_i(t)}\right) \mu_i(t) \\
&= \frac{1}{n}.
\end{aligned} \tag{A.32}$$

A.4 Drift process for the equal weighted portfolio

In a similar manner, the drift process for the equal weighted portfolio can be derived using Equation (A.24) in Theorem A.2.1. Firstly $D_{ij}\mathbf{S}(\mu(t))$ is evaluated before completing the expression in Equation (A.24). In the derivations below, the time subscript t is dropped for simplicity. Therefore, μ_i implies the expression $\mu_i(t)$ for example.

$$\begin{aligned}
D_i\mathbf{S}(\mu) &= D_i \left(\mu_1^{1/n} \cdots \mu_n^{1/n}\right) \\
&= \frac{1}{n}\mu_1^{1/n} \cdots \mu_{i-1}^{1/n} \cdot \mu_i^{1/n} \cdot \mu_i^{-1} \cdot \mu_{i+1}^{1/n} \cdots \mu_n^{1/n} \\
&= \frac{1}{n}\frac{1}{\mu_i}\mathbf{S}(\mu).
\end{aligned}$$

Similarly then,

$$D_{ij}\mathbf{S}(\mu) = \frac{1}{n}\frac{1}{n}\frac{1}{\mu_i}\frac{1}{\mu_j}\mathbf{S}(\mu). \tag{A.33}$$

Substituting Equation (A.33) into Equation (A.24),

$$\begin{aligned}
d\Theta(t) &= \frac{-1}{2\mathbf{S}(\mu)} \sum_{i,j=1}^{n} \frac{1}{n}\frac{1}{n}\frac{1}{\mu_i}\frac{1}{\mu_j}\mathbf{S}(\mu)\mu_i\mu_j\tau_{ij}dt \\
&= \frac{-1}{2} \sum_{i,j=1}^{n} \frac{1}{n}\frac{1}{n}\tau_{ij}dt.
\end{aligned}$$

The expression $\frac{1}{n}$ is an equal weight which is represented here as π_i and this can, therefore be rewritten as,

$$\begin{aligned} d\Theta(t) &= \frac{-1}{2}\sum_{i,j=1}^{n} \pi_i \pi_j \tau_{ij} dt \\ &= \gamma^*_{\pi}(t) dt \end{aligned}$$

as a result of Equation (3.6), Equation (A.17) and the fact that $\tau^{\mu(t)}_{\pi\pi} = \pi(t)\tau^{\mu}(t)\pi^T(t)$ as shown in Equation (A.21).

This page has been intentionally left blank

B Code extracts

B.1 Setting up the main portfolio class

At a high level each portfolio, whether cap weighted, equal weighted or optimised, can be generated using the parameter p from Equation (5.4). Other portfolios can also be generated easily in the code by modifying the trade object (discussed later). As such, many of the portfolios share the same need to calculate metrics such as the covariance matrix or the portfolio generating function. The main starting point, therefore, is a class called *portfolio*. This class defines a portfolio, the portfolio type, holds all the necessary data and the output generated.

Listing B.1 shows the initialisation of this class which involves setting up the parameters and building empty *Pandas* dataframes to hold the output as the simulation is performed.

```python
import pandas as pd
import numpy as np
from dateutil.relativedelta import relativedelta

from trade import tradeObject
from return_metrics import returns
import metric_functions as mfunc
from generating_functions import genFunc



class portfolio:
    """
    Provides interface to simulate and track portfolio performance
    """

    def __init__(self, stk_data, members, rfr, start_cash=10e6, wgt_method='equal',
                 rebal_period='monthly', trade_cost=0, calc_metrics=False, dwp_p=0.01, cov_lag=60,
                 verbose=False):
        """
        @Input:
```

```
            stk_data: dict with at least PX_LAST, MARKET_CAP and ADJ_PX dataframes. Each dataframe has dates on rows and stocks on columns
            members: dataframe with members of index on each date
            rfr: one column dataframe with risk free rate
            start_cash: starting cash value of portfolio
            wgt_method: weighting method- 'equal', 'mcap', or 'dwp' (dwp allows for a blend of cap and equal)
            rebal_period: frequency of rebalance - 'weekly', 'monthly', 'quarterly', 'annually'
            calc_metrics: should we calculate things like excess growth rate?
            dwp_p: the p paramter to use which allows for blended portfolio
            cov_lag: lag of covariance matrix calculation
            verbose: print statements while processing?
        """

        self.stk_data = stk_data
        self.rebal_period = rebal_period
        self.wgt_method = wgt_method
        self.calc_metrics = calc_metrics
        self.members = members
        self.trade_cost = trade_cost
        self.verbose = verbose

        # initialize some dfs
        self.__initialize(start_cash)

        # initialize trade object
        self.tobj = tradeObject()

        # connect a returns object to this portfolio object
        self.returns = returns(self, rfr)

        # connect a generating function class
        self.dwp_p = dwp_p
        self.genfunc = genFunc(wgt_method, port_obj=self)

        self.stk_cov = None
        self.cov_lag = cov_lag


    def __initialize(self, starting_cash):
        """
        Initialize some dfs
        """

        self.portfolio = pd.DataFrame(0, index = self.stk_data['PX_LAST'].index, columns=['Stocks', 'Cash', 'Total'])

