Imperial College London

Department of Mathematics

Open Markets and Rank Jacobi Processes

Author: Pedro Urbina Rodríguez (CID: 02357738)
Supervisor: Dr. David Itkin

Declaration

The work contained in this thesis is my own work unless otherwise stated.

Acknowledgements

I want to express my deepest gratitude to Dr. David Itkin. His invaluable guidance and insightful contributions have played a key role in the completion of this project. Our countless discussions have proven to be among the most enriching experiences throughout my academic journey.

I am also profoundly grateful to the dedicated faculty who have been integral to my academic progress over the past five years of my double degree program, spanning the Autonomous University of Madrid, the National University of Singapore, and Imperial College London. Their fervor and love for knowledge has left an enduring imprint on my academic and personal development.

Lastly, I would like to express heartfelt appreciation to my family and friends. Their unwavering support and encouragement, particularly during the development of this thesis, have been invaluable. Their belief in my capabilities has fueled my determination and resilience, and for that, I am eternally thankful.

Abstract

This thesis delves into the world of Stochastic Portfolio Theory, focusing on the application of two distinct investment strategies within the open market setting. The open market here refers to a subset of an equity market, comprising the top N capitalized stocks. We explore both a mixed approach, which includes the open market and the market portfolio, and a pure approach, which solely considers the open market.

Through rigorous theoretical examination and empirical validation, we extend and compare the performance of these two strategies, notably under the assumption of market weights being modelled as rank Jacobi processes. The thesis also aims to bridge the gap between theory and practice, offering a comprehensive comparison of these strategies and their respective advantages and shortcomings.

On the theoretical side, we derive explicit formulas for the growth optimal strategy without assumptions on the market weights, and introduce the Rank Jacobi processes as models for these weights. From a practical standpoint, we develop robust software for backtesting these investment strategies using historical financial data.

The results indicate contrasting behaviours and performances of the two strategies, with different implications for leverage and risk. The insights derived could significantly impact long-term financial planning and provide valuable guidance for investors. Future directions for research include diversifying market weight models, enhancing estimation techniques, refining leverage usage, and improving the parameter estimation process.

This thesis contributes to our understanding of Stochastic Portfolio Theory, providing new insights into the dynamics of financial markets and offering a novel perspective on investment strategies in the open market.

Contents

List of Figures

Chapter 1

Introduction

This thesis delves into the realm of Stochastic Portfolio Theory, a mathematical framework that attempts to understand and stduy the behaviour of financial markets. This field has gained significant attention due to its potential implications for investment strategies, portfolio optimization, and risk management. The theory explores the impact of uncertainties, such as price fluctuations and market volatilities, on investment outcomes, emphasizing the need for informed decision-making.

In the financial context, an open market is a fluid subset of a broader equity market, encompassing a fixed count N of the highest capitalized stocks at a specific moment. The makeup of this subset is continually changing, reflecting the change in companies' market capitalizations. This flexibility makes open markets a more realistic representation of actual stock market behaviour compared to static portfolios. Thus, open markets create a diverse, yet manageable, investing environment that evolves in line with real-world market changes.

The core of our investigation revolves around two investment approaches: one which incorporates the open market plus a market portfolio, developed in [1], and another that focuses solely on investments in the open market. We aim to extend and compare the results of these two strategies under different restrictions and delve into the case of market weights being modelled as Rank Jacobi processes, a class of multivariate stochastic processes.

1.1 Motivation and Significance of the Study

The motivations behind this study are rooted in both theoretical curiosity and practical application. Investment strategies play a crucial role in financial planning and portfolio management. Improving our understanding and refining these strategies is a continuous endeavour that underlies many aspects of financial mathematics. A comprehensive comparison of two distinct investment strategies — investing solely in top capitalization stocks (the pure open market approach) versus a combination of top capitalization stocks and the market portfolio (mixed open market approach) — is poised to provide valuable insights. This comparison has the potential to profoundly impact long-term investment strategies, including those integral to pension fund management, by facilitating enhanced performance and improved risk mitigation. Hence, the implications of this research extend significantly into both academic realms and practical applications, particularly those concerning long-term financial planning. The insights gained can inform the construction and management of portfolios serving as valuable guidance for long-term investors such as pension fund managers.

Comprehensive theoretical research on open markets has been meticulously explored in studies such as [2], and the optimal investment strategies considering both top capitalization stocks and the market portfolio have been investigated in previous research, including [1]. However, an extensive study that focuses on the calibration and implementation of growth optimal strategies, exclusively when the investment is restricted to top capitalization stocks, remains a largely unexplored territory. Venturing into this uncharted domain and exploring the implications of this restriction form the key theoretical motivations for this study. The novel insights derived from this research could broaden our understanding of optimal strategies within the open market setting, and offer a new perspective on the dynamics of financial markets.

From a practical standpoint, the motivation of this study extends beyond theoretical exploration. The real test of any theoretical result lies in its applicability to real-world scenarios. Testing these theoretical findings against actual market data is critical to assess their practical value. This empirical analysis closes the bridge between theory and practice. It ensures that the advancements of this study are not confined to the theoretical realm, but can potentially translate into tangible improvements in financial planning and investment strategies.

In addition, comparing the performance of the two strategies offers further insights into their efficacy, their relative advantages and disadvantages, and their suitability for different market conditions. This analysis can inform investment decision-making processes, providing practitioners with an evidence-based understanding of the potential impact of different investment strategies.

1.2 Objectives and Goals

The principal aim of this research is to delve deep into the heart of Stochastic Portfolio Theory, with a particular focus on examining and contrasting two distinct investment strategies. One strategy involves the combination of the open market and the market portfolio, while the other solely considers investments in the open market. By offering a thorough comparison of these two approaches, we aim to provide an analytical perspective on their respective merits and shortcomings. Our exploration extends to the performance of these strategies under different constraints, offering a more nuanced understanding of their behaviour in varying investment environments.

An important aspect of our objectives involves a concentrated study on market weights modelled as rank Jacobi processes. This forms a significant part of our theoretical investigation and opens up a challenging scenario for the comparison of the two investment approaches. By evaluating the dynamics of these strategies under the rank Jacobi model, we aim to establish a theoretical foundation that would enhance our understanding of their operation within this particular setup.

On the practical side, a key goal of this thesis is the development of robust software that allows us to carry out backtesting of investment strategies. By moving from a purely theoretical realm to a practical application of our findings, we aim to assess the viability and effectiveness of investment strategies in real-world settings. This part of the project involves software development, data cleaning, parameter estimation, and intensive backtesting over different time periods, facilitating an in-depth analysis of the strategies' performance over time. We also intend to explore different methods for estimating and smoothing the parameter a, further refining the accuracy and reliability of our model.

1.3 Outline of the Thesis

The introductory chapter of this thesis, Chapter 1, provides the necessary foundation for understanding the topic at hand. It gives a broad overview of the research, laying down the context and background of the study. The primary motivations driving this research are presented, along with a clear articulation of the objectives and goals.

Moving forward, the Literature Review (Chapter 2) and Mathematical Framework (Chapter 3) chapters delve into the theoretical underpinnings of the research. The Literature Review provides an examination of the existing body of knowledge, presenting a thorough review of relevant research in the area of Mathematical Finance, Stochastic Portfolio Theory, and open markets. Following this, the Mathematical Framework provides a detailed exploration of the mathematical concepts and techniques required to comprehend the subsequent chapters. It introduces multidimensional Itô's Lemma, stochastic processes, and rank Jacobi processes, among others.

The Theoretical Development chapter (Chapter 4) builds on the foundation laid by the preceding chapters, moving into a more in-depth exploration of the topic. The core theoretical developments of the study, which include the computation of the growth optimal strategies in different market situations and the exploration of functionally generated portfolios, are presented here. This section expounds on the intricate mathematical models formulated and employed throughout this study, focusing on the two different open market settings, and the rank Jacobi model.

Subsequent to the theoretical exploration, the Backtesting chapter (Chapter 5) shifts the focus to the empirical side of the study. It elaborates on the backtesting methodology used, detailing the data cleaning process, parameter estimation, and the computation of optimal strategies. The software development process and the consequent results are examined, including fixed and moving investment universe approaches and the final software that enables this empirical evaluation.

The final chapter of the thesis is the Conclusion chapter (Chapter 6), which wraps up the thesis by summarizing the findings, reflecting on the research process, and considering potential avenues for future work.

Chapter 2

Literature Review

The field of financial mathematics is vast and complex, spanning several distinct but interconnected theories and models. The underpinnings of modern portfolio theory and investment management are largely attributed to the pioneering work of Harry Markowitz and the development of the Capital Asset Pricing Model (CAPM). These theories have laid the foundation for subsequent advances, leading to sophisticated techniques in portfolio management such as Stochastic Portfolio Theory (SPT). Yet, these developments have not evolved in isolation, but within the broader context of evolving financial markets.

Specifically, the notion of open markets has presented an alternative approach to traditional investment opportunities, prompting the need to understand and optimize investment strategies within these contexts. This review provides a comprehensive exploration of these topics and sets the stage for the subsequent analysis of the rank Jacobi Process application to portfolio management.

This chapter begins by examining the roots of modern portfolio theory with Markowitz's seminal work, delving into the fundamental principles, mathematical models, and limitations of his theory. Stochastic Portfolio Theory (SPT) is explored next, presenting the principles, mathematical framework, and comparative analysis with Markowitz's theory. This is followed by a discussion of the concept, characteristics, and implications of open markets on investment strategies.

2.1 Markowitz's Portfolio Theory

Harry Markowitz's Modern Portfolio Theory (MPT) is a pivotal development in financial economics that brought a substantial shift in the understanding and implementation of investment approaches. This groundbreaking theory was first introduced through Markowitz's 1952 paper, "Portfolio Selection" [3].

One of the primary insights from Markowitz's theory is the power of diversification. Diversification postulates that by owning a range of different assets that do not perfectly move together, investors can lessen the risk inherent to individual investments, without necessarily sacrificing potential returns [4]. This proposition has been widely analyzed and confirmed in subsequent financial literature [5, 6, 7].

The risk-return trade-off is another cornerstone of MPT. It suggests that the potential for higher returns usually comes with higher risk, an observation that has formed an integral part of portfolio management and investment decision-making processes. Lastly, the concept of the efficient frontier represents the set of optimal portfolios offering the highest expected return for a defined level of risk [8]. The efficient frontier serves as a valuable benchmark for evaluating portfolio performance.

The mathematical model underpinning MPT focuses on calculating the expected returns, variances, and covariances of the portfolio's constituent assets. The aim is to identify portfolio weights that minimize portfolio variance (risk) for a given level of expected return, or alternatively, maximize expected return for a given level of risk. This optimization problem forms a standard approach to portfolio construction and has been widely adopted in various financial contexts.

Despite its significant contributions, Markowitz's theory has also been subject to criticisms and limitations. One key critique is its assumption of normally distributed returns, which often fails to accurately represent real-world markets that exhibit skewness, kurtosis, and fat tails in return distributions [9]. Additionally, MPT's heavy reliance on historical data to predict future returns and risk has been questioned, given the dynamic and often unpredictable nature of financial markets. Furthermore, MPT is inherently a single-period model, which can limit its applicability over multi-period investment horizons [10]. These limitations have underscored the need for more robust investment theories and models, such as Stochastic Portfolio Theory.

2.2 Stochastic Portfolio Theory

Stochastic Portfolio Theory (SPT), introduced by Robert Fernholz in [11, 12], proposes an innovative method for portfolio selection that captures the dynamic and uncertain nature of financial markets. Unlike focusing on individual securities, SPT concentrates on the relative performance and capitalization of assets in the market, modelled as stochastic processes. SPT is a descriptive theory, rather than a normative theory, as it tries to describe all the events which occur in equity markets [13]. STP, as most of the field of Mathematical Finance, descends from the original Markowitz's Portfolio Theory described in the previous section.

Assume we have a market of $d \geq 2$ stocks. The stock prices or capitalizations in STP are modelled by non negative semimartingales, a particular kind of stochastic process. The stock price of asset i is denoted S_i for $i = 1, \ldots, d$. It is a common assumption that the total market capitalization $S_1 + \cdots + S_d$ is strictly positive at all times. It is a usual practice to switch to this numeraire, and thus measure the relative market capitalizations of the stocks instead of their absolute capitalization S_i.

Mathematically, we denote the market weight vector of d assets as $X(t) =$

$(X_1(t), X_2(t), ..., X_d(t))$, where

$$X_i(t) = \frac{S_i(t)}{S_1(t) + \cdots + S_d(t)}, \quad i = 1, \ldots, d.$$

$X(t)$ is a stochastic process which represents the relative capitalization of the stocks with respect to the whole market capitalization. These market weights are assumed to evolve according to a stochastic differential equation.

Markowitz's Modern Portfolio Theory (MPT) and SPT diverge in several critical respects. Where MPT is static and focuses on a single period, SPT is dynamic and examines the evolution of assets' relative capitalizations over time. Unlike MPT, which assumes normally distributed returns and constant correlations, SPT doesn't rely on these assumptions. We refer the interested reader to [12, 13, 14, 15, 16] for an in-depth discussion on Stochastic Portfolio Theory.

2.3 Open Markets

In the context of Mathematical Finance, there has been a common assumption that markets are static entities with a fixed number of stocks being offered. This assumption, while helpful, has been shown not to be true in many real-world markets. It is a usual phenomenon that the amount of tradeable stocks changes significantly over time. Many stock exchanges have seen huge ups and downs in the number of stocks listed [2].

Addressing the dynamic nature of real-world markets requires a paradigm shift from conventional models. The concept of open markets appears to address this problem. In essence, an investor in an open market confines its attention to a predetermined number of top-performing stocks by market capitalization, a list that evolves in response to the shifting market trends. This way, we are able to capture the dynamic nature of the subset of stocks to be considered. This also helps model the fact that it is usually more difficult to trade the smallest capitalization stocks than the largest capitalization stocks.

This setting has gained some popularity recently, thanks to works such as [2], [17] and [1]. In [2], a theory of open markets is developed and growth optimal strategies, arbitrage, and functional generation of portfolios are discussed in this context.

Given that tractability is still a complex issue in open markets, the authors in [1] are able to overcome this problem by relaxing the open market constraints. In the original open market, the investor is only allowed to invest in a subset consisting of the N top capitalization stocks, among the total $d > N$ stocks. This is the setting we have called the pure open market setting. Their idea is to relax this constraint by additionally allowing investment in the complete market portfolio, and thus indirectly investment in the lowest capitalization stocks (in a fixed proportion). This is the setting we have called the mixed open market setting. This idea is a quite practical relaxation of the constraints, as there are usually many ways to invest in a proxy of the market portfolio, even though it might be complicated to invest in the lowest capitalization stocks. Under this new setting and an assumption on the correlation between large and small stocks, the authors are able to derive the growth optimal strategy in their setting. This work serves as the foundation for our own research.

Chapter 3

Mathematical Framework

The mathematical infrastructure that underpins any theoretical work acts as the scaffolding, providing the fundamental concepts, techniques, and structures that facilitate in-depth exploration and understanding of complex phenomena. In the context of Stochastic Portfolio Theory, our edifice is built on the powerful mathematics of stochastic processes, Itô calculus, and related methodologies that give us the necessary insight into the intricate workings of financial markets. This chapter unfolds the theoretical constructs that underlie our research, establishing a solid foundation for the detailed investigations that follow.

Our first subject of interest, stochastic processes, is a broad mathematical concept encapsulating a collection of random variables representing the evolution of a system of random values over time. Itô calculus, a cornerstone in the realm of stochastic differential equations, is another pivotal mathematical tool that grants us the ability to capture and predict the movements and variations within financial markets. We study the dynamics of the market weights and introduce the rank Jacobi processes, which play a vital role in our work as models of the market weights. We formally introduce the concept of investment strategies, and wealth processes, along with a study of its dynamics. We finally formally introduce the concept of mixed and pure open markets.

3.1 Stochastic Processes and Itô Calculus

Stochastic processes form the core mathematical framework of financial mathematics, enabling the modelling of random phenomena evolving over time. A stochastic process can be understood as a family of random variables which are indexed by a certain set, most commonly $\mathbb{R}^+$ interpreted as time. In the realm of finance, particularly in portfolio theory, various types of stochastic processes, such as Brownian motion and geometric Brownian motion, have been employed to model uncertainties inherent to financial markets.

The dynamics of financial quantities, such as asset prices, are often captured by stochastic differential equations (SDEs), which are differential equations wherein one or more of the terms is a stochastic process. A general form of a one-dimensional

SDE is:

$$dS(t) = \mu(t, S)dt + \sigma(t, S)dW \tag{3.1.1}$$

where $S(t)$ is the stochastic process, $W = W(t)$ is a standard Brownian motion, $\mu(t, S)$ represents the drift term (trend of the process), and $\sigma(t, S)$ signifies the diffusion term (volatility of the process). We call the process which can be written in the form 3.1.1 Itô processes.

