On Portfolio Construction through Functional Generation

Alexander Vervuurt

The Queen's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2016

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Michael Monoyios, for taking me on as his student. This thesis would not have been possible without his advice, enthusiasm and meticulousness throughout the years.

I am also greatly indebted to Ioannis Karatzas for his collaboration and for inviting me to Columbia University. He taught me a lot about writing scientific articles, and introduced me to Adrian Banner and Robert Fernholz, with whom I have had numerous constructive and inspiring discussions since. I thank Vassilios Papathanakos for his involved reading of and detailed feedback on several papers.

Importantly, I would like to thank Johannes Ruf for his support during the early stages of my DPhil; he was always available for advice and ideas.

Yves-Laurent Kom Samo helped me learn programming skills and statistical inference techniques, from which I have gained a lot. I thank Samuel Cohen for his very useful comments, which helped greatly with the proof of Theorem 4.1.2, and Pietro Siorpaes for useful discussions regarding cyclical monotonicity.

My deepest thanks go to my dear friend Adrien Hitz for his continued interest in my research, for challenging me and my viewpoints, and for his thorough reading of an earlier version of this thesis.

This thesis would not have been possible without the care and support of my family and friends. I am forever grateful to Luz for her love, dedication and energy.

Finally, I gratefully acknowledge studentships from the Engineering and Physical Sciences Research Council, Nomura and the Oxford-Man Institute of Quantitative Finance.

Abstract

One of the main research questions in financial mathematics is that of portfolio construction: how should one systematically invest their wealth in a financial market? This problem has been tackled in numerous ways, typically through the modeling of market prices and the optimization of an investment objective. A recent approach to portfolio construction is that offered by Stochastic Portfolio Theory, in which a relatively general market model is assumed, and the portfolio selection criterion is to outperform a benchmark with probability one. In order to achieve this, Robert Fernholz developed the method of *functional generation*, which allows one to explicitly construct and study portfolios that depend deterministically on the currently observable prices.

The typical example of such a strategy is the diversity-weighted portfolio, which we extend in the first chapter of this work with a negative-parameter variation. We show that several modifications of this portfolio outperform the market index in theory, under certain assumptions on the market, and we perform an empirical study that confirms this.

In our second chapter, we develop a data-driven portfolio construction method that goes beyond functional generation, allowing for the inclusion of factors other than current prices. We empirically show that this Bayesian nonparametric approach, which utilizes Gaussian processes, leads to drastically improved performance compared to benchmark portfolios.

Next, we establish a formal equivalence between the method of functional generation and the mathematical field of optimal transport. Our results fortify known relations between the two, and extend this connection to *additive* functional generation, a recent variation of the method.

In Chapter 4, we apply our results to derive new properties and characterizations of functionally-generated wealth processes in very general market models. Finally, we develop methods for incorporating defaults into functional generation, improving its real-world implementability.

Contents

Introduction

Stochastic Portfolio Theory (SPT) is an area of financial mathematics that studies properties of continuous-time market models with weak no-arbitrage assumptions. It was initiated and developed to a large extent by Fernholz (2002). We introduce the essential definitions and notation of the field in Section 0.1 below, and present an extensive literature review in Section 0.2. Investments in SPT markets have been studied primarily through the method of functional generation, which is used to construct portfolios that depend deterministically on the currently observable prices in a market. These relatively simple portfolios can be shown to theoretically outperform the market index with probability one and over sufficiently long time horizons. These results hold in models which satisfy certain conditions, most of which are motivated by observations in actual markets. The computation and implementation of these strategies does not require any explicit model postulation or parameter estimation, nor prediction of any quantities or processes, which makes them very attractive for real-world use.

In this dissertation, we apply, modify, study and extend the method of functional generation. More precisely, in the four chapters of this work, we

1. apply functional generation to define new portfolios, studying their theoretical and empirical performance[1] relative to known ones;

2. modify functional generation to construct a data-driven method for portfolio construction, which we empirically show to yield excellent results;

3. study theoretical properties of functional generation by establishing connections with the mathematical field of optimal transport;

4. extend results for functionally-generated wealth processes to generalized settings and models with defaults, thus improving their real-world implementability.

[1]Wharton Research Data Services (WRDS) was used in preparing the data for Chapter 1 and Chapter 2. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers.

0.1 Definitions & notation

In the first two chapters of this work, we place ourselves in the standard Itô market model of Stochastic Portfolio Theory, which we define in Section 0.1.1 below. We refrain from imposing any market model in the first part of Chapter 3, and assume only a general continuous semimartingale model in Section 3.3 of Chapter 3, and Sections 4.1–4.4 of Chapter 4. Finally, we return to the Itô model in Section 4.5. We do not assume the existence of a local martingale deflator (see Definition 4.4.1), let alone that of an equivalent local martingale measure.

We introduce the definitions necessary for describing investments in our market models in Section 0.1.2, where we also formally define the central investment objective of SPT: almost-sure outperformance of the market index. Following this, in Section 0.1.3 we define the class of portfolios that has been used to construct explicit relative arbitrages, which is the main theme of this dissertation: *functionally-generated portfolios*. A variation of this, where portfolios only distinguish between firms by their *rank*, is exhibited in Section 0.1.4. We define trading strategies in Section 0.1.5, which give an alternative way of describing investments. Section 0.1.6 then defines the method of additive functional generation of trading strategies.

We assume a basic knowledge of measure-theoretic probability theory and stochastic integration. In particular, no prior knowledge of Stochastic Portfolio Theory is presumed.

0.1.1 Itô market model

In the 'classical' market model of Stochastic Portfolio Theory, stock capitalizations are modeled as Itô processes. Namely, the dynamics of the $n \geq 2$ capitalization processes $X_i(\cdot)$, $i = 1, \dots, n$, are given by

$$\mathrm{d}X_i(t) = X_i(t)\left(b_i(t)\,\mathrm{d}t + \sum_{j=1}^{d} \sigma_{ij}(t)\,\mathrm{d}W_j(t)\right), \quad t \geq 0, \quad i = 1, \dots, n\,; \tag{0.1}$$

here $W_1(\cdot), \dots, W_d(\cdot)$ are independent standard Brownian motions with $d \geq n$, and the initial capitalizations are positive: $X_i(0) > 0$, $i = 1, \dots, n$. We assume all processes to be defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\mathbb{P}$ the 'physical' measure, and adapted to a filtration $\mathbb{F} = \{\mathcal{F}(t)\}_{0 \leq t < \infty}$ that satisfies the usual conditions and contains the filtration generated by the "driving" Brownian motions.

The processes of *rates of return* $b_i(\cdot)$, $i = 1, \ldots, n$ and of *volatilities* $\sigma(\cdot) = (\sigma_{ij}(\cdot))_{1 \le i \le n, 1 \le j \le d}$ are $\mathbb{F}$-progressively measurable and assumed to satisfy the integrability condition

$$\sum_{i=1}^{n} \int_0^T \Big(|b_i(t)| + \sum_{j=1}^{d} (\sigma_{ij}(t))^2 \Big) \, \mathrm{d}t < \infty, \quad \mathbb{P}\text{-a.s.}, \quad \forall\, T \in (0, \infty).$$

Often, we will also assume the following *non-degeneracy* condition:

$$\exists\, \varepsilon > 0 \text{ such that } \xi' \sigma(t) \sigma'(t) \xi \ge \varepsilon ||\xi||^2, \quad \forall\, \xi \in \mathbb{R}^n, \quad t \ge 0\,; \quad \mathbb{P}\text{-a.s.} \qquad (\mathrm{ND}^{\varepsilon})$$

This condition states that the eigenvalues of the volatility matrix are bounded away from zero; in a sense, there is a minimal amount of volatility in the market at all times. In particular, this excludes the existence of a riskless asset in the market — we only consider investment in risky assets.

0.1.2 Portfolios, wealth and relative arbitrage

We study investments in equity markets using *portfolios*, considering only finite horizons $[0, T]$ for some $T > 0$. These are $\mathbb{R}^n$-valued and $\mathbb{F}$-progressively measurable processes $\pi(\cdot) = \big(\pi_1(\cdot), \cdots, \pi_n(\cdot)\big)'$, where $\pi_i(t)$ stands for the proportion of wealth invested in stock i at time t.

We restrict ourselves to *long-only* portfolios. These invest solely in the stocks, namely

$$\pi_i(t) \ge 0, \quad i = 1, \ldots, n, \qquad \text{and} \qquad \sum_{i=1}^{n} \pi_i(t) = 1, \quad \forall\, t \ge 0;$$

in particular, there is no investment in any money market. The corresponding wealth process $V^{\pi}(\cdot)$ of an investor implementing $\pi(\cdot)$ is seen to evolve as follows (we normalize the initial wealth to 1):

$$\frac{\mathrm{d}V^{\pi}(t)}{V^{\pi}(t)} = \sum_{i=1}^{n} \pi_i(t) \, \frac{\mathrm{d}X_i(t)}{X_i(t)}\,, \qquad V^{\pi}(0) = 1. \tag{0.2}$$

An example of a very simple portfolio is the *equally-weighted portfolio*, which invests an equal proportion of wealth in each stock at any given point in time:

$$\pi_i(t) = \frac{1}{n}\,, \quad i = 1, \ldots, n, \quad \forall\, t \ge 0. \tag{0.3}$$

We shall measure performance, for the most part, with respect to the market index. This is the wealth process $V^\mu(\cdot)$ that results from a buy-and-hold portfolio, given by the vector process $\mu(\cdot) = \big(\mu_1(\cdot), \cdots, \mu_n(\cdot)\big)'$ of *market weights*

$$\mu_i(t) := \frac{X_i(t)}{\sum_{j=1}^n X_j(t)}, \quad i = 1, \ldots, n. \tag{0.4}$$

Throughout this thesis, we assume the market weights vector takes values in the set

$$\Delta_+^n := \Big\{x \in (0,1)^n \,:\, \sum_{i=1}^n x_i = 1\Big\}, \tag{0.5}$$

which has closure $\bar{\Delta}_+^n$.

To ease notation, in Chapters 3 and 4 we will write the short

$$V(t) := \frac{V^\pi(t)}{V^\mu(t)}$$

for the market-relative wealth corresponding to a portfolio $\pi(\cdot)$; that is, we drop the superscript to denote this. Here, 'market-relative' refers to 'with the market portfolio $\mu(\cdot)$ as the numéraire', or, equivalently, 'expressed as a multiple of the wealth corresponding to the market index'. By an application of Itô's formula, the process $V(\cdot)$ can be seen to satisfy the SDE

$$\frac{\mathrm{d}V(t)}{V(t)} = \sum_{i=1}^n \pi_i(t)\, \frac{\mathrm{d}\mu_i(t)}{\mu_i(t)}\,, \qquad V(0) = 1. \tag{0.6}$$

The following defines the measure of performance that is the main portfolio selection criterion in SPT:

0.1.1 Definition (relative arbitrage)**.** A *relative arbitrage* with respect to a portfolio $\rho(\cdot)$ over the time-horizon $[0,T]$, for a real number $T > 0$, is a portfolio $\pi(\cdot)$ such that

$$\mathbb{P}\big(V^\pi(T) \geq V^\rho(T)\big) = 1 \qquad \text{and} \qquad \mathbb{P}\big(V^\pi(T) > V^\rho(T)\big) > 0.$$

An equivalent way to express this notion, is to say that *the portfolio* $\pi(\cdot)$ *outperforms portfolio* $\rho(\cdot)$ over the time-horizon $[0,T]$. When $\rho(\cdot) = \mu(\cdot)$, we simply call $\pi(\cdot)$ a relative arbitrage, and say it (almost-surely) outperforms the market.

We call a relative arbitrage *strong*, if in fact $\mathbb{P}\big(V^\pi(T) > V^\rho(T)\big) = 1$; and sometimes we express this by saying that the portfolio $\pi(\cdot)$ outperforms the portfolio $\rho(\cdot)$ *strongly* over the time-horizon $[0,T]$. $\square$

In Chapter 1, we shall use the reverse order-statistics notation. For any $\theta \in \{\vartheta \in \mathbb{R}^n : \vartheta_i \neq \vartheta_j \text{ for } i \neq j\}$, this is defined recursively by

$$\begin{aligned} \theta_{(1)} &= \max_{1\le i\le n}\{\theta_i\} \\ \theta_{(k)} &= \max\big(\{\theta_1,\dots,\theta_n\} \setminus \{\theta_{(1)},\dots,\theta_{(k-1)}\}\big), \quad k = 2,\dots,n. \end{aligned} \tag{0.7}$$

In the case that $\theta_i = \theta_j$ for some $i < j$, ties are resolved lexicographically, always in favor of the lowest index i. Thus we have

$$\theta_{(1)} \ge \theta_{(2)} \ge \ldots \ge \theta_{(n)}.$$

Furthermore, for any two vectors $a, b \in \mathbb{R}^n_+$ we write the element-wise fraction as

$$\frac{a}{b} := \left(\frac{a_1}{b_1},\dots,\frac{a_n}{b_n}\right). \tag{0.8}$$

For the Itô model (0.1), we introduce the $\mathbb{R}^{n\times n}$-valued *covariation* process

$$a(\cdot) = \sigma(\cdot)\sigma'(\cdot)$$

and, writing e_i for the i^{th} unit vector in $\mathbb{R}^n$, the *relative covariances*

$$\tau^\mu_{ij}(t) := \big(\mu(t) - e_i\big)' a(t)\big(\mu(t) - e_j\big), \qquad 1 \le i, j \le n.$$

Finally, we define the *excess growth rate* $\gamma^*_\pi(\cdot)$ of a portfolio $\pi(\cdot)$ as

$$\gamma^*_\pi(t) := \frac{1}{2}\bigg(\sum_{i=1}^n \pi_i(t)a_{ii}(t) - \sum_{i,j=1}^n \pi_i(t)a_{ij}(t)\pi_j(t)\bigg).$$

In the Itô model, the non-degeneracy condition (ND$^\varepsilon$) implies, on the strength of Lemma 3.4 in Fernholz and Karatzas (2009) (originally proved in the Appendix of Fernholz et al. (2005)), that for any long-only portfolio $\pi(\cdot)$, we have with probability one:

$$\gamma^*_\pi(t) \ge \frac{\varepsilon}{2}\big(1 - \pi_{(1)}(t)\big), \quad \forall\, t \ge 0\,. \tag{0.9}$$

0.1.3 Functionally-generated portfolios

A particular class of portfolios, called *functionally-generated portfolios*, was introduced and studied by Fernholz (1999a). It consists of portfolios $\pi(\cdot)$ that satisfy $\pi(\cdot) = \pi(\mu(\cdot))$ for a deterministic function $\pi : \Delta^n_+ \to \bar{\Delta}^n_+$ of a certain form; that is, the proportion of wealth that is to be invested in each stock is only allowed to depend deterministically on the current market weights.

0.1.2 Definition (functionally-generated portfolio)**.** Let $\mathbf{G} \in C^2(\Delta_+^n, \mathbb{R}_+)$ be such that $x \mapsto x_i \mathrm{D}_i \log \mathbf{G}(x)$ is bounded on Δ_+^n for $i = 1, \ldots, n$. The portfolio $\pi(\cdot)$ that is defined, for $i = 1, \ldots, n$, by

$$\pi_i(t) := \Big(\mathrm{D}_i \log \mathbf{G}(\mu(t)) + 1 - \sum_{j=1}^{n} \mu_j(t) \mathrm{D}_j \log \mathbf{G}(\mu(t)) \Big) \cdot \mu_i(t), \tag{0.10}$$

is called a *functionally-generated portfolio*, and $\mathbf{G}$ is said to be its *generating function.*

□

Here and throughout the thesis, we write D_i for the partial derivative with respect to the i^{th} variable, and D^2_{ij} for the second partial derivative with respect to the i^{th} and j^{th} variables.[2] Theorem 3.1 of Fernholz (1999a) asserts that the performance of the wealth process corresponding to $\pi(\cdot)$, when measured relative to the market, satisfies the following decomposition:

0.1.3 Theorem (Fernholz's master equation)**.** *Assume the Itô market model* (0.1)*, and let $\pi(\cdot)$ be functionally-generated by some function $\mathbf{G} \in C^2(\Delta_+^n, \mathbb{R}_+)$, as in* (0.10)*. Then*

$$\log\left(\frac{V^\pi(T)}{V^\mu(T)}\right) = \log\left(\frac{\mathbf{G}(\mu(T))}{\mathbf{G}(\mu(0))}\right) + \int_0^T \mathfrak{g}(t)\,\mathrm{d}t\,, \quad \mathbb{P}\text{-a.s.}, \tag{0.11}$$

where the quantity

$$\mathfrak{g}(t) := \frac{-1}{\mathbf{G}(\mu(t))} \sum_{i,j=1}^{n} \mathrm{D}^2_{ij} \mathbf{G}(\mu(t)) \mu_i(t) \mu_j(t) \tau^\mu_{ij}(t) \tag{0.12}$$

is called the drift process *of the portfolio $\pi(\cdot)$.*

The proof of (0.11) consists of a repeated application of Itô's formula; see, for instance, the proof of Lemma 4.3.1 in Chapter 4. A typical example of a functionally-generated portfolio is the equally-weighted portfolio (0.3), which can be seen to be generated by the function

$$\mathbf{G}(x) = \prod_{i=1}^{n} x_i^{1/n}.$$

[2]Note the fact that $\mathbf{G}$ is a function on the $(n-1)$-dimensional space Δ_+^n, while the gradient $\nabla \mathbf{G}$ is n-dimensional. Hence, formally, not all partial derivatives can be defined simultaneously; one way around this is to define each of them through the *superdifferential*, see (3.4).

0.1.4 Rank-based portfolios

Fernholz (2001) studied a variation of functionally-generated portfolios, namely one that does not distinguish between firms by name, but only by *rank*. Such portfolios are called *rank-based portfolios*, and we will use them in Section 1.2.

Let us write $p_t(k)$ for the index of the stock ranked k^{th} at time t (with ties resolved "lexicographically", by choosing the lowest index), so that

$$\mu_{p_t(k)}(t) = \mu_{(k)}(t) \tag{0.13}$$

in the notation of (0.7). We recall the following definition (see, for instance, Definition 4.1.1 of Fernholz (2002)):

0.1.4 Definition (local time)**.** Let $Y(\cdot)$ be a continuous semimartingale. The *local time at the origin* for $Y(\cdot)$ is the process

$$\Lambda^Y(t) := \frac{1}{2}\Big(|Y(t)| - |Y(0)| - \int_0^t \operatorname{sign}\big(Y(s)\big)\, \mathrm{d}Y(s) \Big), \qquad t \geq 0.$$

□

With the above definition in mind, let us denote by

$$\mathfrak{L}^{k,k+1}(t) \equiv \Lambda^{\Xi_k}(t) \tag{0.14}$$

the semimartingale local time accumulated at the origin, over the time-interval $[0, t]$, by the continuous, nonnegative semimartingale

$$\Xi_k(\cdot) \;:=\; \log\big(\mu_{(k)}(\cdot)/\mu_{(k+1)}(\cdot)\big), \qquad k = 1, \ldots, n-1.$$

0.1.5 Definition. Consider the local times at the origin of all continuous nonnegative semimartingales of the form

$$\log\big(\mu_{(k)}(\cdot)/\mu_{(k+r)}(\cdot)\big), \qquad k = 1, \ldots, n-r, \quad r \geq 2.$$

We call these *higher-order collision local times.* □

We shall assume throughout this thesis the following:

0.1.6 Standing Assumption. All higher-order collision local times vanish. □

We refer the reader to Banner and Ghomrasni (2008) for general theory on ranked semimartingales, as well as Ichiba et al. (2011) for sufficient conditions that ensure the evanescence of such higher-order collision local times in specific (namely rank-based) models, as stated in Assumption 0.1.6.

Now suppose $\mathbf{G} : \Delta^n_+ \to \bar{\Delta}^n_+$ is a function such that

$$\mathbf{G}(x_1, \ldots, x_n) = G(x_{(1)}, \ldots, x_{(n)}), \qquad \forall\, x \in \Delta^n_+ \tag{0.15}$$

for some C^2 function $G : \Delta^n_+ \to \bar{\Delta}^n_+$, and such that $x \mapsto x_i \mathrm{D}_i \log G(x)$ is bounded on Δ^n_+. Then let us consider the following portfolio:

$$\pi_{p_t(i)}(t) = \Big(\mathrm{D}_i \log G(\mu_{(\cdot)}(t)) + 1 - \sum_{j=1}^{n} \mu_j(t) \mathrm{D}_j \log G(\mu_{(\cdot)}(t))\Big) \cdot \mu_{(i)}(t), \tag{0.16}$$

for $i = 1, \ldots, n$, where $\mu_{(\cdot)}(t) := \big(\mu_{(1)}(t), \ldots, \mu_{(n)}(t)\big)$ denotes the ranked market weights vector at time t.

Theorem 3.1 of Fernholz (2001) establishes the following analog of Theorem 0.1.3:

0.1.7 Theorem (Master equation for rank-based portfolios)**.** *Let* $\pi(\cdot)$ *be as in* (0.16)*, for a function* $G : \Delta^n_+ \to \bar{\Delta}^n_+$ *satisfying* (0.15)*. Then*

$$\log\left(\frac{V^\pi(T)}{V^\mu(T)}\right) = \log\left(\frac{G(\mu_{(\cdot)}(T))}{G(\mu_{(\cdot)}(0))}\right) + \widetilde{\Gamma}(T), \quad \mathbb{P}\text{-a.s.},$$

with

$$\begin{aligned}\widetilde{\Gamma}(T) := &- \sum_{i,j=1}^{n} \int_0^T \frac{\mathrm{D}^2_{ij} G(\mu_{(\cdot)}(t))}{2G(\mu_{(\cdot)}(t))} \mu_{(i)}(t)\mu_{(j)}(t)\tau^\mu_{p_t(i)p_t(j)}(t)\,\mathrm{d}t \\ &+ \frac{1}{2}\sum_{k=1}^{n-1} \int_0^T \big(\pi_{p_t(k+1)}(t) - \pi_{p_t(k)}(t)\big)\,\mathrm{d}\mathfrak{L}^{k,k+1}(t).\end{aligned}$$

Compare the modified drift process $\widetilde{\Gamma}(\cdot)$ with (0.12), and note the appearance of the local times $\mathfrak{L}^{k,k+1}(\cdot)$.

0.1.5 Trading strategies

An alternative way of describing investments is through the *number of shares* $\eta_i(t)$ to be held in each stock i at time t. This description will be used in Chapters 3 and 4. The market-relative wealth process corresponding to such a strategy is

$$V'(t) := \sum_{i=1}^{n} \eta_i(t)\mu_i(t), \tag{0.17}$$

and we require the following self-financing condition to hold:

0.1.8 Definition. An $\mathbb{R}^n$-valued, predictable process $\eta(\cdot)$ is a *trading strategy* if it is $\mu(\cdot)$-integrable and

$$V'(\cdot) \equiv 1 + \int_0^{\cdot} \sum_{i=1}^n \eta_i(t)\, \mathrm{d}\mu_i(t). \tag{0.18}$$

$\square$

Note that each portfolio $\pi(\cdot)$ defines a trading strategy (and vice versa) through

$$\eta_i(t) = \frac{\pi_i(t)V(t)}{\mu_i(t)}, \qquad i = 1, \ldots, n, \tag{0.19}$$

with $V(\cdot)$ given by (0.6). The defining equations (0.6) and (0.17) coincide: by (0.18),

$$\mathrm{d}V(t) = V(t) \sum_{i=1}^n \pi_i(t) \frac{\mathrm{d}\mu_i(t)}{\mu_i(t)} = \sum_{i=1}^n \eta_i(t)\, \mathrm{d}\mu_i(t) = \mathrm{d}V'(t);$$

hence $V(\cdot) \equiv V'(\cdot)$.

0.1.6 Additive functional generation

Inspired by functional generation as in (0.10), Karatzas and Ruf (2016) defined a similar method for the construction of trading strategies. For this, they defined the 'defect of self-financiability' of a $\mu(\cdot)$-integrable process $\theta(\cdot)$ as

$$Q^\theta(t) := \sum_{i=1}^n \theta_i(t)\mu_i(t) - \sum_{i=1}^n \theta_i(0)\mu_i(0) - \int_0^t \sum_{i=1}^n \theta_i(t)\, \mathrm{d}\mu_i(t).$$

This quantity is identically zero for trading strategies $\theta(\cdot)$, but for general $\mu(\cdot)$-integrable processes that do not satisfy the self-financiability condition (0.18) it is not. Proposition 2.4 of Karatzas and Ruf (2016) shows that any $\mu(\cdot)$-integrable process $\theta(\cdot)$ can be turned into a trading strategy $\eta(\cdot)$ by setting

$$\eta_i(t) = \theta_i(t) - Q^\theta(t) + C, \quad i = 1, \ldots, n, \tag{0.20}$$

for any constant $C \in \mathbb{R}$. Additional functional generation is then defined as follows:

0.1.9 Definition. Let $\mathcal{G} \in C^2(U, \mathbb{R})$, where U is an open subset of Δ_+^n, and set $\theta(\cdot) := \nabla\mathcal{G}(\mu(\cdot))$. Then the trading strategy $\eta(\cdot)$ defined through (0.20), with

$$C := \mathcal{G}(\mu(0)) - \sum_{j=1}^n \mu_j(0)\mathrm{D}_j\mathcal{G}(\mu(0)),$$

is *additively functionally-generated* by $\mathcal{G}$. $\square$

Karatzas and Ruf consider additive generation by functions from a more general class than $C^2(U,\mathbb{R})$. However, by their Example 3.5, the above choice guarantees that $\mathcal{G}$ is regular in the sense of their Definition 3.1 (and, if $\mathcal{G}$ is also concave, it is even Lyapunov — see Definition 3.2 of their paper). Proposition 4.3 of Karatzas and Ruf (2016) then establishes a master equation and a more explicit expression for additively functionally-generated strategies:

0.1.10 Proposition. *A strategy $\eta(\cdot)$ which is additively generated by $\mathcal{G} \in C^2(U,\mathbb{R})$ has the form*

$$\eta_i(t) = \mathrm{D}_i\mathcal{G}(\mu(t)) - \sum_{j=1}^{n} \mu_j(t)\mathrm{D}_j\mathcal{G}(\mu(t)) + \mathcal{G}(\mu(t)) + \Gamma^{\mathcal{G}}(t), \quad i = 1, \dots, n, \tag{0.21}$$

whereas the market-relative wealth corresponding to $\eta(\cdot)$ has the decomposition

$$V(t) = \mathcal{G}(\mu(t)) + \Gamma^{\mathcal{G}}(t). \tag{0.22}$$

Here, the finite-variation process $\Gamma^{\mathcal{G}}(\cdot)$ is defined as

$$\Gamma^{\mathcal{G}}(t) := -\frac{1}{2}\int_0^t \mathrm{D}^2_{ij}\mathcal{G}(\mu(s))\,\mathrm{d}\langle \mu_i, \mu_j\rangle(s);$$

compare with (0.12).

This class of trading strategies will be relevant in Chapter 3.

0.2 Literature review

As was mentioned at the outset of this Introduction, Stochastic Portfolio Theory studies properties of continuous-time market models with weak no-arbitrage assumptions. It was introduced by Fernholz (2002), and was surveyed in Fernholz and Karatzas (2009) and Vervuurt (2015). In particular, the theory refrains from assuming the existence of equivalent local martingale measures, which is equivalent to the assumption of No Free Lunch with Vanishing Risk (NFLVR) by Delbaen and Schachermayer (1994). Instead, most models in Stochastic Portfolio Theory only satisfy the weaker No Unbounded Profit with Bounded Risk assumption (NUPBR, or, equivalently, No Arbitrage of the First Kind), which holds if and only if a local martingale deflator exists, as was shown by Kardaras (2012). The latter plays the role of Radon-Nikodym derivative in the case that an equivalent local martingale measure does exist. Karatzas and Kardaras (2007) proved that NUPBR is a necessary and sufficient condition for the log-utility maximization problem to have a solution. For an overview of these different notions of arbitrage, and their relation to the existence of deflators and equivalent measures, we refer the reader to Fontana (2015).

Methods for theoretically constructing models in which NUPBR holds but NFLVR fails are not new; see for instance the creation of arbitrage possibilities using Bessel processes in Delbaen and Schachermayer (1995). However, the first research to construct markets with properties relevant to SPT was that of Osterrieder and Rheinländer (2006), which was done by applying a nonequivalent measure change to construct markets that are *diverse* — this is a central concept in SPT, see Equation (D^δ) below. A more general and systematic construction was developed by Ruf and Runggaldier: it was shown that any market model that satisfies NUPBR but not NFLVR can be set up using their method. These papers build on the work of Föllmer (1972), who introduced the "exit measure" of a supermartingale. Other ways of constructing these models that have been suggested are through filtration shrinkage, see Protter (2015), and enlargement, see Fontana et al. (2014). An explicit example of a model that allows for arbitrage is the volatility-stabilized model, introduced by Fernholz and Karatzas (2005), which is an Itô model on each stock's log-price process with drift and volatility rates that are inversely proportional to the current price. Shkolnikov (2013) considered the large-market limit of such a model, and Picková (2014) generalized this model to allow for a more complicated dependence of the drift and volatility processes on the market weights.

Another class of models that has received a lot of attention in the SPT community is that of rank-based models, which are Itô models whose drift and volatility processes depend on the stocks' ranks by company capitalization (that is, it is a system of interacting Brownian particles). The simplest such model is the Atlas model, introduced in Banner et al. (2005), which assigns a non-zero and positive drift rate only to the lowest-ranked stock, causing it to "carry the entire market". The dynamics of this model have been studied in depth in Ichiba et al. (2013b), and a mean-field variation of it was proposed by Jourdain and Reygner (2015). A generalization, called the hybrid Atlas model, was introduced in Ichiba et al. (2011), and it was shown that this can accurately model the empirically-observed capital distribution curve in real markets. The latter is the relation between the capitalization and rank by capitalization of firms in a financial market, which seems to exhibit remarkable stability through time; see also Section 5.1 of Fernholz (2002). Furthermore, Shkolnikov (2012) considered the large-market limit of rank-based models, and other results in this area have been proved in Fernholz et al. (2013a,b,c), and Ichiba et al. (2013a,b).

One of the most explored lines of research in Stochastic Portfolio Theory is portfolio construction, which is the theme of this dissertation. The selection criterion for portfolios is to exploit the arbitrages that are present in the market, and performance is typically measured with respect to a benchmark, as in the Benchmark Approach of Platen and Heath (2006). More precisely, write $X_i(t)$ for the capitalization of firm i at time t, and let us introduce the *market weights*

$$\mu_i(t) = \frac{X_i(t)}{\sum_{j=1}^n X_j(t)}, \qquad i = 1, \ldots, n,$$

which each give the proportion of the market that firm i constitutes in terms of capitalization. Then the typical investment objective is to have a terminal wealth that is almost-surely (under the physical measure) greater than that pertaining to the market index, which is a buy-and-hold strategy obtained by holding a proportion $\mu_i(t)$ of wealth in firm i at every time t. This is colloquially referred to as 'outperformance of the market', and a portfolio which achieves this objective is called a *relative arbitrage*. It can be shown that a relative arbitrage exists if all local martingale deflators are *strict* local martingales. For an alternative characterization, see Theorem 8 of Ruf (2011); more explicit conditions on the Itô model parameters were derived in the one-dimensional case by Mijatović and Urusov (2012).

Contrast the SPT portfolio selection approach to that of mean-variance optimization, introduced by Markowitz (1952), and utility maximization, see Rogers (2013), where one aims to *optimize* a certain performance criterion. Nonetheless, the optimization of relative arbitrages has also been attempted and remains an active area of study: Fernholz and Karatzas (2010) solved this problem in a complete Markovian Itô setting, deriving a linear parabolic partial differential equation for the portfolio that outperforms the market from the smallest initial capital. In Fernholz and Karatzas (2011), they extended this result to markets whose parameters exhibit uncertainty, arriving at a non-linear Hamilton-Jacobi-Bellman equation. The same question was answered for outperformance with probability less than one by Bayraktar et al. (2012a). More concrete results were derived in a two-stock setting in Section 4 of Pal and Wong (2013), and for higher dimensions in Wong (2015a). Ruf (2013) generalized the results of Fernholz and Karatzas (2010) to the incomplete Markovian case by considering the problem of optimal relative arbitrage as a European hedging problem, showing that the optimal strategy is a delta hedge. An alternative hedging approach was taken by Chau and Tankov (2015), who characterized the optimal arbitrage as the hedge of a certain claim in markets that are obtained by making the nonequivalent measure change of Ruf and Runggaldier. The problem of hedging American options in the absence of an equivalent local martingale measure was solved by Bayraktar et al. (2012b), who characterized the hedging strategy and optimal stopping time in this case. Furthermore, Kardaras (2015) derived valuation formulas for both European and American exchange options in general semimartingale models with the possibility of default, showing how traditional parity formulas are altered when NFLVR fails. For a background on hedging in markets without an equivalent local martingale measure, we refer the reader to Cox and Hobson (2005).