```

```
        # initialize portfolio cash balance
        self.portfolio.loc[self.portfolio.index[0], 'Cash'] = starting_cash
        self.portfolio.loc[self.portfolio.index[0], 'Total'] = starting_cash

        # df for dividends
        self.divs = pd.DataFrame(0, index=self.stk_data['PX_LAST'].index, columns=['DIV_CASH'])

        # df for transaction costs
        self.trans_cost = pd.DataFrame(0, index=self.stk_data['PX_LAST'].index, columns=['TRANS_COST'])

        # dfs for positions and to hold any required trades
        self.positions = pd.DataFrame(0, index=self.stk_data['PX_LAST'].index, columns=self.stk_data['PX_LAST'].columns)
        self.req_trades = pd.DataFrame()

        # save target weights
        self.target_weights = pd.DataFrame(0, index=self.stk_data['PX_LAST'].index, columns=self.stk_data['PX_LAST'].columns)

        # save excess growth rate and portfolio generating function
        self.eg = pd.DataFrame(columns=['PORT-EG', 'EW-EG', 'AVG-VOL', 'AVG-CORR'])
        self.gen_func = pd.DataFrame(columns=['GEN-FUNC'])
        self.optimal_p = pd.DataFrame(columns=['OPTIMAL-P'])

        # for ranked weights
        self.mc_weights = pd.DataFrame(np.nan, index=self.stk_data['PX_LAST'].index, columns=self.stk_data['PX_LAST'].columns)
        self.mc_ranks = pd.DataFrame(np.nan, index=self.stk_data['PX_LAST'].index, columns=self.stk_data['PX_LAST'].columns)

        # storing mc weights by rank
        self.rank_mcw = pd.DataFrame(np.nan, index=self.stk_data['PX_LAST'].index, columns=range(len(self.stk_data['PX_LAST'].columns)))

        # storing leakage
        self.leak = pd.DataFrame(columns=['LEAK'])
        self.idx_chg = pd.DataFrame(columns=['CHANGE'])
```

Listing B.1: Initialisation of portfolio simluation

B.2 Performing the simulation

Once everything is set up we can start the simulation. The simulation is performed by stepping one day at a time through the raw stock data (see Listing B.2).

```python

def simulateAll(self):
    """
    Perform entire simulation
    """
    if self.verbose:
        print('Starting simulation...')
        print('-------------------------')


    N = len(self.stk_data['PX_LAST'])
    for row in range(1, N): # simulation starts at 1 not 0
        self.performSimulationStep(row)

        if self.verbose:
            print(self.stk_data['PX_LAST'].index[row])

    self.returns.updateReturns()
```

Listing B.2: Portfolio simulation loop

The following list shows the necessary steps for each day. Each of these steps is, at a high level at least, contained in Listing B.3.

1. **Execute any trades from the previous day** We assume the necessary trades are identified at the close of each day and the trades themselves are executed at the close of the following day.
2. **Value the portfolio** Valuation of the portfolio is done at the close of each day (after trades for the day are executed).
3. **Calculate all the metrics we need** This will be metrics such as the excess growth rate, the portfolio generating function, leakage estimation, and the ranked weights (weight for each ranked position).
4. **Determine if we need to rebalance the portfolio** If we do need to rebalance the portfolio, then we need to work out trades that need to be done the next day.