A general form of a multidimensional stochastic differential equation (SDE) is given by

$$dS_i(t) = \mu_i(t, S_i)dt + \sum_j \sigma_{ij}(t, S_i)dW_j, \quad i = 1, \ldots, d, \tag{3.1.2}$$

where $S(t) = (S_1(t), ..., S_d(t))$ is a vector of stochastic processes, $W(t) = (W_1(t), ..., W_d(t))$ is a vector of independent standard Brownian motions, $\mu_i(t, S_i)$ represents the drift term of the i-th process, and $\sigma_{ij}(t, S_i)$ signifies the diffusion term related to the i-th process and the j-th Brownian motion.

To solve such SDEs and manipulate stochastic processes, a branch of mathematics called Itô calculus has been developed. Unlike standard calculus, Itô calculus is tailored to handle functions of stochastic processes by defining integration for stochastic integrands. The cornerstone of Itô calculus is the Itô integral, which is defined as

$$\int f(t, W)dW \tag{3.1.3}$$

for a sufficiently smooth function $f(t, W)$ and a standard Brownian motion $W = W(t)$.

The bridge between deterministic and stochastic calculus is provided by Itô's lemma, which essentially extends the chain rule to stochastic calculus. The one-dimensional version of Itô's lemma can be stated as follows.

Theorem 3.1.1 (One Dimensional Itô's Lemma). *Let $S(t)$ be an Itô process (3.1.1). Let $V(t, S)$ be a function such that both V and its first two derivatives are continuous. Then the process $V(t, S)$ is also an Itô process which satisfies*

$$dV(t, S) = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial S}dS + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(dS)^2 \tag{3.1.4}$$

where, to compute $(dS)^2$, we use the rules $(dt)^2 = dt \cdot dW = 0$ and $(dW)^2 = dt$.

In situations involving multiple correlated assets, we resort to the multidimensional version of Itô's lemma. This version extends Itô's lemma to stochastic processes in higher dimensions, catering to the complexity of multi-asset models.

Theorem 3.1.2 (Multidimensional Itô's Lemma). *Let $S_1(t), \ldots, S_n(t)$ be Itô processes. Let $V(t, S_1, \ldots, S_n)$ be a function such that both V and its first two derivatives are continuous. Then the process $V(t, S_1, \ldots, S_n)$ is also an Itô process which satisfies*

$$dV(t, S_1, \ldots, S_n) = \frac{\partial V}{\partial t}dt + \sum_{i=1}^{n} \frac{\partial V}{\partial S_i}dS_i + \sum_{i,j=1}^{n} \frac{1}{2}\frac{\partial^2 V}{\partial S_i \partial S_j}dS_i dS_j \tag{3.1.5}$$

where, to compute the terms $dS_i dS_j$, we use the rules $(dt)^2 = dt \cdot dW_i = dW_j \cdot dt = dW_i \cdot dW_j = 0$ for all $i \neq j$ and $(dW_i)^2 = dt$.

Equipped with these tools of stochastic processes, SDEs, and Itô calculus, we are capable of modelling the evolution of financial markets. These tools allow us to derive asset price dynamics and understand the mechanics of complex investment strategies in the stochastic environment.

3.2 Dynamics of the Market Weights

One of the key objects of study through this work is the market weights process denoted by $X = (X_1, \ldots, X_d)$ introduced in Section 2.2. We study its dynamics and prove that each of the market weights is indeed an Itô process so that we can later just write $dX(t) = b(t)dt + \sigma(t)dW(t)$. We assume for simplicity of the calculations that the price of the stocks can be modelled by a Geometric Brownian Motion, a particular class of semimartingales.

Proposition 3.2.1. *Let $S = (S_1, \ldots, S_d)$ be the price process of a d-dimensional market. Assume for each $i \in \{1, \ldots, d\}$, S_i is a Geometric Brownian Motion given by*

$$dS_i = S_i(\mu_i dt + \sigma_i dW_i), \quad \mu_i, \sigma_i \in \mathbb{R}. \tag{3.2.1}$$

Each component market weights process $X_k = \frac{S_k}{S_1, \ldots, S_d}$ is an Itô process and its dynamics are given by

$$dX_k = X_k \left(\mu_k + \sum_{i=1}^{d} \sigma_i^2 X_i^2 - \sum_{i=1}^{d} X_i \left(\mu_i - \sigma_i^2 \delta_{ik} \right) \right) dt - X_k \sum_{i=1}^{d} \left(X_i - \delta_{ik} \right) \sigma_i dW_i, \tag{3.2.2}$$

where $\delta_{ik} = 1$ if $i = k$ and 0 otherwise.

Proof. The fact that each X_k is an Itô process is a direct consequence of the Multidimensional Itô's Lemma (Theorem 3.1.2).

In order to find the dynamics of X_k, we define the processes

$$M(S_1, \ldots, S_d) = S_1 + \cdots + S_d, \quad N(M) = M^{-1}.$$

Both functions as well as X_k are twice differentiable and continuous so we can apply Multidimensional Itô's Lemma iteratively. We thus compute

$$dM(S_1, \ldots, S_d) = \left(\sum_{i=1}^{d} \mu_i S_i \right) dt + \sum_{i=1}^{d} \sigma_i S_i dW_i$$

$$(dM)^2 = \left(\sum_{i=1}^{d} \sigma_i^2 S_i^2 \right) dt.$$

We can now compute the dynamics of N using Itô again

$$
\begin{aligned}
dN &= \frac{\partial N}{\partial M}dM + \frac{1}{2}\frac{\partial^2 N}{\partial M^2}(dM)^2 \\
&= -M^{-2}\left(\sum_{i=1}^{d}\mu_i S_i\right)dt - M^{-2}\sum_{i=1}^{d}\sigma_i S_i dW_i + M^{-3}\left(\sum_{i=1}^{d}\sigma_i^2 S_i^2\right)dt \\
&= N\left(\sum_{i=1}^{d}\sigma_i^2 X_i^2 - \sum_{i=1}^{d}\mu_i X_i\right)dt - N\sum_{i=1}^{d}\sigma_i X_i dW_i.
\end{aligned}
$$

We can finally apply Itô to $X_k = S_k \cdot N$ and we get

$$
\begin{aligned}
dX_K &= NdS_k + S_k dN - M^{-2}\sigma_k^2 S_k^2 dt \\
&= \mu_k X_k dt + X_k\sigma_k dW_k + X_k\left[\left(\sum_{i=1}^{d}\sigma_i^2 X_i^2 - \sum_{i=1}^{d}\mu_i X_i\right)dt - \sum_{i=1}^{d}\sigma_i X_i dW_i\right] - \sigma_k^2 X_k^2 dt \\
&= X_k\left(\mu_k + \sum_{i=1}^{d}\sigma_i^2 X_i^2 - \sum_{i=1}^{d}X_i\left(\mu_i - \sigma_i^2\delta_{ik}\right)\right)dt - X_k\sum_{i=1}^{d}\left(X_i - \delta_{ik}\right)\sigma_i dW_i.
\end{aligned}
$$

□

3.3 Rank Jacobi Processes

Itkin and Larsson introduced in [1] a novel concept known as hybrid Jacobi processes, a class of stochastic processes inspired by both Jacobi processes and the work of [18]. These hybrid Jacobi processes, with rank Jacobi processes as a particular instance, are constructed and investigated thoroughly in [1]. In the mixed open market setting, these hybrid Jacobi processes are employed to model market weights and determine growth optimal strategies.

The primary motivation for this model is to represent the stability of the capital distribution curve, one of the few time invariants found in equity markets. Specifically, these hybrid Jacobi processes are leveraged to model the relative market weights denoted as X_i.

Our focus is on a specific subset of these hybrid Jacobi processes, known as rank Jacobi processes. The simplicity of this subclass makes it particularly appealing, enabling us to derive explicit forms for growth optimal strategies introduced in subsequent chapters and estimate all required parameters for backtesting.

We first need to introduce some notation

Notation 3.3.1. Let $v \in \mathbb{R}^d$. We define the tail sum starting from $j = 1, \ldots, d$ by

$$\bar{v}_j = \sum_{i=j}^{d} v_i.$$

We also write $\bar{v}_{(j)} = \sum_{i=j}^{d} v_{(i)}$ for the tail sum of $v_{()}$.

We can now introduce the rank Jacobi processes following Definition 3.1 in [1]

Definition 3.3.2 (Rank Jacobi Process)**.** Let $a = (a_1, \dots, a_d) \in \mathbb{R}^d$, $\sigma > 0$. A rank Jacobi model is the process defined in the standard simplex of $\mathbb{R}^d$ as the weak solution of the following SDE

$$dX_i = \frac{\sigma^2}{2}(a_{r_i(t)} - \bar{a}_1 X_i(t))dt + \sigma \sum_{j=1}^{d} (\delta_{ij} - X_i(t))\sqrt{X_j(t)}dW_j(t), \quad i = 1, \dots, d, \tag{3.3.1}$$

where W is a standard d-dimensional Brownian motion.

We will usually assume $\sigma = 1$ for simplicity. We will also make an assumption to guarantee the existence of the solution to the previously defined SDE.

Assumption 3.3.3. The vector a is such that $\bar{a}_k > 0$ for $k = 2, \dots, d$.

3.4 Investment Strategies and Wealth Processes

We now want to find a way to measure how a given strategy in the open market performs. First, we need to formalize the concept of trading strategy. Following [1], we can parametrize our strategies by an X-integrable process $H = (H_1, \dots, H_d)$ which represents the number of shares we have invested in each asset.

Definition 3.4.1 (Investment Strategy)**.** We say that a process $H = (H_1, \dots, H_d)$ is a investment strategy if it is X-integrable.

We further define the wealth process associated with the strategy H.

Definition 3.4.2 (Wealth Process)**.** Given an investment strategy $H = (H_1, \dots, H_d)$, we define the wealth process associated with strategy H by

$$V^H = H_1 X_1 + \dots + H_d X_d. \tag{3.4.1}$$

Note that the wealth process is expressed in the numeraire $S_1 + \dots + S_d$ and thus measures the wealth relative to the market. The market portfolio is a common benchmark for evaluating investment performance. If an investment strategy consistently outperforms the market, it is generally considered successful. Also, the efficient market hypothesis posits that it is impossible to consistently outperform the market, after adjusting for risk, because all available information is already incorporated into stock prices. Measuring performance relative to the market allows us to test this hypothesis.

We can now define what we will consider a valid investment strategy.

Definition 3.4.3 (Valid Investment Strategy)**.** We say that a process $H = (H_1, \dots, H_d)$ is a valid investment strategy if it is X-integrable and its associated wealth process V^H is strictly positive for all times t.

When we have a valid investment strategy H, we can also parametrize it in the following way

$$\theta_i = \frac{H_i}{V}, \tag{3.4.2}$$

which represents the number of shares per unit of wealth that the investor holds on asset i. This new parametrization has the property

$$\theta_1 X_1 + \cdots + \theta_d X_d = 1. \tag{3.4.3}$$

This property turns out to be equivalent to the condition that the wealth process V^H is strictly positive for all times t.

In summary, we can parametrize our valid investment strategies in two different ways. We can either use the number of shares we have invested in each asset, denoted by H, or we can use the number of shares per united of wealth that we have invested in each asset, denoted by θ. In order to be valid investment strategies both processes have to be X-integrable and in the first case the associated wealth process $V^H(t)$ has to be positive for all times and in the second case 3.4.3 has to be satisfied. We will often use the form θ when we are discussing valid investment strategies. We will also sometimes drop the word valid as we will be always treating valid investment strategies.

3.5 Dynamics of the Log Wealth

We now want to study the dynamics of the log wealth, denoted by $\log V$. In the following chapter, we will be looking to find the strategy which maximizes the growth of our wealth. The primary reason for focusing on the log wealth, as opposed to the wealth itself, comes down to the properties of the logarithm function. The logarithmic function possesses an advantageous characteristic: it is a monotonic transformation. This feature ensures that it preserves the ordering of returns, thereby retaining the original pattern of return values when applied.

Moreover, the log transformation also turns multiplicative processes into additive ones. This additive property simplifies the calculations. Additionally, in certain settings, optimizing log wealth can be interpreted as maximizing expected utility when an investor has logarithmic utility, which, due to its risk aversion property, is a commonly adopted utility function in economics and finance.

In the following, we will omit the time dependency of different quantities for notational convenience. We will denote $c = \sigma\sigma^\top$ and assume there exists ℓ such that $c\ell = b$. It is important to note that given the fact that the components of X always add up to 1, we have that

$$c\mathbf{1}_d = \mathbf{0}_d, \quad b^\top \mathbf{1}_d = 0. \tag{3.5.1}$$

Proposition 3.5.1. *Suppose that the dynamics of the market weights are given by*

$$dX = bdt + \sigma dW, \quad b \in \mathbb{R}^d, \quad \sigma \in \mathbb{R}^d \times \mathbb{R}^d, \tag{3.5.2}$$

where W is a d dimensional standard Brownian motion. That is, each market weight follows the dynamics

$$dX_i = b_i dt + \sum_{j=1}^{d} \sigma_{ij} dW_j, \quad i = 1, \dots, d. \tag{3.5.3}$$

Given an X-integrable investment strategy $H = (H_1, \dots, H_d)$ such that its associated wealth process $V^H = H^\top X = \sum_{i=1}^{d} H_i(t) X_i(t)$ is strictly positive, the dynamics of the log wealth are given by

$$d \log V = \left(\theta^\top c\ell - \frac{1}{2} \theta^\top c \theta \right) dt + \theta^\top \sigma dW. \tag{3.5.4}$$

Proof. We will use Multidimensional Itôs Lemma. In order to do that we first need to compute the term $(dX)^2$. Applying the standard multiplication rules we obtain

$$\begin{aligned} dX_j dX_k &= (b_j dt + \sigma_{j1} dW_1 + \dots + \sigma_{jd} dW_d)(b_k dt + \sigma_{k1} dW_1 + \dots + \sigma_{kd} dW_d) \\ &= \left(\sum_{i=1}^{d} \sigma_{ji} \sigma_{ki} \right) dt. \end{aligned}$$

We thus obtain

$$(dX)^2 = \sigma \sigma^\top dt = c dt. \tag{3.5.5}$$

Given that $dV = H dX$ and because of the self-financing condition

$$dV(t) = H(t)^\top dX(t) = \sum_{i=1}^{d} H_i(t) dX_i(t). \tag{3.5.6}$$

Thanks to the fact that $V(t)$ is strictly positive, we can write

$$dV(t) = V(t) \frac{H(t)^\top}{V(t)} dX(t) = V(t) \theta(t)^\top dX(t). \tag{3.5.7}$$

We can now apply Itô's Lemma to compute the dynamics of the log wealth

$$\begin{aligned} d \log V(t) &= \frac{1}{V(t)} dV(t) - \frac{1}{2} \frac{1}{V(t)^2} (dV(t))^2 \\ &= \theta(t)^\top dX(t) - \frac{1}{2} \theta(t)^\top dX(t) \theta(t)^\top dX(t) \\ &= \left(\theta(t)^\top b - \frac{1}{2} \theta(t)^\top c \theta(t) \right) dt + \theta(t)^\top \sigma dW(t) \\ &= \left(\theta(t)^\top c\ell - \frac{1}{2} \theta(t)^\top c \theta(t) \right) dt + \theta(t)^\top \sigma dW(t). \end{aligned}$$

□

3.6 Mixed and Pure Open Market Settings

In this section we will explore the concepts of Pure and Mixed Open Markets, along with the strategies which are allowed in these settings.

We start by defining, following [1], the rank and name identifying functions

$$\begin{aligned} r_i(t) = r_i(X(t)) &= \text{the rank occupied by name } i \text{ at time } t, \\ n_k(t) = n_k(X(t)) &= \text{the name that occupies rank } k \text{ at time } t. \end{aligned}$$

We also introduce notation for the ranked trading strategy $\theta_{\mathbf{n}}(t)$ which orders the investments in the assets by rank and the truncated vector of market weights $X_{()}$, which orders the vector $X(t)$ by rank

$$\theta_{\mathbf{n}}(t) = (\theta_{n_1(t)}(t), \ldots, \theta_{n_d(t)}(t)) \tag{3.6.1}$$

$$X_{()}^N = (X_{(1)}, \ldots, X_{(N)}). \tag{3.6.2}$$

We are now in a position to define what we mean by pure and mixed open markets. Given a market made up of d assets, open markets refer to markets where we are allowed to invest in the $N < d$ top capitalization assets. The idea is that in a pure open market, we are only allowed to invest in this N top assets while in a mixed open market, we are allowed to invest not only in the top N assets but also in the complete market portfolio defined by $\theta_{\mathcal{M}} = \mathbf{1}_d$.

The motivation for the mixed open market, introduced in [1], is that there exist investment instruments, such as ETFs, which track the complete d-dimensional market. This means that any investor can easily invest in a proxy of the market portfolio. This contrasts with the difficulty of investing directly in low capitalization stocks, thus motivating the open market setting.