In order to construct explicit relative arbitrages, a powerful tool for portfolio construction has been developed in SPT, namely that of *functional generation*. This was introduced in Fernholz (1999a) and further developed in Fernholz (2002), and defines a class of portfolios that depend deterministically on the current market weights $\mu(t) = \big(\mu_1(t), \ldots, \mu_n(t)\big)$. Namely, for suitable functions $\mathbf{G}$, we call the portfolio (whose components prescribe the proportions of wealth to be held in each stock) defined by

$$\pi_i(t) = \Big(\mathrm{D}_i \log \mathbf{G}(\mu(t)) + 1 - \sum_{j=1}^{n} \mu_j(t) \mathrm{D}_j \log \mathbf{G}(\mu(t))\Big) \cdot \mu_i(t), \qquad i = 1, \ldots, n$$

a *functionally-generated portfolio*; here, D_i denotes the derivative with respect to the i^{th} argument. The strength of these portfolios lies in the fact that, in a general Itô market model, their terminal market-relative wealth $V(T)$ at any point in time T can be shown (see Theorem 3.1 of Fernholz (1999a)) to decompose as follows:

$$\log V(T) = \log\left(\frac{\mathbf{G}(\mu(T))}{\mathbf{G}(\mu(0))}\right) + \int_0^T \mathfrak{g}(t)\,\mathrm{d}t\,, \quad \mathbb{P}\text{-a.s.},$$

for $\mathfrak{g}(\cdot)$ some process depending deterministically on the market weights through $\mathbf{G}$. Note the absence of any stochastic integrals in this equation, which is commonly referred to as 'Fernholz's master equation' — this enables us to bound the terminal market-relative wealth from below for suitable functions $\mathbf{G}$ and under almost-sure descriptive (as opposed to normative, such as NFLVR) assumptions on the market dynamics, showing almost-sure outperformance of the market. Notably, no model calibration or parameter estimation is required for this, since neither the functionally-generated portfolio nor the master equation exhibit any explicit dependency on the market parameters. This makes the portfolios of Stochastic Portfolio Theory very robust and implementable in comparison to those of mean-variance optimization and utility maximization.

The most-studied functionally-generated portfolio to date is the diversity-weighted portfolio, defined in Fernholz et al. (2005) as

$$\pi_i(t) = \frac{(\mu_i(t))^p}{\sum_{j=1}^n (\mu_j(t))^p}, \qquad i = 1, \dots, n,$$

for any parameter $p \in \mathbb{R}$. Using the master equation, Fernholz et al. (2005) showed that the diversity-weighted portfolio with parameter $p \in (0,1)$ almost-surely outperforms the market over sufficiently long time horizons $[0,T]$, meaning that the wealth process corresponding to this portfolio is greater than that pertaining to the market with probability one. For this, they assumed an Itô model with a certain volatility structure and which satisfies the assumption of *diversity*:

$$\exists\, \delta > 0 : \ \mathbb{P}\big(\max_{1 \le i \le n} \mu_i(t) < 1 - \delta, \ \forall\, t \in [0,T] \big) = 1. \tag{D$^\delta$}$$

This condition states that no company can become too large in terms of the proportion of the market capitalization that it constitutes. Also in Fernholz et al. (2005), it was shown that in any such market model, and given an arbitrarily short time horizon, a portfolio exists which almost-surely outperforms the market over that time horizon; this is referred to as short-term relative arbitrage. The same result holds for volatility-stabilized models, by Banner and Fernholz (2008). Condition (D$^\delta$), which was already studied in Fernholz et al. (1998) and Fernholz (1999b), fails in volatility-stabilized models, as was shown in Fernholz and Karatzas (2009). However, Fernholz et al. (2005) showed that Itô models where the above condition holds do exist. Market models for which diversity holds, but which have a varying number of stocks, were proposed and studied in Strong and Fouque (2011), Sarantsev (2014) and Karatzas and Sarantsev (2014).

Several other explicit 'long-term' (as opposed to short-term) relative arbitrages have been constructed in the literature; see Fernholz (2002), Fernholz et al. (2005), Fernholz and Karatzas (2005, 2009), Picková (2014), Pal and Wong (2013), Vervuurt and Karatzas (2015), and Wong (2015b); the last paper studies the universal portfolio of Cover (1991), which is also the topic of Cuchiero et al. (2016). Among these relative arbitrages is the so-called entropy-weighted portfolio. Fernholz and Karatzas (2005) showed that this portfolio outperforms the market almost-surely provided the volatility structure of the Itô market model satisfies a certain condition of 'sufficient intrinsic volatility'. Until recently, it was a major open problem whether this market condition alone is sufficient for the existence of short-term relative arbitrage. In Fernholz et al. (2016), this question was answered negatively, through a counterexample:

an explicit model with sufficient intrinsic volatility was constructed in which there cannot be relative arbitrage over a certain minimal time horizon.

In Chapter 1, we study diversity-weighted portfolios with *negative* parameter $p < 0$, showing that these also outperform the market under certain market conditions. This extends the results of Fernholz et al. (2005); in fact, we show that our portfolios almost-surely outperform their positive-parameter counterpart under sufficient conditions. We employ other techniques to construct variations of the diversity-weighted portfolios which are also relative arbitrages under milder assumptions; namely, we are the first to construct explicit relative arbitrages using rank-based portfolios, whose weights depend on the ranks of the firms by capitalization, and we introduce a novel method of 'mixing', which dynamically combines different portfolios.

To date, the real-world implementation of SPT portfolios has received little attention in the academic community. Namely, the empirical study that we include in Chapter 1 of the performance of the newly defined portfolios is one of the first of its kind; other papers which implement strategies using real-world data are Pal and Wong (2013) and Pal (2016). We take the incorporation of market data much further in Chapter 2, by developing a data-driven portfolio construction algorithm which learns highly-performing strategies from historical data. For this, we take a Bayesian nonparametric approach, and make use of Gaussian processes. Our methodology is a generalization of functional generation, in the sense that it includes all functionally-generated portfolios, but also in the sense that it allows for the incorporation of additional market information other than the current market weights. The latter has been studied theoretically by Strong (2014a), who generalized Fernholz's master equations for portfolios which depend on additional processes that are of finite variation. However, this has yet to be applied in the literature. Our approach circumvents the problem of postulating almost-sure market conditions under which outperformance of the market is guaranteed, and has the benefit of allowing for a user-specified portfolio selection criterion (thus not limiting ourselves to that of SPT). In both our empirical studies, we take into account transaction costs, which affect the real-world performance of all SPT portfolios, since these require continuous rebalancing. The inclusion of transaction costs in Stochastic Portfolio Theory is currently non-existent and, in our opinion, strongly needed; a first heuristic attempt at this was made in Section 6.3 of Fernholz (2002).

In Chapter 3, we move away from the study of market models, and study the method of functional generation itself. In a model-free setting, we establish a formal equivalence to an optimal transport problem, which is an optimization problem of finding a joint distribution with given marginals that minimizes a certain cost — see Villani (2009) for an overview. This extends and fortifies results from Pal and Wong (2016), who initiated this line of research, albeit in a discrete-time setting. In particular, we prove that almost any choice of marginals defines, through the unique and deterministic solution of the optimal transport problem, a functionally-generated portfolio. We also establish a similar result for a new method of portfolio construction, namely that of *additive functional generation*, which was very recently introduced by Karatzas and Ruf (2016).

We then apply these results in Chapter 4 in a general continuous semimartingale model to derive a new lower bound for the terminal relative wealth corresponding to both functionally-generated portfolios and additively-generated strategies. This invokes a limiting result that we prove, which establishes that the wealth process corresponding to discretely rebalancing a portfolio tends to the wealth process of its continuous implementation as the rebalancing frequency tends to infinity. By another application of this result, we prove a generalization of the master equation, arriving at the same result as Karatzas and Ruf (2016), but in an alternative way. Fernholz (2016) generalizes this decomposition of market-relative return to portfolios that are not functionally-generated. An even more general, namely pathwise, version of the master equation was proved by Schied et al. (2016), which holds in a model-free setting and whose proof uses functional pathwise Itô calculus. Following this, in Chapter 4 we show a converse to the master equation: any suitable stochastic process that decomposes multiplicatively or additively can be identified as the wealth process corresponding to a functionally- or additively-generated strategy, respectively.

Finally, in the last part of Chapter 4, we improve the resemblence of the theoretical performance of SPT portfolios in Itô market models to their real-world performance, by incorporating company defaults. Our empirical studies in Chapters 1 and 2 already take these into account, but here we develop two methods for modifying functionally-generated portfolios in such a way that we can study their theoretical performance in the presence of bankruptcies. The first consists of linearly combining existing portfolios in a dynamically changing way, and the second functionally generates a portfolio from a shrinking set of active stocks. In both cases, we demonstrate how the existence of relative arbitrage is maintained, using a new functionally-generated portfolio we call the beta-weighted portfolio. Our approach differs from those of Strong and Fouque (2011), Strong (2014b) and Karatzas and Sarantsev (2014), where the focus lies on studying the model dynamics instead of portfolios.

Chapter 1

Diversity-weighted portfolios with negative parameter

We start this thesis with an extensive demonstration of the explicit construction of portfolios through functional generation, and study their theoretical and empirical properties. To this end, we propose several extensions of *diversity-weighted portfolios*, a widely studied class of functionally-generated portfolios which has been shown to outperform the market portfolio $\mu(\cdot) = \big(\mu_1(\cdot), \ldots, \mu_n(\cdot)\big)$ of (0.4) almost surely under appropriate conditions on the market.

Let us place ourselves in the Itô model of (0.1). The diversity-weighted portfolio with parameter $p \in \mathbb{R}$ is defined in Equation (4.4) of Fernholz et al. (2005) as

$$\pi_i^{(p)}(t) := \frac{(\mu_i(t))^p}{\sum_{j=1}^n (\mu_j(t))^p}, \quad i = 1, \ldots, n. \tag{1.1}$$

One condition that was shown to be sufficient for the existence of relative arbitrage in the model (0.1), under the condition (ND$^\varepsilon$) of Section 0.1.1, is that of *diversity* (see Corollary 2.3.5 and Example 3.3.3 of Fernholz (2002)). This states that no single company's capitalization can take up more than a certain proportion of the entire market.

More formally, a market is said to be *diverse* with parameter $\delta \in (0,1)$ over the time-horizon $[0,T]$, for some real number $T > 0$, if (recall the notation $\mu_{(1)}(t) = \max_i \mu_i(t)$ from (0.7))

$$\mathbb{P}\big(\mu_{(1)}(t) < 1 - \delta\,, \quad \forall\, t \in [0,T]\big) = 1\,. \tag{D$^\delta$}$$

Models of the form (0.1), which satisfy the property (D$^\delta$) and admit local martingale deflators, were explicitly shown to exist in Remark 6.2 of Fernholz et al. (2005). Other constructions have been proposed by Osterrieder and Rheinländer (2006) and

Sarantsev (2014). Fernholz et al. (2005, see Eq. (4.5)) showed that the portfolio (1.1) outperforms the market index $\mu(\cdot)$ strongly, in markets satisfying (ND$^\varepsilon$) and (D$^\delta$), for any $p \in (0,1)$, and over time-horizons $[0,T]$ with[1]

$$T > \frac{2 \log n}{\varepsilon \delta p} \,. \tag{1.2}$$

In Theorem 1.1.1 below, we show that a similar property holds for the diversity-weighted portfolio with *negative* parameter p, in markets satisfying the following *no-failure* condition for some constant $\varphi \in (0, 1/n)$:

$$\mathbb{P}\big(\mu_{(n)}(t) > \varphi, \quad \forall\, t \in [0,T]\big) \;=\; 1 \,. \tag{NF$^\varphi$}$$

We note that this condition implies diversity with parameter $\delta = (n-1)\varphi$.

A rank-based modification of this negative-parameter portfolio which only invests in small-capitalization stocks is then shown, in Section 1.2, to outperform the market, also under this assumption of "no-failure". Furthermore, for certain positive parameters $p \in (0,1)$, such a rank-based portfolio outperforms the market under a milder and more realistic condition, which we call *limited-failure*. This condition postulates a uniform lower bound on market weights only for the $m < n$ largest firms by capitalization, with n the total number of equities in the market.

We present two additional results in Sections 1.3 and 1.4. The first shows that the negative-parameter diversity-weighted portfolio outperforms its positive-parameter version under an additional boundedness assumption on the covariation structure of the market; namely, this postulates that the eigenvalues of the covariance matrix in the underlying Itô model are bounded from *above*, uniformly in time — compare with (ND$^\varepsilon$). Our second result studies a composition of these two diversity-weighted portfolios and shows that, under the condition of diversity, this "mix" almost surely performs better than the market over sufficiently long time-horizons.

We carry out an empirical study of the constructed portfolios in Section 1.5, demonstrating that our adjustments of the diversity-weighted portfolio would have outperformed quite considerably the S&P 500 index, if implemented on the daily index constituents over the 25 year period between January 1, 1990 and December 31, 2014. In this study we incorporate 0.5% proportional transaction costs, delistings due to bankruptcies, mergers and acquisitions, and dividends. We compute the relative

[1]During the 25-year period that we study the S&P 500 constituents, see Section 1.5, we observe a maximum market weight of 5.5%. Hence, using the values $n = 500$, $\delta = 0.89$, $p \sim 1$, and $\varepsilon = 1.0\%$, the minimal time to market outperformance (1.2) is computed as 1397 trading days, or just over 5.5 years.

returns of our portfolios over this period, as well as the Sharpe Ratios (1.54) relative to the market index. We discuss the results and possible future work in Section 1.6.

The work in this chapter is based on collaborative work with Ioannis Karatzas, and has been published as Vervuurt and Karatzas (Ann Finance 11(3-4):411–432, 2015). The largest part of this chapter is an adaptation of this article.

1.1 Outperformance under no-failure

Here is the main result of this chapter.

1.1.1 Theorem. *In the market model* (0.1), *and under the assumptions* ($\mathrm{ND}^{\varepsilon}$) *and* ($\mathrm{NF}^{\varphi}$)*, the diversity-weighted portfolio* $\pi^{(p)}(\cdot)$ *with parameter*

$$p \in \left(\frac{\log n}{\log(n\varphi)}\,,\, 0 \right) \tag{1.3}$$

is a strong arbitrage relative to the market $\mu(\cdot)$ *over the time-horizon* $[0,T]$*, for any*[2]

$$T > \frac{-2n\log(n\varphi)}{\varepsilon(1-p)\big(n-(n\varphi)^p\big)}\,. \tag{1.4}$$

Proof. We apply the theory of functionally-generated portfolios as in Section 0.1.3. For any $p \neq 0$, it is checked that the portfolio (1.1) is generated by the function

$$\mathbf{G}_p : x \mapsto \left(\sum_{i=1}^{n} x_i^p \right)^{1/p}. \tag{1.5}$$

We apply the method of Lagrange multipliers to maximize this function $\mathbf{G}_p$ for $p<0$ subject to $\sum_{i=1}^{n} x_i = 1$; this gives $x_i = 1/n,\ i=1,\ldots,n$. We may therefore write that, under (NF^{φ}), the generating function admits the lower and upper bounds

$$n^{1-p} = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^p \leq \sum_{i=1}^{n} \big(\mu_i(t)\big)^p = \big(\mathbf{G}_p(\mu(t)\big)^p < \sum_{i=1}^{n} \varphi^p = n\varphi^p. \tag{1.6}$$

Hence, for any $T \in (0,\infty)$, we have the lower bound

$$\log\left(\frac{\mathbf{G}_p(\mu(T))}{\mathbf{G}_p(\mu(0))} \right) > \log(n\varphi), \tag{1.7}$$

a negative number. Using assumption (NF^{φ}) and the lower bound from (1.6), we obtain

$$\pi_{(1)}^{(p)}(t) = \frac{(\mu_{(n)}(t))^p}{\sum_{i=1}^{n}(\mu_i(t))^p} < \frac{\varphi^p}{n^{1-p}} = \frac{(n\varphi)^p}{n} < 1, \tag{1.8}$$

[2] See Section 1.6 for a back-of-the-envelope calculation of a typical value of this minimal time to outperformance of the market.

where the last (strict) inequality follows from (1.3). In conjunction with (1.8), the inequality (0.9) gives

$$\int_0^T \gamma^*_{\pi^{(p)}}(t)\,\mathrm{d}t \;\geq\; \frac{\varepsilon}{2}\int_0^T \big(1-\pi^{(p)}_{(1)}(t)\big)\,\mathrm{d}t \;>\; \frac{\varepsilon}{2}\,T\left(1-\frac{(n\varphi)^p}{n}\right). \tag{1.9}$$

Finally, straightforward computation shows that the drift process of the portfolio (1.1), as defined in (0.12), is equal for all $p \in \mathbb{R}$ to

$$\mathfrak{g}_p(\cdot) = (1-p)\,\gamma^*_{\pi^{(p)}}(\cdot). \tag{1.10}$$

We apply now (1.10), (1.7) and (1.9) to Fernholz's master equation (0.11), and conclude that the relative performance of $\pi^{(p)}(\cdot)$ over $[0,T]$, with respect to the market, is given by

$$\begin{aligned}\log\left(\frac{V^{\pi^{(p)}}(T)}{V^{\mu}(T)}\right) &= \log\left(\frac{\mathbf{G}_p(\mu(T))}{\mathbf{G}_p(\mu(0))}\right) + (1-p)\int_0^T \gamma^*_{\pi^{(p)}}(t)\,\mathrm{d}t \\ &> \log(n\varphi) + (1-p)\,\frac{\varepsilon}{2}\,T\left(1-\frac{(n\varphi)^p}{n}\right) \\ &> 0, \qquad \mathbb{P}\text{-a.s.,}\end{aligned} \tag{1.11}$$

provided T satisfies (1.4). Hence the portfolio $\pi^{(p)}(\cdot)$ outperforms the market strongly over sufficiently long time-horizons $[0,T]$, as indicated in (1.4). □

The above proof demonstrates the typical construction of a relative arbitrage in SPT: one bounds each of the terms in the master equation from below, and then finds the minimal value of T for which this lower bound is positive.

1.2 Rank-based variants

Real markets typically do *not* satisfy the (NF$^{\varphi}$) assumption (as stocks can default), so it is desirable to weaken this condition of Theorem 1.1.1. One possible modification of (NF$^{\varphi}$), is to state that "no-failure" only holds for the $\mathfrak{m}$ stocks ranked highest by capitalization, for some fixed $2 < \mathfrak{m} < n$, namely

$$\exists\,\kappa \in (0, 1/\mathfrak{m}) \;\text{ such that }\; \mathbb{P}\big(\,\mu_{(\mathfrak{m})}(t) > \kappa\,,\;\forall\, t \in [0,T]\,\big) \;=\; 1\,. \tag{LF$^{\kappa}$}$$

We call this condition (LF$^{\kappa}$) the *limited-failure* condition, as it postulates that no more than $n-\mathfrak{m}$ companies will "go bankrupt" by time T. In this context, bankruptcy of company i during the time-horizon $[0,T]$ is defined as the event $\{\exists\; t \in [0,T]$ such that $\mu_i(t) \leq \kappa\}$. We also note that (LF$^{\kappa}$) implies the diversity condition with parameter $(\mathfrak{m}-1)\kappa$.

1.2.1 Remark. As was noted by V. Papathanakos (private communication), the diversity condition (D^δ) with parameter $\delta \in (0,1)$ in turn implies (LF^κ) with parameters

$$\mathfrak{m} = \left\lfloor \frac{1}{1-\delta} \right\rfloor \qquad \text{and} \qquad \kappa = \frac{1-(\mathfrak{m}-1)(1-\delta)}{n-(\mathfrak{m}-1)} < \frac{1}{\mathfrak{m}}\,, \tag{1.12}$$

with $\lfloor x \rfloor$ the largest integer less than or equal to $x \in \mathbb{R}$. To see this, we note that, as long as $(k-1)(1-\delta) < 1$, for $1 < k \leq n$, we have the implication

$$\mu_{(k-1)}(t) < 1-\delta \quad \Rightarrow \quad \mu_{(k)}(t) > \frac{1-(k-1)(1-\delta)}{n-(k-1)}\,. \tag{1.13}$$

We identify the inequality on the right hand side of (1.13) as a version of (LF^κ). The largest integer for which this lower bound is positive, is $k = \lfloor 1/(1-\delta) \rfloor$, giving (1.12). □

1.2.1 A diversity-weighted portfolio of large stocks

Under the assumption (LF^κ), one can attempt to construct a relative arbitrage using a variant (introduced in Example 4.2 of Fernholz (2001)) of the diversity-weighted portfolio with parameter $p = r \in \mathbb{R}$ which only invests in the $m = \mathfrak{m}$ highest-ranked stocks (and thus naturally avoids investing in "deaulting" stocks), namely:

$$\pi^{\#}_{p_t(k)}(t) \;=\; \begin{cases} \dfrac{(\mu_{(k)}(t))^r}{\sum_{\ell=1}^m (\mu_{(\ell)}(t))^r}\,, & k = 1, \ldots, m, \\ 0, & k = m+1, \ldots, n. \end{cases} \tag{1.14}$$

Here, $p_t(k)$ is the index of the stock ranked k^{th} at time t, as in (0.13).

When attempting to construct a relative arbitrage with the portfolio (1.14), one encounters a problem. Namely, an application of the rank-based master equation, see Theorem 0.1.7, asserts that the following identity holds for this rank-based portfolio:

$$\begin{aligned} \log\left(\frac{V^{\pi^{\#}}(T)}{V^{\mu}(T)}\right) = \log\left(\frac{\mathcal{G}^{\#}_r(\mu(T))}{\mathcal{G}^{\#}_r(\mu(0))}\right) + (1-r)\int_0^T \gamma^*_{\pi^{\#}}(t)\,\mathrm{d}t \\ - \int_0^T \frac{\pi^{\#}_{p_t(m)}(t)}{2}\,\mathrm{d}\mathfrak{L}^{m,m+1}(t), \end{aligned} \tag{1.15}$$

with $\mathcal{G}^{\#}_r$ the generating function of $\pi^{\#}(\cdot)$, and $\mathfrak{L}^{m,m+1}$ the local time of (0.14); compare with (1.11). Due to the unbounded nature of semimartingale local time, the final term in (1.15) (referred to as "leakage" by Fernholz (2001)) admits no obvious almost sure bound. Thus there is no lower bound on the market-relative performance of $\pi^{\#}(\cdot)$ that holds under reasonable conditions.

Since in real markets this local time term is typically small, we do still expect this portfolio to have a good performance, and therefore include it in our empirical study — see Section 1.5.

1.2.2 A diversity-weighted portfolio of small stocks

Let us then fix $1 \leq m < n$, and define a *small-stock* diversity-weighted portfolio by

$$\pi^\flat_{p_t(k)}(t) = \begin{cases} 0, & k = 1, \ldots, m, \\ \dfrac{(\mu_{(k)}(t))^r}{\sum_{\ell=m+1}^n (\mu_{(\ell)}(t))^r}, & k = m+1, \ldots, n. \end{cases} \tag{1.16}$$

With this new dispensation, the local time term in (1.15) changes sign for $\pi^\flat(\cdot)$ (see equation (1.21) below) and the problems mentioned in the previous subsection disappear. In fact, we can show that the new portfolio in (1.16) outperforms the market under the assumptions (ND$^\varepsilon$), (LF$^\kappa$), for certain positive values of the parameter r; as well as under the assumptions (ND$^\varepsilon$), (NF$^\varphi$), when r is within an appropriate range of negative values.

1.2.2 Proposition. *We place ourselves in the context of the market model* (0.1)*, and under Assumption 0.1.6 and the condition* (ND$^\varepsilon$)*.*

(i) *If the condition* (LF$^\kappa$) *also holds, the small-stock diversity-weighted portfolio of* (1.16) *with positive parameter*

$$r \in \left(-\frac{\log(2)}{\log((m+1)\kappa)}\,,\, 1 \right) \qquad \textit{and} \qquad m = \mathfrak{m} - 2 \tag{1.17}$$

is a strong relative arbitrage with respect to the market $\mu(\cdot)$ over the time-horizon $[0,T]$, provided that

$$\kappa < \frac{1}{2(m+1)} \quad \textit{and} \quad T > T^\flat_+ := \frac{4\left[\log\left(\frac{n-m}{2}\right) - r\log\big((m+1)\kappa\big)\right]}{\varepsilon r(1-r)\big(2-(m+1)^{-r}\kappa^{-r}\big)}\,. \tag{1.18}$$

(ii) *If* (NF$^\varphi$) *also holds, the small-stock diversity-weighted portfolio of* (1.16) *with negative parameter*

$$r \in \left(\frac{\log(n-m)}{\log((m+1)\varphi)}\,,\, 0 \right) \tag{1.19}$$

is a strong arbitrage relative to the market $\mu(\cdot)$ over the time-horizon $[0,T]$, provided

$$T > T^\flat_- := \frac{-2(n-m)\log((m+1)\varphi)}{\varepsilon(1-r)\big(n-m-(m+1)^r\varphi^r\big)}\,. \tag{1.20}$$

Proof. The portfolio (1.16) is generated by the function

$$\mathcal{G}_r(x) := \Bigg(\sum_{\ell=m+1}^{n} \big(x_{(\ell)}\big)^r \Bigg)^{1/r},$$

a rank-based variant of (1.5). In a manner analogous to (1.15), Theorem 0.1.7 implies the following master equation for the performance of our small-stock portfolio:

$$\log\left(\frac{V^{\pi^\flat}(T)}{V^{\mu}(T)}\right) = \log\left(\frac{\mathcal{G}_r(\mu(T))}{\mathcal{G}_r(\mu(0))}\right) + (1-r)\int_0^T \gamma^*_{\pi^\flat}(t)\,\mathrm{d}t \qquad (1.21)$$
$$+ \int_0^T \frac{\pi^\flat_{p_t(m)}(t)}{2}\,\mathrm{d}\mathfrak{L}^{m,m+1}(t).$$

We note the change of sign for the last term, when juxtaposed with the analogue (1.15) of this equation for the large-stock portfolio in (1.14).

Case (i): With $r \in \big(-\log(2)/\log((m+1)\kappa), 1\big)$ as in (1.17), we have

$$\kappa < \frac{1}{2(m+1)} \qquad \Longrightarrow \qquad r > -\frac{\log(2)}{\log\big((m+1)\kappa\big)} > 0\,. \qquad (1.22)$$

Recalling (LF^κ) and $m = \mathfrak{m} - 2$, we have the following bounds

$$\begin{aligned} 2\kappa^r &< \big(\mu_{(\mathfrak{m}-1)}(t)\big)^r + \big(\mu_{(\mathfrak{m})}(t)\big)^r + \sum_{\ell=\mathfrak{m}+1}^{n} \big(\mu_{(\ell)}(t)\big)^r \\ &= \sum_{\ell=m+1}^{n} \big(\mu_{(\ell)}(t)\big)^r = \big(\mathcal{G}_r(\mu(t))\big)^r \qquad (1.23) \\ &\le \sum_{\ell=m+1}^{n} \left(\frac{1}{m+1}\right)^r = (n-m)\,(m+1)^{-r} \end{aligned}$$

for the generating function $\mathcal{G}_r$, which lead to the lower bound

$$\log\left(\frac{\mathcal{G}_r(\mu(T))}{\mathcal{G}_r(\mu(0))}\right) > \log\big((m+1)\kappa\big) - \frac{1}{r}\log\left(\frac{n-m}{2}\right). \qquad (1.24)$$

Note that, using the lower bound in (1.23) and $\mu_{(k)}(t) \le 1/k$, $k = 1, \dots, n$, we obtain

$$\pi^\flat_{(1)}(t) = \frac{\big(\mu_{(m+1)}(t)\big)^r}{\sum_{\ell=m+1}^{n}\big(\mu_{(\ell)}(t)\big)^r} < \frac{(m+1)^{-r}}{2\kappa^r} < 1, \qquad (1.25)$$

where the last bound follows from the inequalities in (1.22).

The non-degeneracy condition (ND^ε) now gives

$$\int_0^T \gamma^*_{\pi^\flat}(t)\,\mathrm{d}t \ge \frac{\varepsilon}{2}\int_0^T \big(1 - \pi^\flat_{(1)}(t)\big)\,\mathrm{d}t > \frac{\varepsilon}{2}\,T\left(1 - \frac{1}{2(m+1)^r\kappa^r}\right) \qquad (1.26)$$

in conjunction with (0.9) and (1.25), whereas the nonnegativity of $\pi^\flat_{p_t(m)}(\cdot)$, coupled with the nondecrease of the local time $\mathfrak{L}^{m,m+1}(\cdot)$, shows that

$$\int_0^T \frac{\pi^\flat_{p_t(m)}(t)}{2}\,\mathrm{d}\mathfrak{L}^{m,m+1}(t) \,\geq\, 0\,. \tag{1.27}$$

We use this, and apply (1.24) and (1.26) to the rank-based master equation (1.21), to obtain

$$\begin{aligned}\log\left(\frac{V^{\pi^\flat}(T)}{V^\mu(T)}\right) &> \log\big((m+1)\kappa\big) - \frac{1}{r}\log\left(\frac{n-m}{2}\right)\\ &\quad + \frac{\varepsilon}{2}(1-r)\,T\left(1-\frac{1}{2(m+1)^r\kappa^r}\right)\\ &> 0\,, \qquad \mathbb{P}\text{-a.s.,}\end{aligned}$$

if and only if $T > T^\flat_+$ as defined in (1.18). We conclude that, under these conditions, $\pi^\flat(\cdot)$ strongly outperforms the market over the horizon $[0,T]$.

Case (ii): With $r \in \big(\log(n-m)/\log(n\kappa), 0\big)$ and under the condition (NF$^\varphi$), we have the following bounds

$$(n-m)\,(m+1)^{-r} \leq \sum_{\ell=m+1}^{n}(\mu_{(\ell)}(t))^r = \big(\mathcal{G}_r(\mu(t))\big)^r < (n-m)\,\varphi^r\,. \tag{1.28}$$

For the first inequality, we have used the simple fact that $\mu_{(\ell)}(t) \leq \mu_{(m+1)}(t) \leq 1/(m+1)$ holds for $\ell = m+1, \cdots, n$. Whereas, from (1.19), we have

$$\pi^\flat_{(1)}(t) = \frac{(\mu_{(n)}(t))^r}{\sum_{\ell=m+1}^{n}(\mu_{(\ell)}(t))^r} < \frac{\varphi^r}{(n-m)\,(m+1)^{-r}} < 1 \tag{1.29}$$

by analogy with (1.8). The inequalities (1.28) lead to the lower bound

$$\log\left(\frac{\mathcal{G}_r(\mu(T))}{\mathcal{G}_r(\mu(0))}\right) > \log\big((m+1)\varphi\big)\,; \tag{1.30}$$

and the non-degeneracy condition (ND$^\varepsilon$), in conjunction with (0.9) and the upper bound (1.29) on the largest portfolio weight $\pi^\flat_{(1)}(\cdot)$, give

$$\int_0^T \gamma^*_{\pi^\flat}(t)\,\mathrm{d}t \,\geq\, \frac{\varepsilon}{2}\int_0^T\big(1-\pi^\flat_{(1)}(t)\big)\,\mathrm{d}t \,>\, \frac{\varepsilon\,T}{2}\left(1-\frac{(m+1)^r\varphi^r}{n-m}\right). \tag{1.31}$$

Again, we use (1.27) and apply (1.30) and (1.31) to the master equation (1.21), to obtain

$$\begin{aligned}\log\left(\frac{V^{\pi^\flat}(T)}{V^\mu(T)}\right) &> \log\big((m+1)\varphi\big) + \frac{\varepsilon\,T}{2}(1-r)\left(1-\frac{(m+1)^r\varphi^r}{n-m}\right)\\ &> 0, \qquad \mathbb{P}\text{-a.s.,}\end{aligned}$$

provided $T > T^\flat_-$ as defined in (1.20). Hence this small-stock, negative-parameter portfolio $\pi^\flat(\cdot)$ outperforms the market over the horizon $[0, T]$. $\square$

To the best of the author's knowledge, the above is the first construction of a rank-based relative arbitrage to date.

1.3 Negative vs. positive parameter

We show now that the diversity-weighted portfolio (1.1) with $p \in (0, 1)$ is outperformed by its negative-parameter counterpart under sufficient conditions. In the following Proposition 1.3.1, besides (ND$^\varepsilon$) we will impose also an *upper* bound on the eigenvalues of the covariance matrix $a(\cdot)$, in the form of the *bounded variance* assumption:

$$\exists\, K > 0 \;\text{ such that: }\; \xi'\sigma(t)\sigma'(t)\xi \le K||\xi||^2, \quad \forall\, \xi \in \mathbb{R}^n,\; t \ge 0\,; \quad \mathbb{P}\text{-a.s.} \qquad (\mathrm{BV}^K)$$

By Lemma 3.5 of Fernholz and Karatzas (2009), the bounded variance condition (BV^K) implies that for long-only portfolios $\pi(\cdot)$ we have the almost sure inequality

$$\gamma^*_\pi(t) \le 2K\big(1 - \pi_{(1)}(t)\big), \quad \forall\, t \ge 0; \tag{1.32}$$

compare with (0.9).

1.3.1 Proposition. *Let us place ourselves in the market model* (0.1), *under the conditions* (ND$^\varepsilon$), (BV^K) *and* (NF$^\varphi$).