Once the full simulation is complete, we calculate the necessary return and risk-adjusted return metrics.

```python
def performSimulationStep(self, row):
    """
    Performs simulation for one time period
    """

    #--------------------------------------------------------
    # carry over positions and portfolio value
    prev_pos = self.positions.iloc[(row-1)]
    self.positions.iloc[row] = prev_pos

    prev_port = self.portfolio.iloc[row-1]
    self.portfolio.iloc[row] = prev_port

    #--------------------------------------------------------
    # execute any trades, essentially done at previous day's close
    if len(self.req_trades) > 0:

        if self.verbose:
            print("Executing trades")

        c_pos = self.positions.iloc[row]
        hist_px = self.stk_data['PX_LAST'].iloc[:row+1]


        port_cash, pos, self.req_trades, trans_cost = self.tobj.executeTrades(c_pos, self.req_trades,
                                                         self.portfolio['Cash'].iloc[row], hist_px,
                                                         self.trade_cost)
        if np.isnan(port_cash):
            raise ValueError("Cash is NaN after trade execution")

        # adjust portfolio
        self.positions.iloc[row] = 0
        self.positions.update(pd.DataFrame(pos).T)
        self.portfolio['Cash'].iloc[row] = port_cash

        # save transaction costs
        self.trans_cost.iloc[row]['TRANS_COST'] = trans_cost


    #--------------------------------------------------------
    # value the portfolio with new holdings
    port_cash, port_divs, port_stk, del_pos = self.__valuePortfolio(self.positions.iloc[row], self.stk_data['PX_LAST'].iloc[:(row+1)],
                                                     self.stk_data['ADJ_PX'].iloc[:(row+1)])
    if np.isnan(port_cash):
```

```
                raise ValueError("Cash is NaN after valuation")

        # remove any delistings from positiosn -- would have been added to cash in valuePortfolio method
        for d in del_pos:
            self.positions[d].iloc[row] = 0

        # adjust cash and stock value
        self.portfolio['Cash'].iloc[row] += port_cash
        self.portfolio['Stocks'].iloc[row] = port_stk
        self.portfolio['Total'].iloc[row] = self.portfolio['Cash'].iloc[row] + self.portfolio['Stocks'].iloc[row]

        # add dividends
        self.divs['DIV_CASH'].iloc[row] = port_divs

        # -----------------------------------------------------
        # work out excess growth rate
        current_date = self.stk_data['PX_LAST'].index[row]
        prev_date = self.stk_data['PX_LAST'].index[row-1]
        current_members = self.__getMembers(current_date)

        # -------------------------------------------
        # Calculate ranked weights
        hist_mkt_cap = self.stk_data['MARKET_CAP'].iloc[:row]
        self.__calcRankedWeights(hist_mkt_cap, current_members, current_date)

        if self.calc_metrics and (current_date.month != prev_date.month):


            c_pos = self.positions.iloc[row]
            hist_data = self.stk_data['ADJ_PX'].iloc[:row]
            hist_px = self.stk_data['PX_LAST'].iloc[:row]

            # calculate covariance matrix
            self.__calcCovar(row, c_pos, hist_data, hist_px, lag=self.cov_lag)

            # excess growth rate
            eg, eg_ew, avg_vol, avg_corr = self.__calcMetrics(row, c_pos, hist_data, hist_px, lag=self.cov_lag)
            self.eg = self.eg.append(pd.DataFrame(np.asarray([eg, eg_ew, avg_vol, avg_corr]).reshape(1,4),
                                                  index=[self.stk_data['PX_LAST'].index[row]],
                                                  columns=['PORT-EG', 'EW-EG', 'AVG-VOL', 'AVG-CORR']))

            # -------------------------------------------
```

```
            # Calculate portfolio generating function
            port_gen_func = self.__calcPortGenFunc(hist_mkt_cap, current_members, current_date)
            self.gen_func = self.gen_func.append(pd.DataFrame(np.asarray([port_gen_func]),
                                          index=[self.stk_data['PX_LAST'].index[row]],
                                          columns=['GEN-FUNC']))


            # get leakage
            if len(self.gen_func.index) >= 2:
                prev_date = self.gen_func.index[-2]
                prev_memb = self.__getMembers(prev_date)
                old_port_gen_t = self.__calcPortGenFunc(hist_mkt_cap, prev_memb, prev_date, save=False)
                old_port_gen_0 = self.__calcPortGenFunc(hist_mkt_cap.loc[hist_mkt_cap.index<=prev_date], prev_memb, prev_date, save=False)

                new_port_gen_0 = self.__calcPortGenFunc(hist_mkt_cap.loc[hist_mkt_cap.index<=prev_date], current_members, current_date, save=False)


                leak = (old_port_gen_0-old_port_gen_t) - (new_port_gen_0-port_gen_func)
                self.leak = self.leak.append(pd.DataFrame(np.asarray([leak]),
                              index=[self.stk_data['PX_LAST'].index[row]],
                              columns=['LEAK']))

                # save changes in number of members
                chg = len(set(prev_memb).symmetric_difference(set(current_members)))/2
                self.idx_chg = self.idx_chg.append(pd.DataFrame(np.asarray([chg]),
                              index=[self.stk_data['PX_LAST'].index[row]],
                              columns=['CHANGE']))


        #--------------------------------------------------------
        # check if we need to rebalance
        rebal = self.__needRebalance(row)

        #--------------------------------------------------------
        # work out trades if we need to rebalance

        if rebal:


            if self.verbose:
```

```
            print("Rebalancing")

        # need to rebalance so get members
        active_date = self.stk_data['PX_LAST'].index[row]
        index_memb = self.__getMembers(active_date)

        weight_metric = self.stk_data['MARKET_CAP'].iloc[row]

        # determine required trades
        c_pos = self.positions.iloc[row]
        self.req_trades, t_wgt = self.tobj.detPortfolioTrades(index_memb, c_pos,
                                                              self.portfolio['Total'].iloc[row],
                                                              self.stk_data['PX_LAST'].iloc[row],
                                                              method=self.wgt_method,
                                                              weight_metric=weight_metric,
                                                              dwp_p=self.dwp_p,
                                                              reduce_leakage_stocks=self.reduce_leakage_stocks,
                                                              cov_mat = self.stk_cov,
                                                              prev_w = self.prev_w)

        # save target weights
        t_wgt = pd.DataFrame(t_wgt).T
        t_wgt.index = [c_pos.name]
        self.target_weights.update(t_wgt)
        self.prev_w = t_wgt
```