It also turns out to be convenient mathematically. If we have an investment strategy θ which does not satisfy the positive wealth process condition (3.4.3), [1] notes that we can always shift by a multiple $1 - X^\top\theta$ of the market portfolio to obtain a new investment strategy

$$\theta' = \theta + (1 - X^\top\theta)\mathbf{1}_d \tag{3.6.3}$$

This new strategy is also X-integrable and satisfies condition 3.4.3. This means that any X-integrable process can be converted into a valid trading strategy.

This allows us to define mixed open market strategies.

Definition 3.6.1 (Mixed Open Market Strategy, Definition 2.5 in [1])**.** Fix $N \leq d$. A valid trading strategy θ is called a mixed open market strategy if it admits the representation

$$\theta_{\mathbf{n}}(t) = \begin{pmatrix} h(t) \\ 0 \end{pmatrix} + (1 - h(t)^\top X_{()}^N(t))\mathbf{1}_d \tag{3.6.4}$$

for some N-dimensional process h such that the d-dimensional process $h_{r_i(t)}(t)1_{r_i(t)\leq N}$, $i = 1, \ldots, d, t \geq 0$ is X-integrable. We denote $\mathcal{O}_m^N$ the set of all such strategies.

h_k represents the number of shares per unit of wealth in which we invest in asset k at time t for $k \in \{1, \ldots, N\}$, while we finance this strategy by shorting or longing the market portfolio $\mathbf{1}_d$.

On the contrary, on the pure open market, we do not allow investment in the market portfolio $\mathbf{1}_d$. One of the main objectives of this work is to understand how this restriction affects the overall result of the optimal trading strategy.

Definition 3.6.2 (Pure Open Market Strategy)**.** Fix $N \leq d$. A valid trading strategy θ is called a pure open market strategy if it admits the representation

$$\theta_{\mathbf{n}}(t) = \begin{pmatrix} h(t) \\ 0 \end{pmatrix} \tag{3.6.5}$$

for some N-dimensional process h such that the d-dimensional process $h_{r_i(t)}(t) 1_{r_i(t) \leq N}$, $i = 1, \ldots, d, t \geq 0$ is X-integrable. We denote $\mathcal{O}_p^N$ the set of all such strategies.

We note that in this case, it is not sufficient that the process $h_{r_i(t)}(t) 1_{r_i(t) \leq N}$, $i = 1, \ldots, d, t \geq 0$ is X-integrable, but the strategy θ also has to be a valid trading strategy and thus satisfy the condition 3.6.6, which boils down to

$$h_1(t)X_1(t) + \cdots + h_N(t)X_N(t) = 1, \quad \text{for all } t \geq 0. \tag{3.6.6}$$

We also define the ranked covariation and the ranked vector l, which will be very useful in subsequent chapters.

$$\kappa_{kl} = c_{n_k(t), n_l(t)}(t), \quad \rho_k(t) = \ell_{n_k(t)}(t), \quad t \geq 0, \quad k, l = 1, \ldots, d. \tag{3.6.7}$$

Finally, we define notation which will be useful in subsequent chapters.

Notation 3.6.3. Given $N, d \in \mathbb{N}$, $N < d$ and a vector $v = (v_1, \ldots, v_d) \in \mathbb{R}^d$ we denote by v^N the truncated vector which only contains the first N components, that is, $v^N = (v_1, \ldots, v_N) \in \mathbb{R}^N$.

Notation 3.6.4. Given $N, d \in \mathbb{N}$, $N < d$ and a matrix $a = (a_{ij}) \in \mathcal{M}_{d \times d}, i, j \in \{1, ..., d\}$ we denote by a^N the truncated matrix which only contains the N by N top-left square matrix, that is, $a^N = (a_{ij}) \in \mathcal{M}_{N \times N}(\mathbb{R}), i, j \in \{1, ..., N\}$.

Chapter 4

Theoretical Development

The theoretical development forms the backbone of our exploration in the realm of Stochastic Portfolio Theory, focusing on open markets. As the bedrock of our study, this chapter provides an in-depth understanding of the key constructs that govern portfolio optimization and wealth growth. We study the Growth Optimal Strategies and the rank Jacobi Model, drawing their implications in different market settings.

In this chapter, we embark on a meticulous dissection of Growth Optimal Strategies in diverse market scenarios: closed markets, pure open markets, and mixed open markets. An integral part of our theoretical discourse involves delving into the intricacies of the Rank Jacobi Model and applying it to our previously obtained growth optimal strategies. One of the motivations behind employing this model is its capacity to incorporate or model the stability of the capital distribution curve, which is a significant real-world characteristic of financial markets. They also allow us to explicitly study growth optimal strategies, and they have some robustness properties as studied in [1]. Moreover, we also study the asymptotic robust growth and the functional generation of the growth optimal portfolio in the pure open market setting and compare it with the mixed open market setting.

This chapter also bridges the gap between the theoretical constructs and their practical implementations. The theories and strategies established here form the foundation for the practical applications and backtesting procedures described in the subsequent chapters. The process of understanding how these strategies work in the theoretical realm prepares us for observing their performance in real-world scenarios. As we transition from this chapter to the next, we carry forward the insights gained here, providing a solid groundwork for our empirical analysis and interpretation of the results.

4.1 Growth Optimal Strategies in Closed and Mixed Open Markets

This section studies the growth optimal strategies in the Closed and Mixed Open Markets. Following [19, 1] we define a growth optimal strategy in terms of semimartingale decomposition of its associated log wealth process

$$\log V^\theta = M^\theta + A^\theta, \tag{4.1.1}$$

where M^θ is a local martingale and A^θ is a finite variation process.

We now define the concept of growth optimal strategy following [1].

Definition 4.1.1 (Growth Optimal Strategy)**.** Let $\mathcal{O}$ be a set of strategies. A strategy $\hat{\theta}$ is said to be growth optimal if $A^{\hat{\theta}} - A^\theta$ is a non-decreasing process for every strategy $\theta \in \mathcal{O}$, where A^θ is the finite variation process of the semimartingale decomposition of the wealth process associated with strategy θ.

In our case, we have obtained that

$$d\log V = \left(\theta^\top c\ell - \frac{1}{2}\theta^\top c\theta\right) dt + \theta^\top \sigma dW, \tag{4.1.2}$$

which means that we want to maximize the finite variation term or drift

$$A^{\hat{\theta}} = \left(\theta^\top c\ell - \frac{1}{2}\theta^\top c\theta\right) dt. \tag{4.1.3}$$

4.1.1 Growth Optimal Strategy in the Closed Market

The closed market setting is the usual setting. Here, we are allowed to invest in all the d assets in the market. We do not have the restriction of only being allowed to invest in the top N assets.

Because of condition 3.5.1, we always have that the rank of our covariation matrix c is at most $d-1$. We assume, for simplicity, that indeed the rank of c is $d-1$, which turns out to be true in most practical settings.

Proposition 4.1.2. *Let the covariation matrix c have a rank of $d-1$. A growth optimal strategy in the closed market setting is given by*

$$\hat{\theta} = c^+ b = \ell, \tag{4.1.4}$$

where $c^+ = V\Sigma^+ U^\top$, V, Σ, U are given by the always existing singular value decomposition $c = U\Sigma V^\top$, and $\Sigma^+ = diag(\lambda_1^{-1}, \dots, \lambda_{d-1}^{-1}, 0)$, where $\lambda_1, \dots, \lambda_{d-1}$ are the non-zero eigenvalues of the matrix c.

Proof. The problem of finding the growth optimal strategy in the closed market reduces to an unconstrained optimization problem. This problem is defined by

$$\max_{\theta \in \mathcal{O}} \{\theta(t)^\top b - \frac{1}{2}\theta(t)^\top c\theta(t)\},$$

where $\mathcal{O}$ is the set of all investment strategies in the closed market, that is, all the strategies from Definition 3.4.1. In the context of our optimization problem, we are able to employ pointwise maximization since the vectors and matrix within our function are continuous and independent at each moment in time. This approach allows us to independently determine the function's value at each time point and find the maximum separately, thus establishing an array of optima throughout the given time period. This approach maximizes instantaneous growth, suitable for our analysis.

Unfolding the vector and matrix notation we obtain

$$\theta(t)^\top b - \frac{1}{2}\theta(t)^\top c\theta(t) \ = \sum_{i=1}^{d} b_i\theta_i - \frac{1}{2}\sum_{i,j=1}^{d} c_{ij}\theta_i\theta_j.$$

We simply define

$$f : \mathbb{R}^d \to \mathbb{R}, \quad f(x_1, \ldots, x_d) = \sum_{i=1}^{d} b_i x_i - \frac{1}{2}\sum_{i,j=1}^{d} c_{ij}x_i x_j.$$

We want to find the critical points of f so we take partial derivatives, for $i = 1, \ldots, d$

$$\frac{\partial f}{\partial x_i} = b_i - \frac{1}{2}\left(\sum_{j=1}^{d} c_{ij}x_j + \sum_{j=1}^{d} c_{ji}x_j\right) = b_i - \frac{1}{2}(c_{:i}x + c_{:i}^\top x).$$

We have thus obtained that

$$\nabla f(x) = b - \frac{1}{2}(c + c^\top)x.$$

We note that

$$c = \sigma\sigma^\top \Rightarrow c^\top = (\sigma\sigma^\top)^\top = \sigma\sigma^\top = c,$$

so we can finally write

$$\nabla f(x) = b - cx.$$

Imposing the first-order condition we find that the growth optimal strategy must satisfy

$$\nabla f(x) = b - cx = 0 \Rightarrow cx = b.$$

It is also easy to show f is concave, as the Hessian matrix is given by $-c$ and c is definite positive every time the matrix σ has full rank.

The problem now is that the matrix c is not invertible because we have already noted that $c\mathbf{1}_d = 0$ given that the components of X add up to 1. This problem is due to the dimensionality reduction we performed when switching from market capitalizations S_i to the relative market capitalizations X_i, we lost one degree of freedom.

Despite this problem, we can use the pseudoinverse given by

$$c^+ = V\Sigma^+U^\top,$$

where the matrices V, Σ, U are given by the always existing singular value decomposition $c = U\Sigma V^\top$ and $\Sigma^+ = diag(\lambda_1^{-1}, \dots, \lambda_{d-1}^{-1}, 0)$, where $\lambda_1, \dots, \lambda_{d-1}$ are the non-zero eigenvalues of the matrix c.

The solutions to our problem are thus defined by

$$x = c^+ b + (I - c^+ c)w, \quad w \in \mathbb{R}^d.$$

If we choose $w = 0$ we obtain

$$\hat{\theta} = c^+ b = \ell.$$

□

We now turn or attention to the two open market settings.

4.1.2 Growth Optimal Strategy in the Mixed Open Market

The mixed market setting is the setting developed and studied in [1]. Here, we are allowed to invest in only the market portfolio $\mathbf{1}_d$ and in the top N assets by capitalization in a d-dimensional market. Given that we will be comparing the pure open market setting with the mixed open market setting we introduce in this section one of the main results in [1].

First, we need to make an assumption on the instantaneous covariation between large capitalization assets and small capitalization assets in order to find explicit solutions for the growth optimal strategy problem.

Assumption 4.1.3 (Assumption 2.9 in [1])**.** There exist progressively measurable processes $f_1, \dots, f_d$ and g such that , up to $\mathbb{P} \otimes dt$-nullsets, $\sum_{k=N+1}^d = 1$ and

$$\kappa_{kl} = -f_k f_l \text{ for all } k = 1, \dots, N, \quad l = N+1, \dots, d. \tag{4.1.5}$$

Proposition 4.1.4 (Theorem 2.10 in [1])**.** *Fix $N < d$ and let Assumption 4.1.3 be satisfied. Then the optimal strategy in the mixed open market is determined by Equation* (3.6.4) *with*

$$h(t) = \hat{h}_m(t) = \rho^N(t) - \xi(t)\mathbf{1}_N \tag{4.1.6}$$

where $\rho^N(t) = (\rho_1(t), \dots, \rho_N(t))^\top$ and $\xi(t) = \sum_{i=N+1}^d f_l(t)\rho_l(t)$.

Proof. In the open market setting, the drift of the log wealth of our allowed strategies is now given by

$$h(t)^\top (\kappa(t)\rho(t))^N - \frac{1}{2} h(t)^\top \kappa^N(t) h(t), \tag{4.1.7}$$

where $h(t)$ is the one which appears in Equation (3.6.4) and we have just substituted the terms in the drift of the log wealth in expression 4.1.3 by their ranked equivalents.

This turns out again to be an unconstrained, concave optimization problem. A similar first-order condition analysis yields that the growth optimal strategy in the mixed open market setting must satisfy

$$(\kappa(t)\rho(t))^N = \kappa^N(t) h(t). \tag{4.1.8}$$

Fix $k \in \{1, \dots, N\}$. Given that $\kappa \mathbf{1}_d = 0$ we have that $(\kappa \mathbf{1}_d)_k = 0$ and we can write

$$(\kappa^N \mathbf{1}_N)_k = \sum_{i=1}^{N} \kappa_{ki} = - \sum_{i=N+1}^{d} \kappa_{ki} = \sum_{i=N+1}^{d} f_k f_i g = g f_k.$$

If we fix $t > 0$ we have

$$\begin{aligned}
(\kappa\rho)_k^N = \sum_{i=1}^{d} \kappa_{ki}\rho_i &= \sum_{i=1}^{N} \kappa_{ki}\rho_i + \sum_{i=N+1}^{d} \kappa_{ki}\rho_i \\
&= \sum_{i=1}^{N} \kappa_{ki}\rho_i - \sum_{i=N+1}^{d} f_k f_i g \rho_i \\
&= (\kappa^N \rho^N)_k - (\kappa^N \mathbf{1}_N)_k \sum_{i=N+1}^{d} f_i \rho_i \\
&= (\kappa^N (\rho^N - \xi \mathbf{1}_N)_k).
\end{aligned}$$

We thus obtain

$$(\kappa\rho)^N = (\kappa^N (\rho^N - \xi \mathbf{1}_N)).$$

If we define $\hat{h}_m(t) = \rho^N - \xi \mathbf{1}_N$, we have found a solution to equation 4.1.8 which completely characterizes our growth optimal strategy in the mixed open market setting. □

4.2 Growth Optimal Strategy in the Pure Open Market

We now shift our focus to the pure open market setting, which is the primary setting examined in this work. This model was previously developed in an abstract way in [2]. Our goal is to adapt this pure open market concept into a more practical framework that allows for comparison with other scenarios, such as the mixed open market and closed market settings.

In this configuration, only strategies of the form outlined in Definition 3.6.2 are permitted. We restrict investment to the top N stocks, determined by capitalization (which we will often refer to simply as "the top N stocks"). This differs from the mixed market setting, where investment in the open market portfolio is also allowed.

The class of allowed trading strategies is now strictly smaller, i.e. $\mathcal{O}_p^N \subset \mathcal{O}_m^N$, as the market portfolio $\mathbf{1}_d \in \mathcal{O}_m^N$ and every pure open market strategy is also a mixed open market strategy. This in particular means that we expect the performance of the growth optimal strategy in the pure open market to be worse than that of the mixed open market. However, we are interested in understanding how big this effect of removing from the investment universe the market portfolio is.

4.2.1 Computation of the Growth Optimal Strategy in the Pure Open Market

We start the section by computing the growth optimal strategy in the pure open market setting. One of the main terms involved in this calculation will be the inverse of κ^N. We thus need to start by assuming that the matrix κ^N is invertible.