The diversity-weighted portfolio $\pi^{(p^-)}(\cdot)$ with negative parameter

$$p^- \in \left(\frac{\log n}{\log(n\varphi)}\,,\, 0 \right)$$

is then a strong arbitrage relative to the diversity-weighted portfolio $\pi^{(p^+)}(\cdot)$ with positive parameter

$$p^+ \in \left(\max\left\{0, 1 - \frac{\varepsilon(n - (n\varphi)^{p^-})(1 - p^-)}{4K(n-1)}\right\}, 1 \right), \tag{1.33}$$

over any horizon $[0, T]$ of length

$$T > \frac{-2\log(n\varphi)}{C}\,. \tag{1.34}$$

Here the positive constant C is defined as

$$C := \frac{\varepsilon}{2}\left(1 - \frac{(n\varphi)^{p^-}}{n}\right)\big(1 - p^-\big) - \frac{2K}{n}\,(n-1)(1 - p^+)\,. \tag{1.35}$$

Proof. To simplify notation a bit, we write $\pi^{\pm}(\cdot)$ and $\mathbf{G}_{\pm}$ for $\pi^{(p^{\pm})}(\cdot)$ and $\mathbf{G}_{p^{\pm}}$, respectively. Note that for $p^+ > 0$ the inequalities in (1.6) reverse, i.e.,

$$n\varphi^{p^+} < \big(\mathbf{G}_+(\mu(t))\big)^{p^+} \leq n^{1-p^+}, \tag{1.36}$$

which again gives the lower bound of (1.7) for $p = p^+$. Using (1.32) with the observation $\pi^+_{(1)}(t) \geq 1/n$, we get that

$$\int_0^T \gamma^*_{\pi^+}(t)\,\mathrm{d}t \;\leq\; 2K\int_0^T \big(1-\pi^+_{(1)}(t)\big)\,\mathrm{d}t \;\leq\; 2KT\big(1-(1/n)\big). \tag{1.37}$$

Hence, recalling (1.10), we see that by virtue of (1.36) and (1.37) the master equation (0.11) for $\pi(\cdot) = \pi^+(\cdot)$ leads to the upper bound

$$\log\left(\frac{V^{\pi^+}(T)}{V^{\mu}(T)}\right) < -\log(n\varphi) + 2(1-p^+)KT\big(1-(1/n)\big), \quad \mathbb{P}\text{-a.s.} \tag{1.38}$$

Combining (1.11) and (1.38), we find that

$$\begin{aligned}\log\left(\frac{V^{\pi^-}(T)}{V^{\pi^+}(T)}\right) &= \log\left(\frac{V^{\pi^-}(T)}{V^{\mu}(T)}\right) - \log\left(\frac{V^{\pi^+}(T)}{V^{\mu}(T)}\right)\\ &> 2\log(n\varphi) + CT\\ &> 0, \quad \mathbb{P}\text{-a.s.,}\end{aligned}$$

provided that

$$CT > -2\log(n\varphi) > 0\,. \tag{1.39}$$

Here, the constant C is given by (1.35). An easy calculation shows that (1.33) implies $C > 0$, whereas the last inequality in (1.39) comes from $\varphi < 1/n$. It follows that the first inequality in (1.39) is equivalent to the condition posited in (1.34), and that in this case $\pi^-(\cdot)$ outperforms $\pi^+(\cdot)$ strongly over the time-horizon $[0,T]$. □

Proposition 1.3.1 shows that, as long as the diversity-weighted portfolio $\pi^{(p^+)}(\cdot)$ is "sufficiently similar" to the market portfolio $\mu(\cdot)$ (and thus "far enough" from $\pi^{(p^-)}(\cdot)$), it is outperformed strongly, over sufficiently long time-horizons, by the diversity-weighted portfolio $\pi^{(p^-)}(\cdot)$ with negative parameter — provided, of course, that the aforementioned conditions on the volatility structure and non-failure of stocks hold.

1.3.2 Remark. Compared to diversity (D^{δ}), the stronger no-failure condition (NF^{φ}) implies a possibly shorter minimal time horizon than that of (1.2), over which the diversity-weighted portfolio with parameter $p^+ \in (0,1)$ is guaranteed to strongly outperform the market. Namely, the fact that $\mathbf{G}_{p^+}(x) \geq 1,\ \forall x \in \Delta^n_+$, together with (1.36), implies that

$$\max\left\{1, n\varphi^{p^+}\right\} \leq \left(\mathbf{G}_{p^+}(\mu(t))\right)^{p^+} \leq n^{1-p^+}.$$

Recall that (NF^φ) implies diversity (D^δ) with parameter $\delta = (n-1)\varphi$. It follows from (1.2), and using an argument similar to that in the proof of Theorem 1.1.1, that the positive-parameter diversity-weighted portfolio is a strong arbitrage relative to the market over horizons $[0, T]$, with

$$T > \min\left\{\frac{2\log n}{\varepsilon\delta p}, \frac{-2\log\left(n\delta/(n-1)\right)}{\varepsilon\delta(1-p)}\right\}.$$

This is an improvement of (1.2) if and only if $\varphi > n^{-1/p^+}$. $\square$

1.4 Mixing

Since the no-failure assumption (NF^φ) does not hold in real markets, it is of interest to find variants of (1.1) which exhibit similar performance, but require weaker assumptions. The rank-based variant in Proposition 1.2.2 is one attempt at this; Proposition 1.4.1 below is another, and establishes a relative-arbitrage result for a dynamic linear combination of positive- and negative-parameter diversity-weighted portfolios. This novel method of combining portfolios retains their advantageous properties, and could be applied more generally when constructing portfolios.

1.4.1 Proposition. *Define the portfolio*

$$\widehat{\pi}(t) \;=\; \mathfrak{p}(t)\pi^+(t) + (1-\mathfrak{p}(t))\pi^-(t), \tag{1.40}$$

with $\pi^\pm(\cdot) = \pi^{(p^\pm)}(\cdot)$ *diversity-weighted portfolios as defined in* (1.1) *with* $p^+ \in (0,1)$ *and* $p^- < 0$, *and the mixing proportion* $\mathfrak{p}(\cdot)$ *given by*

$$\mathfrak{p}(t) = \frac{\mathbf{G}_{p^+}(\mu(t))}{\mathbf{G}_{p^+}(\mu(t)) + \mathbf{G}_{p^-}(\mu(t))} \in (0,1). \tag{1.41}$$

In the market model (0.1), *with assumptions* (ND^ε) *and* (D^δ), *the portfolio* $\widehat{\pi}(\cdot)$ *is a strong arbitrage relative to the market* $\mu(\cdot)$, *over time-horizons* $[0, T]$ *with*

$$T \;>\; \mathfrak{T} \;:=\; \frac{2(1+n^{1/p^- -1})\log\left(n^{1/p^+ -1} + n^{1/p^- -1}\right)}{\varepsilon\delta(1-p^+)}\,. \tag{1.42}$$

Proof. The portfolio (1.40) is generated by the function $\widehat{\mathbf{G}} = \mathbf{G}_+ + \mathbf{G}_-$, with $\mathbf{G}_\pm := \mathbf{G}_{p^\pm}$ the generating functions of $\pi^\pm(\cdot) := \pi^{(p^\pm)}(\cdot)$ as in (1.5); let $\mathfrak{g}_\pm(\cdot) := \mathfrak{g}_{p^\pm}(\cdot)$ be their drift processes as in (1.10).

By (0.12), the drift process of the composite portfolio $\widehat{\pi}(\cdot)$ is then

$$\begin{aligned}
\widehat{\mathfrak{g}}(t) &= \frac{-1}{2\widehat{\mathbf{G}}(\mu(t))} \sum_{i,j=1}^{n} \mathrm{D}_{ij}^2 \widehat{\mathbf{G}}(\mu(t))\mu_i(t)\mu_j(t)\tau_{ij}^\mu(t) \\
&= \mathfrak{p}(t)\, \frac{-1}{2\mathbf{G}_+(\mu(t))} \sum_{i,j=1}^{n} \mathrm{D}_{ij}^2 \mathbf{G}_+(\mu(t))\mu_i(t)\mu_j(t)\tau_{ij}^\mu(t) \\
&\quad + (1-\mathfrak{p}(t))\, \frac{-1}{2\mathbf{G}_-(\mu(t))} \sum_{i,j=1}^{n} \mathrm{D}_{ij}^2 \mathbf{G}_-(\mu(t))\mu_i(t)\mu_j(t)\tau_{ij}^\mu(t) \\
&= \mathfrak{p}(t)\mathfrak{g}_+(t) + (1-\mathfrak{p}(t))\mathfrak{g}_-(t) \\
&= (1-p^+)\mathfrak{p}(t)\gamma_{\pi^+}^*(t) + (1-p^-)(1-\mathfrak{p}(t))\gamma_{\pi^-}^*(t)\,;
\end{aligned}$$

the final step follows from (1.10). We note that

$$\gamma_{\pi^-}^*(t) \geq 0,$$

as this holds for any long-only portfolio by Lemma 3.3 of Fernholz and Karatzas (2009), which together with the observation that $\mathfrak{p}(t) < 1$ allows us to obtain the bound

$$\widehat{\mathfrak{g}}(t) \geq (1-p^+)\mathfrak{p}(t)\gamma_{\pi^+}^*(t). \tag{1.43}$$

We note the simple bounds

$$1 = \sum_{i=1}^{n} \mu_i(t) \leq \sum_{i=1}^{n} \big(\mu_i(t)\big)^{p^+} = \big(\mathbf{G}_+(\mu(t)\big)^{p^+} \leq n^{1-p^+}, \tag{1.44}$$

and that the lower bound in (1.6) holds even without (NF^φ), so now

$$0 \;\leq\; \left(\sum_{i=1}^{n} \big(\mu_i(t)\big)^{p^-}\right)^{1/p^-} \;=\; \mathbf{G}_-(\mu(t)) \;\leq\; n^{1/p^- -1}\,. \tag{1.45}$$

Using the lower bound from (1.44), and the upper bound from (1.45), we assert that

$$\mathfrak{p}(t) = \left(1 + \frac{\mathbf{G}_-(\mu(t))}{\mathbf{G}_+(\mu(t))}\right)^{-1} \geq \frac{1}{1+n^{1/p^- -1}} > 0. \tag{1.46}$$

One can easily see that in the positive-parameter diversity-weighted portfolio, one's proportion of wealth invested relative to the market portfolio is diminished for large-cap stocks, and increased for small-capitalization stocks; therefore

$$\pi_{(1)}^+(t) = \frac{(\mu_{(1)}(t))^p}{\sum_{i=1}^n (\mu_i(t))^p} \leq \mu_{(1)}(t). \tag{1.47}$$

The non-degeneracy condition ($\mathrm{ND}^{\varepsilon}$) implies (0.9), which by (1.47) and the diversity assumption (D^{δ}) leads to

$$\gamma^*_{\pi^+}(t) \geq \frac{\varepsilon}{2}\big(1-\pi^+_{(1)}(t)\big) \geq \frac{\varepsilon}{2}\big(1-\mu_{(1)}(t)\big) > \frac{\varepsilon}{2}\,\delta\,. \tag{1.48}$$

Finally, applying (1.43) to the master equation (0.11), and then using the bounds (1.44), (1.45) and (1.46), (1.48), we conclude that

$$\begin{aligned}\log\left(\frac{V^{\widehat{\pi}}(T)}{V^{\mu}(T)}\right) &\geq \log\left(\frac{\mathbf{G}_+(\mu(T))+\mathbf{G}_-(\mu(T))}{\mathbf{G}_+(\mu(0))+\mathbf{G}_-(\mu(0))}\right) + (1-p^+)\int_0^T \mathfrak{p}(t)\gamma^*_{\pi^+}(t)\,\mathrm{d}t \\ &> -\log\left(n^{1/p^+-1}+n^{1/p^--1}\right) + \frac{\varepsilon}{2}\left(1-p^+\right)\frac{\delta\,T}{1+n^{1/p^--1}} \\ &> 0\,, \quad \mathbb{P}\text{-a.s.}\end{aligned} \tag{1.49}$$

under the condition that T satisfies (1.42). $\square$

We have shown that the long-only portfolio (1.40) strongly outperforms the market over sufficiently long time-horizons, under the assumptions of non-degeneracy and diversity. This is a property that the portfolio $\pi^+(\cdot)$ also has on its own, as proved in the Appendix of Fernholz et al. (2005).

We remark also that, since the "threshold" $\mathfrak{T}$ of (1.42) is strictly decreasing in p^-, and the lower bound (1.49) is strictly increasing in p^- for fixed T, our result becomes stronger the closer the negative parameter p^- gets to the origin. As we take $p^- \uparrow 0$, we recover the well-known result that $\widehat{\pi}(\cdot) = \pi^{(p^+)}(\cdot)$ is a strong arbitrage relative to the market.

1.4.2 Remark. The statement of Proposition 1.4.1 can be strengthened by weakening the diversity condition (D^{δ}) to that of *weak diversity* over the horizon $[0,T]$. This notion is defined in Equation (4.2) of Fernholz et al. (2005) as

$$\exists\ \delta \in (0,1) \quad \text{such that} \quad \mathbb{P}\left(\frac{1}{T}\int_0^T \mu_{(1)}(t)\,\mathrm{d}t < 1-\delta\right) = 1\,.$$

Under this weaker assumption, it is straightforward to see that the following modification of (1.48) will hold, thus maintaining the validity of the proof:

$$\int_0^T \gamma^*_{\pi^+}(t)\,\mathrm{d}t \geq \frac{\varepsilon}{2}\int_0^T (1-\pi^+_{(1)}(t))\,\mathrm{d}t \geq \frac{\varepsilon}{2}\int_0^T (1-\mu_{(1)}(t))\,\mathrm{d}t > \frac{\varepsilon}{2}\,\delta\,T.$$

$\square$

1.4.3 Remark. We should note here that, since both the (NF$^\varphi$) and (LF$^\kappa$) conditions imply diversity, by Lemma 8.1 and Example 8.3 of Fernholz et al. (2005) it follows that *short-term* relative arbitrage exists. That is, using the "mirror portfolios" (in which a short position in an underperforming portfolio is combined with a long position in the market, in such a way as to create a long-only portfolio) of Fernholz, Karatzas and Kardaras, one can construct a portfolio which outperforms the market over *any* time-horizon $[0, T]$.

1.4.4 Remark (Two portfolios with a threshold). We mention briefly two other variants of the negative-parameter portfolio (1.1), which exhibit similar behavior for mid- and upper-range market weights (in the sense that they invest in stocks which decrease in value relative to the market, and sell stock when a company's market weight increases), but start selling stock once a firm's market weight falls below a certain threshold. The idea behind this threshold is that it represents a value below which the investor fears bankruptcy of the firm, and wishes to liquidate their position in the firm so as to minimize losses. The two portfolios we propose have weights

$$\gamma_i(t) = \frac{\mu_i(t)^k e^{-\mu_i(t)/\theta}}{\sum_{j=1}^n \mu_j(t)^k e^{-\mu_j(t)/\theta}}\,, \qquad i = 1, \ldots, n \tag{1.50}$$

and

$$B_i(t) = \frac{\mu_i(t)^\alpha (1 - \mu_i(t))^\beta}{\sum_{j=1}^n \mu_j(t)^\alpha (1 - \mu_j(t))^\beta}\,, \qquad i = 1, \ldots, n\,. \tag{1.51}$$

Here, k, θ and α, β are all positive constants; the threshold values below which the portfolio weights become *increasing* functions of market weights are θk and $\alpha/(\alpha + \beta)$, respectively. So far, our results regarding these portfolios remain restricted to empirical ones (see Section 1.5). We establish theoretical results for the portfolio (1.51) with $\alpha = \beta = 1$ in Example 4.2.2 of Chapter 4; a result for the portfolio $\gamma(\cdot)$ might follow in a way similar to that of Proposition 1 in Banner and Fernholz (2008).

1.5 Empirical results

We test the validity and applicability of our theoretical results by conducting an empirical study of the performance of our portfolios using historical market data. More precisely, we back-test by investing according to the portfolios studied here in the $n = 500$ daily constituents of the S&P 500 index for $T = 6301$ consecutive trading days between 1 January 1990 and 31 December 2014. We obtain these data from the Compustat and CRSP data sets (these data sets are obtainable from the website `http://wrds-web.wharton.upenn.edu/wrds/`). We have incorporated dividends and delistings such as mergers, acquisitions and liquidations. For another empirical analysis using this data set, see Section 2.2.

We aim to simulate as realistically as possible the wealth evolution of an investor implementing our portfolios without any additional information, and as such impose on each trade proportional transaction costs equal to 0.5% of the total absolute value traded. In our methodology, we replace the old index constituent by the new one at each constituent change, and transaction costs are only determined by the change in the number of shares held in a constituent index; as a consequence, we do not incur the full transaction costs due to selling the entire old position and establishing the new one, but only the costs due to the *change* in the investor's position. This is an imperfection in our approach, and its resolution would improve the real-world resemblence of the performance of all portfolios.

We rebalance the portfolio using a simple Total Variance criterion, that is, we only rebalance when the total variance "distance"

$$\mathbf{TV}(\widetilde{\pi}(t), \pi(t)) \;=\; \sum_{i=1}^{n} \widetilde{\pi}_i(t) \left|\widetilde{\pi}_i(t) - \pi_i(t)\right| \tag{1.52}$$

between the target portfolio $\pi(\cdot)$ and the portfolio $\widetilde{\pi}(\cdot)$ becomes larger than a certain threshold, which we determine empirically and in-sample — that is, we compute the total return for a range of threshold values for each portfolio, and select the optimal one for that particular strategy. Here, $\widetilde{\pi}(t)$ is the vector of proportions of wealth invested in stocks obtained when the investor does *not* rebalance at time t, that is

$$\widetilde{\pi}_i(t) = \bar{\pi}_i(t-1)\frac{\mathcal{V}^{\bar{\pi}}(t-1)}{\mathcal{V}^{\bar{\pi}}(t)}\frac{X_i(t)}{X_i(t-1)}\,, \qquad i = 1, \ldots, n, \quad t = 2, \ldots, T,$$

where $\bar{\pi}(t-1)$ is the portfolio that was actually implemented in the previous time step. We have written $\mathcal{V}^{\pi}(\cdot)$ for the discrete-time wealth process corresponding to a portfolio $\pi(\cdot)$, defined through

$$\mathcal{V}^{\pi}(0) = 1, \qquad \frac{\mathcal{V}^{\pi}(t+1)}{\mathcal{V}^{\pi}(t)} = \sum_{i=1}^{n} \pi_i(t)\frac{X_i(t+1)}{X_i(t)}\,, \quad t = 0, 1, \ldots, T. \tag{1.53}$$

We identify $X_i(t+1)/X_i(t)$ as the daily return of stock i from day t to $t+1$.

We use `R` for our analysis — the code is available upon request. We summarize our findings in Table 1.1, which displays the average annual relative returns in excess of the market (denoted "Market-RR"), and the Sharpe Ratios of our portfolios over the entire 25-year period that we use; the latter is computed as

$$\text{SharpeRatio}(\pi(\cdot)) = \frac{\text{Mean}(R^\pi)}{\text{StdDev}(R^\pi)} \cdot \sqrt{\frac{T}{25}} \,. \tag{1.54}$$

Here, $R^\pi = \{R^\pi(t),\, t = 1, \ldots, T\}$ are the daily returns of the portfolio π, namely

$$R^\pi(t) = \frac{\mathcal{V}^\pi(t)}{\mathcal{V}^\pi(t-1)} - 1, \quad t = 2, \ldots, T,$$

and the sample standard deviation StdDev is defined as follows for any sequence of numbers $x_1, \ldots, x_k \in \mathbb{R}$ with average $\overline{x}$:

$$\big(\text{StdDev}(x_1, \ldots, x_k)\big)^2 := \frac{1}{k-1} \sum_{i=1}^{k} (x_i - \overline{x})^2.$$

Moreover, we compute the following measure of performance for each portfolio:

$$\widetilde{\gamma}_\pi := \frac{\gamma_\pi}{\text{StdDev}(R^\pi)} \cdot \frac{1}{25} \,,$$

that is, $\widetilde{\gamma}_\pi$ is the total growth rate divided by the sample standard deviation. The former is estimated as

$$\gamma_\pi = \log\left(\frac{\mathcal{V}^\pi(T)}{\mathcal{V}^\pi(1)}\right).$$

Figure 1.1 showcases the wealth processes corresponding to our portfolios.

From Table 1.1 and Figure 1.1, we can see that all portfolios outperform the market by quite a margin (however, only after the year 2000), both in terms of the market-relative return, as well as the Sharpe Ratio. We also note that the simulated realizations of wealth processes are much more volatile than that of the market; the portfolios seem to exploit market growth much better, but also lose value more quickly when the market does poorly. Moreover, the negative-parameter portfolios we study appear to perform better than their positive-parameter versions over the period studied. The reader will notice that the positive-parameter portfolio $\pi^{(p)}(\cdot)$ with $p = 0.5$, and the mixing portfolio $\widehat{\pi}(\cdot)$, give identical results, which is because the weight (1.41) of the positive-parameter diversity-weighted portfolio is always very near 1 (namely $1 - \mathfrak{p}(t) \sim$ 1e-11). Finally, we wish to stress that the above portfolio and threshold parameters were optimized only heuristically, and therefore there is considerable room for improvement (that is, through more systematic optimization)

Table 1.1: Some measures of performance for the studied portfolios when traded between 1 January 1990 and 31 December 2014 on the constituents of the S&P 500 index, over which the market had an average annual return of 9.7190%. We write $\pi^{\mathrm{E}}(\cdot)$ for the equally-weighted portfolio $\pi_i^{\mathrm{E}}(\cdot) = 1/n,\ i = 1, \ldots, n$. The parameter "**TV** Threshold" determines over which value of **TV** as in (1.52) we rebalance, which we determine in-sample by trial-and-error.

Portfolio	**TV** Threshold	Market-RR	Sharpe Ratio	$\widetilde{\gamma}_\pi$
$\mu(\cdot)$	N/A	0%	0.60477	8.1678
$\pi^{\mathrm{E}}(\cdot)$	0.0005	2.2331%	0.70428	9.7127
$\pi^{(p)}(\cdot),\ p = 0.5$	0.0022	1.6125%	0.76766	10.949
$\pi^{(p)}(\cdot),\ p = -0.5$	0.0015	4.9655%	0.78558	10.910
$\pi^{\#}(\cdot),\ r = -0.5,\ m = 470$	0.0025	2.5778%	0.76979	10.873
$\pi^{\flat}(\cdot),\ r = 0.5,\ m = 30$	0.0001	0.9578%	0.64713	8.8215
$\pi^{\flat}(\cdot),\ r = -0.5,\ m = 30$	0.0100	1.7645%	0.67412	9.2114
$\widehat{\pi}(\cdot),\ p^{+} = 0.5,\ p^{-} = -0.5$	0.0022	1.6125%	0.76766	10.949
$\gamma(\cdot),\ k = 0.65,\ \theta = 1\text{e-}4$	0.0020	3.6336%	0.68160	9.0796
$B(\cdot),\ \alpha = 1\text{e-}4,\ \beta = 2$	0.0002	1.8701%	0.67919	9.2920

— the performance of the portfolios is quite sensitive to these parameters, as well as the level of transaction costs. Our empirical study is therefore merely a demonstration of potential outperformance of the market with the portfolios studied.

In Figure 1.1, note the jump in value of the negative-parameter diversity-weighted portfolio in early 1993; this is due to a large position in a small-cap stock that has a very large return over a short period of time. This opportunity is missed by the rank-based portfolios $\pi^{\#}(\cdot)$ and $\pi^{\flat}(\cdot)$ because of their different **TV** thresholds, and the effect is maintained when perturbing the value of p around -0.5. Nonetheless, such an event is not representative of the portfolio's performance, and it increases the portfolio's total return and Sharpe Ratio without reflecting any superior quality of the portfolio. Excluding the one firm which causes the negative-parameter diversity-weighted portfolio's spike diminishes its annualized market-relative return to just below 2%, and its Sharpe Ratio to 0.65.

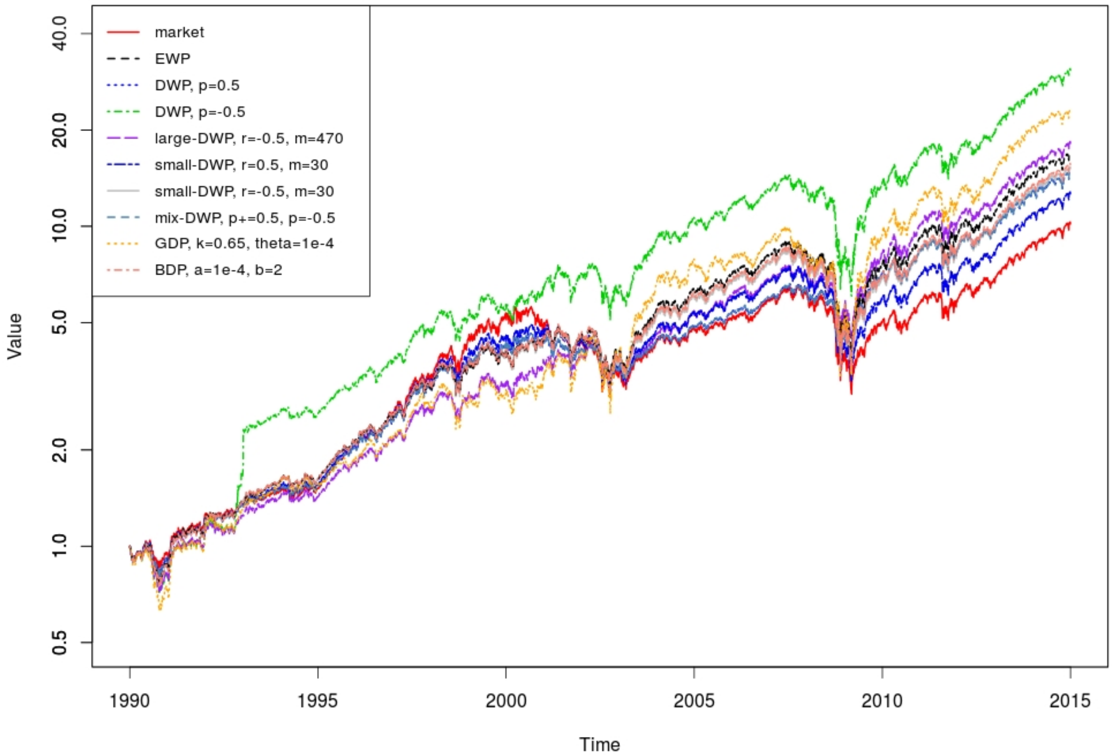

Figure 1.1: The wealth processes corresponding to the portfolios in Table 1.1, namely: the market portfolio $\mu(\cdot)$; the diversity-weighted portfolios $\pi^{(p)}(\cdot)$ with $p = 0$ (i.e. the equally-weighted portfolio), $p = 0.5$, and $p = -0.5$; the rank-based diversity-weighted portfolios $\pi^{\#}(\cdot)$ with $r = -0.5$, $m = 470$ and $\pi^{\flat}(\cdot)$ with $r = 0.5$, $m = 30$, and $r = -0.5$, $m = 30$; the mixed portfolio $\widehat{\pi}(\cdot)$ from (1.40), with $p^{+} = 0.5$, $p^{-} = -0.5$; and the portfolios (1.50) and (1.51), with $k = 0.65$, $\theta =$ 1e-4 and $\alpha =$ 1e-4, $\beta = 2$, respectively.

1.6 Discussion

We point out that all of the results in this chapter require certain assumptions on the volatility structure of the market, as well as on the behavior of the market weights. This demonstrates the fact that functional generation can be applied to construct relative arbitrages from certain conditions on the market; it is a tool that can be used to translate these market properties into suitable portfolios.

The first such condition that we considered, the no-failure condition (NF$^{\varphi}$), does *not* hold in real markets, since typically some companies do fail. One could assume (or predict) that it does hold during certain periods of time, though, and apply the corresponding portfolios during these periods. The smallest observed market weight during the 25-year period that we perform our data study on is $3.92 \times 10^{-4}\%$, hence we may say that (NF^{φ}) holds when there are no defaults with $\varphi = 2.0 \times 10^{-4}\%$. With this parameter for the no-failure condition, and $n = 500$, $p = -0.5$, and $\varepsilon = 1.0\%$, the minimal time to outperformance (1.4) of the negative-parameter diversity-weighted portfolio is computed as 983 trading days, or almost 4 years.

We theoretically incorporate defaults into functional generation in Section 4.5, and algorithmically in Chapter 2, improving its applicability to portfolio construction in the real world. The weaker (LF^{κ}) assumption is an improvement on the (NF^{φ}) condition, and can be argued to hold for m not too close to n — although, to the author's knowledge, there is no mechanism holding this in place, as there is for the diversity assumption (D^{δ}), which can be seen as a consequence of anti-trust regulation. In this regard, see Strong and Fouque (2011), as well as Karatzas and Sarantsev (2014), in which the number of companies is allowed to fluctuate — due to both splits and mergers of companies.

It has been raised by V. Papathanakos (private communication) that from a practitioner's point of view, there are several restrictions on the implementation of negative-parameter diversity-weighted portfolios on a large scale. One of these is that one typically demands that no position consists of more than, say, 1% of the outstanding shares of a security; our portfolios strongly invest in small stocks. Another constraint is related to liquidity: regular rebalancing in a predictable way (which is implied by all functionally-generated portfolios) incurs very high transaction costs. It would be useful to model these phenomena and their influence on portfolio performance.

More generally, it would be of great interest to develop a theory of transaction costs in the framework of Stochastic Portfolio Theory, which would allow one to improve upon, or even optimize over, rebalancing rules given a certain portfolio. A first attempt at this was made by (Fernholz, 2002, Section 6.3), where R. Fernholz estimates the turnover in a diversity-weighted portfolio.

Another idea is to replace the almost sure assumptions (D^{δ}), (LF^{κ}), and (NF^{φ}), by *probabilistic versions* of these assumptions, where the corresponding bounds on market weights hold with a probability close to but smaller than 1. It would be interesting to see whether *probabilistic* relative arbitrages could be constructed, which have a certain "likelihood of outperforming" the market — see Bayraktar et al. (2012a), who study the problem of hedging a given claim with a certain probability, for a first study in this direction.

Moreover, it would be of interest to develop methods for finding the optimal relative arbitrage within the class of functionally-generated portfolios; an attempt at this was made in Pal and Wong (2016), and for more general strategies in Fernholz and Karatzas (2010) and Fernholz and Karatzas (2011). Limitations on the existence of relative arbitrages with respect to certain portfolios have been established, in discrete time, in the recent paper Wong (2015a).

Finally, it is very desirable to incorporate additional information about the index constituents into the portfolio construction; investors typically use more than just the current capitalization of a firm to determine their position in it. We proceed to do exactly that in the next Chapter.

Chapter 2

A data-driven approach to portfolio construction

In Stochastic Portfolio Theory, the generation of portfolios has been performed using functions of only one argument: the current market weights. This choice of portfolio class has many far-reaching and surprising consequences, as was demonstrated in Chapter 1. Namely, under descriptive assumptions on the price dynamics (and, notably, without the assumption of No Arbitrage), functionally-generated portfolios that outperform the market with probability one can be constructed. The tool that enables this construction is Fernholz's master equation (0.11).

In real financial markets, much more information is available than just the current market weights, and investors consider both historical information and other factors in their investment decisions. A first attempt at incorporating additional information into functionally-generated portfolio construction was made by Strong (2014a), who derived a modified master equation for portfolios that are allowed to depend on the current values of processes other than the market weights, provided these are of finite variation. Namely, assuming the Itô model (0.1) for the stock prices, define the process $\mathbf{x}(\cdot) = (\mu(\cdot), F(\cdot))^T$, with $F(\cdot)$ a continuous, $\mathbb{R}^k$-valued, $\mathbb{F}$-progressively measurable process of finite variation, and let $\mathcal{H} \in C^{2,1}(\mathbb{R}^n \times \mathbb{R}^k, \mathbb{R}_+)$. By an application of Theorem 3.1 of Strong (2014a), for any portfolio $\pi(\cdot)$ defined by

$$\pi_i(t) := \Big(\mathrm{D}_i \log \mathcal{H}(\mathbf{x}(t)) + 1 - \sum_{j=1}^{n} \mu_j(t) \mathrm{D}_j \log \mathcal{H}(\mathbf{x}(t))\Big) \cdot \mu_i(t)\,, \quad i = 1, \ldots, n,$$

(compare with (0.10)) the following modified master equation holds:

$$\log\left(\frac{V^{\pi}(T)}{V^{\mu}(T)}\right) = \log\left(\frac{\mathcal{H}(\mathbf{x}(T))}{\mathcal{H}(\mathbf{x}(0))}\right) + \int_0^T \tilde{\mathfrak{g}}(t)\,\mathrm{d}t \\ - \int_0^T \sum_{l=1}^{k} \mathrm{D}_{n+l} \log \mathcal{H}(\mathbf{x}(t))\,\mathrm{d}F_l(t)\,, \qquad \mathbb{P}\text{-a.s.} \tag{2.1}$$

Here (compare with (0.11) and (0.12)) the modified drift process is

$$\tilde{\mathfrak{g}}(t) := \frac{-1}{2\mathcal{H}(\mathbf{x}(t))} \sum_{i,j=1}^{n} \mathrm{D}^2_{ij}\mathcal{H}(\mathbf{x}(t))\mu_i(t)\mu_j(t)\tau^{\mu}_{ij}(t)\,.$$

Besides the above, an empirical approach to generating portfolios from beta factors instead of market weights has been developed by Oderda (2016).[1]

Nonetheless, the actual construction of explicit portfolios that outperform the market under realistic conditions on the price dynamics remains very difficult; Fernholz's original master equation and Strong's extension are only tools to *check*, given a set of market assumptions and a portfolio, whether a certain strategy is a relative arbitrage. In particular, the master equation does not *suggest* any portfolios (nor does the extension (2.1) suggest which information stream $F(\cdot)$ to use, or what model to put on it), and as such the number of functionally-generated portfolios that are known to outperform the market almost surely is relatively small. No explicit relative arbitrage has yet been constructed that depends on anything other than the current market weights.

The above problems are the motivation behind the work in this chapter: we expand the class of functionally-generated portfolios to allow for dependence on additional information, and develop a novel data-driven methodology that statistically infers such investment strategies from historical market information. This we do through a Bayesian nonparametric framework, using Gaussian processes (GP's) as priors on our portfolio functions — see Section 2.1 for the details of our method. Here, the quality of a portfolio is determined through a user-specified criterion, such as the Sharpe Ratio (recall (1.54)), and practically any additional characteristics of stocks can be incorporated: examples are moving averages of historical prices, trading volumes, balance sheet variables, economic forecasts, analyst recommendations, and relevant news.

[1] In fact, Section 7.4 of Fernholz (2002) already mentions incorporating the price-to-book ratio into SPT portfolio construction. However, this idea is not developed further; in particular, no modified master equation is derived.