Listing B.3: Portfolio simulation

Listing B.4 shows the functions that calculate the excess growth rate and stores the ranked weights for each position in the index. The covariance matrix is simply the EWMA covariance calculated using the Pandas EWM function (not shown in the listing).

```


def __calcMetrics(self, row, current_positions, hist_data, hist_px, lag=3*260):
    """
    Function to calculate various metrics such as excess growth rate, etc
    """

    if row <= lag+1:
        return np.nan, np.nan, np.nan, np.nan


    port_members = current_positions[current_positions>0].index
```

```

    # -----------------------------------------------------------
    # get returns -- select lag+1 since first pct_change will be NA and be dropped
    stk_rets = hist_data.iloc[(-lag-2):][port_members].pct_change()

    # drop any na columns
    stk_rets.dropna(axis=0, how='all', inplace=True)
    stk_rets.dropna(axis=1, how='any', inplace=True)

    # anything over +-50% in one day becomes zero
    stk_rets[abs(stk_rets) >= 0.5] = 0

    # drop columns where more than half of returns are zero
    prop_zero = stk_rets.apply(lambda x: len(x[x==0])/len(x), axis=0)
    prop_zero = prop_zero[prop_zero <= 0.5]
    stk_rets = stk_rets[prop_zero.index]

    N = len(stk_rets.columns)

    # -----------------------------------------------------------
    # work out current weights
    pos_val = current_positions[port_members]*hist_data.iloc[-1][port_members]/100
    port_wgts = pos_val/pos_val.sum()


    # work out equal weight excess growth rate
    wgts = np.ones((1,N))/N
    X1 = np.dot(self.vol.T,wgts.T)
    X2 = np.dot(np.dot(wgts,self.stk_cov),wgts.T)
    ew_eg = 0.5*(X1-X2)


    # convert to np array
    port_wgts = port_wgts[stk_rets.columns] # some columns may have been dropped
    port_wgts = port_wgts.values.reshape(1,-1)

    # some portfolio weights may be NaN if delisted for example
    port_wgts[np.isnan(port_wgts)] = 0
    port_wgts = port_wgts/port_wgts.sum()

    X1 = np.dot(self.vol.T,port_wgts.T)
    X2 = np.dot(np.dot(port_wgts,self.stk_cov),port_wgts.T)
    port_eg = 0.5*(X1-X2)

    # work out average volatility and correlations
    avg_vol = self.vol.mean()
```

```

    stk_corr = mfunc.calcCorr(stk_rets, lag=lag)
    triu_corr = np.triu(stk_corr,1) # take only one diagonal
    triu_corr[triu_corr==0] = np.nan # set zeros to nan
    avg_corr = np.nanmean(triu_corr) # ave ignoring nan

    return port_eg[0][0], ew_eg[0][0], avg_vol, avg_corr


def __calcRankedWeights(self, mkt_data, current_members, current_date):


    # work out market cap weights
    mkt_cap = mkt_data.iloc[-1]
    mkt_cap = mkt_cap[current_members] # only for those in index
    mkt_cap.dropna(inplace=True)
    mc_wgts = mkt_cap/mkt_cap.sum()

    # save mkt cap weights
    self.mc_weights.loc[current_date, mc_wgts.index] = mc_wgts

    # save ranks
    mc_wgts_rank = mc_wgts.rank(ascending=False, method='first')
    self.mc_ranks.loc[current_date, mc_wgts_rank.index] = mc_wgts_rank

    # ranked market cap weights
    temp = mc_wgts_rank.dropna()
    temp = temp.apply(lambda x: int(x-1)).to_dict() # change to int and rebase to start at 0
    rank_wgts = mc_wgts.rename(index=temp).dropna()

    # some stocks may have come in and don't have ranking from prior day
    rank_wgts = rank_wgts.loc[set(rank_wgts.index) & set(self.rank_mcw.columns)]

    self.rank_mcw.loc[current_date, rank_wgts.index] = rank_wgts




def __calcCovar(self, row, current_positions, hist_data, hist_px, lag=3*260):
    """
    Calculate covariance matrix
    """

    if row <= lag+1:
        return

    port_members = current_positions[current_positions>0].index
```

```python

    # ---------------------------------------------------------
    # get returns -- select lag+1 since first pct_change will be NA and be dropped
    stk_rets = hist_data.iloc[(-lag-2):][port_members].pct_change()

    # drop any na columns
    stk_rets.dropna(axis=0, how='all', inplace=True)
    stk_rets.dropna(axis=1, how='any', inplace=True)

    # anything over +-50% in one day becomes zero
    stk_rets[abs(stk_rets) >= 0.5] = 0

    # drop columns where more than half of returns are zero
    prop_zero = stk_rets.apply(lambda x: len(x[x==0])/len(x), axis=0)
    prop_zero = prop_zero[prop_zero <= 0.5]
    stk_rets = stk_rets[prop_zero.index]


    # ---------------------------------------------------------
    # work out annualized stk vol and covar
    self.stk_cov = mfunc.calcCovar(stk_rets, lag=lag)

    self.vol = np.diag(self.stk_cov)
```