Theorem 4.2.1 (Growth Optimal Strategy in the Pure Open Market)**.** *Fix $N < d$ and suppose that the matrix κ^N is invertible. Then the optimal strategy in the pure open market is determined by Equation* (3.6.5) *with*

$$\hat{h}_p(t) = (\kappa^N)^{-1}(\kappa\rho)^N - \frac{(X_{()}^N)^\top((\kappa^N)^{-1}(\kappa\rho)^N) - 1}{(X_{()}^N)^\top(\kappa^N)^{-1}X_{()}^N}(\kappa^N)^{-1}X_{()}^N. \tag{4.2.1}$$

Proof. In the open market setting, the drift of the log wealth of our allowed strategies is given by

$$h(t)(\kappa(t)\rho(t))^N - \frac{1}{2}h(t)^\top\kappa^N(t)h(t),$$

where $h(t)$ is the one which appears in Equation (3.6.5). As expressed in 3.6.2, we also need that the strategy θ to be a valid trading strategy and thus satisfy the condition 3.4.3, which is equivalent to h satisfying 3.6.6. The nature of the optimization problem now is different, as we have gone from an unrestricted optimization problem in the case of the mixed open market to a constrained optimization problem. The problem is now

$$\max_{h(t)^\top \in \mathcal{O}_p^N} \left\{ h(t)(\kappa(t)\rho(t))^N - \frac{1}{2}h(t)^\top\kappa^N(t)h(t) \right\} \text{ subject to } h(t)^\top X_{()}^N(t) = 1.$$

In order to solve this restrained optimization problem we use the method of Lagrange multipliers. We drop the time dependence for clarity. The Lagrangian function is defined by

$$\begin{aligned} L(h,\lambda) = L(h_1,\dots,h_N,\lambda) &= h^\top(\kappa\rho)^N - \frac{1}{2}h^\top\kappa^N h - \lambda(h^\top X_{()}^N - 1) \\ &= \sum_{i=1}^N\sum_{j=1}^d h_i\rho_j\kappa_{ij} - \frac{1}{2}\sum_{i,j=1}^N h_ih_j\kappa_{ij} - \lambda\left(\sum_{i=1}^N h_iX_{(i)} - 1\right). \end{aligned}$$

We can now impose the First Order Condition on the gradient of the Lagrangian function $\nabla L(h_1,\dots,h_N,\lambda) = 0$ to get

$$\begin{aligned} &\frac{\partial L}{\partial\lambda} = 0 \Rightarrow h^\top X_{()}^N = 1, \\ &\frac{\partial L}{\partial h_k} = 0 \Rightarrow \sum_{j=1}^d \rho_j\kappa_{kj} - \frac{1}{2}\left(\sum_{i=1}^N h_i\kappa_{ik} + \sum_{i=1}^N h_i\kappa_{ki}\right) - \lambda X_{(k)} = 0 \\ &\qquad\qquad \Rightarrow (\kappa\rho)_k - \frac{1}{2}((\kappa^N h)_k + ((\kappa^N)^\top h)_k) - \lambda X_{(k)} = 0, \text{ for all } k = 1,\dots,N. \end{aligned}$$

Given that $c = c^\top$, we also have that $\kappa^N = (\kappa^N)^\top$ and the two conditions can be summarized as

$$(\kappa\rho)^N - \kappa^N h - \lambda X_{()}^N = 0 \text{ and } h^\top X_{()}^N = 1.$$

By the first condition we immediately get, thanks to the fact that κ^N is invertible, that

$$h = (\kappa^N)^{-1}((\kappa\rho)^N - \lambda X_{()}^N). \tag{4.2.2}$$

In order to compute λ we use $h^\top X_{()}^N = 1$, which implies

$$((\kappa^N)^{-1}(\kappa\rho)^N)^\top X_{()}^N - \lambda((\kappa^N)^{-1}X_{()}^N)^\top X_{()}^N = 1.$$

So we get

$$\lambda = \frac{(X_{()}^N)^\top(\kappa^N)^{-1}(\kappa\rho)^N}{(X_{()}^N)^\top(\kappa^N)^{-1}X_{()}^N} \tag{4.2.3}$$

Combining 4.2.2 and 4.2.3, and observing that the objective function is concave for the same reason as in the closed market, we get the resulting growth optimal strategy

$$\hat{h}_o(t) = (\kappa^N)^{-1}(\kappa\rho)^N - \frac{(X_{()}^N)^\top((\kappa^N)^{-1}(\kappa\rho)^N) - 1}{(X_{()}^N)^\top(\kappa^N)^{-1}X_{()}^N}(\kappa^N)^{-1}X_{()}^N.$$

□

It is important to observe, that if we grant Assumption 4.1.3, the obtained pure open market growth optimal strategy is very similar to its mixed open market equivalent, with an additional correction term. As computed in the previous section, $(\kappa^N)^{-1}(\kappa\rho)^N = \rho^N - \xi\mathbf{1}_N$, which implies that

$$\hat{h}_o(t) = \rho^N - \xi\mathbf{1}_N - \frac{(X_{()}^N)^\top\rho^N + \xi\bar{X}_{(N+1)} - (1+\xi)}{(X_{()}^N)^\top(\kappa^N)^{-1}X_{()}^N}(\kappa^N)^{-1}X_{()}^N. \tag{4.2.4}$$

This means that

$$\hat{h}_o(t) = \hat{h}_m(t) - \frac{(X_{()}^N)^\top\rho^N + \xi\bar{X}_{(N+1)} - (1+\xi)}{(X_{()}^N)^\top(\kappa^N)^{-1}X_{()}^N}(\kappa^N)^{-1}X_{()}^N. \tag{4.2.5}$$

This means that restricting our strategy to the pure open market affects the growth optimal strategy by adding a correction term. This correction term is financed with the money originally invested in the whole market portfolio. In subsequent chapters, we will study whether this correction term has a significant impact on the strategy, as it is difficult to understand it analytically.

We also tried studying the particular case when $N = 2$ but this did not give us any particular insight or simple formula for the growth optimal strategy. We thus turned our attention to modelling the market weights by the rank Jacobi processes in order to deepen our understanding of this difference.

4.3 Growth Optimal Strategy in the Pure Open Market: Rank Jacobi Model

We now want to study the growth optimal strategy in the pure open market when we assume that the market weights X_i can be modelled by the rank Jacobi processes introduced in Section 3.3. This will allow us to get much more explicit formulas which will allow us to backtest the performance of the strategy in the subsequent chapter, thanks to the ability to estimate the parameter a in the rank Jacobi model.

4.3.1 Preliminary Results

We start by proving some small results which will be useful for our later computations.

Proposition 4.3.1. *Assume that the market weights X_i are rank Jacobi processes defined in 3.3.2. Then the covariation matrix c is given by*

$$c_{ij}(t) = \sigma^2 X_i(t)(\delta_{ij} - X_j(t)). \tag{4.3.1}$$

Proof. We omit time dependence for clarity:

$$\begin{aligned}
c_{ij}dt = dX_i dX_j &= \\
&= \sigma^2 \left(\sum_{k=1}^{d} (\delta_{ik} - X_i \sqrt{X_k} dW_k \right) \left(\sum_{l=1}^{d} (\delta_{jl} - X_j \sqrt{X_l} dW_l) \right) dt \\
&= \sigma^2 \sum_{m=1}^{d} (\delta_{im} - X_i) \sqrt{X_m} (\delta_{jm} - X_j) \sqrt{X_m} dt \\
&= \sigma^2 \sum_{m=1}^{d} X_m (\delta_{im}\delta_{jm} + X_i X_j - X_i \delta_{jm} - X_j \delta_{im}) dt.
\end{aligned}$$

We study two cases separately:

- If $i = j$:

$$\begin{aligned}
c_{ii} &= \sigma^2 \sum_{m=1}^{d} X_m (\delta_{im} + X_i^2 - X_i \delta_{im} - X_i \delta_{im}) \\
&= \sigma^2 \sum_{m=1}^{d} X_m X_i^2 + X_m \delta_{im} (1 - 2X_i) \\
&= \sigma^2 \left(X_i^2 \sum_{m=1}^{d} X_m + X_i (1 - 2X_i) \right) \\
&= \sigma^2 (X_i - X_i^2) \\
&= \sigma^2 X_i (1 - X_i)
\end{aligned}$$

- If $i \neq j$:

$$\begin{aligned} c_{ii} &= \sigma^2 \sum_{m=1}^{d} X_m (X_i X_j - X_i \delta_{jm} - X_j \delta_{im}) \\ &= \sigma^2 \left(X_i X_j - X_i X_j - X_j X_i \right) \\ &= -\sigma^2 X_i X_j. \end{aligned}$$

Therefore, we obtain

$$c_{ij}(t) = \sigma^2 X_i(t)(\delta_{ij} - X_j(t)).$$

□

We are also interested in the ranked covariation matrix. Proposition 3.12 in [1], tells us about the dynamics of the ranked processes $X_{(k)}$. The dynamics are the same except for the appearance of two extra terms consisting of local time processes defined in [20]

$$\begin{aligned} dX_{(i)} =& \frac{\sigma^2}{2}(a_{r_i(t)} - \bar{a}_1 X_{(i)}(t))dt + \sigma \sum_{j=1}^{d} (\delta_{ij} - X_{(i)}(t)) \sqrt{X_{(j)}(t)} dW_j(t) \\ &+ \frac{1}{4} dL_{k-1,k} - \frac{1}{4} dL_{k,k+1}, \quad i = 1, \dots, d. \end{aligned}$$

This implies, in particular, that the ranked covariation matrix is given by

$$\kappa_{ij}(t) = \sigma^2 X_{(i)}(t)(\delta_{ij} - X_{(j)}(t)). \tag{4.3.2}$$

We now set $\sigma = 1$ for simplicity.

Proposition 4.3.2. *Assume that the market weights X_i are rank Jacobi processes. Then the vector $\rho = \ell_{()}$, where ℓ satisfies $c\ell = b$ and b is the drift X_i is given by*

$$\rho_k(t) = \frac{1}{2} \frac{a_k}{X_{(k)}(t)}, \quad k = 1, \dots, d. \tag{4.3.3}$$

Proof. The drift term of the market weights, when assuming they are rank Jacobi processes is given by

$$b_i(t) = \frac{1}{2}(a_{r_i(t)} - \bar{a}_1 X_i(t)).$$

Given Proposition 4.3.1, we can compute

$$(c(t)\ell(t))_i = X_i(t)\ell_i(t) - X_i(t) \sum_{j=1}^{d} X_j(t)\ell_j(t). \tag{4.3.4}$$

We want $b_i(t) = (c(t)l(t))_i$. So the equation to solve is

$$\frac{1}{2}(a_{r_i(t)} - \bar{a}_1 X_i(t)) = X_i(t)\ell_i(t) - X_i(t) \sum_{j=1}^{d} X_j(t)\ell_j(t). \tag{4.3.5}$$

If we choose

$$\ell_i(t) = \frac{1}{2}\frac{a_{r_i(t)}}{X_i(t)},$$

eq. (4.3.5) is satisfied. We finally note that

$$\rho_k(t) = \ell_{n_k(t)}(t) = \frac{1}{2}\frac{a_{r_{n_k(t)}(t)}}{X_{n_k(t)}(t)} = \frac{1}{2}\frac{a_k}{X_{(k)}(t)}.$$

□

Proposition 4.3.3. *Assume that the market weights X_i are rank Jacobi processes and that $X_{(N+1)} > 0$. Then the matrix κ^N is invertible and its inverse is given by*

$$(\kappa^N)^{-1} = diag((X^N_{()})^{-1}) + \bar{X}^{-1}_{(N+1)}\boldsymbol{1}_{N\times N}, \tag{4.3.6}$$

where $diag(v)$ represents the matrix whose diagonal terms are those of vector v and $\boldsymbol{1}_{N\times N}$ is the identity matrix of rank N.

Proof. We first note that $X_{(N+1)} > 0$ implies that $X_{(i)} > 0$ for $i = 1, \dots, N$ and $\bar{X}_{(N+1)} > 0$.

Thanks to eq. (4.3.2), we can write

$$\kappa^N = \mathrm{diag}(X^N_{()}) - (X^N_{()})^\top$$

and we define $D = \mathrm{diag}(X^N_{()})$, $u = -X^N_{()}$ and $v = X^N_{()}$ so that we can write $\kappa_N = D + uv^\top$.

We want to apply the Sherman-Morrison formula [21]. We first need to check that $1 + v^\top D^{-1}u \neq 0$ to prove the existence of the inverse. D is invertible due to the fact that $X_{(i)} > 0$ for $i = 1, \dots, N$.

$$1 + v^\top D^{-1}u = 1 - (X^N_{()})^\top D^{-1}X^N_{()} = 1 - \sum_{i=1}^{N} X_{(i)} = \bar{X}_{(N+1)} > 0.$$

Therefore, we can apply the Sherman-Morrison formula to compute the inverse of κ^N

$$\begin{aligned}(\kappa^N)^{-1} = (D + uv^\top)^{-1} &= D^{-1} - \frac{1}{1 + v^\top D^{-1}u} D^{-1}uv^\top D^{-1} \\ &= D^{-1} + \frac{1}{1 - (X^N_{()})^\top D^{-1}X^N_{()}} D^{-1}X^N_{()}(X^N_{()})^\top D^{-1} \\ &= \mathrm{diag}((X^N_{()})^{-1}) + (\bar{X}_{(N+1)})^{-1}D^{-1}X^N_{()}\mathbf{1}^\top_N \\ &= \mathrm{diag}((X^N_{()})^{-1}) + (\bar{X}_{(N+1)})^{-1}\mathbf{1}_{N\times N}\end{aligned}$$

□

4.3.2 Computation of Growth Optimal Strategy in the Pure Open Market

We now have all the ingredients to compute the growth optimal strategy in the pure open market when we assume that the market weights are to be rank Jacobi models. We will use the following notation throughout the calculations, which is just a combination of previous notation.

Notation 4.3.4. Let $x \in \mathbb{R}^d$ be a d-dimensional vector. For any $k \in \{1, \dots, d\}$ we denote the sum of the first k components of the vector x by

$$\bar{x}^k = \sum_{i=1}^{k} x_i.$$

Theorem 4.3.5. *Assume that the market weights X_i are rank Jacobi processes and let Assumption 4.1.3 and Assumption 3.3.3 be satisfied. Fix $N < d$ and assume $X_{(N+1)} > 0$. Then the optimal strategy in the pure open market is determined by Equation* (3.6.5) *with*

$$\hat{h}_p(t) = \rho^N + \frac{1}{\bar{X}^N_{()}} \left(1 - \frac{1}{2}\bar{a}^N_1\right) \mathbf{1}_N. \tag{4.3.7}$$

Proof. The proof builds upon the previous results in Section 4.3.1 and Theorem 4.2.1.

From Proposition 4.1.4 we have that

$$(\kappa^N)^{-1}(\kappa\rho)^N = \rho^N - \xi \mathbf{1}_N, \text{ where } \xi = (\bar{X}_{(N+1)})^{-1} \sum_{N+1}^{d} \rho_i X_{(i)} = \frac{1}{2} \frac{\bar{a}_{N+1}}{\bar{X}_{(N+1)}}$$

We can also compute

$$(\kappa^N)^{-1}(X^N_{()}) = \operatorname{diag}((X^N_{()})^{-1})(X^N_{()}) + (\bar{X}_{(N+1)})^{-1}(1 - \bar{X}_{(N+1)})\mathbf{1}_N = (\bar{X}_{(N+1)})^{-1}\mathbf{1}_N,$$

and thus

$$(X^N_{()})^\top (\kappa^N)^{-1}(X^N_{()}) = (\bar{X}_{(N+1)})^{-1}(1 - (\bar{X}_{(N+1)})) = (\bar{X}_{(N+1)})^{-1} - 1.$$

We also have that

$$\begin{aligned}
(X^N_{()})^\top ((\kappa^N)^{-1}(X^N_{()})) &= \sum_{i=1}^{d} \rho_i X_{(i)} - (\bar{X}_{(N+1)})^{-1} \sum_{i=N+1}^{d} \rho_i X_{(i)} + \sum_{i=N+1}^{d} \rho_i X_{(i)} \\
&= \frac{1}{2}\bar{a}_1 - (\bar{X}_{(N+1)})^{-1} \frac{1}{2} \bar{a}_{N+1} \\
&= \frac{1}{2}\left(\bar{a}_1 - (\bar{X}_{(N+1)})^{-1} \bar{a}_{N+1}\right),
\end{aligned}$$

and thus

$$\frac{(X_{()}^N)^\top((\kappa^N)^{-1}(X_{()}^N))-1}{(X_{()}^N)^\top(\kappa^N)^{-1}X_{()}^N} = \frac{\frac{1}{2}\left(\bar{a}_1-(\bar{X}_{(N+1)})^{-1}\bar{a}_{N+1}\right)-1}{(\bar{X}_{(N+1)})^{-1}-1}$$
$$= -\frac{1}{2}\frac{\bar{a}_{N+1}+(2-\bar{a}_1)}{1-(\bar{X}_{(N+1)})}$$
$$= -\frac{1}{2}\frac{\bar{a}_{N+1}(1-\bar{X}_{(N+1)}))+\bar{X}_{(N+1)}(2-\bar{a}_1^N)}{1-\bar{X}_{(N+1)}}.$$

We can finally compute

$$\begin{aligned}
\hat{h}_o &= (\kappa^N)^{-1}(\kappa\rho)^N - \frac{(X_{()}^N)^\top((\kappa^N)^{-1}(\kappa\rho)^N)-1}{(X_{()}^N)^\top(\kappa^N)^{-1}X_{()}^N}(\kappa^N)^{-1}X_{()}^N \\
&= (\kappa^N)^{-1}(\kappa\rho)^N + \frac{1}{2}\frac{\bar{a}_{N+1}(1-\bar{X}_{(N+1)}))+\bar{X}_{(N+1)}(2-\bar{a}_1^N)}{(1-\bar{X}_{(N+1)})\bar{X}_{(N+1)}}\mathbf{1}_N \\
&= \rho^N - \frac{1}{2}\frac{\bar{a}_{N+1}}{\bar{X}_{(N+1)}}\mathbf{1}_N + \frac{1}{2}\frac{\bar{a}_{N+1}}{\bar{X}_{(N+1)}}\mathbf{1}_N + \frac{1}{2}\frac{2-\bar{a}_1^N}{1-\bar{X}_{(N+1)}}\mathbf{1}_N \\
&= \rho^N + \frac{1}{\bar{X}_{()}^N}\left(1-\frac{1}{2}\bar{a}_1^N\right)\mathbf{1}_N.
\end{aligned}$$

□

This result implies that at time t, the growth optimal strategy in the pure open market invests the following proportion of our wealth in the k-th ranked asset

$$\pi_k(t) := X_{(k)}(t)(\hat{h}_p(t))_k = \frac{1}{2}a_k + \frac{X_{(k)}(t)}{\sum_{i=1}^N X_{(i)}}\left(1-\frac{1}{2}\sum_{i=1}^N a_i\right). \tag{4.3.8}$$

It is also interesting to observe that the resulting optimal strategy is determined by the vector ρ^N, that is, the ordered and truncated optimal strategy in the closed market and the market portfolio in the pure open market $\mathbf{1}_N$. One might be led to think that the reason for this is the strong condition we assumed on the form of the correlation κ, whose components have to be of the form $\kappa_{kl]} = -f_k f_l g$. However, this happens not to be the case. It is a feature of the rank Jacobi model.