An important added advantage is that our method allows for the easy inclusion of market imperfections, such as transaction costs (whose theoretical incorporation into SPT is very challenging and underdeveloped) and the possibility of bankruptcy, thus improving the real-world implementability of its results. The latter is particularly relevant, since functionally-generated portfolios typically invest heavily in small-capitalization stocks, which have a higher rate of default than their large-capitalization counterparts. For a theoretical incorporation of defaults into functionally-generated portfolio construction, we refer the reader to Section 4.5.

The contribution of the research in this chapter lies primarily in that we develop the first application of machine learning (or statistical inference) techniques to Stochastic Portfolio Theory. The only other paper so far which applies a statistical technique to augment SPT portfolios is Audrino et al. (2007), in which a so-called generalized tree-structured threshold model is invoked to forecast market diversity and thus improve the performance of diversity-weighted portfolios. We have no theoretical guarantees for our statistical inference to reach a certain level of optimality; the study of the efficacy of our method is limited entirely to empirical results. We perform two types of experiments to this end: a consistency test which learns on rolling windows, and a stress test that studies the robustness of learned portfolios in light of the recent financial crisis. For this, we use the daily closing prices of the Standard & Poor's 500 index constituents over a 23-year period. We present our results in Section 2.2, and end with a discussion in Section 2.3; we conclude that our approach learns portfolios, in an effective way, which perform well according to the user-specified criterion. Furthermore, our results confirm that negative-parameter diversity-weighted portfolios have the potential to outperform the market index, as was found in Chapter 1, and we compute the optimal parameter p.

This chapter is based on collaborative research with Yves-Laurent Kom Samo, which has been published as Kom Samo and Vervuurt (Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence 2016). In the current chapter, we omit some of the technical details of the exact algorithmic implementation of our method, for which we refer the reader to Section 3 of the article.

2.1 Bayesian nonparametric approach

Let $\mathcal{X} \subset \mathbb{R}^c$, for some $c \geq 1$, and let $\mathbf{x}(t) = \big(\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\big)^{\mathrm{T}}$ be the $(n \times c)$-dimensional process of trading characteristics (i.e. any stock or market observables) of all the stocks, so that $\mathbf{x}_i(t) \in \mathcal{X}$ for all $i = 1, \ldots, n$, $t \geq 0$. Note that we do not restrict ourselves to finite-variation processes, thus going beyond the framework of Strong (2014a). We extend the class of functionally-generated portfolios by considering long-only portfolios of the form

$$\pi_i^f(t) = \frac{f(\mathbf{x}_i(t))}{\sum_{j=1}^n f(\mathbf{x}_j(t))}, \qquad i = 1\ldots, n, \tag{2.2}$$

for continuous functions $f : \mathcal{X} \to \mathbb{R}_+$. This definition retains the property of functional generation that stocks with similar characteristics receive similar portfolio weights, since the C^0 function f is identical for all firms $i = 1, \ldots, n$. Note that (2.2) is a generalization of the standard definition indeed, since (0.10) can be obtained by setting $\mathcal{X} = \Delta_+^n \times \{1, \ldots, n\}$, $\mathbf{x}_i(\cdot) = (\mu(\cdot), i)^{\mathrm{T}}$, and

$$f : \Delta_+^n \times \{1, \ldots, n\} \to \mathbb{R}_+, \quad (u, i) \mapsto \Big(\mathrm{D}_i \log \mathbf{G}(u) + 1 - \sum_{j=1}^n u_j \mathrm{D}_j \log \mathbf{G}(u)\Big) \cdot u_i.$$

In particular, (2.2) includes the entropy-weighted and equally-weighted portfolios, as well as the diversity-weighted portfolio with parameter $p \in \mathbb{R}$, the latter of which corresponds to setting $\mathcal{X} = [0, 1]$, $\mathbf{x}_i(\cdot) = \mu_i(\cdot)$, and

$$f : [0, 1] \to \mathbb{R}_+, \quad u \mapsto u^p. \tag{2.3}$$

Our objective now is to learn the function f, given a set of characteristics $\mathcal{X}$, and from some data $\mathcal{D}$ which is of the form

$$\mathcal{D} = \big\{\mathbf{x}(t)\big\}_{t=1,\ldots,T}.$$

We do this by making the function f *random*; that is, we let $\{f(x)\}_{x\in\mathcal{X}}$ be a stochastic process indexed by its argument $x \in \mathcal{X}$ (such a process is commonly called a *random field*, if $c > 1$), whose exact form is to be inferred from the data. We take a Bayesian approach, meaning we employ Bayes' theorem to formally write the posterior distribution of f given $\mathcal{D}$ as

$$p(f|\mathcal{D}) \propto p(\mathcal{D}|f) \cdot p(f). \tag{2.4}$$

This expression will make more sense later on, when we discretize our space and the above becomes a prior on a multivariate random variable — see (2.10). Hence in order to define our method, we need to specify the *prior* distribution $p(f)$ and the *likelihood* $p(\mathcal{D}|f)$.

2.1.1 Prior model

In learning a suitable trading strategy from data given an investment objective, we aim for a flexible method that learns from a large class of candidate portfolio functions f, in order to uncover possibly highly non-linear dependencies on the stock characteristics. This leads us to the use of a nonparametric model for the functions f. In particular, we choose a *Gaussian process* prior for the function $\log f$, that is:

$$\log f \sim \mathcal{GP}\big(0, k\big), \tag{2.5}$$

with $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ the *kernel* (or covariance function) of the Gaussian process; the second-order statistics completely define the Gaussian process. This means the following: for any sequence of arguments $\mathbf{x}_1, \dots, \mathbf{x}_m \in \mathcal{X}$, $m \in \mathbb{N}$, the m-dimensional random variable $\big(\log f(\mathbf{x}_1), \dots, \log f(\mathbf{x}_m)\big)$ is a multivariate Gaussian:

$$\big(\log f(\mathbf{x}_1), \dots, \log f(\mathbf{x}_m)\big) \sim \mathcal{N}\big(\mathbf{0}, \mathbf{k}(\mathbf{x}_1, \dots, \mathbf{x}_m)\big), \tag{2.6}$$

with $\mathbf{0}$ the zero vector in $\mathbb{R}^m$, and with covariance matrix

$$\mathbf{k}(\mathbf{x}_1, \dots, \mathbf{x}_m) = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_m) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_m, \mathbf{x}_1) & \cdots & k(\mathbf{x}_m, \mathbf{x}_m) \end{pmatrix}.$$

Gaussian processes are used, among many other things, for supervised learning in the field of machine learning, popular due their flexibility and versatility, but until now it has not been applied in SPT — for an introduction, see (in particular Section 3.1 of) MacKay (1998). Their behavior is completely determined by the kernel k (assuming a mean function of 0), which gives the covariance between two samples from the random function $\log f$, as a function of the arguments with which the samples are drawn. We choose k to be a separable product of Rational Quadratic (RQ) kernels, defined as

$$k_{RQ}(x, y) = k_0 \prod_{i=1}^{c} \left(1 + \frac{|x_i - y_i|^2}{2\alpha_i l_i^2}\right)^{-\alpha_i}, \qquad x, y \in \mathcal{X},$$

where the hyperparameters $k_0, l_i, \alpha_i > 0$, $i = 1, \dots, c$ are given independent log-normal priors, and are to be inferred in the learning process. Note that k_{RQ} is a function of the distances $|x_i - y_i|$, to which it is inversely proportional; this translates the underlying assumption that points which are 'further apart' in terms of the metric k_{RQ} are correlated to a lesser degree. Furthermore, the separability of this kernel allows us to learn a different dependence (through different hyperparameters l_i, α_i, $i = 1, \dots, c$) of f on the different characteristics. Another reason why we choose this kernel is because it is less computationally intensive than for instance the Gaussian kernel.

2.1.2 Likelihood model

Let $\mathcal{P}_{\mathcal{D}}$ be a user-specified performance functional, which maps the portfolio (2.3) corresponding to a candidate investment map f (that is, a sample from the prior $p(f)$) to its historical performance $\mathcal{P}_{\mathcal{D}}(\pi^f)$. Examples of such a performance metric include the Sharpe Ratio — recall (1.54) — and the excess return relative to a benchmark portfolio $\pi^*(\cdot)$, defined as

$$\mathrm{ER}(\pi^f|\pi^*) := \mathcal{V}^{\pi^f}(T) - \mathcal{V}^{\pi^*}(T). \tag{2.7}$$

Here, we have written $\mathcal{V}^{\pi}(\cdot)$ for the discrete-time wealth process corresponding to a portfolio $\pi(\cdot)$, defined in (1.53). We choose the criterion (2.7) as our performance functional, with $\pi^*(\cdot)$ the equally-weighted portfolio (0.3).

2.1.1 Remark. Note that the performance metric

$$\mathbf{1}\big(\mathrm{ER}(\pi^f|\mu) > 0\big),$$

with $\mu(\cdot)$ the market portfolio, corresponds to the SPT objective of strongly outperforming the market — recall Definition 0.1.1. We choose not to use this as our learning objective, but instead aim to outperform the equally-weighted portfolio through (2.7) with $\pi^*(\cdot) \equiv \frac{1}{n}\mathbf{1}$, since this strategy has an empirically higher return than the market portfolio. Furthermore, (2.7) allows us to distinguish candidate portfolios that outperform the market significantly from those which only do so to a moderate degree. □

We now express the investor's view as to what constitutes a good performance through a likelihood model $p(\mathcal{D}|f)$, which we may choose to be a probability distribution on $\mathcal{P}_{\mathcal{D}}(\pi^f)$. This step is crucial to our method: by expressing the investor's 'utility' from a candidate portfolio as the probability density function of the performance of that portfolio, this 'utility' can be interpreted as a likelihood of observing the data $\mathcal{D}$ given the candidate function f. This way, highly-performing candidate portfolios will receive higher weights, and the resulting posterior distribution will 'learn' a portfolio that achieves the user-specified selection criterion. In particular, we choose to set

$$p(\mathcal{D}|f) = \gamma\big(\mathcal{P}_{\mathcal{D}}(\pi^f); a, b\big), \tag{2.8}$$

where γ is the probability density function of a gamma-distributed random variable with mean a and standard deviation b. Our Bayesian method then learns functions f that make a trade-off between how *likely* a map is in light of training data, and how *consistent* it is with prior beliefs. This particular choice of likelihood model is somewhat arbitrary, and serves as a first demonstration of our Bayesian non-parametric method; we leave it for future research to investigate the sensitivity to the choice of prior and likelihood model.

The density function γ reflects the investor's risk appetite, or greediness — by choosing a and b, the investor determines the target performance, as well as the tolerance for values around this value. The positive support of the Gamma distribution avoids (undesirable) negative values of the performance metric. However, a different probability density function could be used instead of γ, which would lead to different results from those presented in Section 2.2 below.

2.1.3 Algorithmic implementation

With the above choice of the prior and likelihood, we have completely specified our Bayesian nonparametric approach. In order to implement it, we discretize the space $\mathcal{X}$ using a Cartesian grid, and approximate the logarithm of the portfolio function f over this discretization. Let us denote the values of f on this grid by the vector $\mathbf{f}$. Having chosen our kernel and (independent) hyperparameter priors as in Section 2.1.1, our prior factorizes as follows:

$$p(\mathbf{f}) = p(\mathbf{f}|k_0, l_1, \alpha_1, \ldots, l_c, \alpha_c) \cdot p(k_0) \prod_{i=1}^{c} p(l_i) p(\alpha_i), \tag{2.9}$$

with $p(\mathbf{f}|k_0, l_1, \alpha_1, \ldots, l_c, \alpha_c)$ as in (2.6). We sample from the posterior

$$p(\mathbf{f}|\mathcal{D}) \propto \gamma\big(\mathcal{P}_{\mathcal{D}}(\pi^f; a, b)\big) \cdot p(\mathbf{f}) \tag{2.10}$$

using a blocked Gibbs sampler which alternates between updating $\mathbf{f}$ conditional on the hyperparameters (the first term in (2.9)), and updating the hyperparameters conditional on $\mathbf{f}$ (the remaining terms in (2.9)) — see Algorithm A.0.1 of Appendix A.

In each experiment, we run $10,000$ iterations, of which the first $5,000$ are discarded as burn-in. These numbers were determined using autocorrelation plots, which showed that this particular number of iterations was sufficient for the sampler to reach equilibrium. We split our data into training and test sets, and study the in-sample and out-of-sample performance of the *mean* of the posterior distribution of f (that is, the means of the posterior distributions on the values $\mathbf{f}$). Note that, by (2.5), the initial mean portfolio (that is, the prior mean) is the equally-weighted portfolio.

2.1.4 Parametric benchmark approaches

In order to determine whether our approach adds any value to the standard SPT approach, we will measure its performance relative to two parametric methods. Namely, we limit ourselves to diversity-weighted portfolios, corresponding to setting $\mathcal{X} = [0,1]$, $\mathbf{x}_i(\cdot) = \mu_i(\cdot)$ for all $i = 1, \ldots, n$, and f as in (2.3).

In the first benchmark method, we take a frequentist approach and learn the portfolio's optimal parameter directly, by noting that $\mathcal{P}_{\mathcal{D}}(p) := \mathcal{P}_{\mathcal{D}}(\pi^f)$ is simply a function of the single variable p, and hence maximizing over $p \in [-8, 8]$. This large range contains all relevant values of p, since $p = -8$ puts practically all weight on the stock of smallest capitalization, and likewise $p = 8$ leads to $\pi_i(t) \approx \mathbf{1}\big(i = \arg\max_j \mu_j(t)\big)$. To avoid getting stuck on local maxima of $\mathcal{P}_{\mathcal{D}}(\pi^f)$, we perform a brute force maximization on the uniform grid with mesh size 0.05.

Our second benchmark is a Bayesian parametric approach: we use the Metropolis-Hastings algorithm, see Algorithm A.0.2 of Appendix A, to sample from the posterior distribution of the parameter p when using a Uniform$(-8, 8)$-prior:

$$p(p|\mathcal{D}) \propto \gamma\big(\mathcal{P}_{\mathcal{D}}(p); a, b\big) \cdot \mathbf{1}(p \in [-8, 8]).$$

We initialize p within $[-8, 8]$, and in each step of the algorithm, we sample a proposal update p^* from a Gaussian centered around the current value p and with standard deviation 0.5. Applying Equation (A.1), we then use the following acceptance probability:

$$\alpha(p, p^*) = \min\left\{1\,,\ \frac{\gamma\left(\mathcal{P}_{\mathcal{D}}(p^*); a, b\right)}{\gamma\left(\mathcal{P}_{\mathcal{D}}(p); a, b\right)}\,\mathbf{1}(p^* \in [-8, 8])\right\}.$$

As in the nonparametric case, we run $10,000$ iterations of the algorithm and discard the first $5,000$ as burn-in; the learned portfolio is then the diversity-weighted portfolio with p equal to the mean of the learned parameter. In terms of the portfolio function, this corresponds to the following:

$$f : \mu \mapsto \mu^{\mathbb{E}(p|\mathcal{D})}.$$

2.2 Empirical results

As in Section 1.5, the universe of stocks we consider in our two experiments is the constituents of the S&P 500 index, accounting for changes in index constituents. We rebalance our portfolios on a daily basis, and use the portfolio selection criterion (2.7) with $\pi^*(\cdot)$ the equally-weighted portfolio, as mentioned earlier. At the end of each trading day, we determine our target portfolio for the next day, which is acquired at the open of the next trading day. We incur a proportional transaction cost of 0.1% of the value traded on each day. The returns we use account for corporate events such as dividends, defaults, M&A's, etc. Our data sources are the CRSP and Compustat databases, and we use data from 1 January 1992 to 31 December 2014.

In our first experiment, we aim to illustrate that the approach we propose in this paper *consistently* and considerably outperforms SPT alternatives over a wide range of market conditions. We learn investment strategies using the methods described in Section 2.1 by training on 10 consecutive years worth of data and testing on the following 5 years. Starting from 1 January 1992 for the first training data set and 1 January 2002 for the first test set, we iteratively roll both training and test data sets by one year, which leads to a total of 9 pairs of training and test subsets. We study the performance of the following portfolios:

- MARKET: the market portfolio;
- EWP: the equally-weighted portfolio;
- DWP*: the diversity-weighted portfolio whose exponent p^* is learned by maximizing the evaluation functional (the frequentist parametric approach of Section 2.1.4);
- DWP: the diversity-weighted portfolio whose exponent p is learned using the Metropolis-Hastings MCMC algorithm (the Bayesian parametric approach of Section 2.1.4);
- CAP: the Gaussian process approach with $\mathbf{x}_i(t) = \mu_i(t)$, that is, using the market weights as the sole trading characteristic;
- CAP+ROA: the Gaussian process approach with one additional stock feature, namely $\mathbf{x}_i(t) = \big(\mu_i(t), \mathrm{ROA}_i(t)\big)$, where $\mathrm{ROA}_i(t)$ is the last-known *return-on-assets* of firm i at time t.

The last-known return-on-assets $\mathrm{ROA}_i(t)$ is defined as the ratio between the last-known net income and the last-known total assets of company i on day t. We note that this quantity does not change every day (it is reported on a quarterly basis), but this does not affect our analysis. The rationale behind using the return-on-assets as an additional characteristic is to capture not only how big a company is, but also how well it performs relative to its size.

As was mentioned before, we use the excess return with respect to the equally-weighted portfolio (so (2.7) with $\pi_i^*(t)$ as in (0.3)) as our performance metric $\mathcal{P}_{\mathcal{D}}$.

Furthermore, we choose the parameters of the gamma distribution $\gamma(\cdot; a, b)$ that defines our likelihood through (2.8) as $a = 7.0$ and $b = 0.5$; this corresponds to postulating that, starting from \$1, the ideal portfolio should lead to a terminal wealth that is \$7.0 higher than the terminal wealth of the equally-weighted portfolio. We are purposefully greedy in our choice of these parameters, in order to move away from the initial mean portfolio (the equally-weighted portfolio).

Table 2.1: Results of our first experiment on the consistency of our learning algorithms to varying market conditions. Returns are averaged over the 9 rolling windows and annualized, and IS (respectively OOS) stands for in-sample (respectively out-of-sample).

PORTFOLIO	IS return (%)	OOS return (%)
MARKET	8.56 ± 1.62	6.23 ± 2.07
EWP	10.56 ± 1.67	8.99 ± 1.85
DWP*	11.94 ± 2.01	12.51 ± 1.12
DWP	11.91 ± 1.99	12.50 ± 1.11
CAP	26.54 ± 2.38	22.05 ± 2.89
CAP+ROA	56.18 ± 7.35	25.14 ± 2.58

The results of our first experiment are summarized in Table 2.1, which displays the yearly in-sample and out-of-sample returns plus-minus two standard errors, averaged over the 9 rolling windows. It can be seen that all of our strategies do indeed outperform the benchmark (EWP), in-sample (i.e., on the training set) but also out-of-sample (i.e., on the test set). Moreover, the performance is greatly improved by considering nonparametric models, even when the only characteristic considered is the market weight. Finally, it can be seen that adding more trading characteristics does indeed add value, although the increase in performance is not as pronounced as the increase obtained from moving to a nonparametric approach. Crucially, the CAP+ROA portfolio *considerably* and *consistently* outperforms the benchmark EWP, both in-sample and out-of-sample.

In our second experiment, we aim to illustrate that our approach is robust to a financial crisis. To do so, we train our model on data from 1 January 1992 to 31 December 2005, and test the learned strategy between 1 January 2006 and 31 December 2014, which includes the 2008 financial crisis. We study the same investment strategies as in the first experiment.

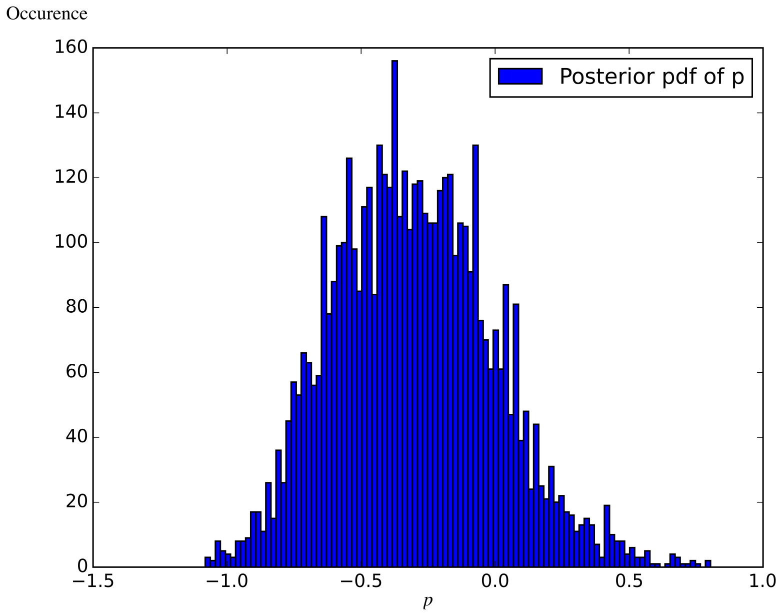

Figure 2.1: Histogram showing the posterior distribution of the parameter p of the diversity-weighted portfolio DWP in our second experiment.

The posterior distribution over the exponent p of the DWP-portfolio in the Bayesian parametric method is illustrated in Figure 2.1, and the optimal parameter of the DWP* portfolio is computed as $p^* = -0.45$, confirming our findings in Section 1.3 that the negative-parameter diversity-weighted portfolio has the potential to outperform its positive-parameter counterpart. The mean posterior investment maps $\log f$ that are learned in the nonparametric approaches are illustrated in Figure 2.3. Note that the learned mean portfolio map of the portfolio CAP invests heavily in small-capitalization stocks, but avoids putting a lot of weight on companies of very small capitalization — we conjecture that this is so as to avoid bankruptcies. In the two-dimensional case of the CAP+ROA portfolio, a similar dependence of $\log f$ on the market weights is maintained, in that both very small- and large-capitalization stocks are avoided; however, we see that the added dimension of return-on-assets (ROA) allows the portfolio function to differentiate highly-performing stocks from underperforming ones — note the blue area at the top right corner of the bottom of Figure 2.3.

In Table 2.2, we provide in-sample and out-of-sample average yearly returns as well as out-of-sample Sharpe Ratios. Once again, it can be seen that: i) all learned portfolios do indeed outperform the benchmark EWP, both in-sample and out-of-sample, ii) nonparametric methods outperform parametric methods, and iii) adding the return-on-assets as an additional characteristic does indeed add value. These

Table 2.2: Results of our second experiment on the robustness of the proposed approaches to financial crises. Returns are annualized, and IS (respectively OOS) stands for in-sample (respectively out-of-sample), while we denote the Sharpe Ratio by SR.

Portfolio	IS return (%)	OOS return (%)	OOS SR
MARKET	9.60	7.90	0.47
EWP	13.46	9.60	0.51
DWP*	14.62	11.74	0.56
DWP	14.62	11.38	0.55
CAP	16.49	18.01	0.60
CAP+ROA	37.54	18.33	0.72

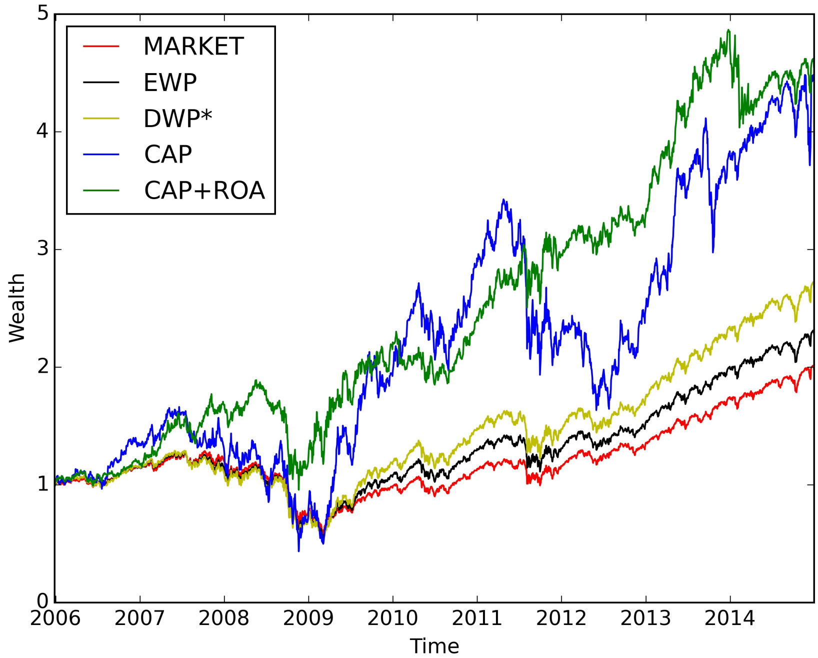

Figure 2.2: Time series of out-of-sample wealth processes in our second experiment.

conclusions hold true not only in absolute terms (i.e. returns), but also after adjusting for risk (i.e. the Sharpe Ratio), despite the fact that our learning objective (2.7) does not take into account the volatility of a portfolio. A more granular illustration of how our method performs during the 2008 financial crisis can be seen in the time series of total wealth, illustrated in Figure 2.2 — we omit the DWP portfolio, since its wealth process is practically identical to that of the DWP* portfolio. It turns out that adding the return-on-assets does not only improve the out-of-sample return, but that it also has a 'stabilizing effect' in that the volatility of the wealth process is considerably reduced; this is reflected in the Sharpe Ratio, which is 20% higher for the CAP+ROA-portfolio compared to CAP.

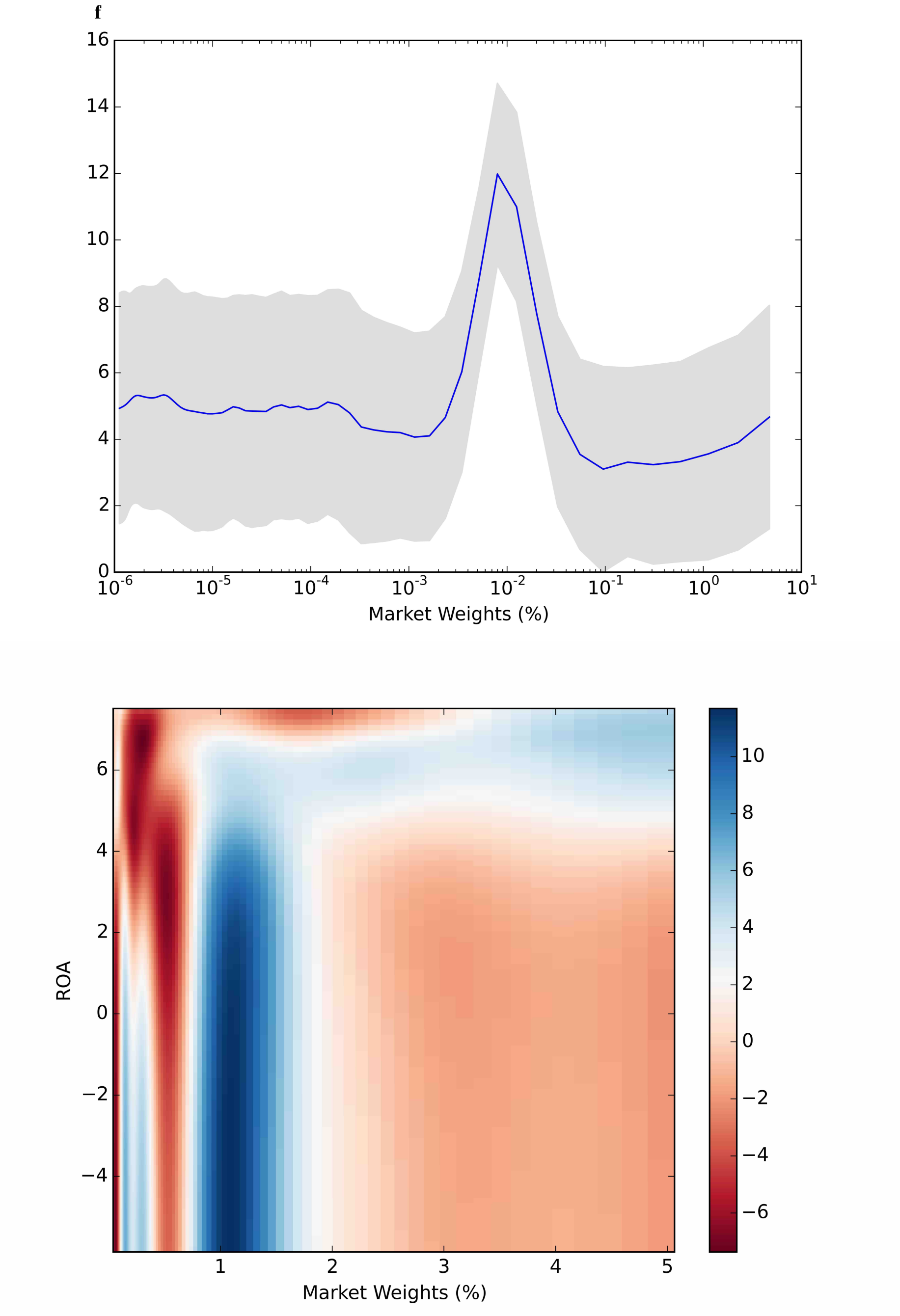

Figure 2.3: Logarithms of the mean posterior investment maps **f** for the CAP portfolio (top) and the CAP+ROA portfolio (bottom) in our second experiment. In the case of the CAP portfolio, the grey band corresponds to 2 posterior standard deviations.

Finally, it is also worth stressing that the shape of the learned investment map in the two-dimensional case (see the bottom of Figure 2.3) suggests that the investment strategy uncovered by our Bayesian nonparametric approach can hardly be replicated using a parametric model. It would be near impossible to derive analytical results pertaining to such a highly non-linear portfolio function within the SPT framework.

2.3 Discussion

We have developed and demonstrated a novel data-driven method for constructing portfolios that depend on market characteristics that go beyond the current market weights, thus extending the class of functionally-generated portfolios. By being Bayesian and sampling from a Gaussian process, we consider a very broad class of candidate portfolios; we have empirically shown this added flexibility to considerably and consistently improve performance. The latter is, notably, specifiable by the investor, allowing us to go beyond the SPT objective of almost sure outperformance of the market.

An important advantage of our empirical approach is that it allows one to circumvent the postulation of almost-sure assumptions on the market, such as diversity (D^{δ}), no-failure (NF^{φ}) and limited-failure (LF^{κ}); one could argue that these can never be assumed to hold with probability one, since one only observes a single realization of market prices, and market mechanisms (such as regulations) might change. The downside is, of course, that one loses the theoretical guarantees for performance (particularly out-of-sample). A compromise is the Bayesian parametric approach (portfolio DWP of Section 2.1.4), which optimizes the parameters of portfolios for which lower bounds on the market-relative return exist. It would be of interest to extend this methodology beyond diversity-weighted portfolios; one possibility is to optimize rank-based portfolios, for instance those of Section 1.2. Furthermore, one might have a model for the behavior of the market weights or other stock characteristics, and estimates of their future values — one would ideally incorporate this knowledge into the Bayesian method.

Despite being relatively simple theoretically, our method is not trivial to implement algorithmically. It requires a fair amount of engineering, for instance in 'tweaking' the hyperparameters in order to avoid overfitting; while one might get extremely good results in-sample, this does not automatically translate into a good out-of-sample performance of the learned portfolios. We have determined the hyperparameters that avoid these problems heuristically, that is, by trial-and-error. As such, there is much room for improvement; that is, for optimization. Furthermore, it would be interesting to investigate the sensitivity of our methodology to the prior and likelihood models; we have merely given a demonstration of the working of our approach for particular choices of these. For instance, one could change the prior mean function, setting it equal to a portfolio that is known to perform well theoretically (such as the diversity-weighted portfolio).

Although the Gaussian process in our model was approximated to be piecewise constant on a grid, there is no theoretical or practical obstacle in using an alternative approximation such as sparse Gaussian processes — we refer the reader to Quiñonero Candela and Rasmussen (2005) — or string Gaussian processes — see Kom Samo and Roberts (2015b). Our method may be extended to learn even subtler patterns using the non-stationary general purpose kernels of Kom Samo and Roberts (2015a). Our work may also be changed to allow for investment strategies that allow for short-selling. Finally, it would be interesting to develop an *online* extension of our work, that is, to modify our method in such a way that new daily information can be incorporated into the posterior distribution $p(f|\mathcal{D})$, without having to re-run the entire algorithm.

Chapter 3

Functional generation & optimal transport

We now move away from the explicit constructions of portfolios of Chapter 1 and Chapter 2, and establish a theoretical relation between the method of functional generation and the mathematical area of optimal transport. The latter is a very active field of research, with applications in various branches of mathematics — see Villani (2009). Its central object of study is the optimal transport problem, which aims to minimize the cost associated to a measure with given marginals (see Section 3.1.1 below), as well as extensions of this problem.