Listing B.4: Function that calculates excess growth rate and ranked cap weights

The portfolio valuation is performed in a separate function, shown in Listing B.5.

```python
def __valuePortfolio(self, positions, prices, adj_prices):

    # work out impact of any corporate actions

    # delistings
    latest_px = prices.iloc[-1]
    na_px = latest_px[latest_px.isnull()].index
    na_pos = set(positions[positions>0].index) & set(na_px)

    del_cash = 0
    del_pos = []
    for p in list(na_pos):
        del_cash += prices[p].iloc[-2]*positions[p]/100
        del_pos.append(p)
        positions[p] = 0

    # general
    price_chg = prices.iloc[-1]/prices.iloc[-2]
    adjprice_chg = adj_prices.iloc[-1]/adj_prices.iloc[-2]

```

```
    # work out residual and assign as dividend cash
    resid = round(adjprice_chg/price_chg - 1,4)
    p = positions[positions>0].index
    div_cash = prices[p].iloc[-2]*positions[p]*resid[p]/100
    div_cash = div_cash.sum()


    # work out stock value
    stk_val = positions.multiply(prices[positions.index].iloc[-1]/100).sum()

    ca_cash = div_cash + del_cash

    return ca_cash, div_cash, stk_val, del_pos
```

Listing B.5: Portfolio valuation function

The function in Listing B.6 takes in a date and returns the list of stocks in the index on the the latest date prior to the given date.

```
def __getMembers(self, active_date):

    """
    Gets the spy members for active month
    """

    # need previous months month and year
    mem_date = active_date.replace(day=1)+relativedelta(days=-1)
    m = mem_date.month
    y = mem_date.year

    cols = pd.DataFrame(self.members.columns)
    cols['month'] = cols[0].apply(lambda x: x.month)
    cols['year'] = cols[0].apply(lambda x: x.year)

    spx_col = cols[(cols['month']==m) & (cols['year']==y)][0].iloc[0]
    mems = self.members[spx_col].dropna().tolist()

    return mems
```

Listing B.6: Function that returns a list of index constituents on a specific date

The function in Listing B.7 calculates the portfolio generating function.

```

def __calcEWPgenFunc(self, mkt_data, current_members):
    """
    Calculate the portfolio log generating function for equal weight portfolio
    """
    # get mkt cap of index members
```

```python
    mkt_cap = mkt_data.iloc[-1]
    mkt_cap = mkt_cap[current_members] # only for those in index
    mc_wgts = mkt_cap/mkt_cap.sum()

    # calculate the portfolio generating function for ew portfolio
    port_gen_func = sum(np.log(mc_wgts).dropna())*(1/len(mc_wgts.dropna()))

    return port_gen_func
```

Listing B.7: Portfolio generating function calculation

B.3 Determining trades and executing

Listing B.8 shows the trade class. Each portfolio has a trade object that: (1) determines what trades are necessary, and (2) executes a given set of trades. When determining trades, the function is given the current nominal positions for all the stocks and the weight methodology (cap weight, equal weight, etc.). The function then works out the target weights based on the methodology (here is where any other methodology can be included), and then determines the nominal positions resulting from those weights and the trades necessary to transform the portfolio.