Proposition 4.3.6. *Fix $t > 0$. Let Assumption 4.1.3 be satisfied. It is not always possible to write $(\kappa^N)^{-1}X_{()}^N$ in the form*

$$(\kappa^N)^{-1}X_{()}^N = a\rho^N + b\mathbf{1}_N \text{ for some } a, b \in \mathbb{R}. \tag{4.3.9}$$

Proof. We only need to find a valid counterexample to prove it. We fix $N = 2$ and $d = 3$. We need to have $\kappa_{kl} = -f_k f_l g$ for $l = 3$ and $k = 1, 2$. Also we must have $f_3 = 1$ by assumption. Finally, we must have $\kappa\mathbf{1}_3 = 0$ and $b^\top\mathbf{1}_3 = 0$. Thus κ and $b_{()}$ must be of the form

$$\kappa = \begin{bmatrix} f_1 g - \kappa_{12} & \kappa_{12} & -f_1 g \\ \kappa_{12} & f_2 g - \kappa_{12} & -f_2 g \\ -f_1 g & f_2 g & g(f_1+f_2) \end{bmatrix}, \quad b_{()} = \begin{bmatrix} b_1 \\ b_2 \\ -b_1-b_2. \end{bmatrix},$$

We want to find $\rho = (\rho_1, \rho_2, \rho_3)^\top$ such that $\kappa\rho = b_{()}$. We set $g = 1$, $\kappa_{12} = 0$, $f_1 = f_2 = 1$, $b_1 = b_2 = \frac{1}{4}$. This parameters imply that the vector $\rho = (\frac{1}{4}, \frac{1}{4}, 0)^\top$.

If we have $X_{(1)} = \frac{1}{2}$ and $X_{(2)} = \frac{1}{4}$ the equation to solve is

$$\begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix} = a \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1/4 \\ 1/4 \end{bmatrix},$$

which has no solution for any $a, b \in \mathbb{R}$. Thus we have found the counterexample

$$\kappa = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \quad b_{()} = \begin{bmatrix} 1/4 \\ 1/4 \\ 1/2 \end{bmatrix}, \quad X_{()} = \begin{bmatrix} 1/2 \\ 1/4 \\ 1/4 \end{bmatrix}.$$

□

4.3.3 Comparison Between Pure and Mixed Open Market Growth Optimal Strategies

We now turn out attention to the comparison between the optimal strategy in the open and mixed open market settings. The corresponding result regarding the optimal growth strategy in the mixed open market is contained in the following theorem from [1].

Theorem 4.3.7 (Theorem 4.1 in [1] with $\gamma = 0$)**.** *Assume that the market weights X_i are rank Jacobi processes. There exists a growth optimal strategy in the mixed open market if and only if $\bar{a}_k \geq 1$ for $k = 2, \ldots, N+1$. In this case, the optimal strategy is given by*

$$\hat{\theta}_{n_k(t)}(t) = \begin{cases} 1 - \frac{1}{2}\bar{a}_1 + \frac{a_k}{2X_{(k)}(t)} & k = 1, \ldots, N, \\ 1 - \frac{1}{2}\bar{a}_1 + \frac{\bar{a}_{N+1}}{2X_{(N+1)}(t)} & k = N+1, \ldots, d, \end{cases} \tag{4.3.10}$$

and it is unique up to $\mathbb{P}_\mu \otimes dt$-nullsets for every $P_\mu \in \mathcal{P}_0$.

To write it in a similar form, the growth optimal strategy obtained in the pure open market is given by

$$\hat{\theta}_{n_k(t)}(t) = \begin{cases} \frac{1}{X_{()}^N} \left(1 - \frac{1}{2}\bar{a}_1^N\right) + \frac{a_k}{2X_{(k)}(t)} & k = 1, \ldots, N, \\ 0 & k = N+1, \ldots, d. \end{cases} \tag{4.3.11}$$

We can see that both strategies seem very similar. The main difference is that the pure open market (by definition) does not invest in the smallest capitalization assets. All this investment is reassigned to the top capitalization stocks by a positive normalization factor $\frac{1}{X_{()}^N}$ on the first term while also summing only the first N components of the vector a, instead of all of them.

This last fact will be important in subsequent chapters, as it means that for implementing the growth optimal strategy in the pure open market we only need to estimate the truncated vector a^N instead of the full vector a, which is coherent with the fact that we should not need information on the $d - N$ last assets if we are not investing in them. However, this fact will also affect drastically the behaviour of our strategy, as we will discuss in the backtesting chapter.

4.4 Asymptotic Growth Rate of the Growth Optimal Portfolio

One of the main motivations of the open market setting and the rank Jacobi processes is to maximize the growth rate of our initial wealth under long time horizons. The Asymptotic Growth Rate is a crucial metric in our study, representing the rate at which a portfolio's value expands over an extended period. It measures the exponential growth rate of the geometric mean return, capturing the long-term expected return of an investment. In the context of finance and investment strategies, the focus is often on sustainable, long-term portfolio growth, making this metric particularly relevant.

In our investigation of the open market setting and rank Jacobi processes, the Asymptotic Growth Rate serves as a benchmark for evaluating the long-term performance of different investment strategies. The objective is to optimize our strategies to achieve the highest possible Asymptotic Growth Rate, as even minor differences in growth rates can lead to substantial disparities in final wealth over time. Hence, this metric plays a pivotal role in guiding our choice of the optimal investment strategy that ensures maximum and sustainable growth.

We now define the asymptotic growth rate following [1].

Definition 4.4.1 (Asymptotic Growth Rate)**.** Given a wealth process V^θ associated to a given strategy θ, we define (following [1]) the asymptotic growth rate as

$$g(V^\theta) = \sup\left\{\alpha \in \mathbb{R} : \liminf_{T\to\infty} \frac{1}{T}\log V^\theta(T) \geq \alpha \text{ almost surely}\right\}. \tag{4.4.1}$$

Our objective now is to compute the asymptotic growth rate of the growth optimal strategy in the pure open market. In order to achieve this, we start by assuming the rank Jacobi model for the weights. We now write $Y_k = X_{(k)}$ and thus $Y = X_{()}$. We start by computing the drift of the log wealth in order to later compute the asymptotic growth rate.

Proposition 4.4.2. *Assume the rank Jacobi model for the market weights $Y_k = X_{(k)}$. Then the drift of the log wealth process generated by the growth optimal strategy in the pure open market $\hat{\theta}$ is given by*

$$drift(\log V^{\hat{\theta}}) = \frac{1}{2}(1-\bar{a}_1) + \frac{1}{8}\sum_{k=1}^{N}\frac{a_k^2}{Y_k} - \frac{1}{8}\frac{1}{\bar{Y}^N}(\bar{a}_1^N - 2)^2. \tag{4.4.2}$$

Proof. We define the function $F(Y_1,\dots,Y_N) = \hat{\theta}_{\mathbf{n}}(Y_1,\dots,Y_N)$ so that

$$F_k(Y_1,\dots,Y_N) = \frac{1}{2}\left(\frac{a_k}{Y_k} + \frac{1}{\bar{Y}^N}(2-\bar{a}_1^N)\right), \quad k = 1,\dots,N.$$

Thus, the drift of the log-wealth process, as computed in Proposition 3.5.1 but now ordered by rank, is given by

$$drift(\log V^{\hat{\theta}}) = F(Y)^\top \kappa(Y)\rho(Y) - \frac{1}{2}F(Y)^\top\kappa(Y)F(Y).$$

We can now compute the first term:

$$
\begin{aligned}
(\kappa(Y)\rho(Y))_k &= \frac{1}{2}\sum_{j=1}^{d} Y_k(\delta_{kj} - Y_j)\frac{a_j}{Y_j} = \frac{1}{2}(a_k - Y_k\bar{a}_1), \quad k = 1, \dots, N, \\
F(Y)^\top(\kappa(Y)\rho(Y)) &= \frac{1}{4}\sum_{k=1}^{N}\left(\frac{a_k}{Y_k} + \frac{1}{\bar{Y}^N}(2 - \bar{a}_1^N)\right)(a_k - Y_k\bar{a}_1) \\
&= \frac{1}{4}\left(\sum_{k=1}^{N}\frac{a_k^2}{Y_k} - 2\bar{a}_1 + \frac{1}{\bar{Y}^N}(2\bar{a}_1^N - (\bar{a}_1^N)^2\right).
\end{aligned}
$$

We can also compute the second term

$$
\begin{aligned}
(\kappa(Y)F(Y))_k &= \frac{1}{2}\sum_{j=1}^{N} Y_k(\delta_{kj} - Y_j)\left(\frac{a_j}{Y_j} + \frac{1}{\bar{Y}_1^N}(2 - \bar{a}_1^N)\right) \\
&= \frac{1}{2}Y_k\left(\frac{a_k}{Y_k} - \bar{a}_1^N + \frac{1}{\bar{Y}_1^N}(2 - \bar{a}_1^N)(1 - \bar{Y}_1^N)\right) \\
&= \frac{1}{2}Y_k\left(\frac{a_k}{Y_k} - 2 + \frac{1}{\bar{Y}_1^N}(2 - \bar{a}_1^N)\right), \quad k = 1, \dots, N, \\
F(Y)^\top\kappa(Y)F(Y) &= \frac{1}{4}\sum_{k=1}^{N} Y_k\left(\frac{a_k}{Y_k} - 2 + \frac{1}{\bar{Y}_1^N}(2 - \bar{a}_1^N)\right)\left(\frac{a_k}{Y_k} + \frac{1}{\bar{Y}_1^N}(2 - \bar{a}_1^N)\right) \\
&= \frac{1}{4}\left(\sum_{k=1}^{N}\frac{a_k^2}{Y_k} + \frac{1}{\bar{Y}_1^N}\bar{a}_1^N(2 - \bar{a}_1^N) + \frac{1}{\bar{Y}_1^N}(4 - 2\bar{a}_1^N) - 4\right) \\
&= -\frac{1}{2}\left(-\frac{1}{2}\sum_{k=1}^{N}\frac{a_k^2}{Y_k} - 2\left(\frac{1}{\bar{Y}_1^N} - 1\right) + \frac{1}{2}\frac{1}{\bar{Y}_1^N}(\bar{a}_1^N)^2\right).
\end{aligned}
$$

We can finally write

$$
\begin{aligned}
drift(\log V^{\hat{\theta}}) &= \frac{1}{8}\sum_{k=1}^{N}\frac{a_k^2}{Y_k} - \frac{1}{2}\bar{a}_1 - \frac{1}{4}\frac{1}{\bar{Y}_1^N}((\bar{a}_1^N)^2 - 2\bar{a}_1^N + 2 - \frac{1}{2}(\bar{a}_1^N)^2) + \frac{1}{2} \\
&= \frac{1}{2}(1 - \bar{a}_1) + \frac{1}{8}\sum_{k=1}^{N}\frac{a_k^2}{Y_k} - \frac{1}{8}\frac{1}{\bar{Y}^N}(\bar{a}_1^N - 2)^2.
\end{aligned}
$$

□

Definition 4.4.3 (Standard and Ordered Simplex)**.** We define the standard simplex in $\mathbb{R}^d$ and the ordered simplex in $\mathbb{R}^d$ respectively as

$$
\Delta^{d-1} = \{x \in [0,\infty)^d : x_1 + \cdots + x_d = 1\}, \quad \nabla^{d-1} = \{y \in \Delta^{d-1} : y_1 \geq \cdots \geq y_d\}.
$$

We also define

$$
\Delta_+^{d-1} = \{x \in \Delta^{d-1} : x_d > 0\}, \quad \nabla_+^{d-1} = \{y \in \nabla^{d-1} : y_d > 0\}.
$$

By Remark 3.13 in [1], we know that the density function of the ranked market weights is given by

$$q(y) = Q_a^{-1} \prod_{k=1}^{d} y_k^{a_{k-1}}, \quad y \in \nabla^{d-1}, \tag{4.4.3}$$

where Q_a is a normalizing constant.

We can thus define the probability measure β by

$$d\beta(y) = q(y)dy. \tag{4.4.4}$$

We are now prepared to compute the asymptotic growth rate of the growth optimal strategy in the pure open market setting.

Theorem 4.4.4 (Asymptotic Growth Rate Pure Open Market Growth Optimal Strategy)**.** *Assume the rank Jacobi model for the market weights $Y_k = X_{(k)}$. Then the asymptotic growth rate of the growth optimal strategy in the pure open market $\hat{\theta}_p$ is given by*

$$\hat{\lambda}_p = \frac{1}{2}(1-\bar{a}_1) + \frac{1}{8}\int_{\nabla^{d-1}} \left(\sum_{k=1}^{N} \frac{a_k^2}{Y_k} - \frac{1}{\bar{Y}_1^N}(\bar{a}_1^N - 2)^2 \right) q(Y)dY. \tag{4.4.5}$$

Proof. Thanks to Proposition 4.4.2, we can define

$$f(Y_1,\ldots,Y_N) := drift(\log V^{\hat{\theta}}) = \frac{1}{2}(1-\bar{a}_1) + \frac{1}{8}\sum_{k=1}^{N}\frac{a_k^2}{Y_k} - \frac{1}{8}\frac{1}{\bar{Y}^N}(\bar{a}_1^N - 2)^2.$$

By Proposition 3.5.1, reorganizing the terms by rank and assuming $V_0 = 1$, we can write

$$\begin{aligned} \log V_T^{\hat{\theta}} &= \int_0^T \left(\hat{\theta}_{\mathbf{n}}(t)^\top \kappa(t)\rho(t) - \frac{1}{2}\hat{\theta}_{\mathbf{n}}(t)^\top \kappa(t)\hat{\theta}_{\mathbf{n}}(t) \right) dt + \int_0^T \hat{\theta}_{\mathbf{n}}(t)\sigma_{()}(t)dW \\ &= \int_0^T f(Y_1,\ldots,Y_N)dt + \int_0^T \hat{\theta}_{\mathbf{n}}(t)\sigma_{()}(t)dW. \end{aligned}$$

Thanks to the rank ergodic property of Corrollary 3.10 in [1] we can write

$$\begin{aligned} \hat{\lambda}_p &= \sup\left\{ \alpha \in \mathbb{R} : \liminf_{T\to\infty} \frac{1}{T}\log V^{\theta}(T) \geq \alpha \text{ almost surely} \right\} \\ &= \lim_{T\to\infty}\frac{1}{T}\log V^{\theta}(T) \\ &= \lim_{T\to\infty}\frac{1}{T}\int_0^T f(Y_1,\ldots,Y_N)dt + \lim_{T\to\infty}\frac{1}{T}\int_0^T \hat{\theta}_{\mathbf{n}}(t)\sigma_{()}(t)dW \\ &= \int_{\nabla^{d-1}} f(Y_1,\ldots,Y_N)q(Y)dY \\ &= \int_{\nabla^{d-1}} \left(\frac{1}{2}(1-\bar{a}_1) + \frac{1}{8}\sum_{k=1}^{N}\frac{a_k^2}{Y_k} - \frac{1}{8}\frac{1}{\bar{Y}_1^N}(\bar{a}_1^N - 2)^2 \right) q(Y)dY \\ &= \frac{1}{2}(1-\bar{a}_1) + \frac{1}{8}\int_{\nabla^{d-1}} \left(\sum_{k=1}^{N}\frac{a_k^2}{Y_k} - \frac{1}{\bar{Y}_1^N}(\bar{a}_1^N - 2)^2 \right) q(Y)dY. \end{aligned}$$

□

On the other hand, the asymptotic growth rate of the growth optimal strategy in the mixed open market can be found in Theorem 5.4 of [1], and it is given by

$$\hat{\lambda}_m = -\frac{1}{8}(\bar{a}_1)^2 + \frac{1}{8}\int_{\nabla^{d-1}} \left(\sum_{k=1}^{N} \frac{a_k^2}{Y_k} - \frac{(\bar{a}_{N+1})^2}{\bar{Y}_{N+1}} \right) q(Y)dY. \tag{4.4.6}$$

The asymptotic growth rates of the growth optimal strategies are difficult to compare analytically. So the hope is to compare them numerically in the subsequent Backtesting chapter.

4.5 Functional Generation of the Growth Optimal Portfolio

This section introduces the concept of functionally generated portfolios. In a similar manner to [1], who is able to demonstrate that the growth optimal strategy in the mixed open market is functionally generated, we obtain the same result for the case of the pure open market setting.