As before, let us denote the market weights vector at any time $t \in \mathbb{R}_{\geq 0}$ by $\mu(t) = \big(\mu_1(t), \ldots, \mu_n(t)\big)$, and assume this takes values in the set Δ^n_+ of (0.5), which has closure $\bar{\Delta}^n_+$. Note that we do not allow for defaults. We take a model-free approach, meaning that market weights $\mu(\cdot)$ and portfolio weights $\pi(\cdot)$ are defined as *Borel-measurable paths* in Δ^n_+ and $\bar{\Delta}^n_+$, respectively. In this chapter, we will invoke results from convex analysis, which requires us to slightly modify the conventional definition (0.10) of functionally-generated portfolios that we have used so far. Namely, we will employ the following definition:

3.0.1 Definition. A portfolio $\pi(\cdot)$, whose elements $\pi_i(t)$ prescribe the proportion of wealth to be invested in firm i at time t, is said to be *(multiplicatively) functionally-generated* if $\pi(\cdot) \equiv \pi\big(\mu(\cdot)\big)$ for some function

$$\pi : \Delta^n_+ \to \bar{\Delta}^n_+, \tag{3.1}$$

which has the property that there exists a concave function $\Phi : \Delta^n_+ \to \mathbb{R}_{\geq 0}$ such that

$$\left\langle \frac{\pi(\mu)}{\mu}, \nu \right\rangle \geq \frac{\Phi(\nu)}{\Phi(\mu)}, \qquad \forall\, \mu, \nu \in \Delta^n_+; \tag{3.2}$$

we write $\langle a, b\rangle$ for the inner product between two vectors $a, b \in \mathbb{R}^n$, and $\pi(\mu)/\mu := \big(\pi_1(\mu)/\mu_1, \ldots, \pi_n(\mu)/\mu_n\big)$, as in (0.8). By definition, this is equivalent to

$$\pi(\mu)/\mu \in \frac{1}{\Phi(\mu)}\partial\Phi(\mu) := \left\{\frac{\rho}{\Phi(\mu)} : \rho \in \partial\Phi(\mu)\right\}, \qquad \forall\, \mu \in \Delta^n_+, \tag{3.3}$$

where $\partial\Phi(\mu)$ is the superdifferential of Φ at $\mu \in \Delta^n_+$. Namely, the superdifferential $\partial\Upsilon$ of a function $\Upsilon : \mathcal{W} \to \mathcal{Z}$ is defined as

$$\partial\Upsilon := \big\{w \in \mathcal{W} : \Upsilon(u) + \langle w, v-u\rangle \geq \Upsilon(v),\ \forall\, u, v \in \mathcal{W}\big\}. \tag{3.4}$$

□

3.0.2 Remark. This definition of functionally-generated portfolios comes from Pal and Wong (2016), and is slightly different from the usual definition (0.10) in that it requires the 'generating function' Φ to be concave instead of C^2. However, if the function Φ is assumed to be C^2, the above Definition 3.0.1 reduces to the concave case of (0.10); namely, let the portfolio function π satisfy (3.2). Then fix values $\mu \in \Delta^n_+$, $1 \leq i \leq n$, and $h \in \mathbb{R} \setminus \{0\}$ sufficiently small so that $\mu + h(e_i - \mu) \in \Delta^n_+$, with e_i the i^{th} unit vector in $\mathbb{R}^n$. Note that, by (3.2), we have, similarly to Proposition 6(ii) of Pal and Wong (2016),[1]

$$\begin{aligned}
&1 + h\left(\frac{\pi_i(\mu)}{\mu_i} - 1\right) = 1 + \left\langle \frac{\pi(\mu)}{\mu}, h(e_i - \mu)\right\rangle \geq \frac{\Phi(\mu + h(e_i-\mu))}{\Phi(\mu)} \\
\iff\ &\frac{1}{h}\log\left(1 + h\left(\frac{\pi_i(\mu)}{\mu_i} - 1\right)\right) \geq \frac{\log\Phi(\mu + h(e_i-\mu)) - \log\Phi(\mu)}{h} \\
\Rightarrow\ &1 + \mathrm{D}_{e_i-\mu}\log\Phi(\mu) \leq \frac{\pi_i(\mu)}{\mu_i} \leq 1 - \mathrm{D}_{\mu-e_i}\log\Phi(\mu),
\end{aligned} \tag{3.5}$$

where the last inequalities follow by taking $h \downarrow 0$ and $h \uparrow 0$, respectively, and where we have applied L'Hôpital's rule to derive

$$\lim_{h\to 0}\frac{1}{h}\log\left(1 + h\left(\frac{\pi_i(\mu)}{\mu_i} - 1\right)\right) = \lim_{h\to 0}\frac{\frac{\pi_i(\mu)}{\mu_i} - 1}{1 + h\left(\frac{\pi_i(\mu)}{\mu_i} - 1\right)} = \frac{\pi_i(\mu)}{\mu_i} - 1.$$

When $\Phi \in C^2$, the directional derivatives $\mathrm{D}_{e_i-\mu}\log\Phi(\mu)$ and $-\mathrm{D}_{\mu-e_i}\log\Phi(\mu)$ coincide, and we obtain (0.10); hence our alternative Definition 3.0.1 is consistent with the classical definition of Fernholz (1999a).

[1] We use e_i to denote the i^{th} unit vector, and $\mathrm{D}_\nu\Phi$ for the directional derivative of Φ in the direction $\nu \in \mathbb{R}^n$, which exists for any concave Φ by Theorem 23.1 of Rockafellar (2015).

In this chapter, we further develop recently made connections between functional generation as defined above and optimal transport, an area of mathematics that has received a lot of interest of late — for an introduction, see Section 3.1 below, as well as Villani (2009). Pal and Wong (2016) were the first to make such a connection, by engineering an optimal transport problem which, provided it has a deterministic solution, defines a functionally-generated portfolio. We engineer an alternative optimal transport problem in Section 3.1 below, and strengthen the connection between this problem and functional generation by showing *uniqueness* of the optimal solution in Section 3.2. In particular, this allows us to establish a formal *equivalence* between functionally-generated portfolios and the optimal transport problem, thus elevating the correspondence between them from a mathematical nicety to a tool for constructing functionally-generated portfolios (or FGP's): each pair of initial and terminal distributions of the optimal transport problem implies an optimal transport map, which in turn defines a functionally-generated portfolio. In Section 3.3, we extend this correspondence to additively generated portfolios, a new class of portfolios developed very recently by Karatzas and Ruf (2016). However, since these strategies depend on the entire history of market weights, unlike in (3.1), the results here are not as explicit. We mention other observations and make suggestions for future research in Section 3.4. Some of the results in this chapter will be used in Chapter 4 to derive model-free lower bounds on the wealth of functionally-generated strategies relative to the market.

3.1 Engineering the optimal transport problem

We start by recalling some necessary definitions from optimal transport and convex analysis, and use these to derive the relevant optimal transport problem in Section 3.1.3. In the latter, we will deviate from Pal and Wong (2016) in order to arrive at an optimal transport problem that has a unique and deterministic solution, as we show in Section 3.2.

Interestingly, we need not introduce a model for the market weights to be able to study functionally-generated portfolios from an optimal transport perspective. Namely, the transport will occur between a distribution on the market weights, to a distribution on some auxiliary weights which determine the portfolio weights — since the portfolio function π is unchanged in time, no model is needed for the time-evolution of market weights.

3.1.1 Introduction to optimal transport

Let us start by giving a brief introduction to optimal transport — we refer the reader to Villani (2009) for a more detailed account. Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces, and call the measurable function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ the *cost function*. Let $\mathcal{P} : \mathcal{X} \to [0, 1]$ and $\mathcal{Q} : \mathcal{Y} \to [0, 1]$ be some given Borel probability measures. A *coupling* of $(\mathcal{P}, \mathcal{Q})$ is a probability measure $\mathcal{R} : \mathcal{X} \times \mathcal{Y} \to [0, 1]$ with marginals $\mathcal{P}$ and $\mathcal{Q}$; let us write $\Pi(\mathcal{P}, \mathcal{Q})$ for the set of all such couplings. The Monge-Kantorovich optimal transport problem is the following: find $\mathcal{R}^* \in \Pi(\mathcal{P}, \mathcal{Q})$ solving

$$\int_{\mathcal{X}\times\mathcal{Y}} c(x, y) \,\mathrm{d}\mathcal{R}^*(x, y) = \inf_{\mathcal{R}\in\Pi(\mathcal{P},\mathcal{Q})} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y) \,\mathrm{d}\mathcal{R}(x, y) =: \mathcal{C}(\mathcal{P}, \mathcal{Q}). \tag{3.6}$$

If it exists, such a measure $\mathcal{R}^*$ is called the *optimal coupling*, and it is *deterministic* (and is said to solve the Monge problem) if there exists a function $T : \mathcal{X} \to \mathcal{Y}$, called the *transport map*, such that $T(X) \sim \mathcal{Q}$ if $X \sim \mathcal{P}$, and

$$\int_{\mathcal{X}} c(x, T(x)) \,\mathrm{d}\mathcal{P}(x) = \int_{\mathcal{X}\times\mathcal{Y}} c(x, y) \,\mathrm{d}\mathcal{R}^*(x, y).$$

One of the most central results in optimal transport is Kantorovich duality, which characterizes optimal couplings for a large class of cost functions c. Let us define the following local optimality property of couplings, which states that the total cost over almost every sequence in $\mathcal{X} \times \mathcal{Y}$ is locally minimal; this definition is an adaptation from Definition 5.1 of Villani (2009):

3.1.1 Definition (c-cyclical monotonicity)**.** Let $\Gamma \subset \mathcal{X} \times \mathcal{Y}$. If for any sequence

$$\big\{(x^1, y^1), \ldots, (x^m, y^m)\big\} \subset \Gamma, \ m \in \mathbb{N},$$

it holds that

$$\sum_{k=1}^{m} c\big(x^k, y^k\big) \le \sum_{k=1}^{m-1} c\big(x^{k+1}, y^k\big) + c\big(x^1, y^m\big), \tag{3.7}$$

then Γ is *c-cyclically monotone*. We say that a coupling $\mathcal{R}$ is c-cyclically monotone if it is concentrated on a c-cyclically monotone set; that is, there exists a c-cyclically monotone $\Gamma \subset \mathcal{X} \times \mathcal{Y}$ such that $\mathcal{R}(\Gamma) = 1$. □

Kantorovich duality then states that this local optimality is in fact global (see Theorem 5.10 of Villani (2009) for a proof):

3.1.2 Theorem (Kantorovich duality)**.** *Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a lower semicontinuous cost function such that $c(x, y) \geq a(x) + b(y)$ on all of $\mathcal{X} \times \mathcal{Y}$ for some real-valued upper semicontinuous functions $a \in L^1(\mathcal{P})$ and $b \in L^1(\mathcal{Q})$. Assume $\mathcal{C}(\mathcal{P}, \mathcal{Q}) < \infty$. Then a coupling $\mathcal{R} \in \Pi(\mathcal{P}, \mathcal{Q})$ is optimal if and only if it is c-cyclically monotone.*

This duality establishes a 'monotonicity principle', in that it connects the optimality of a coupling to the geometry of certain sets. By Theorem 4.1 of Villani (2009), an optimal coupling is guaranteed to exist provided c is as in Theorem 3.1.2.

3.1.2 Convex analytic cyclical monotonicity

We present a standard result from convex analysis, which characterizes all functions that are in the superdifferential of a convex (or concave) function — see Section 24 of Rockafellar (2015). By viewing this result as an analog of Kantorovich duality, as was done by Pal and Wong (2016), we will be able to make a link between FGP's and optimal transport.

3.1.3 Definition (cyclical monotonicity)**.** Let $R \subseteq \mathbb{R}^n$ and $\rho : R \to \mathbb{R}^n$. If for any sequence

$$\{x^1, \ldots, x^m\} \subset R, \ m \in \mathbb{N},$$

it holds that

$$\sum_{k=1}^{m-1} \langle \rho(x^k), x^{k+1} - x^k \rangle + \langle \rho(x^m), x^1 - x^m \rangle \geq 0, \tag{3.8}$$

then ρ is *cyclically monotone.* $\square$

Compare the above with Definition 3.1.1. Similarly to Kantorovich duality, we have the following result:

3.1.4 Theorem (Theorem 24.8 of Rockafellar (2015))**.** *Let $R \subseteq \mathbb{R}^n$ and $\rho : R \to \mathbb{R}^n$. There exists a closed proper concave function $f : R \to \mathbb{R}_{\geq 0}$ such that ρ is in the superdifferential of f, if and only if ρ is cyclically monotone.*

In their Proposition 4, Pal and Wong (2016) show that the proof of the above result can be used to deduce the following: a function $\pi : \Delta^n_+ \to \bar{\Delta}^n_+$ is a functionally-generated portfolio — recall (3.2) — if and only if it is what they call 'multiplicatively cyclically monotone' (or MCM):

3.1.5 Definition (multiplicative cyclical monotonicity)**.** Let $\{\mu^1, \dots, \mu^m\}$ with $m \in \mathbb{N}$ be any sequence in Δ^n_+. A portfolio map $\pi : \Delta^n_+ \to \bar{\Delta}^n_+$ is *multiplicatively cyclically monotone* if it holds that

$$\sum_{k=1}^{m-1} \log \left\langle \pi(\mu^k), \frac{\mu^{k+1}}{\mu^k} \right\rangle + \log \left\langle \pi(\mu^m), \frac{\mu^1}{\mu^m} \right\rangle \geq 0. \tag{3.9}$$

□

3.1.6 Remark. The above definition is equivalent to that of Pal and Wong (2016), which formulates that (3.9) holds over *cycles* of market weights, which are sequences starting and ending at the same point. Namely, setting $\mu^{m+1} := \mu^1$, we recover their Definition 2. We refrain from studying cycles, since the set of these has measure zero in a continuous-time setting — this will be relevant in Sections 4.1 and 4.2. Note that in the above, there is no time component; we simply study *sequences* in the space Δ^n_+ of market weights (the positive simplex). We will, however, introduce a time-dependency in Section 4.1, by studying continuous-time *paths* of market weights. □

By choosing the cost function c and spaces $\mathcal{X}$ and $\mathcal{Y}$ appropriately, we can make Definitions 3.1.1 and 3.1.5 coincide, thus relating FGP's to an optimal transport problem. This is the "engineering" of the problem referred to in the section title.

3.1.3 Construction for continuous portfolio functions

This subsection is a pedagogic demonstration to show the reader how one arrives at the correct optimal transport problem. An important secondary benefit is that it clarifies to what degree one is free to choose the triplet $(c, \mathcal{X}, \mathcal{Y})$ that defines the optimal transport problem.

If an equivalence between some optimal transport problem and FGP's is to hold, then it must hold for the subclass of *continuous* FGP's. Rewriting (3.7) as

$$\sum_{k=1}^{m-1} \left(c\big(x^{k+1}, y^k\big) - c\big(x^k, y^k\big) \right) + c(x^1, y^m) - c(x^m, y^m) \geq 0,$$

and comparing with (3.9), we see that it is necessary to set, for all k,

$$\log \left\langle \pi(\mu^k), \frac{\mu^{k+1}}{\mu^k} \right\rangle = c\big(x^{k+1}, y^k\big) - c\big(x^k, y^k\big)\,. \tag{3.10}$$

Suppose now that π is a continuous FGP. Then by (3.3) — see also Proposition 6(ii)–(iii) of Pal and Wong (2016) — the portfolio is given as in the long-only case of Fernholz's original definition (0.10) of functionally-generated portfolios:

$$\pi_i : \mu \mapsto \mu_i\big(1 + \mathrm{D}_{e_i - \mu} \log \Phi(\mu)\big)\,, \qquad i = 1, \dots, n. \tag{3.11}$$

Substituting (3.11) into (3.10), we get that for all k

$$c\big(x^{k+1}, y^k\big) - c\big(x^k, y^k\big) = \log\left(\sum_{i=1}^{n}\big(1 + \mathrm{D}_{e_i-\mu^k}\log\Phi(\mu^k)\big)\mu_i^{k+1}\right).$$

Now, adding and subtracting several terms, we can transform this into the following:

$$\begin{aligned} c\big(x^{k+1}, y^k\big) - c\big(x^k, y^k\big) &= \log\left(\sum_{i=1}^{n}\big(1 + \mathrm{D}_{e_i-\mu^k}\log\Phi(\mu^k)\big)\mu_i^{k+1}\right) \\ &\quad - \log\left(\sum_{j=1}^{n}\big(1 + \mathrm{D}_{e_j-\mu^k}\log\Phi(\mu^k)\big)\right) \\ &\quad + \log\left(\sum_{j=1}^{n}\big(1 + \mathrm{D}_{e_j-\mu^k}\log\Phi(\mu^k)\big)\right) \\ &\quad - \log\Bigg(\underbrace{\sum_{i=1}^{n}\big(1 + \mathrm{D}_{e_i-\mu^k}\log\Phi(\mu^k)\big)\mu_i^{k}}_{=\sum_{i=1}^{n}\pi_i(\mu^k)=1}\Bigg) \\ &= \log\left(\sum_{i=1}^{n}\frac{1 + \mathrm{D}_{e_i-\mu^k}\log\Phi(\mu^k)}{\sum_{j=1}^{n} 1 + \mathrm{D}_{e_j-\mu^k}\log\Phi(\mu^k)}\,\mu_i^{k+1}\right) \\ &\quad - \log\left(\sum_{i=1}^{n}\frac{1 + \mathrm{D}_{e_i-\mu^k}\log\Phi(\mu^k)}{\sum_{j=1}^{n} 1 + \mathrm{D}_{e_j-\mu^k}\log\Phi(\mu^k)}\,\mu_i^{k}\right); \end{aligned}$$

identifying $\pi_i(\mu)/\mu_i$ in the numerators and denominators, we conclude that we must have

$$\begin{aligned} c\big(x^{k+1}, y^k\big) - c\big(x^k, y^k\big) &= \log\left(\sum_{i=1}^{n}\frac{\pi_i(\mu^k)/\mu_i^k}{\sum_{j=1}^{n}\pi_j(\mu^k)/\mu_j^k}\,\mu_i^{k+1}\right) \\ &\quad - \log\left(\sum_{i=1}^{n}\frac{\pi_i(\mu^k)/\mu_i^k}{\sum_{j=1}^{n}\pi_j(\mu^k)/\mu_j^k}\,\mu_i^{k}\right). \end{aligned}$$

Hence we choose

$$\mathcal{X} = \Delta_+^n, \qquad \mathcal{Y} = \bar{\Delta}_+^n \subseteq \mathbb{R}_{\geq 0}^n \setminus \{\mathbf{0}\}\,, \tag{3.12}$$

(this particular choice for $\mathcal{Y}$ will prove crucial later on, as we shall see in Theorem 3.2.2) and

$$c(x, y) := \log\langle x, y\rangle\,. \tag{3.13}$$

One can see that one is limited in the choice of the triplet $(c, \mathcal{X}, \mathcal{Y})$ by (3.10); changing $\mathcal{Y}$, for instance, would change the cost function c accordingly. An example of such an alternative is Equation (6) of Pal and Wong (2016). We deviate from their choice in order to arrive at an optimal transport problem that allows for a unique and deterministic optimal coupling — see Theorem 3.2.2 below.

With this selection of $(c, \mathcal{X}, \mathcal{Y})$, values $x \in \mathcal{X}$ correspond to market weights, and values $y \in \mathcal{Y}$ correspond to auxiliary weights $w \in \bar{\Delta}^n_+$, which are related to portfolio weights π through

$$w_i = \frac{\pi_i/\mu_i}{\sum_{j=1}^n \pi_j/\mu_j}, \qquad i = 1, \ldots, n.$$

Note also that in the above demonstration, these weights are a deterministic function of the market weights, as required of an FGP. In Corollary 3.2.3, we establish these results in general for all functionally-generated portfolios, and show that this deterministic function corresponds to the transport map of the optimal transport problem defined by (3.12) and (3.13).

3.2 Uniqueness of the optimal transport problem

We show a formal equivalence between FGP's and an optimal coupling of the problem (3.6), with $\mathcal{X}$ and $\mathcal{Y}$ as in (3.12), cost function (3.13), and any suitable pair of measures $\mathcal{P}$ and $\mathcal{Q}$ (this will be made clear later on). Theorem 2 of Pal and Wong (2016) gives us the implication 'π is FGP $\Rightarrow$ optimal coupling solves Monge problem'; this comes down to engineering an optimal transport problem in which the relation between $\mathcal{P}$ and $\mathcal{Q}$ is a change of variables defined through the portfolio function π — see Corollary 3.2.3 below. However, their result does not guarantee the existence of a transport map for a given optimal transport problem, with which a portfolio might be constructed. Here we show that the optimal coupling is unique and deterministic, thus guaranteeing the existence of a transport map. We stress again the model-independence of our result, as well as the absence of any time component, which is not required.

Instead of applying Kantorovich duality, we follow Villani's local approach to solving the Monge problem, which we state as follows:

3.2.1 Theorem (Theorem 10.28 of Villani (2009))**.** *Let M be an m-dimensional Riemannian manifold, $\mathcal{X}$ a closed subset of M with* $\dim(\partial\mathcal{X}) \leq m - 1$, *and $\mathcal{Y}$ an arbitrary Polish space. Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a continuous cost function, bounded below, and let $\mathcal{P}, \mathcal{Q}$ be such that the optimal cost* (3.6) *is $\mathcal{C}(\mathcal{P}, \mathcal{Q}) < \infty$. Assume that:*

(i) c is superdifferentiable everywhere;

(ii) $\nabla_x c(x, \cdot)$ is injective where defined;

(iii) $\mathcal{P} \ll \mathrm{Leb}$ *and* c *is Lipschitz.*[2]

Then there exists a unique (in law) optimal coupling $\mathcal{R}^*$ *of* $(\mathcal{P}, \mathcal{Q})$*, and it is deterministic. Therefore, there is a unique transport map* T *solving the Monge problem.*

We now show that the optimal coupling of the problem we have engineered in Section 3.1 is unique and deterministic, by considering a sequence of sub-problems and applying the above Theorem 3.2.1 to each of them, and then carefully taking the limit to recover the original optimal transport problem.

3.2.2 Theorem. *Let* $\mathcal{X} \times \mathcal{Y}$ *be as in* (3.12)*, and* c *as in* (3.13)*. Assume* $\mathcal{C}(\mathcal{P}, \mathcal{Q}) < \infty$ *and* $\mathcal{P} \ll \mathrm{Leb}$*, as well as* $\min_i \log x_i \in L^1(\mathcal{P})$*. Then the optimal transport problem* $(c, \mathcal{X}, \mathcal{Y})$ *has a unique optimal coupling, which is deterministic.*

Proof. We prove the statement in three steps. Since

$$\begin{aligned} c(x,y) = \log\left(\sum_{i=1}^n x_i y_i\right) &\geq \log\big(\min_{1\leq i\leq n} x_i\big) + \log\left(\sum_{i=1}^n y_i\right) \\ &= \min_{1\leq i\leq n} \log x_i, \qquad \forall (x,y) \in \mathcal{X} \times \mathcal{Y}, \end{aligned} \tag{3.14}$$

by our assumption $\min_i \log x_i \in L^1(\mathcal{P})$ and Theorem 4.1 of Villani (2009) it follows that an optimal coupling exists for our problem. Let $\mathcal{R}^*$ be such an optimal coupling; we aim to show that it is unique and deterministic.

Step 1. Define the sequence of sets

$$\mathcal{X}_m := \left\{ x \in \Delta_+^n \;\middle|\; \min_i x_i \geq \frac{1}{m} \right\} \subset \mathcal{X}, \qquad m \geq n, \tag{3.15}$$

where we note that $\mathcal{X} = \Delta_+^n \subset \mathbb{R}^n$ is an $(n-1)$-dimensional Riemannian manifold, and all the $\mathcal{X}_m$ are closed, with $\dim(\partial \mathcal{X}_m) = n-2$. Write $\mathcal{P}_m$ for the restriction of $\mathcal{P}$ to $\mathcal{X}_m$, defined by

$$\mathcal{P}_m(A) := \frac{\mathcal{P}(A \cap \mathcal{X}_m)}{\mathcal{P}(\mathcal{X}_m)}, \qquad \forall\, A \in \mathcal{B}(\mathcal{X}).$$

This is well-defined for sufficiently large m, since $\mathcal{X}_m \uparrow \mathcal{X}$. The cost function (3.13) is continuous, and we have $c(x,y) \geq -\log m$ on $\mathcal{X}_m \times \mathcal{Y}$. We check that the three conditions (i)–(iii) of Theorem 3.2.1 are satisfied for all m large:

(i) c is clearly differentiable, and hence superdifferentiable, everywhere;

[2] We have invoked Remark 10.33 of Villani (2009) to modify this assumption from its original formulation.

(ii) $y \mapsto \nabla_x c(x,y) = \left(\frac{y_1}{\sum_j x_j y_j}, \dots, \frac{y_n}{\sum_j x_j y_j}\right)$ is injective on $\mathcal{Y}$ (it is crucial here that we have chosen $\mathcal{Y} = \Delta^n_+$ instead of $\mathbb{R}^n_{\geq 0}$);

(iii) the last condition is satisfied since $\mathcal{P} \ll$ Leb by assumption, and c is Lipschitz; namely,

$$\begin{aligned}
|c(x',y') - c(x,y)| &\leq \sup_{(\tilde{x},\tilde{y}) \in \mathcal{X}_m \times \mathcal{Y}} ||\nabla c(\tilde{x},\tilde{y})|| \cdot ||(x',y') - (x,y)|| \\
&= \sup_{(\tilde{x},\tilde{y}) \in \mathcal{X}_m \times \mathcal{Y}} \frac{\sqrt{||\tilde{x}||^2 + ||\tilde{y}||^2}}{\langle \tilde{x}, \tilde{y} \rangle} \cdot ||(x',y') - (x,y)|| \\
&= \mathcal{K}_m \cdot ||(x',y') - (x,y)||, \qquad \forall\, (x,y), (x,y) \in \mathcal{X}_m \times \mathcal{Y},
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{K}_m &= \frac{\sqrt{(n-1)/m^2 + (1-(n-1)/m)^2 + 1}}{1/m} \\
&= \sqrt{n(n-1) - 2nm + 2m(m+1)}\,.
\end{aligned}$$

Theorem 3.2.1 then implies that there exists a unique optimal coupling $\mathcal{S}^*_m$ of $(\mathcal{P}_m, \mathcal{Q})$, which is deterministic. Write $T_m : \mathcal{X}_m \to \mathcal{Y}$ for the corresponding (unique) transport map.

Step 2. By the restriction property of optimal couplings, see Theorem 4.6 of Villani (2009), we have that the restriction $\mathcal{R}^*_m$ of $\mathcal{R}^*$ to $\mathcal{X}_m \times \mathcal{Y}$, defined by

$$\mathcal{R}^*_m(A) := \frac{\mathcal{R}^*(A \cap (\mathcal{X}_m \times \mathcal{Y}))}{\mathcal{R}^*(\mathcal{X}_m \times \mathcal{Y})}\,, \qquad \forall\, A \in \mathcal{B}(\mathcal{X} \times \mathcal{Y}), \tag{3.16}$$

is an optimal coupling of $(\mathcal{P}_m, \mathcal{Q})$. Hence by Step 1 it follows that we must have

$$\mathcal{R}^*_m = \mathcal{S}^*_m$$

for all m. Now by the Kantorovich duality (see Theorem 3.1.2 above), optimality of $\mathcal{R}^*_m$ is equivalent to the existence of a c-cyclically monotone set $\Gamma_m \subset \mathcal{X}_m \times \mathcal{Y}$ with $\mathcal{R}^*_m(\Gamma_m) = 1$; that is, for any sequence

$$\{x^1, \dots, x^l\} \subset \{x \in \mathcal{X}_m \,|\, \exists\, y \in \mathcal{Y} \text{ s.t. } (x,y) \in \Gamma_m\},$$

$l \in \mathbb{N}$, it holds that

$$\sum_{k=1}^{l} c\big(x^k, T_m(x^k)\big) \leq \sum_{k=1}^{l-1} c\big(x^{k+1}, T_m(x^k)\big) + c\big(x^1, T_m(x^l)\big). \tag{3.17}$$

For any $m' > m$, write $T_m^{m'}$ for the restriction of $T_{m'}$ to $\mathcal{X}_m$, and write $\mathcal{R}_m^{*m'}$ for the restriction of $\mathcal{R}_{m'}^*$ to $\mathcal{X}_m \times \mathcal{Y}$, defined as in (3.16). Again, by the restriction property of optimal couplings and uniqueness, we have $\mathcal{R}_m^{*m'} = \mathcal{R}_m^*$. We note that $T_m^{m'}$ satisfies equation (3.17) on the set $\Gamma_{m'} \cap (\mathcal{X}_m \times \mathcal{Y})$, which has $\mathcal{R}_m^*$-measure 1:

$$\begin{aligned}\mathcal{R}_m^*(\Gamma_{m'} \cap (\mathcal{X}_m \times \mathcal{Y})) &= \mathcal{R}_m^{*m'}(\Gamma_{m'} \cap (\mathcal{X}_m \times \mathcal{Y})) \\ &= \frac{\mathcal{R}_{m'}^*(\Gamma_{m'} \cap (\mathcal{X}_m \times \mathcal{Y}))}{\mathcal{R}_{m'}^*(\mathcal{X}_m \times \mathcal{Y})} = 1,\end{aligned}$$

since $\mathcal{R}_{m'}^*(\Gamma_{m'}) = 1$. Hence, by uniqueness of T_m, we conclude that

$$T_m^{m'} = T_m.$$

Step 3. Define the function $T : \mathcal{X} \to \mathcal{Y}$ pointwise by

$$T : x \mapsto T_{\lceil (\min_i x_i)^{-1} \rceil}(x),$$

that is, by the value that all functions T_m assign to x for m large enough; by Step 2, these functions agree on this value. Let the deterministic coupling $\widetilde{\mathcal{R}}$ of $(\mathcal{P}, \mathcal{Q})$ be the one with transport map T. By uniqueness, the restrictions of $\widetilde{\mathcal{R}}$ and $\mathcal{R}^*$ to $\mathcal{X}_m \times \mathcal{Y}$ coincide for all m. Therefore we have that

$$\mathcal{R}^*(A) = \lim_{m\to\infty} \mathcal{R}^*(A \cap (\mathcal{X}_m \times \mathcal{Y})) = \widetilde{\mathcal{R}}(A), \qquad \forall\, A \in \mathcal{B}(\mathcal{X} \times \mathcal{Y}),$$

by the monotone-convergence property of measures. This shows that $\mathcal{R}^*$ is deterministic, and is the unique optimal coupling. □

This theorem extends Pal and Wong (2016), in that it proves the converse to their Theorem 2 that each FGP is the solution of an optimal transport problem, thus implying equivalence. Namely, they do not establish the existence of a deterministic solution to the transport problem, which bars them from showing that any pair of measures $\mathcal{P}, \mathcal{Q}$ defines a functionally-generated portfolio through the optimal coupling's transport map. Unlike Theorem 2 of Pal and Wong (2016), we do not need the assumption that $\log(\pi_i(\mu)/\mu_i)$ is coordinate-wise bounded below, but only require the condition $\min_i \log x_i \in L^1(\mathcal{P})$. Thus, we conclude the following:

3.2.3 Corollary. *Let $\mathcal{X} \times \mathcal{Y}$ be as in* (3.12)*, and c as in* (3.13)*. Assume $\mathcal{P}$ and $\mathcal{Q}$ are such that $\mathcal{C}(\mathcal{P}, \mathcal{Q}) < \infty$ and $\mathcal{P} \ll$ Leb, as well as $\min_i \log x_i \in L^1(\mathcal{P})$.*

If $\pi = (\pi_1, \ldots, \pi_n) : \Delta_+^n \to \bar{\Delta}_+^n$ is an FGP, then

$$T : \Delta_+^n \to \bar{\Delta}_+^n, \quad x \mapsto \left(\frac{\pi_1(x)/x_1}{\sum_{j=1}^n \pi_j(x)/x_j}, \ldots, \frac{\pi_n(x)/x_n}{\sum_{j=1}^n \pi_j(x)/x_j} \right) \tag{3.18}$$

is the transport map of the unique optimal coupling of the optimal transport problem $(c, \mathcal{X}, \mathcal{Y})$ *with* $\mathcal{P}$ *and* $\mathcal{Q}$ any *measures such that if* $\mu \sim \mathcal{P}$*, then* $T(\mu) \sim \mathcal{Q}$.

Conversely, our optimal transport problem $(c, \mathcal{X}, \mathcal{Y})$ *has, for any pair* $\mathcal{P}$*,* $\mathcal{Q}$*, a unique transport map* $T = (T_1, \ldots, T_n) : \Delta^n_+ \to \bar{\Delta}^n_+$*, and the portfolio defined through*

$$\pi : \Delta^n_+ \to \bar{\Delta}^n_+, \quad \mu \mapsto \left(\frac{\mu_1 T_1(\mu)}{\sum_{j=1}^n \mu_j T_j(\mu)}, \ldots, \frac{\mu_n T_n(\mu)}{\sum_{j=1}^n \mu_j T_j(\mu)} \right) \tag{3.19}$$

is functionally-generated.

Proof. Given a functionally-generated portfolio π, it is MCM by Proposition 4 of Pal and Wong (2016): for any sequence of market weights $\{\mu^1, \ldots, \mu^m\} \subset \Delta^n_+$, $m \in \mathbb{N}$, and writing $\mu^{m+1} := \mu^1$ for convenience (and as in Definition 5.1 of Villani (2009)), it holds that

$$\begin{aligned}
& \sum_{k=1}^m \log \left\langle \pi(\mu^k), \frac{\mu^{k+1}}{\mu^k} \right\rangle \geq 0 \\
\iff & \sum_{k=1}^m \log \left(\left\langle T(\mu^k), \mu^{k+1} \right\rangle \left\langle \pi(\mu^k), \mu^k \right\rangle \right) \geq 0 \\
\iff & \sum_{k=1}^m \log \left\langle T(\mu^k), \mu^{k+1} \right\rangle - \sum_{k=1}^m \log \left\langle T(\mu^k), \mu^k \right\rangle \geq 0.
\end{aligned}$$

That is, the deterministic coupling defined by the transport map T as in (3.18) is c-cyclically monotone. Using (3.14), we may apply Kantorovich duality to conclude that $\mathcal{R}$ is an optimal coupling of the optimal transport problem; by Theorem 3.2.2 it is unique.