In the main portfolio object these trades can be checked and on the following day the execute trades function can be called and the trade object returns the new positions and the net cash position after execution. Including bid/offer prices can be done at this step.

```python
import pandas as pd
import numpy as np
from scipy.optimize import minimize, LinearConstraint

class tradeObject():
    """
    Handles execution of trades
    Determines if a rebalance required
    """

    def __init__(self):
        pass

    def detPortfolioTrades(self, ticker_list, current_positions, portfolio_value,
     current_px, method='equal',
                           weight_metric='None', dwp_p=0.01,
    reduce_leakage_stocks=10, cov_mat=None,
                           prev_w=None):
```

```
    """
    Determines trades required
    @Input:
        ticker_list: list of tickers
        current_positions: pandas Series objects with positions (nominal)
        current_px: pandas Series object with current prices -- assumes in
cents per share
        portfolio_value: value of portfolio in ZAR
        method: method to weight tickers by (equal, mc, dwp)
        weight_metric: must be provided if method not equal and must be
pandas Series object with weights
        apply_active: randomly apply active weights
        active_max: maximum multiple to apply to for active weight
        active_min: minimum multiple to apply to for active weight
        is_never: only allocate for new members leave existing members as is
    @Output:
        trades: list of required trades
        wgt: target weights for tracking
    """


    ticker_list = list(set(ticker_list))

    # mc wgts is default
    # drop nans in mc wgts
    weight_metric = weight_metric[ticker_list]
    wgt = pd.Series(weight_metric/weight_metric.sum())
    wgt = wgt.loc[~np.isnan(wgt)]

    # refresh ticker list in case any changes
    ticker_list = wgt.index

    # work out desired weights
    if method == 'equal':
        wgt = pd.Series(1/len(ticker_list), name='Weights', index=ticker_list
)
    elif method == 'dwp':
        wgt = wgt**dwp_p
        wgt = wgt/wgt.sum()
    elif method == 'mcap':
        pass
    else:
        raise ValueError("Weighting method {0} not recognised".format(method)
)


    if len(wgt[wgt<0]) > 0:
        raise ValueError("Some weights are negative!")
```

```python

        # some checks
        if wgt.sum() < 0.98 or wgt.sum() > 1.01:
            raise ValueError("Weights deviate from 1 significantly",wgt.sum())

        # workout desired nominal
        target_nom = pd.Series(0, index=current_positions.index, name=current_positions.name)
        target_nom[ticker_list] = 100*(wgt*portfolio_value).divide(current_px[ticker_list])
        target_nom = np.floor(target_nom)

        # determine trades required

        trades = pd.DataFrame(columns=['ticker', 'side', 'size'])

        # get difference in nominals
        nom_diff = target_nom[current_positions.index] - current_positions
        nom_diff = nom_diff[nom_diff!=0]

        # add to trades df
        trades['ticker'] = nom_diff.index
        trades['size'] = nom_diff[trades['ticker']].values
        trades.loc[trades['size']>0, 'side']= 'L'
        trades.loc[trades['size']<0, 'side']= 'S'
        trades['size'] = abs(trades['size'])

        return trades, wgt



    def executeTrades(self, current_positions, trades, port_cash, prices, costs=0):
        """
        Executes a list of trades
        @Input:
            current_positions: pd Series of current positions in portfolio
            trades: pd dataframe of trades, columns must include ticker, side and size
            port_cash: current portfolio cash position in ZAR
            prices: pd dataframe of all price history up until this point in cents per share
        @Output:
            net_cash: net cash position after trades
            new_positions: positions after all trades executed
            trades: pd dataframe of any leftover trades -- for now assume none leftover
        """
```

```
        new_positions = current_positions.copy()

        trade_prices = prices[trades['ticker']].iloc[-1]

        # any trades with nan prices gets removed
        isnan = trade_prices.loc[trade_prices.isna()]
        if len(isnan) > 0:
            trades = trades[~trades['ticker'].isin(isnan.index)]

        # sales
        sales = trades.loc[trades['side']=='S']
        if len(sales) > 0:
            new_positions[sales['ticker']] -= sales['size'].values
            port_cash += (1-costs/100)*sum(trade_prices[sales['ticker']]*sales['size'].values/100)
            trans_cost = (costs/100)*sum(trade_prices[sales['ticker']]*sales['size'].values/100)

        # buys
        buys = trades.loc[trades['side']=='L']
        if len(buys) > 0:
            new_positions[buys['ticker']] += buys['size'].values
            port_cash -= (1+costs/100)*sum(trade_prices[buys['ticker']]*buys['size'].values/100)
            trans_cost = (costs/100)*sum(trade_prices[sales['ticker']]*sales['size'].values/100)


        return port_cash, new_positions, pd.DataFrame(), trans_cost
```

Listing B.8: The trade class

B.4 Optimising the portfolio

To optimise the portfolio, no other modifications need to be made to any of the objects or functions shown previously (the portfolio class in particular). Instead we need only change the parameter *dwp_p* shown in Listing B.3 just before any rebalance. The trade object will then receive this new parameter and work out the necessary trades.

Therefore, instead of running the full simulation (Listing B.2), we call each day's simulation step individually (Listing B.3) and determine the parameter *dwp_p* (and passing it) just before calling the simulation step.

This can obviously be done in any number of ways and for any number of portfolios with some small modifications. That is, the structure presented here is not limited to the analysis we did in this thesis, one could modify the trade object to achieve any kind of portfolio. Listing B.9 shows our implementation. There are three portfolio objects (implementation of the portfolio class): cap weight, equal weight, and optimal. We step through each day and for the equal and cap weighted portfolios we can just perform the simulation step (lines 19 and 20).