Functionally generated portfolios were first used by Fernholz in [22]. One of the key features of functionally generated portfolios is that we can find out their generated wealth process without computing a stochastic integral. Following [1] we define functionally generated strategies as follows.

Definition 4.5.1 (Functionally Generated Strategy)**.** Let θ be a strategy, let $G : \mathbb{R}^d \to (0, \infty)$ be a function that is continuous on Δ_+^{d-1} and such that $G(X)$ is a semimartingale. We say that θ is a functionally generated strategy with generating function G and drift process Γ if we have the representation

$$\log V^\theta(T) = \log G(X(T)) + \Gamma(T) \tag{4.5.1}$$

for some finite variation process Γ with $\Gamma(0) = 0$. We will denote the strategy θ as θ^G in this case.

We now show that the growth optimal strategy in the pure open market is functionally generated.

Proposition 4.5.2. *Assume the rank Jacobi model for the market weights $Y_k = X_{(k)}$. Then the growth optimal strategy in the pure open market $\hat{\theta}$ is functionally generated by*

$$G_o(Y) = \left(\prod_{k=1}^{N} Y_k^{\frac{1}{2}a_k} \right) \left(\sum_{k=1}^{N} Y_k \right)^{1-\frac{1}{2}\bar{a}_1^N}. \tag{4.5.2}$$

Proof. From the proof of Proposition 3.5.1 we have that

$$\log V_T^\theta = \int_0^T \theta(t)^\top dX(t) - \frac{1}{2}\int_0^T \theta(t)^\top c\theta(t)dt$$

Let us denote $u = \log G_0(X)$ where $G_0(X)$ is the generator function we are looking for. If we assume that we can write $du(X) = \nabla u(X)dX$ we get the following representation

$$\log V_T^\theta = u(X) + \Gamma(T) = \int_0^T du(X) + \Gamma(T) = \int_0^T \nabla u(X)dX + \Gamma(T).$$

Thus, we only need to find $u(X)$ such that $\nabla u(X) = \hat{\theta}(X)$. If we choose u to be

$$u(Y) = \frac{1}{2}\left(\sum_{k=1}^N a_k \log Y_k + (2 - \bar{a}_1^N) \log \sum_{k=1}^N Y_k\right),$$

we have that

$$\frac{\partial u(Y)}{\partial Y_i} = \frac{1}{2}\left(\frac{a_i}{Y_i} + (2 - \bar{a}_1^N)\frac{1}{\bar{Y}_1^N}\right) \Rightarrow \nabla u(X_{()}) = \hat{\theta}(X_{()}).$$

Therefore, our pure open market growth optimal strategy is functionally generated by

$$G_o(Y) = e^{u(Y)} = \left(\prod_{k=1}^N Y_k^{\frac{1}{2}a_k}\right)\left(\sum_{k=1}^N Y_k\right)^{1-\frac{1}{2}\bar{a}_1^N}.$$

□

In the mixed open market setting, the growth optimal strategy is functionally generated by

$$G_m(Y) = \left(\prod_{k=1}^N Y_k^{\frac{1}{2}a_k}\right)\left(\bar{Y}_{N+1}\right)^{\frac{1}{2}\bar{a}_{N+1}}, \tag{4.5.3}$$

which very closely resembles the obtained formula in the pure open market.

Chapter 5

Backtesting

This chapter aims to practically apply and analyze the theoretical results we have garnered so far. Specifically, we will compare the performance of two growth optimal strategies within the open market: the pure and mixed open market strategies. Our intention is to gain a nuanced understanding of the investment philosophies driving these theoretical strategies, and the distinctions between them.

To evaluate our investment strategy, we opted to conduct backtesting using historical financial data. Backtesting is a method where we simulate how our investment approach would have fared if applied in past market conditions.

Our data will be sourced from the Wharton Research Data Service (WRDS), with a specific focus on the databases of the Center for Research in Security Prices (CRSP). The backtesting procedure commences with data cleaning and the selection of the appropriate investment universe. Subsequently, we have adopted the rank Jacobi model for market weights, which facilitates the implementation of our strategies. This step allows us to estimate all necessary parameters, including the variable a.

Preliminary backtesting revealed high leverage in our strategies, prompting us to explore methods to mitigate this. Final tests encompass a series of backtests using both monthly and daily data, varying start and end periods, and a multitude of different model hyperparameters. These include various estimations of the parameter a, leverage scale factors, and different values for N, among other variables.

5.1 Backtesting Implementation

In this section, we detail the methodological journey of the backtesting implementation, which was made possible through a custom software that I developed specifically for this research. The source code for this software is freely available at `https://github.com/pedrou2000/mathematical_finance` for review and reuse. Starting with the process of database selection and data cleaning, we move on to deliberations regarding the choice of our investment universe and methodologies for estimating the parameter a in the rank Jacobi model. This will pave the way to the core of our backtesting execution and supplementary strategies aimed at reducing leverage. Each step in this process enhances our comprehension of the study's outcomes and reinforces the robustness of our findings.

Database Selection and Cleaning

Embarking on the backtesting process, the initial step was data selection and cleaning. Our choice of the database was the Wharton Research Data Service (WRDS), with a particular focus on the Center for Research in Security Prices (CRSP).

CRSP, a research centre affiliated with the University of Chicago's Booth School of Business, provides a comprehensive collection of security price, return, and volume data for the NYSE, AMEX and NASDAQ stock markets. Recognized for its robustness and accuracy, the CRSP database is widely used in academic, commercial, and government analyses that require precise, reliable historical stock market data.

Choosing CRSP was a decision driven by the extensive, high-quality data that it offers. It comprises a broad range of information about stocks, such as their prices, trading volumes, dividends, and other related data. The meticulous data cleaning and validation processes that CRSP employs ensure the reliability of their information. The cleaning process was further facilitated by a useful code repository by Ruf [23]. This tutorial aided our understanding of the database structure and alerted us to potential issues and limitations of the raw database.

The requisite data for backtesting our strategy encompasses the returns of each stock and the market capitalization of the assets. CRSP offers two data varieties, monthly and daily. Our preliminary experiments employed monthly data, though daily data were also scrutinized. We adhered to Ruf's [23] guidelines in notebooks 5 and 6 to execute data cleaning. The key cleaning phases for the returns dataframe incorporated:

- **Preliminary Cleaning Steps**: Initial tasks involved flagging problematic returns and dealing directly with intermediate, missing, and problematic returns.

- **Pivoting the Data**: We orchestrated our final dataframe to present dates as rows and different assets or stocks as columns.

- **Cleaning the Time Series Beginnings and Ends**: In line with real-world investment preferences for well-traded securities, we purged time series beginnings without observed valid returns. Similarly, missing returns at the ends of series were removed, with missing delisting returns set to a predetermined value, specifically -30%.

- **Flagging Extreme Returns**: Returns exceeding 100% or falling below -50% were marked for caution.

Upon executing these data-cleaning steps, the refined dataframes were stored for subsequent usage. The primary dataframes to be used include the market capitalizations and returns for each time step, which enable us to derive the market weights, estimate the parameter a, and calculate the weights of our investment strategies.

Investment Universe

In the context of your project, the investment universe refers to the selected range of stocks considered for implementing and testing the growth optimal strategies.

A visualization of the number of stocks over the years indicates that, since 1965, at least 2000 stocks have been tracked, and from 1975 onward, the number increases to a minimum of 3000. Consequently, when configuring an investment universe comprising $d = 3000$ or $d = 2000$ stocks, the backtesting process must initiate in 1975 or 1965 respectively.

In contrast, the daily data, owing to computational constraints, could only accommodate data cleaning from 2013 to 2018. Hence, backtesting for daily data, with $d = 3000$, was restricted to this five-year window.

Our preliminary strategy for constructing the investment universe involved selecting the top d stocks at the outset of 1975 and maintaining these stocks unchanged for the analysis. This strategy, was not without limitations - by the end of the study period in 2023, only around 300 stocks had survived. This simplicity, however, was instrumental in familiarizing us with the dataframes and backtesting procedure and paved the way for the progressive adoption of more complex investment universes.

For the final backtesting stages, we opted for a dynamic investment universe, one that could see its composition change at any given time period. For a given date, this universe incorporated only the top d stocks, ranked by market capitalization at that date. The range of strategies that could be successfully employed with this approach was somewhat limited, as it precluded the possibility of maintaining a long position in a stock over an extended time frame. If at any point a stock's capitalization dropped out of the top d, tracking the returns it generated became unfeasible. Nonetheless, this approach was a good fit for our growth optimal investment strategies which are less concerned with the specific identities of stocks and focused instead on their respective ranks. Here, the rank of a stock corresponds to its position when stocks are arranged in descending order by market capitalization.

Estimating parameter a

To implement and backtest the growth optimal strategies described earlier, we embraced the rank Jacobi model for the market weights. Through this model, we needed only to estimate the vector parameter $a = (a_1, \ldots, a_d)$ of the rank Jacobi model, simplifying the application of our growth optimal strategies.

However, estimating the parameters $a_1, \ldots, a_d$ turned out to be a complex task. To address this, we examined various methods, including the First Moment Matching approach, as outlined in Corollary 9 of [24]. Additionally, we utilized a method introduced in [25], which estimates a by the following equations:

$$\begin{aligned} \bar{a}_1 &= 0 \\ \bar{a}_k &= (\mathbb{E}[\log X_{(k-1)} - \log X_{(k)}])^{-1}, \quad k = 2, \dots, d. \end{aligned} \tag{5.1.1}$$

To compute these expectations, we first established an initial estimation period of 20 years, approximating the expectation by averaging values over this period. The outcomes derived from both methods exhibited strong similarity, leading us to select the latter approach for the final estimation of parameter a. A shorter time period for the estimation (5 years instead of 20) yielded very similar results. Once we had the vector $\bar{a}$, computing the vector a became straightforward.

The preliminary estimations of the vector $\bar{a}$ were significantly noisy. Consequently, we implemented various strategies to smooth out the estimated curve, separating the intrinsic signal from the noise. These strategies encompassed the use of moving averages, fitting polynomials of differing degrees, and fitting an exponential decay curve. The results of these approaches are depicted in Figure 5.1. The black dots represent noisy estimations directly obtained from equation 5.1.1, while the colored curves depict smoothed estimations. The x-axis represents the index of the array $\bar{a} = (\bar{a}_1, \dots, \bar{a}_d)$.

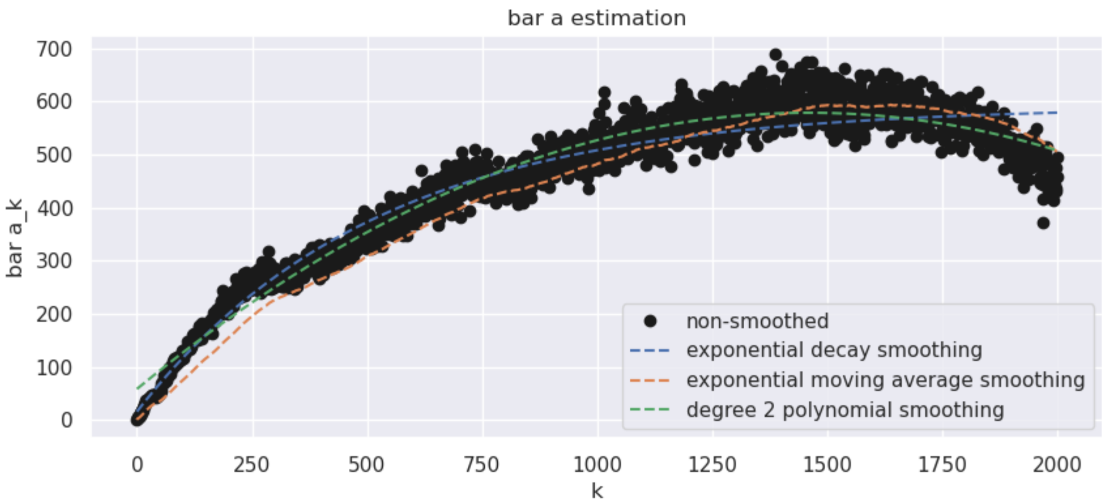

Figure 5.1: Estimation of $\bar{a}$ parameter.

On scrutinizing these estimations, we observed that the apparent decay for the last values of $\bar{a}$ is likely noise. This conclusion is supported by the displacement of the pattern around $k = 1500$ as d increases. For our final estimations, we opted for the exponential decay smoothing, which does not demonstrate this decay, as shown in Figure 5.1. Furthermore, the results from final backtesting showed no substantial impact from these minor intermediate decisions.

As illustrated in Figure 5.2, we successfully eliminated noise in the estimation of a. It is important to note that the value of a_d is significantly larger compared to the other values in the vector a, as $a_k = \bar{a}_k - \bar{a}_{k+1}$ for $k = 1, \dots, d-1$ and $\bar{a}_d = a_d$. In our analysis, $a_d \approx 600$, while the other values of a, denoted as $a^{d-1} = (a_1, \dots, a_{d-1})$, are

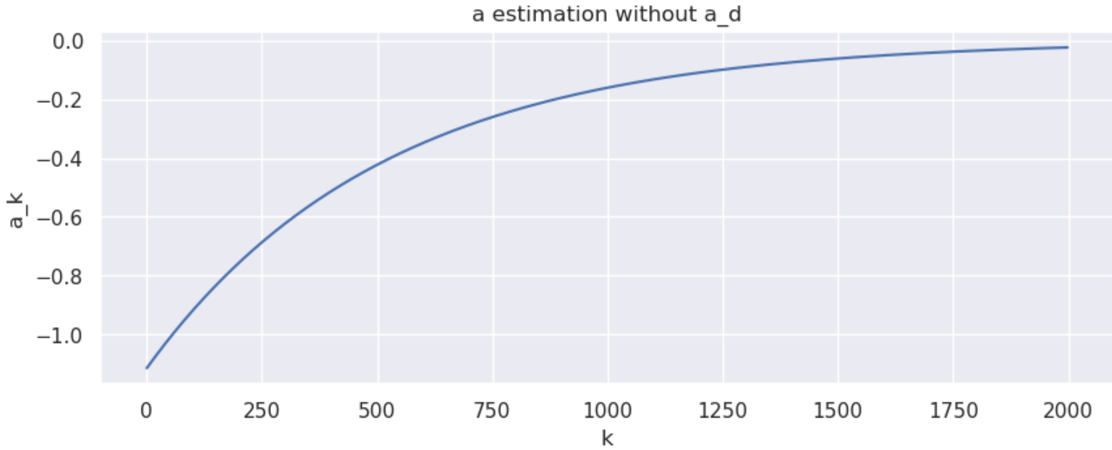

Figure 5.2: Estimation of the parameter a without including the last component a_d.

much smaller and negative, all in the interval $[-2, 0]$, as shown in Figure 5.2. Thus we had to eliminate this last component a_d from the final plot. The fact that a_d is so big will play a very important role in our final backtesting. It has been argued that a_d encapsulates the growth rate not only of the d-th ranked asset but also all the other assets which are left out from the study. These very small capitalization shares are known to have a bigger propensity to become larger [26], and thus the extraordinarily big value of a_d.

Backtesting of the Strategies

For final backtesting, we calculated the investment weights of the growth optimal strategies in both the pure and mixed open market scenarios. We applied the results derived from our previous study to achieve this.

In the context of the pure open market growth optimal strategy, we allocated our wealth in each time period to the k-th ranked stock according to the following proportions

$$\pi_k^p(t) = \hat{\theta}^p_{n_k(t)}(t)(X_{(k)}(t)) = \begin{cases} \frac{1}{2}\left(a_k + \frac{X_{(k)}(t)}{\bar{X}^N_{(1)}(t)}(2 - \bar{a}_1^N)\right), & k = 1, \dots, N, \\ 0, & k = N+1, \dots, d. \end{cases} \tag{5.1.2}$$

For the mixed open market growth optimal strategy, our wealth allocation to the k-th ranked stock in each time period was defined as follows

$$\pi_k^m(t) = \hat{\theta}^m_{n_k(t)}(t)(X_{(k)}(t)) = \begin{cases} \frac{1}{2}\left(a_k + X_{(k)}(t)(2 - \bar{a}_1)\right), & k = 1, \dots, N, \\ \frac{1}{2}X_{(k)}(t)\left(\frac{\bar{a}_{N+1}}{\bar{X}_{(N+1)}(t)} + (2 - \bar{a}_1)\right), & k = N+1, \dots, d. \end{cases} \tag{5.1.3}$$

It's crucial to understand that these proportions may be negative or exceed 1.

Thus, even if the quantities $\pi_k(t) = \hat{\theta}_{n_k(t)}(t)(X_{(k)}(t))$ satisfy

$$\sum_{k=1}^{d} \pi_k(t) = \sum_{k=1}^{d} \hat{\theta}_{n_k(t)}(t)(X_{(k)}(t)) = 1$$

in both scenarios, this does not imply that $\hat{\theta}_{n_k(t)}(t)(X_{(k)}(t))$ lies in the interval $[0, 1]$. Indeed, in many cases, they do not: these strategies are highly leveraged.