For the converse, let $\mathcal{R}$ be an optimal coupling of our optimal transport problem. By Theorem 3.2.2, it is deterministic and the unique optimal coupling of the optimal transport problem, with some transport map T. We apply Kantorovich duality to conclude that $\mathcal{R}$ is c-cyclically monotone: for every sequence $\{x^1, \ldots, x^m\}$ in Δ^n_+, with $m \in \mathbb{N}$, and writing $x^{m+1} := x^1$, it holds that

$$\begin{aligned}
& \sum_{k=1}^m \log \left\langle x^{k+1}, T(x^k) \right\rangle \geq \sum_{k=1}^m \log \left\langle x^k, T(x^k) \right\rangle \\
\iff & \sum_{k=1}^m \log \left(\frac{\left\langle x^{k+1}, T(x^k) \right\rangle}{\left\langle x^k, T(x^k) \right\rangle} \right) \geq 0 \\
\iff & \sum_{k=1}^m \log \left\langle \frac{\pi(x^k)}{x^k}, x^{k+1} \right\rangle \geq 0.
\end{aligned}$$

Hence π defined by (3.19) satisfies MCM, which implies that π is functionally-generated by Proposition 4 of Pal and Wong (2016). $\square$

We conclude the following: given any distribution $\mathcal{P} \ll \text{Leb}$ and any 'objective distribution' $\mathcal{Q}$, the unique transport map of our optimal transport problem defines a functionally-generated portfolio through (3.19). Note that finding the generating function of that portfolio is a non-trivial problem; in the case that $\pi \in C^0\big(\Delta^n_+, \bar{\Delta}^n_+\big)$, it amounts to solving (for Φ) the following system of linear PDE's:

$$\frac{T_i(\mu)}{\sum_{j=1}^n \mu_j T_j(\mu)} = 1 + \mathrm{D}_{e_i-\mu} \log \Phi(\mu), \qquad i = 1, \ldots, n. \tag{3.20}$$

This follows by an application of (3.11). Nevertheless, the existence of such a function Φ is guaranteed by the fact that π is functionally-generated.

Finally note that, given an FGP $\pi : \Delta^n_+ \to \bar{\Delta}^n_+$, the optimal cost of the corresponding optimal transport problem is, by Corollary 3.2.3, equal to

$$\mathcal{C}(\mathcal{P}, \mathcal{Q}) = \int_{\Delta^n_+} \log \langle x, T(x) \rangle \, \mathrm{d}\mathcal{P}(x) = -\int_{\Delta^n_+} \log \left(\sum_{j=1}^n \pi_j(x)/x_j \right) \mathrm{d}\mathcal{P}(x).$$

Hence, the exact cost depends on the choice of $\mathcal{P}$.

3.2.4 Remark. Before proceeding to study other investment rules, let us interpret the above optimal transport problem. Recall the cost function $c(\mu, w) = \log \langle \mu, w \rangle$; using the relation (0.19) for the strategy (i.e. number of shares η) corresponding to a portfolio, and equation (3.18), we have that

$$w_i := T_i(\mu) = \frac{\pi_i/\mu_i}{\sum_{j=1}^n \pi_j/\mu_j} = \frac{\eta_i/\sum_{k=1}^n \eta_k \mu_k}{\sum_{j=1}^n \eta_j/\sum_{k=1}^n \eta_k \mu_k} = \frac{\eta_i}{\sum_{j=1}^n \eta_j};$$

that is, the vector of weights $w \in \bar{\Delta}^n_+$ gives the proportions of the total number of shares that are held in each stock i. Therefore we may rewrite the cost function (3.13) as

$$c(\mu, w) = \log \langle \mu, w \rangle = \log \left(\frac{\sum_{i=1}^n \eta_i \mu_i}{\sum_{i=1}^n \eta_i} \right) = \log \left(\frac{\sum_{i=1}^n \eta_i X_i}{\sum_{i=1}^n \eta_i \bar{X}} \right) - \log n,$$

where $\bar{X} := \frac{1}{n} \sum_{j=1}^n X_j$. Hence one interpretation of the quantity that is minimized in the optimal transport problem is the following: it is the ratio of the absolute dollar value of the strategy to its hypothetical dollar value were the market equally distributed (i.e. $\mu = \frac{1}{n}\mathbf{1}$). In a sense, the optimal transport problem selects portfolios that underperform their hypothetical implementation in a market in which all stock prices are replaced by their arithmetic average $\bar{X}$. Corollary 3.2.3 then yields that for a given distribution $\mathcal{P}$ of μ, the way to maximize this 'underperformance' is through a functionally-generated portfolio — that is, a portfolio function satisfying (3.2). Note that this interpretation is completely model-free; we have made no assumptions on the behavior of the market weights whatsoever. □

3.3 Additive functional generation

We now emulate the analysis of Section 3.2, extending the correspondence between optimal transport and functional generation to another class of investment rules, namely that of additively generated strategies. These were introduced and studied very recently in Karatzas and Ruf (2016) — see Section 0.1.6 of the Introduction, and Definition 0.1.9 in particular. Again, we derive the appropriate optimal transport problem for which additively functionally-generated strategies define the transport map (see Lemma 3.3.2), which in Proposition 3.3.3 we show to be the unique optimal coupling. The converse is shown in Lemma 3.3.4; however, since the strategies depend on the entire history of market weights, we do not obtain an explicit formula as in the classical case — recall (3.19).

In this section and the next, we do need a model for the market weights — albeit a very general one — in order to be able to define stochastic integrals with respect to them. We place ourselves in a general continuous-time framework with a finite horizon $[0, T]$, and introduce a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and filtration $\mathbb{F} = \{\mathcal{F}(t)\}_{t\in[0,T]}$. We assume that $\mathbb{F}$ satisfies the usual conditions, and on this probability space we consider the market weights vector $\mu(\cdot)$, which we assume to be a continuous semimartingale. As we wrote in Section 0.1.2 of the Introduction, a portfolio $\pi(\cdot)$ is a $\bar{\Delta}^n_+$-valued, $\mathbb{F}$-progressively measurable process, and expresses the *proportions of wealth* to be invested at each point in time.

3.3.1 Redefining additive functional generation

Recall from (0.21) and (0.22) that a trading strategy $\eta(\cdot)$ which is additively generated by a C^2 function $\Psi : \Delta^n_+ \to \mathbb{R}$ satisfies

$$\eta_i(t) = \mathrm{D}_{e_i-\mu(t)}\Psi(\mu(t)) + V(t), \qquad i = 1, \ldots, n, \tag{3.21}$$

where the market-relative wealth $V(\cdot)$ equals $V'(\cdot)$ of (0.17).

Before we relate additive functional generation to optimal transport, let us take a step back and look at the classical definition of functional generation (called 'multiplicative functional generation' by Karatzas and Ruf (2016)) from a different perspective. Using the relation (0.19), we can rewrite Fernholz's expression (3.11) for a multiplicatively generated portfolio as

$$\frac{\eta_i(t)}{V(t)} = \varphi_i(\mu(t)) := 1 + \mathrm{D}_{e_i - \mu(t)} \log \Phi(\mu(t)); \tag{3.22}$$

that is, the ratio of the number of shares held in each firm to the total relative wealth is a deterministic function φ of the market weights. Likewise, for additively generated strategies (3.21) we have that

$$1 + \eta_i(t) - V(t) = \psi_i(\mu(t)) := 1 + \mathrm{D}_{e_i - \mu(t)} \Psi(\mu(t)); \tag{3.23}$$

that is, the *difference* between the number of shares held in each firm and the relative wealth is a deterministic function ψ of the market weights. This observation sheds light on which approach we should take; as in the multiplicative case, we wish to construct an optimal transport problem of which (a normalization of) the deterministic function ψ will be the optimal transport map (recall (3.18)), and choose the cost function $\check{c}$ so that $\check{c}$-cyclical monotonicity will coincide with a cyclical monotonicity property of $\eta(\cdot)$.

First of all, we slightly modify the definition of additively generated portfolios, and take it to be the following (compare with (3.2) and (3.3)):

3.3.1 Definition. A trading strategy $\eta(\cdot)$ is *additively generated* if

$$\psi(t) := \eta(t) + (1 - V(t))\mathbf{1} = \psi(\mu(t)) \tag{3.24}$$

for some deterministic function $\psi : \Delta^n_+ \to \mathbb{R}^n$, and if there exists a concave function $\Psi : \Delta^n_+ \to \mathbb{R}$ such that

$$\psi(\mu) \in \partial\Psi(\mu), \qquad \forall\, \mu \in \Delta^n_+,$$

where $\partial\Psi(\mu)$ denotes the superdifferential of Ψ at $\mu \in \Delta^n_+$; equivalently, by (3.4),

$$\Psi(\mu) + \langle \psi(\mu), \nu - \mu \rangle \geq \Psi(\nu), \qquad \forall\, \mu, \nu \in \Delta^n_+. \tag{3.25}$$

□

The justification for the above definition is the following, and is similar to (3.5): if ψ satisfies (3.25), then for fixed $\mu \in \Delta_+^n$, $1 \le i \le n$, and $h \in \mathbb{R} \setminus \{0\}$ sufficiently small so that $\mu + h(e_i - \mu) \in \Delta_+^n$, we have that

$$\begin{aligned} \Psi(\mu) + \langle \psi(\mu), h(e_i - \mu)\rangle &\ge \Psi(\mu + h(e_i - \mu)) \\ \iff \langle \psi(\mu), e_i - \mu\rangle &\ge \frac{\Psi(\mu + h(e_i - \mu)) - \Psi(\mu)}{h} \\ \Rightarrow 1 + \mathrm{D}_{e_i - \mu}\Psi(\mu) \le \psi_i(\mu) &\le 1 - \mathrm{D}_{\mu - e_i}\Psi(\mu), \end{aligned}$$

where the last inequalities follow by taking $h \downarrow 0$ and $h \uparrow 0$, respectively, and where we have used the fact that

$$\langle \psi(\mu(t)), \mu(t)\rangle = \big\langle \eta(t) + \big(1 - \langle \eta(t), \mu(t)\rangle\big)\mathbf{1}, \mu(t)\big\rangle = 1, \qquad \forall t \ge 0.$$

When $\Psi \in C^1$, the directional derivatives coincide, and we have

$$\psi_i(\mu) = 1 + \mathrm{D}_{e_i - \mu}\Psi(\mu), \qquad i = 1, \ldots, n; \tag{3.26}$$

that is, we recover (3.23), and our Definition 3.3.1 is consistent with that of Karatzas and Ruf (2016).

3.3.2 Relation to optimal transport

As in Section 3.1, we wish to engineer an optimal transport problem that is related to additively generated strategies. Without any additional work, there is equivalence between additive functional generation and ‘additive’ cyclical monotonicity, which we will abbreviate as ACM (and which is simply cyclical monotonicity in convex analysis — see Definition 3.1.3 above), in a similar way that equivalence holds between multiplicative functional generation and multiplicative cyclical monotonicity.

Namely, let us recall cyclical monotonicity of ψ from (3.8): for any sequence $\{\mu^1, \ldots, \mu^m\} \subset \Delta_+^n$, $m \in \mathbb{N}$, it requires

$$\sum_{k=1}^{m-1} \big\langle \psi(\mu^k), \mu^{k+1} - \mu^k\big\rangle + \big\langle \psi(\mu^m), \mu^1 - \mu^m\big\rangle \ge 0. \tag{3.27}$$

As mentioned before, we now wish to choose a cost function $\check{c}$ in such a way that $\check{c}$-cyclical monotonicity coincides with (3.27). This turns out to be very straightforward: since $\check{c}$-cyclical monotonicity (3.7) states that

$$\sum_{k=1}^{m-1} \big(\check{c}\big(x^{k+1}, y^k\big) - \check{c}\big(x^k, y^k\big)\big) + \check{c}(x^1, y^m) - \check{c}(x^m, y^m) \ge 0,$$

over all sequences $\{(x^1, y^1), \dots, (x^m, y^m)\}$, $m \in \mathbb{N}$, comparing with (3.27) leads us to require for all k that

$$\big\langle \psi(\mu^k), \mu^{k+1} - \mu^k \big\rangle = \check{c}\big(x^{k+1}, y^k\big) - \check{c}\big(x^k, y^k\big) \,.$$

Hence we should clearly choose

$$\check{c}(x, y) = \langle x, y \rangle \tag{3.28}$$

and

$$\mathcal{X} = \Delta^n_+,$$

and we set

$$\mathcal{Y} = \big\{ y \in \mathbb{R}^n : \sum_{i=1}^n y_i = 1 \big\};$$

notice the relative simplicity in comparison with multiplicative generation of Section 3.2. Again, we have chosen $\mathcal{Y}$ so that uniqueness of the optimal coupling follows — see Proposition 3.3.3 below. Since $\check{c}$ satisfies the conditions of Theorem 3.1.2, a direct application of Kantorovich duality yields that $\check{c}$-cyclical monotonicity of a coupling implies optimality of that coupling for the optimal transport problem. Thus we conclude the following:

3.3.2 Lemma. *Fix $t \in [0, T]$, and let the strategy $\eta(\cdot)$ be additively generated by $\Psi : \Delta^n_+ \to \mathbb{R}^n$, with corresponding $\psi : \Delta^n_+ \to \mathbb{R}^n$ as in (3.23). Define the optimal transport problem*

$$\check{C}(\mathcal{P}, \mathcal{Q}) := \inf_{\mathcal{R} \in \Pi(\mathcal{P}, \mathcal{Q})} \int_{\mathcal{X} \times \mathcal{Y}} \check{c}(x, y) \, \mathrm{d}\mathcal{R}(x, y), \tag{3.29}$$

with $\mathcal{P} = \mathrm{Law}\big(\mu(t)\big)$, cost function $\check{c}$ as in (3.28), and $\mathcal{X} \times \mathcal{Y} = \Delta^n_+ \times \{y \in \mathbb{R}^n : \sum_i y_i = 1\}$. Then the function $\check{T} : \mathcal{X} \to \mathcal{Y}$ defined by

$$\check{T}(\mu) := \psi(\mu) + \frac{1}{n}\big(1 - \langle \psi(\mu), \mathbf{1} \rangle\big)\mathbf{1} \tag{3.30}$$

is a transport map of (3.29), with $\mathcal{Q}$ the law of $\check{T}(\mu(t))$. In the case that $\Psi \in C^1(\Delta^n_+, \mathbb{R})$, this simplifies to

$$\check{T}_i(\mu) = \mathrm{D}_{e_i - \mu}\Psi(\mu) + \frac{1}{n}\Big(1 - \sum_{j=1}^n \mathrm{D}_{e_j - \mu}\Psi(\mu)\Big), \qquad i = 1, \dots, n. \tag{3.31}$$

Proof. Additive generation of $\eta(\cdot)$ implies cyclical monotonicity of ψ, by Definition 3.3.1 and Theorem 3.1.4, which immediately implies cyclical monotonicity of $\check{T}$. (The second term in (3.30) is merely for normalization.) By construction, this is equivalent to $\check{c}$-cyclical monotonicity of the deterministic coupling with transport map $\check{T}$. Finally, invoking Kantorovich duality, optimality of the coupling follows. The equivalence of the definition in (3.30) with that of (3.31) is then implied by (3.26). $\square$

Compare (3.31) to the multiplicative case; by (3.18) and (3.11), the transport map T of the optimal transport problem corresponding to a portfolio that is multiplicatively generated by some $\Phi \in C^1(\Delta^n_+, \bar{\Delta}^n_+)$ satisfies

$$T_i(\mu) = \frac{1 + \mathrm{D}_{e_i - \mu}\Phi(\mu)}{n + \sum_{j=1}^n \mathrm{D}_{e_j - \mu}\Phi(\mu)}\,, \qquad i = 1, \ldots, n.$$

Interestingly, the converse to Lemma 3.3.2 does *not* hold; given an optimal coupling of (3.29), which is unique and necessarily deterministic as we show in Proposition 3.3.3 below, one cannot explicitly construct an additively generated strategy, contrary to the case of multiplicative generation.

3.3.3 Proposition. *Assume* $\mathcal{P} \ll \mathrm{Leb}$*. Then the optimal transport problem* (3.29) *has a unique optimal coupling, which is deterministic.*

Proof. As in Theorem 3.2.2, we check that all the conditions of Theorem 3.2.1 are satisfied. Similarly to (3.15), introduce the sequence of sets

$$\mathcal{Y}_m := \big\{ y \in \mathcal{Y} : \min_i y_i \geq -m \big\}, \qquad m \in \mathbb{N}; \tag{3.32}$$

note that $\mathcal{Y}_m \uparrow \mathcal{Y}$. Then it holds that $\check{c}$ takes values in $(-m, (n-1)m+1)$ on $\mathcal{X} \times \mathcal{Y}_m$ for all $m \geq 1$, so the optimal cost $\check{C}(\mathcal{P}, \mathcal{Q})$ is finite, and an optimal coupling exists by Theorem 4.1 of Villani (2009). We check that the three conditions (i)–(iii) of Theorem 3.2.1 are satisfied on the sets $\mathcal{Y}_m$ of (3.32) for all $m \geq 1$:

(i) $\check{c}$ is clearly differentiable, and hence superdifferentiable, on $\mathcal{X} \times \mathcal{Y}_m$;

(ii) $y \mapsto \nabla_x \check{c}(x, y) = y$ is injective on $\mathcal{X} \times \mathcal{Y}_m$;

(iii) $\mathcal{P} \ll \mathrm{Leb}$ by assumption, and $\check{c}$ is Lipschitz on $\mathcal{X} \times \mathcal{Y}_m$ since all of its partial derivatives are bounded.

The rest now follows as in the proof of Theorem 3.2.2. $\square$

We proceed by showing the existence of an additively functionally-generated strategy that is defined through the transport map of the problem (3.29).

3.3.4 Lemma. *Assume $\mathcal{P} \ll \mathrm{Leb}$. Given the unique transport map $\check{T} = (\check{T}_1, \ldots, \check{T}_n)$ of the optimal transport problem (3.29), and any constants $\alpha > 0$, $\beta \in \mathbb{R}$, define the function*

$$\psi : \Delta_+^n \to \mathbb{R}^n, \quad \mu \mapsto \alpha \check{T}(\mu) + \beta \mathbf{1}. \tag{3.33}$$

Then there exists a concave function $\Psi : \Delta_+^n \to \mathbb{R}$ such that $\psi(\mu) \in \partial\Psi(\mu)$ for all $\mu \in \Delta_+^n$. Furthermore, if $\Psi \in C^2(\Delta_+^n, \mathbb{R})$ there exists a strategy $\eta(\cdot)$ for which we have

$$\eta(t) + (1 - V(t))\mathbf{1} = \psi(\mu(t)), \qquad \forall\, t \in [0, T]. \tag{3.34}$$

Proof. Kantorovich duality (3.1.2) implies that the deterministic coupling satisfies $\check{c}$-cyclical monotonicity: writing $x^{m+1} := x^1$, for every sequence $\{x^1, \ldots, x^m\}$ in $\mathcal{X}$, $m \in \mathbb{N}$, we have

$$\begin{aligned} &\sum_{k=1}^{m} \left\langle \check{T}(x^k), x^{k+1} \right\rangle - \sum_{k=1}^{m} \left\langle \check{T}(x^k), x^k \right\rangle \geq 0 \\ \Rightarrow &\sum_{k=1}^{m} \alpha \left\langle \check{T}(x^k), x^{k+1} - x^k \right\rangle \geq 0 \\ \Rightarrow &\sum_{k=1}^{m} \left\langle \psi(x^k), x^{k+1} - x^k \right\rangle \geq 0, \end{aligned}$$

where we use the fact that the components of each x^k sum up to one, since $x^k \in \Delta_+^n$. Hence cyclical monotonicity of ψ follows, which by Theorem 3.1.4 implies the existence of a concave function $\Psi : \Delta_+^n \to \mathbb{R}$ with the desired property. Finally, let the trading strategy $\eta(\cdot)$ be defined through Definition 0.1.9 with $\mathcal{G} = \Psi$; by an application of Theorem 3.6(i) of Karatzas and Ruf (2016), this is possible, and $\eta(\cdot)$ satisfies (3.34). □

We have shown the existence of an additively generated strategy that is related to the transport map of our optimal transport problem. Combining (3.23) with (3.33), we see that the linear system of partial differential equations that defines the generating function Ψ of the additively generated strategy (3.34) is

$$\alpha \check{T}_i(\mu) + \beta = 1 + \mathrm{D}_{e_i - \mu(t)} \Psi(\mu(t)), \qquad i = 1, \ldots, n.$$

compare with (3.20). Again, this has a solution because $\eta(\cdot)$ is additively generated.

Considering $\eta(\cdot)$ itself, it is a solution to the linear system (3.34), which can be rewritten as

$$\eta_i(t) - \sum_{j=1}^{n} \eta_j(t)\mu_j(t) = \psi_i(\mu(t)) - 1, \qquad i = 1, \ldots, n.$$

However, note that this linear system alone is not sufficient to determine an explicit form for $\eta(\cdot)$; in particular, given a solution $\eta(\cdot)$, we have that

$$\big(\eta_i(t)+\Theta(t)\big)-\sum_{j=1}^{n}\big(\eta_j(t)+\Theta(t)\big)\mu_j(t)=\eta_i(t)-\sum_{j=1}^{n}\eta_j(t)\mu_j(t), \qquad i=1,\dots,n$$

for any adapted process $\Theta(\cdot)$, implying that $\eta(\cdot)+\Theta(\cdot)$ is also a solution. Equation (0.18) resolves this issue, in that it augments the system with the requirement of self-financiability (or 'admissibility') of the strategy. However, it does introduce dependency of $\eta(\cdot)$ (and of the corresponding portfolio $\pi(\cdot)$, defined through (0.19)) on the *entire history* of market prices, explaining why we do not retrieve a formula of the form (3.11), which only depends on currently observable prices. See also Equations (4.1) and (4.4) of Karatzas and Ruf (2016) for formulae for additively generated strategies.

3.4 Discussion

We have established connections between functional generation and optimal transport problems that go beyond Pal and Wong (2016): we proved a formal equivalence between the two, and have extended the relation to *additive* functional generation. We apply these results in Chapter 4 to obtain a model-free lower bound on the market-relative wealth process of a functionally-generated portfolio — see Corollary 4.2.1 below.

Let us draw the following generalization from the results of Section 3.2: Note that at each point in time t, an optimal transport problem from a distribution $\mathcal{P}_t$, to some objective distribution $\mathcal{Q}_t$, gives rise to a *unique* deterministic portfolio map by Theorem 3.2.2, generated by some concave $\Phi_t : \Delta^n_+ \to \mathbb{R}_{\geq 0}$. In general, this function will not be time-invariant, since the optimal coupling will be different at each point in time, and, assuming $\Phi_t \in C^1$, it will generate at each point in time a different portfolio

$$\pi_i(t)=\mu_i(t)\big(1+\mathrm{D}_{e_i-\mu(t)}\log\Phi_t(\mu(t))\big).$$

Saying that $\pi(\cdot)$ is multiplicatively functionally-generated corresponds to requiring $\Phi_t \equiv \Phi$ for some fixed function Φ; the objective measure $\mathcal{Q}_t$ then becomes the change of variables from $\mathcal{P}_t$ using Φ. However, one might take the approach of fixing the objective distribution $\mathcal{Q}_t \equiv \mathcal{Q}$ instead, leading at each point in time to a different portfolio function (that is functionally-generated in the sense of (3.2)), and hence to a portfolio that is *not* functionally-generated in the sense of (0.10). It would be interesting to study the properties and possible applications of such portfolios.

As was pointed out in the introduction to this section, we have excluded the possibility of default by making the choice $\mu(t) \in \Delta^n_+$, $\forall\, t \geq 0$ — this is modified in Section 4.5. Since firms do go bankrupt in real life (which one might model as the corresponding market weight hitting 0), one might desire to require $\mu(t) \in \bar{\Delta}^n_+$ instead. However, our cost function (3.13) is unbounded from below on this set, and so the sequencing argument in the proof of Theorem 3.2.2 ceases to work; hence it is not straightforward to see how equivalence to an optimal transport problem could be extended to this case.

Likewise, one could consider changing $\mathcal{X}$ to $\Delta^n_+ \times \mathbb{R}^d$, where the d additional degrees of freedom could describe characteristics of firms other than their market capitalizations. This would give a possible theoretical relation between the approach of Strong (2014a) (or even our approach in Chapter 2) and an optimal transport problem.

Another idea for future research is to consider outperformance of the market from a *hedging* perspective, and to use known links with (martingale) optimal transport — see Dolinsky and Soner (2014), for instance. This would conceivably have connections to the Skorokhod embedding problem via martingale optimal transport, which have been established in Beiglböck et al. (2016) and remain an area of active study. Yet another possibility would be to apply causal transport, which allows one to introduce a filtration on the probability space that the market weights process lives in; see Lassalle (2013).

Chapter 4

Properties & extensions of functionally-generated wealth

In this final chapter, we derive several theoretical results regarding the market-relative wealth processes corresponding to functionally-generated strategies. The first part, Sections 4.1–4.4, considers these wealth processes in the general continuous semimartingale setting of Section 3.3, and in this framework establishes new results as well as generalizing existing ones. In the second part (Section 4.5) of this chapter, we place ourselves in the Itô model of (0.1) (i.e. the same that was employed in Chapters 1 and 2), but extend the SPT framework to include the possibility of defaulting firms, or bankruptcies. This leads to portfolios that can outperform the market even in the presence of disappearing firms, making them more applicable in the real world.

We start the first part with the proof of a general result: in Section 4.1, we show that the wealth process of a discretely-rebalanced portfolio tends to its continuously rebalanced implementation as the rebalancing frequency goes to infinity — showing this is a non-trivial exercise, because the wealth process is defined indirectly through the stochastic differential equation (0.2). In Section 4.2, we then use this mathematical tool, together with some results from Chapter 3, to derive a model-free lower bound on the terminal market-relative wealth functionally-generated portfolios. In another application of our limiting result, we construct a proof of Fernholz's master equation (0.11) in a general continuous semimartingale market model; see Section 4.3. Such a generalization has also been proved in Karatzas and Ruf (2016), where the space of market weights is furthermore extended from Δ_+^n to $\bar{\Delta}_+^n$. Finally, we apply the Optional Decomposition Theorem to show that an appropriate decomposition of any continuous stochastic process implies that it corresponds to the wealth process of a functionally-generated strategy — see Section 4.4. This can be seen as a converse to the master equation.

In Section 4.5, we develop two methods that modify portfolio construction through (multiplicative) functional generation in order to incorporate bankruptcies. We refrain from changing the underlying market model, but instead constrain the set of admissible portfolios, and impose a penalty each time a firm defaults. The first method constructs a portfolio by linearly combining functionally-generated portfolios, each of which is guaranteed to outperform the market in a different scenario of defaulting firms. The second method modifies functional generation by applying a dynamically changing generating function, leading to a modified master equation — see (4.37). In each case, we demonstrate how relative arbitrages can still be constructed.

4.1 Continuous-time limit of the relative wealth process

We first relate Definition 3.1.5 of cyclical monotonicity to discrete rebalancing in continuous time, by associating sequences of market weights with values of the market weights *process* $\mu(\cdot)$ at discrete times — in Theorem 4.1.2, we will show that taking the limit as the rebalancing frequency goes to infinity yields the continuous-time wealth process.

Here and in the following sections, we place ourselves in the continuous-time setting of Section 3.3; in particular, recall our assumption that $\mu(\cdot)$ is a continuous semimartingale. Let $\pi(\cdot)$ be any portfolio. Imagine that instead of implementing $\pi(\cdot)$ continuously, we rebalance it only at a finite number of points in time, $t_1 < \ldots < t_m$, $m \in \mathbb{N}$; that is, on each interval $[t_{k-1}, t_k)$, we hold a fixed number of shares that is determined at time t_{k-1}. If we denote by $V^m(\cdot)$ the market-relative wealth process that results from this, then we have that the number of shares $\eta_i(t_{k-1})$ held in stock i over the interval $[t_{k-1}, t_k)$ satisfies

$$\eta_i(t_{k-1}) = \frac{\pi_i(t_{k-1}) V^m(t_{k-1})}{\mu_i(t_{k-1})} .$$

Hence the relative return over the period $[t_{k-1}, t_k)$ is

$$\frac{V^m(t_k)}{V^m(t_{k-1})} = \frac{\langle \eta(t_{k-1}), \mu(t_k) \rangle}{V^m(t_{k-1})} = \left\langle \pi(t_{k-1}), \frac{\mu(t_k)}{\mu(t_{k-1})} \right\rangle ,$$

and the total log-relative return over $[t_1, t_m]$ is

$$\log \left(\frac{V^m(t_m)}{V^m(t_1)} \right) = \sum_{k=1}^{m-1} \log \left\langle \pi(\mu(t_k)), \frac{\mu(t_{k+1})}{\mu(t_k)} \right\rangle ; \tag{4.1}$$

compare with (3.9). Therefore, one way of interpreting MCM is as follows: over any *cycle* of market weights (i.e. a path $\{\mu(t)\}_{t\in[0,T]}$ with $\mu(t_m) = \mu(t_1)$), any piecewise buy-and-hold implementation of π gives a non-negative log-relative return

$$\log\left(\frac{V^m(t_m)}{V^m(t_1)}\right) \geq 0. \tag{4.2}$$

By setting $t_k = k$, this interpretation is what allows Pal and Wong (2016) to relate MCM to the wealth process of a multiplicatively functionally-generated portfolio, and to arbitrage, in a discrete-time setting.

However, in continuous time, we need a different interpretation of (the continuous-time limit of) cyclical monotonicity, since it does not make sense to make statements about the set of cycles of market weights, which is negligible. Namely, since the wealth process $V(\cdot)$ is only defined up to null sets, see (0.6), making statements on sets of probability zero does not have any consequences for the process itself. A process which disagrees with it only on such sets is indistinguishable from it. In the following, we carefully take the continuous-time limit of the quantity that appears on the right hand side of (4.1), which we interpret in Section 4.2.

Given any portfolio $\pi(\cdot)$, define the process

$$V^m(t) := \prod_{k=1}^{K}\sum_{i=1}^{n}\frac{\pi_i(t_k\wedge t)}{\mu_i(t_k\wedge t)}\,\mu_i(t_{k+1}\wedge t)\,, \tag{4.3}$$

where we define the partition $\{t_k\}_k$ by

$$\begin{aligned} t_k &:= \tilde{t}_k\,, \qquad k = 1,\ldots,K,\\ t_{K+1} &:= T\,, \end{aligned}$$

with the stopping times

$$\begin{aligned} \tilde{t}_1 &:= 0,\\ \tilde{t}_{k+1} &:= \left(\tilde{t}_k + \frac{1}{m}\right)\wedge \inf\left\{t > \tilde{t}_k : \sum_{i=1}^{n}\left|\mu_i(t)-\mu_i(\tilde{t}_k)\right| + \left|\frac{\pi_i(t)}{\mu_i(t)} - \frac{\pi_i(\tilde{t}_k)}{\mu_i(\tilde{t}_k)}\right| > \frac{1}{m}\right\} \end{aligned}$$

for all $k \geq 1$, and

$$K = \inf\left\{k\in\mathbb{N} : \tilde{t}_k \geq T\right\} - 1.$$

This yields a finite partition, since we have assumed $\mu(\cdot)$ to be a continuous semi-martingale. Note that $V^m(\cdot)$ as defined in (4.3) is the wealth process arising from discrete rebalancing at times $\{t_k\}_k$, as in the introduction to this section. Hence, using (4.1) and (4.2), note that a portfolio function $\pi : \Delta^n_+ \to \bar{\Delta}^n_+$ is MCM — see (3.9) and recall $V^m(0) = 1$ — if and only if the following is true: for any path $\{\mu(t)\}_{t\in[0,T]}$ in Δ^n_+ and any $m \in \mathbb{N}$, it holds that

$$V^m(T) \geq 1 \Big/ \sum_{i=1}^{n} \pi_i(T) \frac{\mu_i(0)}{\mu_i(T)} \,. \tag{4.4}$$

In words, this recasting of MCM says the following: the relative performance of any piecewise buy-and-hold implementation of $\pi(\cdot)$ over the period $[0, T]$ is at least one over the return that one would obtain if one could implement the portfolio weights at time T, hold them, and go back in time to the original configuration $\mu(0)$ of market weights. We will give further clarification of this interpretation in Corollary 4.2.1 below.

4.1.1 Definition (Émery topology, Definition 12.4.3 of Cohen and Elliott (2015))**.** Let $Y(\cdot)$ be a semimartingale. The topology of uniform convergence on compacts in probability (or *ucp convergence*) is defined through the norm-like map

$$Y(\cdot) \mapsto ||Y(\cdot)||_{\text{ucp}} := \sum_{n=1}^{\infty} 2^{-n} \mathbb{E}\big[1 \wedge \sup_{t\in[0,n]} |Y(t)|\big].$$

The *Émery norm* is then defined as

$$Y(\cdot) \mapsto ||Y(\cdot)||_{\mathcal{S}} := \sup \left\{ \left|\left| \int_0^{\cdot} H(t)\, \mathrm{d}Y(t) \right|\right|_{\text{ucp}} : H(\cdot) \text{ simple predictable}, |H(\cdot)| \leq 1 \right\}.$$

Convergence in the topology thus defined, i.e. $||Y^m(\cdot) - Y(\cdot)||_{\mathcal{S}} \to 0$ as $m \to \infty$, will be denoted by

$$Y^m(\cdot) \xrightarrow{\mathcal{S}} Y(\cdot).$$

□

We now take the limit in which the partition size goes to zero, and obtain the following result:

4.1.2 Theorem. *Let* $V(\cdot)$ *denote the relative wealth process* (0.6) *of any continuous portfolio* $\pi(\cdot)$*. Then there exists a sequence of stopping times* $\sigma_k \to \infty$ *such that*

$$\sup_{t\in[0,T]} \mathbb{E}\big[\big|V^m(t \wedge \sigma_k) - V(t \wedge \sigma_k)\big|^p\big]^{1/p} \to 0 \tag{4.5}$$

as $m \to \infty$ *for all* k*. We will write this as*

$$V^m(\cdot) \to V(\cdot).$$

Proof. By construction, $V^m(\cdot)$ is a continuous semimartingale. By construction of t_k and continuity of $\pi(\cdot)$ and $\mu(\cdot)$, we have that, for every $t \in (t_j, t_{j+1}]$,

$$\begin{aligned}
\left|\sum_{i=1}^n \frac{\pi_i(t_j)}{\mu_i(t_j)}\big(\mu_i(t) - \mu_i(t_j)\big)\right| &\leq \sum_{i=1}^n \left|\frac{\pi_i(t_j)}{\mu_i(t_j)}\right| |\mu_i(t) - \mu_i(t_j)| \\
&\leq \sum_{i=1}^n \left(\left|\frac{\pi_i(t_j)}{\mu_i(t_j)} - \frac{\pi_i(t)}{\mu_i(t)}\right| + \left|\frac{\pi_i(t)}{\mu_i(t)}\right|\right) |\mu_i(t) - \mu_i(t_j)| \\
&\leq \sum_{i=1}^n \left(\frac{1}{m} + \left|\frac{\pi_i(t)}{\mu_i(t)}\right|\right) \cdot \frac{1}{m} \\
&\overset{\text{a.s.}}{\to} 0, \qquad (4.6)
\end{aligned}$$

as well as

$$\left|\sum_{i=1}^n \frac{\pi_i(t_j)}{\mu_i(t_j)} - \sum_{i=1}^n \frac{\pi_i(t)}{\mu_i(t)}\right| \leq \frac{n}{m} \overset{\text{a.s.}}{\to} 0, \qquad (4.7)$$

as $m \to \infty$.