For the optimised portfolio, however, we need to first work out our prediction, determine the optimal *dwp_p* parameter, and pass this to the portfolio object before performing the simulation step for the optimised portfolio. The prediction itself can be a separate function or class and any data necessary can just be passed from the portfolio objects to this prediction object or function.

```
1 def simulate(self, opt_p_method, verbose=True):
2
3     """
4     opt_p_method = '0-1', 'std_dev', 'none'
5     """
6
7     start_time = time.time()
8     N = len(self.ew_port.stk_data['PX_LAST'])
9     prev_date = None
10    self.p_df = pd.DataFrame(columns=['p'])
11    for row in range(1, N): # simulation starts at 1 not 0
12
13        active_date = self.ew_port.stk_data['PX_LAST'].index[row]
14        if verbose:
15            print(active_date)
16
17        # ----------------------------------------------------------------------
18        # perform simulation steps
19        self.ew_port.performSimulationStep(row)
20        self.mc_port.performSimulationStep(row)
21
22        # ----------------------------------------------------------------------
23        # if new month then work out new prediction
24        if prev_date is None or prev_date.month != active_date.month:
25
26            # make predicton
27            pred_rel, pred_err = self.pred_obj.predict(row)
28
29            if pred_rel is not None:
30
31                optimal_p = self._get_opt_p(opt_p_method, pred_rel, pred_err, row
    , period_actual=3,
```

```
                                              current_p=self.optimal_port.dwp_p)


                # only switch if difference more than trade cost
                if opt_p_method == '0-1':
                    if optimal_p != self.optimal_port.dwp_p and abs(pred_rel) >=
    (self.threshold*self.trade_cost+0)/100:
                        self.optimal_port.dwp_p = optimal_p
                else:
                    self.optimal_port.dwp_p = optimal_p

                print(f"Optimal p: {self.optimal_port.dwp_p}")

        self.optimal_port.performSimulationStep(row)

        # save optimal p
        temp_df = pd.DataFrame([self.optimal_port.dwp_p], index=['p'],
                               columns=[active_date]).T
        self.p_df = self.p_df.append(temp_df)

        prev_date = active_date

    # ----------------------------------------------------------------------
    self.ew_port.returns.updateReturns()
    self.mc_port.returns.updateReturns()
    self.optimal_port.returns.updateReturns()

    self._get_return_df()

    run_time = time.time()
    sim_time = run_time-start_time
    print(f"Simulation complete: {round(sim_time/60,2)} minutes")
```
Listing B.9: Optimised portfolios

List of *tsfresh* features

C

The list of features extracted using *tsfresh* are provided in Tables C.1 to C.4 and can also be found at `https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html`.

Table C.1.: List of features extracted using *tsfresh*[a].

Short Code	Description
abs_energy(x)	Returns the absolute energy of the time series which is the sum over the squared values.
absolute_maximum(x)	Calculates the highest absolute value of the time series x.
absolute_sum_of_changes(x)	Returns the sum over the absolute value of consecutive changes in the series x.
agg_autocorrelation(x, param)	Descriptive statistics on the autocorrelation of the time series.
agg_linear_trend(x, param)	Calculates a linear least-squares regression for values of the time series that were aggregated over chunks versus the sequence from 0 up to the number of chunks minus one.
approximate_entropy(x, m, r)	Implements a vectorized Approximate entropy algorithm.
ar_coefficient(x, param)	This feature calculator fits the unconditional maximum likelihood of an autoregressive AR(k) process.
augmented_dickey_fuller(x, param)	Does the time series have a unit root?
autocorrelation(x, lag)	Calculates the autocorrelation of the specified lag.
benford_correlation(x)	Useful for anomaly detection applications.
binned_entropy(x, max_bins)	First bins the values of x into max_bins equidistant bins.
c3(x, lag)	Uses c3 statistics to measure non linearity in the time series.
change_quantiles(x, ql, qh, isabs, f_agg)	First fixes a corridor given by the quantiles ql and qh of the distribution of x.
cid_ce(x, normalize)	This function calculator is an estimate for a time series complexity. (A more complex time series has more peaks, valleys etc.).
count_above(x, t)	Returns the percentage of values in x that are higher than t.

[a]Source: `https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html`

Table C.2.: List of features extracted using *tsfresh*[a] (Continued).

Short Code	Description
count_above_mean(x)	Returns the number of values in x that are higher than the mean of x
count_below(x, t)	Returns the percentage of values in x that are lower than t.
count_below_mean(x)	Returns the number of values in x that are lower than the mean of x.
cwt_coefficients(x, param)	Calculates a Continuous wavelet transform for the Ricker wavelet, also known as the "Mexican hat wavelet".
energy_ratio_by_chunks(x, param)	Calculates the sum of squares of chunk i out of N chunks expressed as a ratio with the sum of squares over the whole series.
fft_aggregated(x, param)	Returns the spectral centroid (mean), variance, skew, and kurtosis of the absolute fourier transform spectrum.
fft_coefficient(x, param)	Calculates the fourier coefficients of the one-dimensional discrete Fourier Transform for real input by fast fourier transform.
first_location_of_maximum(x)	Returns the first location of the maximum value of x.
first_location_of_minimum(x)	Returns the first location of the minimal value of x.
fourier_entropy(x, bins)	Calculate the binned entropy of the power spectral density of the time series (using the welch method).
friedrich_coefficients(x, param)	Coefficients of polynomial h(x), which has been fitted to the time series.
has_duplicate(x)	Checks if any value in x occurs more than once.
has_duplicate_max(x)	Checks if the maximum value of x is observed more than once.
has_duplicate_min(x)	Checks if the minimal value of x is observed more than once.
index_mass_quantile(x, param)	Calculates the relative index i of time series x where q% of the mass of x lies left of i.
kurtosis(x)	Returns the kurtosis of x (calculated with the adjusted Fisher-Pearson standardized moment coefficient G2).
large_standard_deviation(x, r)	Does time series have large standard deviation?
last_location_of_maximum(x)	Returns the relative last location of the maximum value of x.
last_location_of_minimum(x)	Returns the last location of the minimal value of x.
lempel_ziv_complexity(x, bins)	Calculate a complexity estimate based on the Lempel-Ziv compression algorithm.