We rebalanced our entire portfolio at each time step to maintain these proportions precisely. Our strategies focus solely on stock ranks, making the dynamic investment universe ideally suited for our needs, even though it might not be ideal for strategies that depend on stock names.

To conduct the final backtesting, we independently computed the returns of our optimal strategies for each time period. We calculated the returns for each period by multiplying the stock returns for that period by the proportion $\hat{\theta}_{n_k(t)}(t)(X_{(k)}(t))$ invested in that stock, and then adding all these returns. Starting with an initial wealth of 1, we computed the cumulative product of these returns to obtain the final cumulative return over the period. Simultaneously, we also computed other relevant statistics, such as the Sharpe ratio and the maximum drawdown. However, the main limitation of this approach is that if the losses at a given time period are so substantial as to result in a negative balance, the final cumulative returns become meaningless.

We introduced an alternative backtesting method to address this issue. Instead of focusing on returns for each time period, we directly tracked the logarithm of wealth using an intermediate derivation found in the proof of Proposition 3.5.1. Mathematically, this is represented as

$$\log V(t) = \pi_k(t)^\top \frac{dX_{()}(t)}{X_{()}(t)} - \frac{1}{2}\pi_k(t)^\top d[X_{()}, X_{()}](t)\pi_k(t) \tag{5.1.4}$$

which, in numerical terms, translates to

$$\log V(T) = \sum_{t=0}^{T} \left(\pi_k(t-1)^\top \frac{X_{()}(t) - X_{()}(t-1)}{X_{()}(t-1)} - \frac{1}{2}\pi_k(t-1)^\top d[X_{()}, X_{()}](t)\pi_k(t-1) \right) \tag{5.1.5}$$

where $d[X_{(i)}, X_{(j)}](t) = (X_{(i)}(t) - X_{(i)}(t-1))(X_{(j)}(t) - X_{(j)}(t-1))$.

Leverage Reduction

Our initial results demonstrated high degrees of leverage, which rendered backtesting ineffective as negative wealth amounts were almost invariably achieved due to end-of-period returns dipping below -100%. As evident in Figure 5.3, the original growth optimal strategy shorts the top 500 stocks over 150 times its original wealth. To address this issue, we devised a straightforward approach to minimize the leverage of the mixed open market growth optimal strategy. This strategy retained the form of the original approach, producing valid backtesting results.

The key realization for reducing the leverage of the mixed open market growth optimal strategy was that, given the specific form of the parameter a, the strategy invariably shorts the top N stocks to purchase the market portfolio $\mathbf{1}_d$. This phenomenon is illustrated in Figure 5.3, where the blue line represents the original mixed open market growth optimal strategy.

Figure 5.3: Weights (π_k^p) of the mixed open market growth optimal strategy with $N = 1000$ and $d = 2000$

Our proposed leverage reduction strategy involves rescaling the weights of the top stocks by a factor $\alpha \in [0, 1]$. This results in a new strategy $\hat{\theta}'$ such that $\hat{\theta}'_{(k)}(t) = \alpha\hat{\theta}_{(k)}(t)$ for all $k = 1, \dots, N$. For the new strategy $\hat{\theta}'$ to remain a valid mixed open market strategy, we adjusted the latter components of the strategy according to

$$\hat{\theta}'_{(k)}(t) = \alpha\hat{\theta}(k)(t) + \frac{1-\alpha}{\bar{X}(N+1)} \text{ for } k = N+1, \dots, d.$$

We call α the leverage scaler parameter. Figure 5.3 illustrates the outcome of utilizing various α values. It is crucial to note that the shape of the strategy remains consistent across all leverage-scaled strategies.

The adjustment in leverage facilitated more meaningful backtesting outcomes, which we will examine in the subsequent section.

Asymptotic Growth Rate Estimation

We also tried to numerically estimate the asymptotic growth rates of both the pure and mixed open market growth optimal strategies. To this end, we used some basic Montecarlo estimations of the complicated integrals which appear in the equations 4.4.5 and 4.4.6.

Thanks to Theorem 4.4.4 we have that

$$\hat{\lambda}_o = \frac{1}{2}(1-\bar{a}_1) + \frac{1}{8}\sum_{k=1}^{N} a_k^2 \int_{\nabla^{d-1}} \frac{1}{y_k} q(y)dy - \frac{1}{8}(\bar{a}_1^N - 2)^2 \int_{\nabla^{d-1}} \frac{1}{\bar{y}_1^N} q(y)dy, \quad (5.1.6)$$

so we need to estimate

$$\int_{\nabla^{d-1}} \frac{1}{y_k} q(y) dy \quad \text{and} \quad \int_{\nabla^{d-1}} \frac{1}{\bar{y}_1^N} q(y) dy.$$

Thanks to Proposition 4 in [24], we know that if we have $\bar{a}_1 = 0$ we have that the log gaps $Z_k = \log Y_{k-1} - \log Y_k$ are independent and satisfy

$$Z_k \sim \exp(\bar{a}_k), \quad \forall k \in \{2, \ldots, d\}.$$

If we define $E_1 := 1$ and $E_k := \exp(-(Z_2 + \ldots Z_k)) \quad \forall k \in \{2, \ldots, d\}$ a simple calculation shows that we can write

$$Y_d = \frac{E_d}{\sum_{i=1}^d E_i}, \quad Y_k = \frac{E_k}{\sum_{i=1}^k E_i}(1 - \bar{Y}_{k+1}), \quad \forall k \in \{2, \ldots, d\}$$

From these relations, we can now sample Z_k, which allows us to sample E_k which thus allows us to sample Y_k. This way, we were able to compute an estimation of the integrals which appear in equations 4.4.5 and 4.4.6 using Montecarlo estimations.

These estimations gave huge positive values in the case of the mixed open market growth optimal strategy and huge negative values in the case of the pure open market growth optimal strategy. This goes in line with the results we will explore in the following section, where the pure open market performs very poorly, while the mixed open market strategy performs well at the expense of a huge amount of leverage. We should note that by adopting this method, our benchmark is the complete market portfolio $\mathbf{1}_d$, which is not a valid strategy within the pure open market setting. As such, this benchmark might not be entirely appropriate for the evaluation of a pure open market strategy, potentially leading to such unsatisfactory results.

5.2 Backtesting Results

In this section, we will delve into the key findings from our backtesting of growth optimal strategies. Our analysis commences with a thorough examination of these strategies' behaviours within both pure and mixed open markets. Subsequently, we reveal the ultimate outcomes in these respective markets as well as the performance of leverage-adjusted strategies in the mixed open market.

For the purposes of this analysis, we maintain a consistent value of $d = 2000$, while allowing N to vary from 1 to d. This will facilitate an understanding of the impact of this hyperparameter. The initial wealth is established at 1. The backtesting is carried out using monthly data.

Our backtesting data is bifurcated into two distinct datasets: the first, spanning the financial data from 1965 to 2000, is used for estimating the a parameter; the second, encompassing the period from 2000 to 2023 is deployed for the actual backtesting. This approach circumvents a pitfall we previously encountered wherein we mistakenly estimated a using the same data for backtesting, which resulted in the lookahead bias.

We also run the backtesting using daily data and monthly data starting from 1970 instead of 2000. The relative performance of the growth optimal strategies with respect to the market portfolio was very similar in these different settings, so we will focus here on understanding the results in the previously mentioned one.

5.2.1 Comparative Analysis of Growth Optimal Strategies

In this subsection, we undertake a comparative analysis of the growth optimal strategies within the pure and mixed open market settings. Insightful visualizations of these strategies are presented in Figure 5.4 and Figure 5.5.

Figure 5.4 offers a graphical representation of the weights π_k assigned in the pure open market growth optimal strategy. Remarkably, the investment strategy's form remains consistent even when the value of N increases. This strategy demonstrates a clear propensity towards heavy investment in top-ranked stocks, financing this strategy by short-selling those ranked lower. The utilization of leverage is striking – investing thrice the initial wealth in the foremost stock when $N = 200$. It is noteworthy that leverage exhibits an upward trend as N expands. This pattern is prevalent across the entire range of N values, barring the exceptional case when $N = d$.

Figure 5.4: Growth Optimal Strategies in Pure Open Market.

Figure 5.5 depicts the weights π_k in the mixed open market growth optimal strategy. Intriguingly, this strategy is diametrically opposite to its counterpart in the pure open market. Here, the highest-ranked stocks are short-sold to purchase the entire market portfolio $\mathbf{1}_d$. Essentially, we are shorting the top N ranked stocks to invest in the remaining $d - N$ stocks. The driving factor behind this approach is the considerably high value of a_d, the final component of the vector parameter a. The deployment of leverage in this strategy is also substantial. For instance, when $N = 1400$, the top 500 stocks are shorted over 150 times the initial wealth to long the smaller ones. This leverage increases in tandem with N.

One intriguing observation is the sudden alteration in the form of the pure open market growth optimal strategy when $N = d$, morphing to mirror the mixed open market strategy. This abrupt transformation can be attributed to the influence of

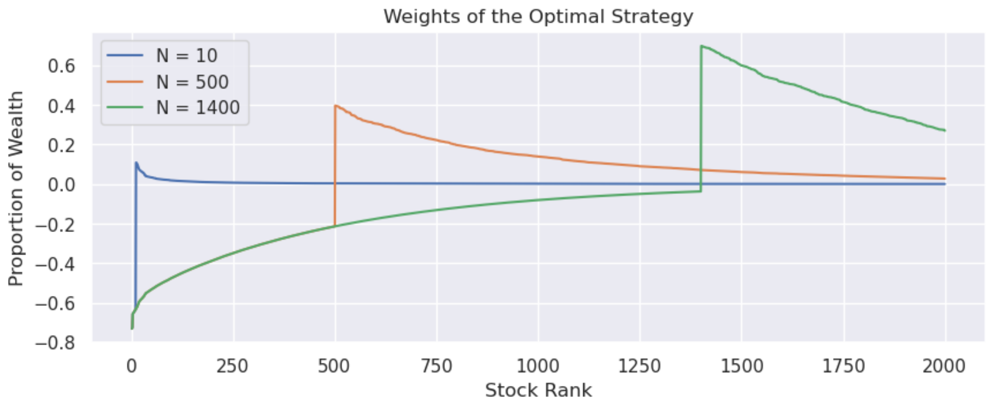

Figure 5.5: Growth Optimal Strategies in Mixed Open Market.

the parameter a_d, which according to Equation (3.6.5), does not impact the strategy unless $N = d$. This observation reinforces our argument that the form of the mixed open market strategy is predominantly determined by the substantial parameter a_d. A preliminary and informal analysis where we manually set a_d to be a_{d-1} reveals that the mixed open market strategy behaves similarly to the pure open market strategy, longing the highest-ranked stocks while shorting the lower-ranked ones.

5.2.2 Results of the Pure Open Market Growth Optimal Strategy

We backtested the growth optimal strategy in the pure open market for different values of the hyperparameter N. As noted in the previous section, this strategy is extremely leveraged. This means that the strategies are not practically implementable in real world settings, but we can nevertheless do a theoretical backtesting. The leverage resulted in a huge variance in the monthly returns, which made the initial wealth of 1 become negative for values of N bigger than 60. As noted before, the leverage increases as N increases so this effect kept getting worse.

Figure 5.6 shows the final wealth achieved after backtesting the pure open market growth optimal strategy when varying N. Only values of N smaller than 100 are included in the plot due to the negative wealth problem produced by extreme leverage. We can see that the performance of the strategy is quite poor, except for when N is less than 10. In this case, the performance is similar to the returns of the market portfolio, which in this period obtained a final wealth of 4.1. The Sharpe Ratio obtained for $N \leq 10$ is around 0.1 and smaller or even negative otherwise. Figure 5.7 shows the monthly average returns throughout the backtesting when run for different Ns. The average monthly return decreases as N gets bigger, pointing out the fact that it is not a good strategy to short small stocks to buy big ones.

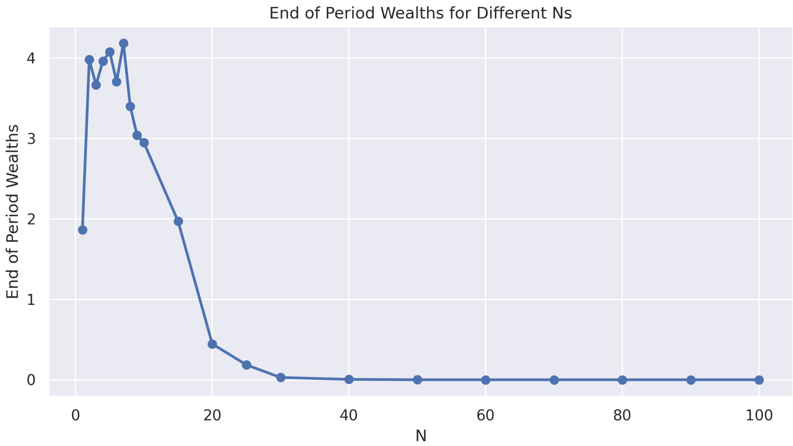

Figure 5.6: Final Wealth Backtesting of Pure Open Market Growth Optimal Strategy.

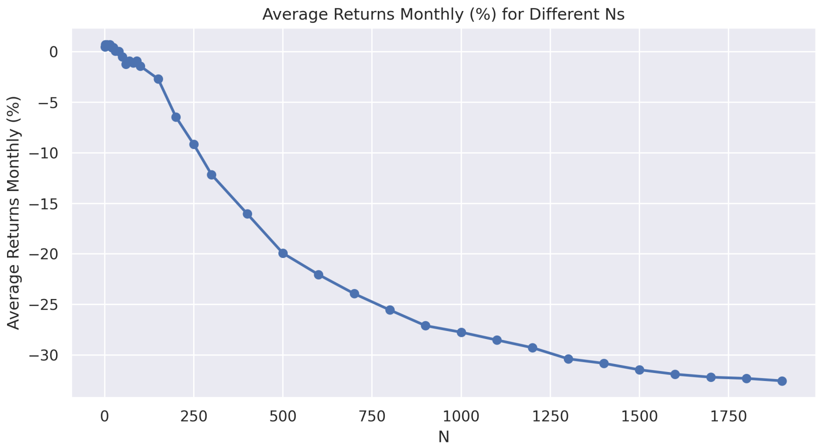

Figure 5.7: Average Backtesting Returns of Pure Open Market Growth Optimal Strategy.

5.2.3 Results of the Mixed Open Market Growth Optimal Strategy and Leverage Scaling

Similarly to the pure open market setting, we backtested the growth optimal strategy in the mixed open market for different values of the hyperparameter N. As noted in the previous section, this strategy is extremely leveraged. This resulted in a huge variance in the monthly returns, which made the initial wealth of 1 become negative for values of N bigger than 15. As noted before, the leverage increases as N increases so this effect kept getting worse.

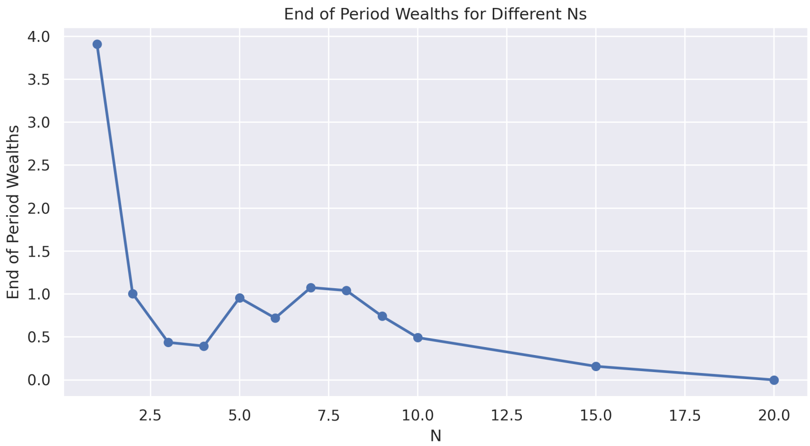

Figure 5.8: Final Wealth Backtesting of Mixed Open Market Growth Optimal Strategy.

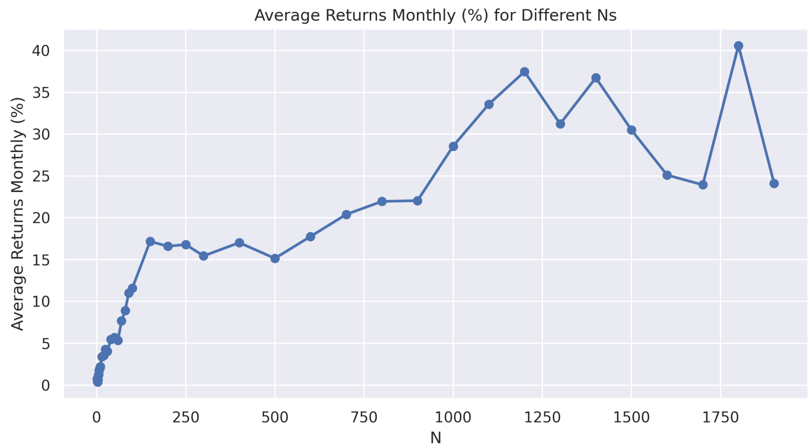

Figure 5.9: Average Backtesting Returns of Mixed Open Market Growth Optimal Strategy.