Now, conditioning on the partition interval $(t_j, t_{j+1}]$ that t is in, we write

$$V^m(t) = \sum_{j=1}^K \mathbf{1}_{(t_j, t_{j+1}]}(t) V^m(t_j) \sum_{i=1}^n \frac{\pi_i(t_j)}{\mu_i(t_j)} \mu_i(t); \qquad (4.8)$$

note that $V^m(t_j)$ is the total return up to time t_j. Hence, we conclude that

$$\begin{aligned}
\int_0^\cdot \frac{\mathrm{d}V^m(t)}{V^m(t)} &= \int_0^\cdot \sum_{j=1}^K \frac{\mathbf{1}_{(t_j, t_{j+1}]}(t) V^m(t_j) \sum_{i=1}^n \frac{\pi_i(t_j)}{\mu_i(t_j)} \,\mathrm{d}\mu_i(t)}{\mathbf{1}_{(t_j, t_{j+1}]}(t) V^m(t_j) \sum_{i=1}^n \frac{\pi_i(t_j)}{\mu_i(t_j)} \,\mu_i(t)} \\
&= \int_0^\cdot \sum_{j=1}^K \frac{\sum_{i=1}^n \frac{\pi_i(t_j)}{\mu_i(t_j)} \mathbf{1}_{(t_j, t_{j+1}]}(t) \,\mathrm{d}\mu_i(t)}{\underbrace{\sum_{i=1}^n \frac{\pi_i(t_j)}{\mu_i(t_j)} \big(\mu_i(t) - \mu_i(t_j)\big) \mathbf{1}_{(t_j, t_{j+1}]}(t)}_{\overset{\text{a.s.}}{\to} 0 \text{ as } m \to \infty, \text{ by (4.6)}} + 1} \\
&\overset{\mathcal{S}}{\to} \int_0^\cdot \sum_{i=1}^n \frac{\pi_i(t)}{\mu_i(t)} \,\mathrm{d}\mu_i(t) = \int_0^\cdot \frac{\mathrm{d}V(t)}{V(t)}
\end{aligned}$$

as $m \to \infty$, where we have used (4.6) and (4.7), and convergence follows by the Stochastic Dominated Convergence Theorem — see Theorem 12.4.10, and in particular Corollary 12.4.11, of Cohen and Elliott (2015). Here, convergence in $\mathcal{S}$ stands for convergence in Émery's semimartingale topology, as defined in Definition 4.1.1 above. The last equality is the defining relation (0.2) for $V(\cdot)$.

By Lemma 16.2.7(ii)–(iv), and using Definition 16.2.3 of Cohen and Elliott (2015) and continuity of $V^m(\cdot)$, we conclude that the above implies that

$$\int_0^{\cdot} \frac{\mathrm{d}V^m}{V^m} \to \int_0^{\cdot} \frac{\mathrm{d}V}{V},$$

as well as

$$\left[\int_0^{\cdot} \frac{\mathrm{d}V^m}{V^m} - \int_0^{\cdot} \frac{\mathrm{d}V}{V}\right]^{1/2} \to 0,$$

where the square brackets denote quadratic variation. By the Kunita-Watanabe inequality — see, for instance, Theorem 11.4.1 of Cohen and Elliott (2015); the theorem still holds true, since the quadratic variation of a semimartingale is simply the quadratic variation of its local martingale part — this implies that

$$\begin{aligned}\left|\left[\int_0^{\cdot} \frac{\mathrm{d}V^m}{V^m}\right] - \left[\int_0^{\cdot} \frac{\mathrm{d}V}{V}\right]\right| &= \left|\left[\int_0^{\cdot} \frac{\mathrm{d}V^m}{V^m} + \int_0^{\cdot} \frac{\mathrm{d}V}{V}, \int_0^{\cdot} \frac{\mathrm{d}V^m}{V^m} - \int_0^{\cdot} \frac{\mathrm{d}V}{V}\right]\right| \\ &\leq \left[\int_0^{\cdot} \frac{\mathrm{d}V^m}{V^m} + \int_0^{\cdot} \frac{\mathrm{d}V}{V}\right]^{1/2} \underbrace{\left[\int_0^{\cdot} \frac{\mathrm{d}V^m}{V^m} - \int_0^{\cdot} \frac{\mathrm{d}V}{V}\right]^{1/2}}_{\to 0}.\end{aligned}$$

Hence, using Itô's formula and the isometry property of quadratic variation, as well as the fact that $V^m(0) = V(0) = 1$, we conclude that

$$\begin{aligned}\left|\log V^m(\cdot) - \log V(\cdot)\right| &= \left|\int_0^{\cdot} \frac{\mathrm{d}V^m}{V^m} - \int_0^{\cdot} \frac{\mathrm{d}[V^m]}{2(V^m)^2} - \int_0^{\cdot} \frac{\mathrm{d}V}{V} + \int_0^{\cdot} \frac{\mathrm{d}[V]}{2(V)^2}\right| \\ &= \left|\int_0^{\cdot} \frac{\mathrm{d}V^m}{V^m} - \int_0^{\cdot} \frac{\mathrm{d}V}{V} - \frac{1}{2}\left(\left[\int_0^{\cdot} \frac{\mathrm{d}V^m}{V^m}\right] - \left[\int_0^{\cdot} \frac{\mathrm{d}V}{V}\right]\right)\right| \\ &\leq \left|\int_0^{\cdot} \frac{\mathrm{d}V^m}{V^m} - \int_0^{\cdot} \frac{\mathrm{d}V}{V}\right| + \frac{1}{2}\left|\left[\int_0^{\cdot} \frac{\mathrm{d}V^m}{V^m}\right] - \left[\int_0^{\cdot} \frac{\mathrm{d}V}{V}\right]\right| \\ &\to 0,\end{aligned}$$

by the above and in the sense of (4.5), which thus follows. □

This confirms what one would intuitively expect: the wealth process corresponding to the discretely-rebalanced implementation of any continuous portfolio tends to the continuous-time implementation of that portfolio as the rebalancing frequency increases. Note that, despite the fact that $V^m(\cdot)$ is a simple process, this result does not follow directly from the construction of the stochastic integral for general semimartingales — see, for instance, Section II.4 of Protter (2005). The reason for this is that $V(\cdot)$ is only defined through the stochastic differential equation (0.6), and thus does not have an analytic form in terms of the corresponding portfolio $\pi(\cdot)$.

4.1.3 Remark. One can attempt repeating the above analysis using *trading strategies*. Namely, let us use the analysis of Section 4.1 to derive the discretization $\eta(\cdot)$ of a given trading strategy $\tilde{\eta}(\cdot)$, which by (0.19) has corresponding portfolio

$$\pi_i(t) = \frac{\tilde{\eta}_i(t)\mu_i(t)}{\tilde{V}(t)}, \quad i = 1, \ldots, n,$$

where $\tilde{V}(t) = \langle \tilde{\eta}(t), \mu(t) \rangle$. Namely, we have that

$$\eta_i(t_k) = \frac{\pi_i(t_k)V^m(t_k)}{\mu_i(t_k)} = \frac{V^m(t_k)}{\tilde{V}(t_k)}\tilde{\eta}_i(t_k), \quad i = 1, \ldots, n, \; k = 1, \ldots, K.$$

Hence, analogously to (4.1) and setting $t_1 = 0$, $t_K = T$, we have that

$$V^m(T) = \prod_{k=1}^{K-1} \frac{V^m(t_{k+1})}{V^m(t_k)} = \prod_{k=1}^{K-1} \frac{\langle \tilde{\eta}(t_k), \mu(t_{k+1}) \rangle}{\langle \tilde{\eta}(t_k), \mu(t_k) \rangle}.$$

Therefore one would define the wealth process corresponding to the discrete-time implementation of the trading strategy $\tilde{\eta}(\cdot)$ as

$$V^m(t) := \prod_{k=1}^{K-1} \frac{\langle \tilde{\eta}(t_k \wedge t), \mu(t_{k+1} \wedge t) \rangle}{\langle \tilde{\eta}(t_k \wedge t), \mu(t_k \wedge t) \rangle},$$

and one could then repeat the above proof of Theorem 4.1.2 to get a similar convergence result as the rebalancing frequency increases. □

4.2 A lower bound on functionally-generated wealth

We now use the continuous-time limit of the relative wealth process that was established in Theorem 5 to derive a lower bound on the terminal wealth resulting from a functionally-generated portfolio.

4.2.1 Corollary. *Let the portfolio $\pi(\cdot)$ be multiplicatively generated. Then*

$$V(T) \geq 1 \Big/ \left\langle \frac{\pi(T)}{\mu(T)}, \mu(0) \right\rangle, \qquad \mathbb{P}\text{-a.s.} \tag{4.9}$$

Equivalently,

$$\langle \eta(T), \mu(0) \rangle \geq 1, \qquad \mathbb{P}\text{-a.s.}, \tag{4.10}$$

with $\eta(\cdot)$ defined through (0.19).

Proof. Functional generation is equivalent to cyclical monotonicity, either multiplicative or additive. In the multiplicative case, an application of Theorem 4.1.2 gives convergence $V^m(\cdot) \to V(\cdot)$, implying that the lower bound (4.4) holds for $V(\cdot)$ as well, hence yielding (4.9). Recalling (0.19), we may rewrite this as (4.10). $\square$

The expression (4.10) is an interesting one; it states that independently of how the market evolves through time (provided $\mu(\cdot)$ is a continuous semimartingale), if one were to implement the terminal functionally-generated vector of shares $\eta(T)$ at initial time 0, this would give a relative wealth of at least one.

More explicitly, we can use the above to derive a path-independent lower bound on the relative wealth of continuous multiplicatively generated portfolios; since these satisfy (0.10), we may rewrite (4.9) as

$$V(T) \geq \frac{1}{1 + \sum_{i=1}^n \mu_i(0) \mathrm{D}_{e_i - \mu(T)} \log \Phi(\mu(T))}, \qquad \mathbb{P}\text{-a.s.} \tag{4.11}$$

Note that the right-hand side only depends on the initial and terminal market configurations $\mu(0)$, $\mu(T)$.

Below, we illustrate how this lower bound can be used to yield an example of outperformance of the market. For this, we use a variation of the portfolio (1.51) (namely, with $\alpha = \beta = 1$) — see also (Vervuurt and Karatzas, 2015, Eq. (76)) — and we answer here the open question of finding its generating function, which opens the door to a theoretical study of its properties.

4.2.2 Example (beta-weighted portfolio)**.** Let us consider the following portfolio function, which we shall call the *beta-weighted portfolio:*

$$\pi_i(\mu) := \frac{\mu_i\big(1 - \mu_i\big)}{\sum_{j=1}^n \mu_j\big(1 - \mu_j\big)}, \qquad \mu \in \Delta_+^n,\ i = 1, \dots, n, \tag{4.12}$$

multiplicatively generated by the function

$$\Phi : \Delta_+^n \to \mathbb{R}_+, \quad \mu \mapsto \sqrt{\sum_{i=1}^n \mu_i(1 - \mu_i)}\,. \tag{4.13}$$

One easily computes

$$\mathrm{D}_{e_i - \mu} \log \Phi(\mu) = \frac{\big(1 - \mu_i\big)}{\sum_{j=1}^n \mu_j\big(1 - \mu_j\big)} - 1, \qquad \mu \in \Delta_+^n,\ i = 1, \dots, n.$$

As an example, assume $\mathrm{Leb} \ll \mathbb{P}$, consider the case $\mu(0) = (1 - n\varepsilon)e_1 + \varepsilon\mathbf{1}$, with $\varepsilon > 0$ small, and assume that

$$\mu(T) \in \{\mu \in \Delta_+^n : \mu_1 > \mu_i,\ i = 2, 3, \dots, n\} =: D, \qquad \mathbb{P}\text{-a.s.};$$

that is, we start from an almost completely concentrated market configuration at $t = 0$, and end up with a market in which the firm indexed 1 is still the largest at the terminal time: $p_T(1) = 1$ (recall (0.13)). Then, since $\mathrm{Leb}(D) > 0$, (4.11) implies

$$V(T,\omega) \geq \frac{\sum_{i=1}^n \mu_i(T,\omega)\big(1-\mu_i(T,\omega)\big)}{\sum_{i=1}^n \mu_i(0)\big(1-\mu_i(T,\omega)\big)} > 1, \qquad \mathbb{P}\text{-a.e. } \omega \in \Omega : \mu(T,\omega) \in D, \quad (4.14)$$

as $\varepsilon \downarrow 0$; hence $V(T) \geq 1$ almost surely on D for $\varepsilon > 0$ sufficiently small. The boundedness away from 1 can be seen by viewing the numerator of this fraction as a weighted average of some weights, and realizing that the denominator selects the smallest of those weights. Hence the portfolio (4.12) outperforms the market if it starts from near-complete concentration in one stock, provided dominance of that stock occurs again at time T. This holds independent of the market model. □

4.2.3 Remark. Theorem 2 of Pal and Wong (2016) establishes in discrete time that any portfolio function (3.1) that *surely* (that is, path-by-path) outperforms the market necessarily satisfies MCM, and is therefore multiplicatively generated. In continuous time, such a result can hold at most for almost-sure outperformance $\mathbb{P}(V(T) > 1) = 1$, since the relative wealth process is only defined up to null sets. What makes an analog of their result hard to prove (and, in our opinion, unlikely to exist), is that $V(\cdot)$ relates to MCM only through a limit of $V^m(\cdot)$ as $m \to \infty$; hence a proof by contradiction is required, since properties of $V(\cdot)$ do not imply properties of $V^m(\cdot)$. The negation of MCM gives the existence of a single sequence of market weights that fails the MCM property (3.9), but for a proof by contradiction to work, we would need the MCM property to fail for *every* sequence in a positive-probability set $A \subset \Delta^n_+$, in order to be able to take the continuous-time limit and arrive at a contradiction $\mathbb{P}(V(T) < 1) > 0$. Hence we conjecture that the result of Pal and Wong (2016) does *not* extend to continuous time. Proving this requires different techniques from those applied here, and is left for future research. □

4.3 A master equation for continuous semimartingales

The lower bound (4.11) allows us to show path-independent outperformance of the market for certain pairs of initial and terminal market configurations, as illustrated above. However, ideally one would like to prove *almost-sure* outperformance *regardless* of these configurations — the set of paths with an endpoint in a given set (as in

Example 4.2.2 above) is typically of low measure. This can be achieved by confining ourselves to functionally-generated portfolios that are C^1. Namely, by imitating the proof of Theorem 4.1.2 for the special case that the generating function is twice continuously differentiable, we recover a generalization of Fernholz's well-known master equation — compare with (0.11).

4.3.1 Lemma (Fernholz's master equation)**.** *Let the portfolio $\pi(\cdot)$ be multiplicatively generated by some $\Phi \in C^2(\Delta^n_+, \mathbb{R}_{\geq 0})$, as defined in Definition 0.1.2. Then*

$$\log V(T) = \log\left(\frac{\Phi(\mu(T))}{\Phi(\mu(0))}\right) + \int_0^T \frac{\mathrm{d}\Gamma(t)}{\Phi(\mu(t))}, \qquad \mathbb{P}\text{-a.s.,} \tag{4.15}$$

for[1]

$$\Gamma(t) := -\frac{1}{2}\sum_{i,j=1}^n \int_0^t \mathrm{D}^2_{ij} \log \Phi(\mu(t))\, \mathrm{d}[\mu_i, \mu_j]\,(t). \tag{4.16}$$

Proof. The portfolio $\pi(\cdot)$ satisfies (0.10) with $\mathbf{G} = \Phi$, and thus we have

$$\sum_{i=1}^n \frac{\pi_i(t_k)}{\mu_i(t_k)}\mu_i(t_{k+1}) = 1 + \sum_{i=1}^n \mathrm{D}_i \log \Phi(\mu(t_k))\big(\mu_i(t_{k+1}) - \mu_i(t_k)\big).$$

As in (4.8), conditioning on the partition interval that t is in, we may write

$$\begin{aligned}\log V^m(t) = &\sum_{k=1}^{j-1} \log\Big(1 + \sum_{i=1}^n \mathrm{D}_i \log \Phi(\mu(t_k))\big(\mu_i(t_{k+1}) - \mu_i(t_k)\big)\Big)\\ &+ \log\Big(1 + \sum_{i=1}^n \mathrm{D}_i \log \Phi(\mu(t_k))\big(\mu_i(t) - \mu_i(t_k)\big)\Big), \qquad t \in (t_j, t_{j+1}],\end{aligned}$$

for all $j = 1, \ldots, K$. Hence, applying Itô's formula, we conclude that

$$\begin{aligned}\mathrm{d}\log V^m(t) &= \sum_{j=1}^K \mathbf{1}_{(t_j,t_{j+1}]}(t)\,\mathrm{d}\log\Big(1 + \sum_{i=1}^n \mathrm{D}_i \log \Phi(\mu(t_k))\big(\mu_i(t) - \mu_i(t_k)\big)\Big)\\ &= \sum_{j=1}^K \mathbf{1}_{(t_j,t_{j+1}]}(t)\Bigg[\frac{\sum_{i=1}^n \mathrm{D}_i \log \Phi(\mu(t_k))\,\mathrm{d}\mu_i(t)}{1 + \sum_{i=1}^n \mathrm{D}_i \log \Phi(\mu(t_k))\big(\mu_i(t) - \mu_i(t_k)\big)}\\ &\qquad - \frac{\sum_{i,j=1}^n \mathrm{D}_i \log \Phi(\mu(t_k))\mathrm{D}_j \log \Phi(\mu(t_k))\,\mathrm{d}[\mu_i,\mu_j]\,(t)}{2\big(1 + \sum_{i=1}^n \mathrm{D}_i \log \Phi(\mu(t_k))\big(\mu_i(t) - \mu_i(t_k)\big)\big)^2}\Bigg]\\ &\xrightarrow{\mathcal{S}} \sum_{i=1}^n \mathrm{D}_i \log \Phi(\mu(t))\,\mathrm{d}\mu_i(t) \qquad\qquad (4.17)\\ &\quad - \frac{1}{2}\sum_{i,j=1}^n \mathrm{D}_i \log \Phi(\mu(t))\mathrm{D}_j \log \Phi(\mu(t))\,\mathrm{d}[\mu_i,\mu_j]\,(t)\end{aligned}$$

[1] We write $[\mu_i, \mu_j]$ for quadratic covariation.

as $m \to \infty$, where again we invoke the Stochastic Dominated Convergence Theorem, and use the fact that $\mu(\cdot)$ and $\mathrm{D}_i \log \Phi(\mu(\cdot))$ are continuous. The fact that the denominators above converge to 1 follows, as in the proof of Theorem 4.1.2 (recall (4.6)), by definition of t_k. Finally, we use the chain rule to derive

$$\mathrm{D}^2_{ij} \log \Phi(\mu) = \frac{\mathrm{D}^2_{ij}\Phi(\mu)}{\Phi(\mu)} - \mathrm{D}_i \log \Phi(\mu) \cdot \mathrm{D}_j \log \Phi(\mu), \qquad \mu \in \Delta^n_+,$$

and apply Itô's rule to $\log \Phi(\mu(\cdot))$ to get

$$\begin{aligned}
\mathrm{d} \log \Phi(\mu(t)) &= \sum_{i=1}^n \mathrm{D}_i \log \Phi(\mu(t))\, \mathrm{d}\mu_i(t) + \frac{1}{2} \sum_{i,j=1}^n \mathrm{D}^2_{ij} \log \Phi(\mu(t))\, \mathrm{d}[\mu_i, \mu_j]\,(t) \\
&= \sum_{i=1}^n \mathrm{D}_i \log \Phi(\mu(t))\, \mathrm{d}\mu_i(t) + \frac{1}{2} \sum_{i,j=1}^n \frac{\mathrm{D}^2_{ij}\Phi(\mu(t))}{\Phi(\mu(t))}\, \mathrm{d}[\mu_i, \mu_j]\,(t) \\
&\qquad - \frac{1}{2} \sum_{i,j=1}^n \mathrm{D}_i \log \Phi(\mu(t)) \mathrm{D}_j \log \Phi(\mu(t))\, \mathrm{d}[\mu_i, \mu_j]\,(t).
\end{aligned}$$

Compare with (4.17); since also $V^m(\cdot) \to V(\cdot)$ by Theorem 4.1.2, the result follows. $\square$

The above proof differs from the standard proofs of Fernholz (1999a, 2002) and Fernholz and Karatzas (2009) in that it utilizes the known and simple wealth dynamics for a discretely-rebalanced portfolio, and then takes the continuous-time limit as in Theorem 4.1.2. The result is more general, since $\mu(\cdot)$ is only assumed to be a continuous semimartingale instead of an Itô process. An even more general version of Lemma 4.3.1 was shown in Proposition 4.7 of Karatzas and Ruf (2016), which does not require the function Φ to be C^2, and allows for defaults. Moreover, Schied et al. (2016) derive a pathwise version of the master formula, using Föllmer calculus. The use of decomposition (4.15) of $V(\cdot)$ is exemplified below.

4.3.2 Example (beta-weighted portfolio, continued)**.** Since $\Gamma(\cdot)$ as defined in (4.16) is non-decreasing, by concavity of the generating function (4.13), equation (4.15) implies the alternative lower bound

$$V(T) \geq \frac{\Phi(\mu(T))}{\Phi(\mu(0))}.$$

In our earlier example where $\mu(0) = (1 - n\varepsilon)e_1 + \varepsilon\mathbf{1}$, with $\varepsilon > 0$ small, this leads to the improved (compared to (4.14)) bound

$$V(T) \geq \sqrt{\frac{\sum_{i=1}^n \mu_i(T)(1 - \mu_i(T))}{\sum_{i=1}^n \mu_i(0)(1 - \mu_i(0))}} \to \infty$$

as $\varepsilon \downarrow 0$, for $\mathbb{P}$-almost every $\mu(T) \in \Delta^n_+$.

More generally, let us assume the existence of $\beta, T > 0$ such that

$$\mathbb{P}(\Gamma(T) > \beta) = 1, \tag{4.18}$$

and let us make the assumption of diversity (D^δ) (see Chapter 1), which we recall as

$$\mathbb{P}\big(\mu_{(1)}(t) < 1 - \delta\,, \quad \forall\, t \in [0,T]\big) \;=\; 1\,. \tag{D^δ}$$

Noting the inequalities (recalling (4.13))

$$\sqrt{n\delta(1-\delta)} = \sqrt{(n-1)(1-\delta)\delta + \delta(1-\delta)} \leq \Phi(\mu(t)) \leq \sqrt{1-1/n}\,,$$

equation (4.15) gives that

$$\log V(T) \geq \frac{1}{2}\log\left(\frac{n\delta(1-\delta)}{1-1/n}\right) + \frac{\beta T}{\sqrt{1-1/n}} > 0, \qquad \mathbb{P}\text{-a.s.},$$

provided we have a sufficiently long time horizon

$$T > \frac{\sqrt{1-1/n}}{2\beta}\log\left(\frac{n-1}{n^2\delta(1-\delta)}\right). \tag{4.19}$$

□

The above shows that the beta-weighted portfolio (4.12) is a *strong* relative arbitrage over sufficiently long time horizons and under the assumptions of diversity and a sufficiently increasing drift process $\Gamma(\cdot)$; that is to say, we have $\mathbb{P}\big(V(T) > 1\big) = 1$ in any model that satisfies (4.18) and (D^δ), provided T is as in (4.19).

4.4 Decomposition implies functional generation

In this section, we show that any stochastic process that satisfies suitable — but general — conditions (see Proposition 4.4.2) and which decomposes in a certain way can be associated to the relative wealth process of a functionally-generated strategy. Namely, we show that a multiplicative decomposition (4.20) implies multiplicative functional generation, and that an additive decomposition (4.21) implies additive functional generation. In the multiplicative case, this can be seen as the reverse implication of Lemma 4.3.1.

Recall that $\mu(\cdot)$ denotes the vector of market weights as defined by (0.4); the total market capitalization is used as a numéraire. In order to show the result of this section, we will need the following definition:

4.4.1 Definition. A *local martingale deflator* is a positive local martingale $Z(\cdot)$ satisfying $Z(0) = 1$, with the property that $Z(\cdot)\mu(\cdot)$ is a local martingale. Denote by $\mathcal{Z}$ the set of all local martingale deflators.

The assumption of the existence of a local martingale deflator is equivalent to No Arbitrage of the First Kind (NA1), by Kardaras (2012), and it is what will allow us to relate a general stochastic process to the relative wealth process of a strategy. Namely, a process $W(\cdot)$ is a local martingale when multiplied by any local martingale deflator if and only if it can be written as a stochastic integral with respect to $\mu(\cdot)$. This follows, for instance, by a special case of the Optional Decomposition Theorem as proved in Theorem 1.4 of Karatzas and Kardaras (2015). This is in the formulation of Stricker and Yan (1998), as opposed to the original form of Kramkov (1996), which uses equivalent local martingale measures.

4.4.2 Proposition. *Assume $\mathcal{Z} \neq \emptyset$. Let $W(\cdot)$ be an adapted, continuous process, locally bounded from below, and assume that $Z(\cdot)W(\cdot)$ is a local martingale for all deflators $Z(\cdot) \in \mathcal{Z}$. If there exist some concave $G : \Delta^n_+ \to \bar{\Delta}^n_+$ and some measurable, adapted process $\Lambda(\cdot)$ of finite variation, such that $W(\cdot)$ satisfies*

$$W(t) = G(\mu(t)) \cdot \Lambda(t)\,, \qquad \mathbb{P}\text{-a.s.}, \tag{4.20}$$

then there exists a multiplicatively functionally-generated strategy $\eta(\cdot)$ such that its relative wealth process $V(\cdot)$ satisfies $V(\cdot) \equiv W(\cdot)$, $\mathbb{P}$-a.s.

If, on the other hand,

$$W(t) = G(\mu(t)) + \Lambda(t)\,, \qquad \mathbb{P}\text{-a.s.}, \tag{4.21}$$

then $W(\cdot) \equiv V(\cdot)$ almost surely for $V(\cdot)$ the relative wealth process corresponding to an additively functionally-generated strategy $\eta(\cdot)$.

Proof. As was mentioned earlier, we apply the version of the Optional Decomposition Theorem as proved in (Karatzas and Kardaras, 2015, Theorem 1.4), limiting ourselves to local *martingales* as opposed to local supermartingales. (This limitation forces the non-decreasing 'optional' process in the optional decomposition to be identically zero.) It follows that there exists some trading strategy $\eta(\cdot)$ such that $W(\cdot) \equiv V(\cdot)$ with $V(\cdot)$ satisfying (0.18); that is, being the relative wealth process corresponding to $\eta(\cdot)$.

Both statements now follow from a simple application of Itô's lemma. Namely, from (4.20) it follows that

$$\frac{\mathrm{d}V(t)}{V(t)} - \frac{\mathrm{d}[V](t)}{2V(t)^2} = \mathrm{d}\log V(t) = \mathrm{d}\log G(\mu(t)) + \mathrm{d}\log \Lambda(t);$$

matching the infinite-variation terms and using the defining equation (0.6), we conclude that we must have

$$\sum_{i=1}^{n} \frac{\pi_i(t)}{\mu_i(t)} \, \mathrm{d}\mu_i(t) = \sum_{i=1}^{n} \mathrm{D}_i \log G(\mu(t)) \, \mathrm{d}\mu_i(t).$$

From this it follows that

$$\frac{\pi_i(t)}{\mu_i(t)} = \mathrm{D}_i \log G(\mu(t)) + \rho(t), \qquad i = 1, \dots, n,$$

for some process $\rho(\cdot)$, which we determine by the normalization requirement

$$1 = \sum_{i=1}^{n} \pi_i(t) = \sum_{i=1}^{n} \mu_i(t) \mathrm{D}_i \log G(\mu(t)) + \rho(t).$$

We conclude that $\pi(\cdot) \equiv \pi(\mu(\cdot))$ for a portfolio function π as in (0.10); that is, π is multiplicatively functionally-generated.

In the additive case (4.21), we use (0.6) with (0.19) to write

$$\mathrm{d}V(t) = \sum_{i=1}^{n} \frac{\pi_i(t) V(t)}{\mu_i(t)} \, \mathrm{d}\mu_i(t) = \sum_{i=1}^{n} \eta_i(t) \, \mathrm{d}\mu_i(t) = \mathrm{d}G(\mu(t)) + \mathrm{d}\Lambda(t).$$

Again, we match the infinite-variation terms on both sides of the equation to deduce that

$$\sum_{i=1}^{n} \eta_i(t) \, \mathrm{d}\mu_i(t) = \sum_{i=1}^{n} \mathrm{D}_i G(\mu(t)) \, \mathrm{d}\mu_i(t),$$

and so we have

$$\eta_i(t) = \mathrm{D}_i G(\mu(t)) + \rho(t), \qquad i = 1, \dots, n$$

for some $\rho(\cdot)$. We compute

$$V(t) = \sum_{i=1}^{n} \eta_i(t) \mu_i(t) = \sum_{i=1}^{n} \mu_i(t) \mathrm{D}_i G(\mu(t)) + \rho(t),$$

from which it follows that

$$\eta_i(t) - V(t) = \mathrm{D}_{e_i - \mu(t)} G(\mu(t)), \qquad i = 1, \dots, n;$$

that is, $\eta(\cdot)$ satisfies (3.24) and is therefore additively functionally-generated. $\square$

Indeed, the explicit form of the finite-variation process $\Lambda(\cdot)$ of (4.20) is given in Lemma 4.3.1 when π is C^1. Proposition 4.3 of Karatzas and Ruf (2016) derives its form in the additive case (4.21), proving an analog 'master equation' for additively generated strategies.

4.5 Incorporating defaults

In the field of Stochastic Portfolio Theory, there have been several attempts at incorporating defaults (and, more generally, a varying number of market constituents); defaults are excluded in the 'classical' Itô model of SPT, since the market weights only take values in the strictly positive simplex Δ_+^n — see Section 0.1.1. Extensions that take bankruptcies into consideration have thus far concentrated on the *dynamics of prices*. Some of the first research in this direction was that of Strong (2014b), who proved fundamental theorems of asset pricing in a generalized market of stochastic size with semimartingale prices. In this paper, an example of such a market was given where splits are imposed when a company reaches a market weight of size $1-\delta$, thus imposing the diversity condition, and a relative arbitrage was constructed in this setting. Strong and Fouque (2011) constructed a model with splits and mergers, which are required to occur simultaneously and in pairs, so that the number of companies in the economy remains a constant, and in such a way that total market capital and portfolio wealth are conserved at each regulation event. They explicitly changed (see their Definition 5) the price dynamics by deterministically mapping the vector of prices back into an admissible set each time it hits the boundary of this set, and showed that in their setting, the diversity assumption does *not* imply the existence of relative arbitrage. Karatzas and Sarantsev (2014) then studied, similarly to Strong and Fouque (2011), market models with splits and mergers. However, they allowed for these to occur at different times, causing the size of the market to fluctuate. It was shown that diversity again does *not* imply the existence of a relative arbitrage in this setting.

We deviate from the aforementioned approaches by considering not the dynamics of *prices* in a model with a varying number of stocks (in our case, due to defaults), but by modifying the *portfolios* and their corresponding *wealth processes* in such a way as to give the desired dynamics in a market with defaults — see Definition 4.5.1 below. Namely, we assume the prices in our model to be strictly positive, but introduce stopping times that indicate the occurence of a default for each firm. We then require that no portfolio invest in a defaulted firm, and apply a penalty equal to the portfolio's holding in a firm at each time of default. We demonstrate that one can still construct relative arbitrages in this modified setting.

An alternative to our approach is that of Karatzas and Ruf (2016), who allow for defaults by letting the market weights vector be a continuous semimartingale taking values in the *nonnegative* simplex $\bar{\Delta}_+^n$, and who also study portfolio construction as opposed to price dynamics. However, in their paper the prices can only go to zero *continuously*, which we find too restrictive; one could argue that this allows an investor to continuously liquidate their position before the actual default occurs (for example, by letting their portfolio behave like $\pi_i(\mu) \sim \mu_i$ as $\mu_i \downarrow 0$), thus avoiding a large loss. In contrast, our approach is equivalent to imposing a discrete *jump* to a zero price at the stopping time that represents a default, thus preventing the investor from saving part of their holdings in the corresponding firm. This represents the reality that relatively large firms can go bankrupt without assuming a very small capitalization preceding such an event (the default of Lehman Brothers in 2008 is an example), and that defaults can be very sudden, nullifying all investments in a firm.