[a]Source: `https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html`

Table C.3.: List of features extracted using *tsfresh*[a] (Continued).

Short Code	Description
length(x)	Returns the length of x.
linear_trend(x, param)	Calculate a linear least-squares regression for the values of the time series versus the sequence from 0 to length of the time series minus one.
linear_trend_timewise(x, param)	Calculate a linear least-squares regression for the values of the time series versus the sequence from 0 to length of the time series minus one.
longest_strike_above_mean(x)	Returns the length of the longest consecutive subsequence in x that is bigger than the mean of x.
longest_strike_below_mean(x)	Returns the length of the longest consecutive subsequence in x that is smaller than the mean of x.
matrix_profile(x, param)	Calculates the 1-D Matrix Profile and returns Tukey's Five Number Set plus the mean of that Matrix Profile.
max_langevin_fixed_point(x, r, m)	Largest fixed point of dynamics $argmax\_xh(x) = 0$ estimated from polynomial h(x).
maximum(x)	Calculates the highest value of the time series x.
maximum(x)	Calculates the highest value of the time series x.
mean(x)	Returns the mean of x
mean_abs_change(x)	Average over first differences.
mean_change(x)	Average over time series differences.
mean_n_absolute_max(x, number_of_maxima)	Calculates the arithmetic mean of the n absolute maximum values of the time series.
mean_second_derivative_central(x)	Returns the mean value of a central approximation of the second derivative.
median(x)	Returns the median of x.
minimum(x)	Calculates the lowest value of the time series x.
number_crossing_m(x, m)	Calculates the number of crossings of x on m.
number_cwt_peaks(x, n)	Number of different peaks in x.
number_peaks(x, n)	Calculates the number of peaks of at least support n in the time series x.
partial_autocorrelation(x, param)	Calculates the value of the partial autocorrelation function at the given lag.
percentage_of_reoccurring_datapoints_to_all_datapoints(x)	Returns the percentage of non-unique data points.
percentage_of_reoccurring_values_to_all_values(x)	Returns the percentage of values that are present in the time series more than once.
permutation_entropy(x, tau, dimension)	Calculate the permutation entropy.

[a] Source: `https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html`

Table C.4.: List of features extracted using *tsfresh*[a] (Continued).

Short Code	Description
quantile(x, q)	Calculates the q quantile of x.
query_similarity_count(x, param)	This feature calculator accepts an input query subsequence parameter, compares the query (under z-normalized Euclidean distance) to all subsequences within the time series, and returns a count of the number of times the query was found in the time series (within some predefined maximum distance threshold).
range_count(x, min, max)	Count observed values within the interval [min, max).
ratio_beyond_r_sigma(x, r)	Ratio of values that are more than r * std(x) (so r times sigma) away from the mean of x.
ratio_value_number_to _time_series_length(x)	Returns a factor which is 1 if all values in the time series occur only once, and below one if this is not the case.
root_mean_square(x)	Returns the root mean square (rms) of the time series.
sample_entropy(x)	Calculate and return sample entropy of x.
set_property(key, value)	This method returns a decorator that sets the property key of the function to value.
skewness(x)	Returns the sample skewness of x (calculated with the adjusted Fisher-Pearson standardized moment coefficient G1).
spkt_welch_density(x, param)	This feature calculator estimates the cross power spectral density of the time series x at different frequencies.
standard_deviation(x)	Returns the standard deviation of x.
sum_of_reoccurring_data_points(x)	Returns the sum of all data points, that are present in the time series more than once.
sum_of_reoccurring_values(x)	Returns the sum of all values, that are present in the time series more than once.
sum_values(x)	Calculates the sum over the time series values.
symmetry_looking(x, param)	Boolean variable denoting if the distribution of x looks symmetric.
time_reversal_asymmetry _statistic(x, lag)	Returns the time reversal asymmetry statistic.
value_count(x, value)	Count occurrences of value in time series x.
variance(x)	Returns the variance of x.
variance_larger_than _standard_deviation(x)	Is variance higher than the standard deviation?
variation_coefficient(x)	Returns the variation coefficient (standard error / mean, give relative value of variation around mean) of x.

[a]Source: `https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html`

Colophon

This thesis was typeset with $\LaTeX\,2_\varepsilon$. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at `http://cleanthesis.der-ric.de/`.