Figure 5.8 shows the final wealth achieved after backtesting the mixed open market growth optimal strategy when varying N. Only values of N smaller than 20 are included in the plot due to the negative wealth problem produced by extreme leverage. The strategy performs very poorly except for the case when $N = 1$.

However, a strange phenomenon occurs. In Figure 5.9 we can see that the average monthly return throughout the backtesting increases as N gets bigger, contrary to the final amount of wealth in the period. A closer look into the Sharpe Ratios, which also decrease as N gets bigger points to the fact that even though the average monthly return increases, the variance of the returns increases even further due to the extreme leverage, making the wealth go to 0 or negative when N is big.

However, the fact that the monthly average returns are high motivated us to find ways to reduce the leverage in order to study if the strategy could perform better. As discussed in previous sections, our method for reducing the leverage of the mixed open market growth optimal strategy successfully reduces the leverage while maintaining the form or underlying strategy of investment, as shown in Figure 5.3.

We fixed $N = 100$ and run the backtesting for different leverage scale parameters. The results of this backtesting are shown in Figure 5.10. The market portfolio has a Sharpe Ratio in this period of 0.135, while the different leverage scalers $0.1, 0.01$ and 0.001 have Sharpe Ratios of $0.137, 0.152$ and 0.148 respectively. While the growth optimal strategy with leverage scaler 0.1 achieves three times bigger end-of-period wealth, it is clear that the period crises affect this strategy extremely harshly, resulting in a high variance in the monthly returns which results in a similar Sharpe Ratio as the complete market portfolio $\mathbf{1}_d$.

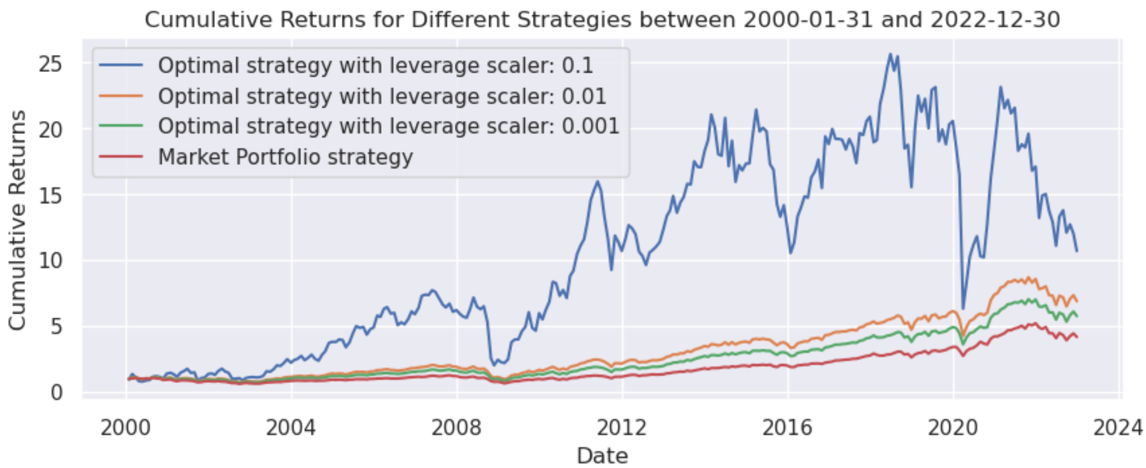

Figure 5.10: Cumulative Returns of Mixed Open Market Growth Optimal Strategy with Different Leverage Scalers.

On the other hand, we did achieve better Sharpe Ratios when the leverage is reduced even further. This seems to point to the fact that it is a good idea to short the 100 biggest stocks in order to buy the market portfolio in the mixed open market setting. This aligns with usually assumed fact that small stocks tend to have higher return rates [26], and taking advantage of this fact can be done without overleveraging. For example, when the leverage scaler is 0.01 we are shorting around 50% of our initial wealth with the first 100 stocks to buy the market portfolio. This quantity decreases to 5% of our initial wealth when the leverage scaler is 0.001. In both cases, we can see that the results are consistently better than the market portfolio.

We also put these strategies to test using the log relative wealth method, as described in Equation (5.1.5). The results, as shown in Figure 5.11, depict the log wealth relative to the performance of the market portfolio, $\mathbf{1}_d$. As anticipated, the performance of the market portfolio remains a constant 0, with the other strategies fluctuating around it. In this context, positive values signify performance superior to the market portfolio, while negative values suggest underperformance.

However, it is crucial to note that this backtesting method may not be as accurate as the initial one. This is because, as shown in Equation (5.1.5), we do not utilize cleaned returns in this case. Instead, we only employ market weights to estimate these returns and calculate the log relative wealth directly. Despite this, the curves of different strategies display behaviors similar to those seen earlier. For instance, the strategy with the leverage scaler of 0.1 undergoes significant fluctuations, whereas those with leverage scalers of 0.01 and 0.001 perform comparably to the market portfolio and without any drastic changes.

Nonetheless, the overall results deviate, with most strategies underperforming the market portfolio, in complete contrast to our previous backtesting. This discrepancy could be attributed to the differences between cleaned data and estimated data, and therefore, any conclusions drawn from this method should be treated with caution.

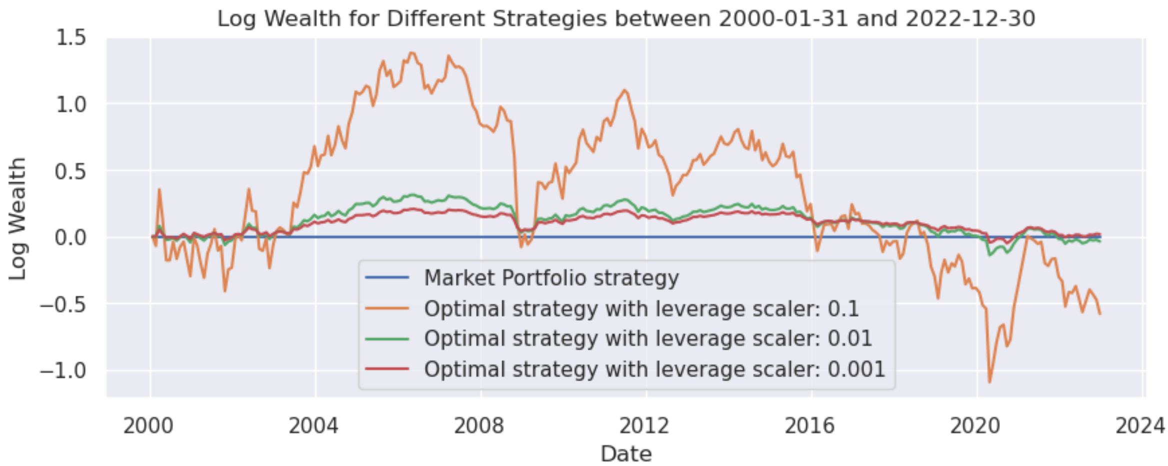

Figure 5.11: Log Relative Wealth of Mixed Open Market Growth Optimal Strategy with Different Leverage Scalers.

5.2.4 Discussion of the Results

Our exploration of the growth optimal strategies in both the pure and mixed open market settings has shed light on intriguing aspects of these strategies. A common theme across both settings is the high degree of leverage inherent in these strategies. While this can sometimes result in high returns, it also dramatically increases the risk and the associated volatility in the monthly returns, as shown in our backtesting results.

In the pure open market setting, the strategy demonstrated a preference for heavily investing in top-ranked stocks and financing this strategy by short-selling those ranked lower. This strategy, although it could generate market portfolio-like average returns for small N, proved to be overall ineffective due to the extreme leverage used and the erroneous bias for large stocks rather than small ones. When the leverage was too high (for large N), the initial wealth frequently became 0 or negative, highlighting the high-risk nature of such an approach.

The erroneous bias towards top-ranked stocks is depicted in Equation (4.3.7), where the last components of a, and especially a_d are not taken into account. This is a feature of the pure open market setting, which suggests that not taking into account the returns of low-ranked stocks is not a good idea. This notion is particularly relevant when we apply the rank Jacobi model. Here, the estimation of a reveals that the last component of this vector is exceptionally large, which drives the underpinnings of the behavior observed in both the pure and mixed open market growth optimal strategies.

In the mixed open market setting, the situation was quite different. The strategy consisted of short-selling the highest-ranked stocks to buy the entire market portfolio. This approach, while equally leveraged, surprisingly resulted in an increase in average monthly returns as N got larger. However, due to the high volatility, the final wealth was often low or negative for large values of N. This paradox suggests that while the strategy itself might be sound, the amount of leverage used is critical.

The application of a leverage scaler proved to be a promising way of handling the high leverage inherent in these strategies. By reducing the leverage, we were able to maintain the same strategic investing approach while mitigating the associated risk. As a result, the strategies with leverage scalers achieved a higher Sharpe Ratio than the market portfolio.

The success of these deleveraged strategies implies that shorting bigger stocks to buy the market portfolio could indeed be a viable investment strategy, provided the leverage used is carefully controlled. However, the potential pitfalls associated with high leverage, including the risk of significant losses, should not be overlooked.

Chapter 6

Conclusion

This thesis embarks on a comprehensive exploration into the intricate domain of Stochastic Portfolio Theory, offering a detailed examination of two distinct investment strategies within open markets. The study effectively bridges the gap between theoretical analysis and empirical validation, providing pivotal insights into the practical applicability of these strategies.

Our work commenced with a theoretical inquiry into the open market investment scenario, encompassing both mixed and pure open market approaches. We successfully extended some results related to the mixed open market found in [1], to the pure open market context, making the abstract results in [2] more practical and implementable. These include an explicit formula for the growth optimal strategy without assumptions on the market weights, the application of rank Jacobi processes as models to deduce an implementable formula for the growth optimal strategy, the computation of the asymptotic growth rate of this strategy, and showing that this strategy is functionally generated.

In the practical realm, this thesis has successfully developed a robust software platform capable of backtesting two growth optimal strategies based on historical financial data. By estimating the parameter a of the rank Jacobi model, we enabled the implementation of these growth optimal strategies, assuming rank Jacobi models for the market weights. Our analysis unveiled stark contrasts in the behaviours of the two strategies: the pure open market approach predominantly invested in top-ranked stocks while short-selling those ranked lower. This initially surprising result was traced back to the disproportionate influence of the final component of the vector parameter a, a_d, which is vastly larger than the other components. The pure open market growth optimal strategy, by not accounting for this last parameter, was rendered excessively risky due to its high leverage and poor investment philosophy, often leading to zero or negative initial wealth, especially for larger N.

Conversely, the mixed open market strategy gravitated towards short-selling the highest-ranked stocks to purchase the entirety of the market portfolio. Despite similarly high leverage, this approach witnessed an increase in average monthly returns as N expanded. Nevertheless, the associated high volatility highlighted the high-risk nature of such an approach. Significantly, the introduction of a leverage scaler to mitigate these risks offered promising results, leading to a higher Sharpe

Ratio than the market portfolio. This outcome indicates the potential practical utility of these strategies when leveraged judiciously. Hence, this study accentuates the need for prudent leverage management and further refinements to optimise these growth optimal strategies.

In conclusion, this research has not only extended some theoretical results in stochastic portfolio theory but also provided practical implications for portfolio management and investment strategies. The extensive exploration of two investment strategies under the lens of the rank Jacobi model unveiled significant differences in their operation and performance, highlighting the pivotal role of the parameter a and the necessity of prudent leverage management.

The journey of this research, from theoretical inquiry to empirical testing, demonstrates the indispensable interplay between theory and practice in financial mathematics. The analytical insights, coupled with the empirical evidence, establish a comprehensive understanding of the growth optimal strategies in open market settings.

6.1 Future Work

This exploration into Stochastic Portfolio Theory and its practical application in open markets, like any evolving domain of study, has uncovered multiple paths for future research. A fascinating direction for future work lies in diversifying the modelling of market weight behaviour. Although the rank Jacobi model served as a beneficial tool in this investigation, the exploration of other models may yield superior insights or additional perspectives. These models could potentially offer a more nuanced understanding of market weight changes, a more precise portrayal of the behaviour of low-ranked stocks, or effectively integrate various macroeconomic and financial indicators.

Building on this, a key area warranting further research involves enhancing the techniques to directly estimate the correlation matrix and drifts. Our study made pragmatic assumptions by using the rank Jacobi model for backtesting. However, a more direct estimation could allow us to backtest the original, more general formula for the growth optimal strategy, avoiding the need for certain model assumptions. This approach could provide a more unfiltered examination of these strategies. Utilizing advanced statistical methods and machine learning techniques, such as sophisticated regression models, time-series analyses, or deep learning algorithms, could lead to improvements in the precision of these estimates. By better capturing the complex dynamics of financial markets, these enhancements could result in the development of more robust investment strategies.

Another vital aspect requiring future exploration is the refinement of our growth optimal strategies concerning the use of leverage. While our current study delineated the potential pitfalls and rewards of leverage, further research could be focused on optimising the level of leverage whilst preserving the core investment strategy. This investigation might include applying advanced risk management techniques or further refining the leverage scaler.

Lastly, future research could be directed towards understanding and modifying the estimation process for the parameter a better. Our objective would be to reduce the disproportionate influence of its last component, a factor which our study showed has a significant bearing on the strategies' performance. Enhancing the robustness of the parameter estimation, and consequently the overall performance of our strategies, could enable us to better exploit this theory for practical financial decisions.

Bibliography

[1] David Itkin and Martin Larsson. Open markets and hybrid jacobi processes, 2021.

[2] Ioannis Karatzas and Donghan Kim. Open markets. *Mathematical Finance*, 31(4):1111–1161, 2021.

[3] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[4] Harry M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press, 1959.

[5] John L. Evans and Stephen H. Archer. Diversification and the reduction of dispersion: An empirical analysis. *Journal of Finance*, 23(5):761–767, 1968.

[6] Edwin Elton and Martin Gruber. Risk reduction and portfolio size: An analytical solution. *The Journal of Business*, 50:415–37, 02 1977.

[7] Meir Statman. How many stocks make a diversified portfolio? *The Journal of Financial and Quantitative Analysis*, 22(3):353–363, 1987.

[8] Robert C. Merton. An analytic derivation of the efficient portfolio frontier. *The Journal of Financial and Quantitative Analysis*, 7(4):1851–1872, 1972.

[9] Benoit B. Mandelbrot. *The variation of certain speculative prices*, pages 371–418. Springer New York, New York, NY, 1997.

[10] Robert C Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413, 1971.

[11] Robert Fernholz. On the diversity of equity markets. *Journal of Mathematical Economics*, 31(3):393–417, 1999.

[12] E. Robert Fernholz. *Stochastic Portfolio Theory*, pages 1–24. Springer New York, New York, NY, 2002.

[13] Ioannis Karatzas and Robert Fernholz. Stochastic portfolio theory: an overview. In Alain Bensoussan and Qiang Zhang, editors, *Special Volume: Mathematical Modeling and Numerical Methods in Finance*, volume 15 of *Handbook of Numerical Analysis*, pages 89–167. Elsevier, 2009.

[14] Robert Fernholz and Brian Shay. Stochastic portfolio theory and stock market equilibrium. *The Journal of Finance*, 37(2):615–624, 1982.

[15] Robert Fernholz, Ioannis Karatzas, and Constantinos Kardaras. Diversity and relative arbitrage in equity markets. *Finance and Stochastics*, 9:1–27, 2005.

[16] Robert Fernholz and Ioannis Karatzas. Relative arbitrage in volatility-stabilized markets. *Annals of Finance*, 1:149–177, 2005.

[17] Robert Fernholz. Numeraire markets, 2018.

[18] Tomoyuki Ichiba, Vassilios Papathanakos, Adrian Banner, Ioannis Karatzas, and Robert Fernholz. Hybrid atlas models. *The Annals of Applied Probability*, 21(2), apr 2011.

[19] Ioannis Karatzas and Constantinos Kardaras. *Portfolio Theory and Arbitrage: A Course in Mathematical Finance*. American Mathematical Society, 2021.

[20] Adrian D. Banner and Raouf Ghomrasni. Local times of ranked continuous semimartingales. *Stochastic Processes and their Applications*, 118(7):1244–1253, 2008.

[21] Jack Sherman and Winifred J. Morrison. Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix. *The Annals of Mathematical Statistics*, 21(1):124 – 127, 1950.

[22] Robert Fernholz. *Portfolio Generating Functions*, pages 344–367. 1999.

[23] Johannes Ruf. Crsp on wrds introduction. `https://github.com/johruf/CRSP_on_WRDS_introduction`, 2023.

[24] David Itkin. Generalized rank dirichlet distributions. *arXiv preprint arXiv:2302.13707*, 2023.

[25] Tomoyuki Ichiba, Vassilios Papathanakos, Adrian Banner, Ioannis Karatzas, and Robert Fernholz. Hybrid Atlas models. *The Annals of Applied Probability*, 21(2):609 – 644, 2011.

[26] Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.