As in Chapters 1 and 2, let us assume the Itô model (0.1) for the capitalizations $X_i(\cdot)$, $i = 1\ldots, n$, and consider investments over some finite time horizon $[0, T]$. Let us enlarge the filtration $\mathbb{F}$ to $\mathbb{G} = \{\mathcal{G}(t)\}_{t\in[0,T]}$, so that $\mathcal{G}(t) \supseteq \mathcal{F}(t)$ for all $t \geq 0$. Now introduce the $\mathbb{G}$-stopping times τ_i, $i = 1, \ldots, n$, each representing the time at which firm i defaults. This particular construction allows for stopping times τ_i that are independent of the stock prices (for instance, they could be exponentially-distributed random variables), but also for stopping times of the form

$$\tau_i = \inf\{t \geq 0 \,|\, X_i(t) < d_i\}, \qquad i = 1, \ldots, n,$$

for some $d_i > 0$, $i = 1, \ldots, n$. We do not make any assumptions about the distributions of τ_i here. We write $\mathcal{N} = \{1, \ldots, n\}$ and $S^c = \mathcal{N} \setminus S$ for all $S \subseteq \mathcal{N}$; using the 'default times' τ_i, we may define the stopping times

$$\tau_S := \bigvee_{i\in S^c} \tau_i, \qquad \forall\, S \subsetneq \mathcal{N}, \qquad \tau_{\mathcal{N}} = 0. \tag{4.22}$$

Note that τ_S is the first time that the set S is equal to the indices of all the non-defaulted (or 'active') stocks, and that $\tau_{\mathcal{N}\setminus\{i\}} = \tau_i$. We make the standing assumption that not all firms will default by time T, i.e.[2]

$$\mathbb{P}(\tau_\emptyset \leq T) = 0.$$

We now define a *stochastic* admissibility set $\mathcal{A}(\cdot)$ of portfolios: writing the (stochastic) set of firms that have not defaulted before time t as

$$\mathcal{I}(t) := \{i \in \mathcal{N} \,|\, \tau_i \geq t\},$$

[2]The event $\{\tau_\emptyset < T\}$ corresponds to all firms defaulting by time T, in which case any portfolio will have value 0, and no meaningful analysis can be done.

we set

$$\mathcal{A}(t) := \{x \in \bar{\Delta}^n_+ \,|\, x_i > 0, \forall\, i \in \mathcal{I}(t),\ \ x_i = 0 \text{ otherwise}\}.$$

We have that $\mathcal{A}(0) = \mathcal{N}$, since $\tau_{\mathcal{N}} = 0$. Note that the process $\mathcal{A}(\cdot)$ takes values among the collection of sets

$$\mathcal{A}_S := \Big\{x_i \in \bar{\Delta}^n_+ \;\Big|\; x_i > 0, \forall\, i \in S,\ x_j = 0, \forall\, j \in S^c\Big\}, \qquad S \subseteq \mathcal{N};$$

that is, for all $(t, \omega) \in [0, T] \times \Omega$, there exists an $S \subseteq \mathcal{N}$ such that $\mathcal{A}(t, \omega) = \mathcal{A}_S$, with S the set of indices of the 'active' stocks. Note the identity

$$\tau_S = \inf\big\{t \in [0, T] \,|\, \mathcal{A}(t) \subseteq \mathcal{A}_S\big\}, \qquad S \subseteq \mathcal{N}.$$

We constrain our study to portfolios which do not invest in defaulted stocks (that is, which are restricted to trade in active stocks), and introduce a new process that measures their wealth when taking into consideration defaults:

4.5.1 Definition. Define a portfolio $\pi(\cdot)$ to be $\mathcal{A}(\cdot)$-*admissible* if it is $\mathbb{G}$-progressively measurable and satisfies $\pi(t) \in \mathcal{A}(t)$ for all $t \geq 0$. For any such portfolio, we introduce the *modified wealth process* $W^\pi(\cdot)$, which incorporates defaults by applying a penalty to the investor's wealth each time a default occurs:

$$W^\pi(t) = V^\pi(t) \cdot \prod_{i=1}^{n} \big(1 - \pi_i(\tau_i \wedge t)\mathbf{1}(\tau_i \leq t)\big). \tag{4.23}$$

□

That is, the investor loses their entire investment in the defaulting stock at each time of default, and continues their investment according to $\pi(\cdot)$ with the remaining wealth. From the above, it follows that we can apply the standard framework to study $V^\pi(\cdot)$, but that we will need to come up with ways to bound the second term in (4.23).

We come up with two methods to achieve this:

- the first constructs a portfolio which is a dynamically-changing linear combination of other portfolios, each of which only invests in a subset of the market constituents — see Section 4.5.1;

- the second studies generating functions that can be extended to $\bar{\Delta}^n_+$, establishing a modified master equation for the modified wealth process (4.23) — see Section 4.5.2.

Note that in the above, we do *not* change the underlying stock price dynamics; this is not necessary, since we do not trade the stocks after they have defaulted. Finally, note that the modified vector of market weights in our model with defaults has components

$$\tilde{\mu}_i(t) := \frac{X_i(t)\mathbf{1}(t \geq \tau_i)}{\sum_{j=1}^n X_j(t)\mathbf{1}(t \geq \tau_j)}, \qquad i \in \mathcal{N}. \tag{4.24}$$

Here, $X_i(\cdot)$, $i = 1, \ldots, n$, are the capitalization processes of (0.1).

4.5.1 Linearly combining scenario-specific portfolios

The construction in this section rests on the following observation:

4.5.2 Observation. On the event that

$$\Big\{ \bigwedge_{i \in A} \tau_i \geq T \Big\},$$

for some $A \subseteq \mathcal{N}$, a portfolio $\pi(\cdot) \in \mathcal{A}_A$ (i.e. one which is 'supported' on A, and therefore does not invest in any defaulting firms) will remain unaffected by any defaults of firms $j \in A^c$, implying that $W^\pi(\cdot) \equiv V^\pi(\cdot)$, by (4.23). □

With the above in mind, let us consider portfolios $\pi^{(k)}(\cdot)$, $k = 2, 3, \ldots, 2^n$, each of which takes values in $\mathcal{A}_{S_k}$; here, $\{S_2, S_3, \ldots, S_{2^n}\} = \mathcal{P}(\mathcal{N}) \setminus \emptyset$ is an indexed collection of all the non-empty subsets of $\mathcal{N}$ — the scenario $k = 1$ corresponds to bankruptcy of all the firms in the market. Write $V^{(k)}(\cdot) := V^{\pi^{(k)}}(\cdot)$ for the (unmodified) wealth process corresponding to the portfolio $\pi^{(k)}(\cdot)$. Let us consider the portfolio $\pi(\cdot)$ of the form

$$\pi(t) = \sum_{k=2}^{2^n} \frac{q_k V^{(k)}(t)\mathbf{1}(t \leq \tau^{(k)})}{\sum_l q_l V^{(l)}(t)\mathbf{1}(t \leq \tau^{(l)})} \, \pi^{(k)}(t), \tag{4.25}$$

where $q_k \geq 0$ are certain weights with $\sum_k q_k > 0$, and we have written

$$\tau^{(k)} := \bigwedge_{i \in S_k} \tau_i, \qquad k = 2, 3, \ldots, 2^n$$

for the first time that a firm that $\pi^{(k)}(\cdot)$ invests in defaults. Note that $\pi(\cdot)$ is $\mathbb{G}$-predictable. Note that

$$\sum_{k : i \in S_k} \bar{q}_k(\tau_i)$$

is a conservative upper bound for the proportion of wealth that the investor can lose due to the default of firm i, $i \in \mathcal{N}$, with

$$\bar{q}_k(t) := \frac{q_k V^{(k)}(t)\mathbf{1}(t \leq \tau^{(k)})}{\sum_l q_l V^{(l)}(t)\mathbf{1}(t \leq \tau^{(l)})}, \qquad k = 2, 3, \ldots, 2^n.$$

Assume now that $\pi^{(k)}(\cdot)$ are functionally generated by some functions $\mathbf{G}^{(k)} : \Delta_+^{|S_k|} \to \bar{\Delta}_+^{|S_k|}$, $k = 2, 3, \ldots, 2^n$. Note that $\pi(\cdot)$ is *not* a functionally-generated portfolio, since it depends on the entire history of prices through $\bar{q}_s(\cdot)$; in a loose sense, its 'generating function' would be the process

$$\prod_{k=2}^{2^n} \left(\mathbf{G}^{(k)}(\mu(t))\right)^{\bar{q}_k(t)}. \tag{4.26}$$

We note that, by an application of (0.2), for all $t \notin \{\tau_i,\ i \in \mathcal{N}\}$ we have

$$\begin{aligned}
\frac{\mathrm{d}V^\pi(t)}{V^\pi(t)} &= \sum_{i=1}^{n} \sum_{k=2}^{2^n} \bar{q}_k(t) \pi_i^{(k)}(t) \frac{\mathrm{d}X_i(t)}{X_i(t)} \\
&= \sum_{k=2}^{2^n} \frac{q_k V^{(k)}(t) \mathbf{1}(t \le \tau^{(k)})}{\sum_l q_l V^{(l)}(t) \mathbf{1}(t \le \tau^{(l)})} \frac{\mathrm{d}V^{(k)}(t)}{V^{(k)}(t)} \\
&= \frac{\mathrm{d}\Big(\sum_k q_k V^{(k)}(t) \mathbf{1}(t \le \tau^{(k)})\Big)}{\sum_l q_l V^{(l)}(t) \mathbf{1}(t \le \tau^{(l)})}.
\end{aligned}$$

By taking into account the 'boundary conditions' at the times $\tau_i,\ i \in \mathcal{N}$ — in other words, enforcing self-financiability — it follows that for all $t \ge 0$ we have

$$V^\pi(t) = \sum_{k=2}^{2^n} \frac{q_k \mathbf{1}(t \le \tau^{(k)})}{\sum_l q_l \mathbf{1}(t \le \tau^{(l)})} V^{(k)}(t) \ge \sum_{k=2}^{2^n} q_k V^{(k)}(t) \mathbf{1}(t \le \tau^{(k)}), \qquad \mathbb{P}\text{-a.s.} \tag{4.27}$$

4.5.3 Remark. The choice of portfolio (4.25) is equivalent to linearly combining the trading strategies $\eta^{(k)}(\cdot)$ corresponding to $\pi^{(k)}(\cdot)$, defined through (0.19). Namely, by (4.25) and (4.27), we have that the strategy $\eta(\cdot)$ defined by $\pi(\cdot)$ satisfies

$$\begin{aligned}
\eta_i(t) &= \frac{\pi_i(t) V^\pi(t)}{X_i(t)} \\
&= \sum_{k=2}^{2^n} \frac{q_k V^{(k)}(t) \mathbf{1}(t \le \tau^{(k)})}{\sum_l q_l V^{(l)}(t) \mathbf{1}(t \le \tau^{(l)})} \frac{\pi_i^{(k)}(t)}{X_i(t)} \frac{\sum_l q_l V^{(l)}(t) \mathbf{1}(t \le \tau^{(l)})}{\sum_m q_m \mathbf{1}(t \le \tau^{(m)})} \\
&= \sum_{k=2}^{2^n} \frac{q_k \mathbf{1}(t \le \tau^{(k)})}{\sum_m q_m \mathbf{1}(t \le \tau^{(m)})} \frac{\pi_i^{(k)}(t) V^{(k)}(t)}{X_i(t)} \\
&= \sum_{k=2}^{2^n} \frac{q_k \mathbf{1}(t \le \tau^{(k)})}{\sum_m q_m \mathbf{1}(t \le \tau^{(m)})} \eta_i^{(k)}(t),
\end{aligned}$$

for each $i \in \mathcal{N}$. One could start from this relation as a definition, and use (0.17) to obtain (4.27), which can be used to show (4.25). We have chosen to study portfolios instead of strategies, because of their explicit form when they are functionally-generated, and the availability of Fernholz's well-known master equation. □

We interpret (4.25) as follows: $\pi(\cdot)$ is a weighted average of portfolios that are supported on distinct subsets of $\mathcal{N}$, each of which correspond to different scenarios of 'surviving firms' at time T. Motivated by (4.27), let us study each process $V^{(k)}(\cdot)$ individually; by Observation 4.5.2, we have that

$$W^{\pi^{(k)}}(t) = V^{(k)}(t), \qquad 0 \le t \le \tau^{(k)}.$$

Thus one can apply Fernholz's master equation to each $\pi^{(k)}(\cdot)$, $k = 2, 3, \dots, 2^n$, over almost every path $\omega \in \big\{\tau^{(k)} \ge T\big\}$. Typically (that is, in all known cases), this will give lower bounds of the form

$$\log\left(\frac{V^{(k)}(\omega, T)}{V^{\mu^{(k)}}(\omega, T)}\right) \ge A_k T - B_k, \qquad \mathbb{P}\text{-a.e. } \omega \in \big\{\tau^{(k)} \ge T\big\}, \tag{4.28}$$

with $A_k, B_k > 0$ some constants, and with k such that $\mathbb{P}\big(\tau^{(k)} \ge T\big) > 0$. Here, we have written $\mu^{(k)}(\cdot)$ for the market portfolio that is restricted to S_k:

$$\mu_i^{(k)}(t) := \frac{X_i(t)\mathbf{1}(i \in S_k)}{\sum_{j \in S_k} X_j(t)}, \qquad i \in \mathcal{N},\ k = 2, 3, \dots, 2^n. \tag{4.29}$$

From (4.28), it follows that for each scenario k for which $q_k > 0$ and $\mathbb{P}\big(\tau^{(k)} \ge T\big) > 0$, we have

$$\log\left(\frac{q_k V^{(k)}(\omega, T)}{V^{\mu^{(k)}}(\omega, T)}\right) \ge A_k T - B_k + \log q_k > 0, \qquad \mathbb{P}\text{-a.e. } \omega \in \big\{\tau^{(k)} \ge T\big\},$$

provided

$$T > \frac{B_k - \log q_k}{A_k}. \tag{4.30}$$

Let us recall the standing assumption that

$$\{\tau_\emptyset \le T\} = \emptyset \quad \mathbb{P}\text{-a.s.} \quad \Rightarrow \quad \bigcup_{k=2}^{2^n} \big\{\mathcal{A}(T) = \mathcal{A}_{S_k}\big\} = \Omega \quad \mathbb{P}\text{-a.s.}$$

Assume that $q_k > 0$ for all k such that $\mathbb{P}\big(\mathcal{A}(T) = \mathcal{A}_{S_k}\big) > 0$, and note that $\big\{\mathcal{A}(T) = \mathcal{A}_{S_k}\big\} \subset \big\{\tau^{(k)} \ge T\big\}$. We use the above to conclude that the probability that the portfolio (4.25) outperforms the market when taking into account defaults is equal to

$$\begin{aligned}
\mathbb{P}\big(W^\pi(T) > W^{\tilde{\mu}}(T)\big) &= \mathbb{P}\Big(\bigcup_{k=2}^{2^n} \big\{W^\pi(T) > W^{\tilde{\mu}}(T)\big\} \cap \big\{\mathcal{A}(T) = \mathcal{A}_{S_k}\big\}\Big) \\
&\ge \mathbb{P}\Big(\bigcup_{k:\, q_k > 0} \big\{q_k V^{(k)}(T) > V^{\mu^{(k)}}(T)\big\} \cap \big\{\mathcal{A}(T) = \mathcal{A}_{S_k}\big\}\Big) \\
&= 1,
\end{aligned}$$

on the condition that

$$T > \max_{k:\, q_k > 0} \frac{B_k - \log q_k}{A_k}, \tag{4.31}$$

by (4.30).

Typically, one would assume $\mathbb{P}(\mathcal{A}(T) = \mathcal{A}_{S_k}) = 0$ for many scenarios k; for instance, for all k such that $|S_k| < n - m$ for some $m < n$, translating the assumption that no more than m firms will default by time T (similarly to the (LF$^\kappa$) condition from Chapter 1). This corresponds to setting those q_k equal to zero in (4.25), which simplifies the results, as is exemplified below.

4.5.4 Example (diversity-weighted portfolio with at most one default)**.** Let us consider the case of the diversity-weighted portfolio with parameter $p \in (0, 1)$. Assume

$$\mathbb{P}(\tau^{(k)} < T) = 0,\ \forall\, k : |S_k| < n - 1, \tag{4.32}$$

meaning that no more than one firm can default over $[0, T]$. Assume diversity (D$^\delta$) with parameter δ holds over all sub-markets $S \subseteq \mathcal{N}$, $|S| \geq n - 1$,[3] and assume non-degeneracy (ND$^\varepsilon$). Consider the diversity-weighted portfolios

$$\begin{aligned} \pi_i^{(0)}(t) &:= \frac{\big(\mu_i(t)\big)^p}{\sum_{j=1}^n \big(\mu_j(t)\big)^p}, && i \in \mathcal{N} \\ \pi_i^{(k)}(t) &:= \frac{\big(\mu_i^{(k)}(t)\big)^p \mathbf{1}(i \neq k)}{\sum_{j \neq k} \big(\mu_j^{(k)}(t)\big)^p}, && i, k \in \mathcal{N}, \end{aligned}$$

and let $S_0 = \mathcal{N}$, $S_k = \mathcal{N} \setminus \{k\}$, $k \in \mathcal{N}$. Then the standard result from Fernholz et al. (2005) (recall the introduction of Chapter 1, in particular Equation (1.2)) gives that these portfolios outperform the market over the horizon $[0, T]$, with T as in (4.30) and

$$\begin{aligned} A_k &= (1 - p)\frac{\varepsilon\delta}{2}, \\ B_k &= \frac{1 - p}{p} \log |S_k|. \end{aligned}$$

Thus, by the above, the portfolio

$$\pi(t) = \sum_{k=0}^n \frac{q_k V^{(k)}(t) \mathbf{1}(t \leq \tau^{(k)})}{\sum_l q_l V^{(l)}(t) \mathbf{1}(t \leq \tau^{(l)})}\, \pi^{(k)}(t)$$

[3] By this, we mean that for any $S \subseteq \mathcal{N}$, $|S| \geq n-1$, it holds that $\max_{i \in S} \mu_i(t) < 1 - \delta$, $\forall\, t \in [0, T]$, $\mathbb{P}$-a.s. A sufficient condition for this is that of 'limited failure' (LF$^\kappa$); recall Remark 1.2.1.

is an arbitrage relative to the market when taking into account defaults, provided T is 'sufficiently large'. Setting $q_0 = q$, and $q_k = \bar{q}$, $k \in \mathcal{N}$, for some $q, \bar{q} > 0$ to be determined, (4.31) implies that we must have

$$T > \max\left\{\frac{2\log(n/q)}{\varepsilon\delta p}, \frac{2\log((n-1)/\bar{q})}{\varepsilon\delta p}\right\}.$$

This lower bound is minimized when the two terms within the maximum are equal; this requires setting

$$n/q = (n-1)/\bar{q}.$$

Using the fact that $q + n\bar{q} = 1$, we conclude that we must choose the weights

$$q = \frac{1}{n}, \qquad \bar{q} = \frac{n-1}{n^2} < q.$$

Then the minimal time horizon required for almost-sure outperformance of the market is

$$T > \frac{4\log n}{\varepsilon\delta p}. \tag{4.33}$$

Comparing this with Eq. (4.5) of Fernholz et al. (2005) (i.e. (1.2)), we see that the introduction of a single default exactly *doubles* the minimal time to outperformance. □

4.5.5 Remark. There are other ways of choosing the weights q_k. One could have an idea about the distributions of τ_i, $i \in \mathcal{N}$,[4] and incorporate that into their choice of weights. For instance, in the setting of Example 4.5.4, suppose the probability of default is very small: $\mathbb{P}(\tau^{(0)} \geq T) \approx 1$. Then

$$\begin{aligned}\mathbb{P}\big(W^{\pi}(T) > W^{\tilde{\mu}}(T)\big) &\geq \mathbb{P}\big(\{qV^{(0)}(T) > V^{\mu}(T)\} \cap \{\tau^{(0)} \geq T\}\big)\\ &= \mathbb{P}\big(\tau^{(0)} \geq T\big)\\ &\approx 1,\end{aligned}$$

provided

$$T > \frac{B_0 - \log q_0}{A_0},$$

which is minimized by choosing $q_0 \approx 1$. Hence, as one would expect intuitively, outperformance of the market with high probability over a shorter time horizon is achieved by putting a lot of weight on the most likely scenario. □

[4] One way in which these distributions could be estimated is through an application of extreme value theory.

4.5.2 Extending generating functions

Recall (4.26); in the previous approach, the portfolio we constructed was the combination of porfolios from different generating functions, each of which corresponded to a different scenario of defaults. We develop here a second approach, in which we use a single generating function, but extend its domain to include defaults. This simplifies the analysis of the wealth process $V(\cdot)$, since it allows us to apply Fernholz's master equation between each two default times, but makes it harder to control for the losses due to defaults, and hence to make bounds for $W(\cdot)$.

Let us first introduce some notation: given a vector $\mu \in \bar{\Delta}^n_+$, define the transformation

$$\mu_i^{S,\varepsilon} := \begin{cases} \dfrac{\mu_i}{\sum_{j\in S}\mu_j} - \dfrac{\varepsilon}{|S|}, & i \in S, \\ \dfrac{\varepsilon}{n-|S|}, & i \notin S, \end{cases} \qquad \mu \in \bar{\Delta}^n_+,\ \emptyset \neq S \subsetneq \mathcal{N},\ \varepsilon \geq 0,$$

so that $\mu^{(k)} = \mu^{S_k,0}$. We will write $\mu^S := \mu^{S,0}$ for elements of $\mathcal{A}_S \subset \bar{\Delta}^n_+ \setminus \Delta^n_+$. Observe that, for all $\mu \in \Delta^n_+$ and $\varepsilon > 0$ sufficiently small, we have that $\mu^{S,\varepsilon} \in \Delta^n_+$. Finally, note that each point in $\bar{\Delta}^n_+ \setminus \Delta^n_+$ is of the form μ^S for some $S \subsetneq \mathcal{N}$.

We shall only consider generating functions $\mathbf{G} : \Delta^n_+ \to \bar{\Delta}^n_+$ that can be uniquely extended to the domain $\bar{\Delta}^n_+$; namely, if the following limit exists, we define

$$\mathbf{G}\big(\mu^S\big) := \lim_{\varepsilon\downarrow 0} \mathbf{G}\big(\mu^{S,\varepsilon}\big), \qquad \mu \in \Delta^n_+,\ \emptyset \neq S \subsetneq \mathcal{N}. \tag{4.34}$$

This lower limit exists for a wide range of generating functions, including the diversity-weighted portfolio; see (4.39) below. Denote by $\pi(\cdot)$ the portfolio generated by this $\mathbf{G}$, and note that in the sense of Strong (2014a), one could alternatively write the (time-changing) generating function as

$$\mathbf{G}(\mu(t), t) = \sum_{k=2}^{2^n} \mathbf{1}(t \leq \tau^{(k)})\, \mathbf{G}^{(k)}(\mu(t)),$$

where $\mathbf{G}^{(k)}(\mu(t)) := \mathbf{G}(\mu^{(k)}(t))$, with $\mu^{(k)}(\cdot)$ as in (4.29). Compare with (4.26); we conclude that this approach is fundamentally different from that in the previous section.

Assume that during a given time period $(\alpha, \beta) \subset [0, T]$, no defaults take place. Then the (constant) set of active stocks is some $S \subseteq \mathcal{N}$, with $|S| = m \leq n$; that is, $\mathcal{A}(t) = \mathcal{A}_S$, $\forall\, t \in (\alpha, \beta)$. Hence during this time period, it is as if we are in a normal market of constant size m. By this reasoning, we see that the 'standard' master equation (0.11) applies, which gives the performance over (α, β) with respect to the 'usual' market portfolio as

$$\log\left(\frac{V^\pi(\beta)}{V^{\tilde{\mu}}(\beta)}\right) - \log\left(\frac{V^\pi(\alpha)}{V^{\tilde{\mu}}(\alpha)}\right) = \log\left(\frac{\mathbf{G}\big(\mu^S(\beta)\big)}{\mathbf{G}\big(\mu^S(\alpha)\big)}\right) + \int_\alpha^\beta \mathfrak{g}^S(t)\,\mathrm{d}t\,, \quad \mathbb{P}\text{-a.s.}, \tag{4.35}$$

with $\tilde{\mu}(\cdot)$ as in (4.24), and where the drift term is equal to

$$\mathfrak{g}^S(t) := \frac{-1}{2\mathbf{G}\big(\mu^S(t)\big)} \sum_{i,j\in S} \mathrm{D}^2_{ij}\mathbf{G}\big(\mu^S(t)\big)\mu_i^S(t)\mu_j^S(t)\tau_{ij}^{\mu^S}(t), \qquad \emptyset \neq S \subsetneq \mathcal{N}. \tag{4.36}$$

We use the above to derive a general master equation for $\pi(\cdot)$. Let us recursively define the k^{th} 'shrinkage time' $\tau_{(k)}$ at which the market size decreases, and the corresponding sets $S_{(k)}$ of indices of active firms, as follows:

$$\begin{aligned}
\tau_{(0)} &:= 0\\
S_{(0)} &:= \mathcal{N}\\
\tau_{(k)} &:= \inf\big\{t \in (\tau_{(k-1)}, T] : \exists\, S \subsetneq S_{(k-1)} \text{ s.t. } \tau_S = t\big\} \wedge T,\\
S_{(k)} &:= \big\{i \in \mathcal{N} : \tau_i > \tau_{(k)}\big\},
\end{aligned}$$

for $k \geq 1$, and with τ_S as in (4.22). Note that

$$\mathcal{A}(t) = S_{(k)}, \qquad \forall\, t \in [\tau_{(k)}, \tau_{(k+1)}),\ k \geq 0,$$

and that for a certain $K < n$, we will have $\mathbb{P}\big(\tau_{(K)} = T\big) = 1$. Then applying (4.35) over each interval $[\tau_{(k)}, \tau_{(k+1)})$, we deduce that the process $V^\pi(\cdot)/V^{\tilde{\mu}}(\cdot)$ satisfies

$$\begin{aligned}
\log\left(\frac{V^\pi(T)}{V^{\tilde{\mu}}(T)}\right) &= \sum_{k=1}^{n-1}\left(\log\left(\frac{V^\pi(\tau_{(k)})}{V^{\tilde{\mu}}(\tau_{(k)})}\right) - \log\left(\frac{V^\pi(\tau_{(k-1)})}{V^{\tilde{\mu}}(\tau_{(k-1)})}\right)\right)\\
&= \sum_{k=1}^{n-1}\left(\log\left(\frac{\mathbf{G}\big(\mu^{S_{(k-1)}}(\tau_{(k)})\big)}{\mathbf{G}\big(\mu^{S_{(k-1)}}(\tau_{(k-1)})\big)}\right) + \int_{\tau_{(k-1)}}^{\tau_{(k)}} \mathfrak{g}^{S_{(k-1)}}(t)\,\mathrm{d}t\right)
\end{aligned} \tag{4.37}$$

with the drift terms $\mathfrak{g}^{S_{(k-1)}}(\cdot)$ as in (4.36). Finally, recalling (4.23) we derive the following adjusted master equation for the modified relative wealth:

$$\begin{aligned}
\log\left(\frac{W^\pi(T)}{W^{\tilde{\mu}}(T)}\right) &= \sum_{k=1}^{n-1}\left(\log\left(\frac{\mathbf{G}\big(\mu^{S_{(k-1)}}(\tau_{(k)})\big)}{\mathbf{G}\big(\mu^{S_{(k-1)}}(\tau_{(k-1)})\big)}\right) + \int_{\tau_{(k-1)}}^{\tau_{(k)}} \mathfrak{g}^{S_{(k-1)}}(t)\,\mathrm{d}t\right)\\
&\quad + \sum_{i\in\mathcal{N}} \mathbf{1}(\tau_i \leq T)\log\left(\frac{1-\pi_i(\tau_i\wedge t)}{1-\mu_i(\tau_i\wedge t)}\right).
\end{aligned} \tag{4.38}$$

We now show how this can be used to construct an arbitrage.

4.5.6 Example (diversity-weighted portfolio with at most one default: revisited)**.** As in Example 4.5.4, assume again non-degeneracy ($\mathrm{ND}^{\varepsilon}$) and diversity ($\mathrm{D}^{\delta}$) for all sub-markets of size at least $n-1$, as well as (4.32). Note that for the diversity-weighted portfolio's generating function, see also (1.5),

$$\mathbf{G} : \Delta_{+}^{n} \to \Delta_{+}^{n}, \quad x \mapsto \left(\sum_{i=1}^{n} x_i^p \right)^{1/p},$$

the limit (4.34) is trivial; $\mathbf{G}$ extends to $\bar{\Delta}_{+}^{n}$, and we have that

$$\mathbf{G}(\mu^S) = \left(\sum_{i \in S} (\mu_i^S)^p \right)^{1/p}, \qquad \forall\, \mu^S \in \mathcal{A}_S. \tag{4.39}$$

Note the inequalities

$$1 \leq \mathbf{G}(\mu^S) \leq |S|^{1/p-1}, \qquad \forall\, \mu^S \in \mathcal{A}_S,$$

as well as

$$\int_{\alpha}^{\beta} \mathfrak{g}^S(t)\,\mathrm{d}t \geq (1-p)\frac{\varepsilon}{2} \int_{\alpha}^{\beta} \big(1 - \pi_{(1)}(t)\big)\,\mathrm{d}t \geq \frac{\varepsilon}{2}(1-p)\delta T, \qquad \forall\, S \subseteq \mathcal{N};$$

the latter follows by (1.10) and (0.9). We have used here the reverse order-statistics notation (0.7). Noting finally that

$$\sum_{i \in \mathcal{N}} \mathbf{1}(\tau_i \leq t) \log\left(\frac{1 - \pi_i(\tau_i \wedge t)}{1 - \mu_i(\tau_i \wedge t)} \right) \geq \sum_{i \in \mathcal{N}} \mathbf{1}(\tau_i \leq t) \log\big(1 - \pi_{(1)}(\tau_i \wedge t)\big) \geq \log \delta,$$

an application of (4.38) gives that

$$\begin{aligned} \log\left(\frac{W^{\pi}(T)}{W^{\tilde{\mu}}(T)} \right) &\geq -\frac{1-p}{p} \log(n^2 - n) + \frac{\varepsilon}{2}(1-p)\delta T + \log \delta \\ &> 0 \end{aligned}$$

for all

$$T > \frac{2(1-p)\log(n^2-n) - \log \delta}{\varepsilon \delta p (1-p)}.$$

Compare with (4.33) and (1.2). □

4.6 Discussion

We have directly applied the results that relate optimal transport to functional generation in Corollary 4.2.1; translating the property of cyclical monotonicity, which holds for finite sequences, to a lower bound on the continuous-time wealth process was made possible by the limiting result of Theorem 4.1.2. Another application of this theorem enabled us to extend the central tool of SPT, namely Fernholz's master equation for functionally-generated portfolios, to more general market models. It would be nice to merge this with the approach of Section 4.5, yielding a single master equation that holds in continuous-semimartingale markets with defaults, similarly to (Karatzas and Ruf, 2016, Proposition 4.7). Finally, we were able to show the 'inverse' to this, proving that any process that decomposes as in the master equation necessarily corresponds to the wealth process of a functionally-generated strategy. However, we were unable to extend Pal and Wong (2016)'s discrete-time result that every portfolio function (3.1) that leads to a relative arbitrage is necessarily multiplicatively functionally-generated; this open question seems to call for the application of different techniques.

In Section 4.5, we have suggested two ways of incorporating defaults in the construction of portfolios: one by linearly combining functionally-generated portfolios, and the other by modifying the method of functional generation to one that only uses the sub-universe of active stocks at each point in time. Our examples show that relative arbitrages can still be constructed in such a modified setting, and it would be interesting to extend this result to include other market 'imperfections', such as mergers and acquisitions, splits, and IPO's; or, more generally, a varying number of stocks in the market. Furthermore, one could consider rank-based portfolios (studied in Section 1.2) for investment in markets with defaults. For instance, if one assumes that no more than $m < n$ firms can go bankrupt, then a large-stock rank-based portfolio like (1.14) will automatically avoid investing in a defaulted stock, deeming constructions of the above type unnecessary. Finally, one could implement a backtest of these modified portfolios, as in Sections 1.5 and 2.2, and compare their empirical performance with their unmodified counterparts. One could then also empirically optimize the weights q_k of (4.25), or even apply the data-driven approach of Chapter 2 to portfolios of the form (4.25) or those generated by extended generating functions (4.34).

Discussion

We have taken different approaches to studying functional generation in this thesis. First, we considered this methodology from an application-focused point of view, constructing new portfolios in Chapter 1 and automating the method using machine learning techniques in Chapter 2 — in both chapters we aimed to improve the performance of functionally-generated portfolios. Section 4.5 also followed an applied approach, making a more realistic study of functionally-generated wealth processes by including the possibility of default. Chapter 3 and the rest of Chapter 4, on the other hand, studied functional generation together with its wealth processes from a theoretical perspective, focusing on the method itself rather than the portfolios that it generates.

In terms of making Stochastic Portfolio Theory more applicable, several open problems remain. Many market imperfections have until now not been considered in the theory, the most crucial ones being transaction costs and market impact. Since these are present in every single trade that an investor performs, and since these can alter the performance of a portfolio significantly, their omission greatly reduces the degree to which the master equation resembles the real-world performance of functionally-generated portfolios. One way in which these could be incorporated, is to create a 'No-Trading' region of a certain width around the intended strategy, and to rebalance each time the boundary of this region is hit. One could attempt computing the optimal width theoretically, using optimal control techniques. However, this problem is typically very high-dimensional because of the large number of stocks in the market. Therefore, one might prefer determining the optimal width of the No-Trading region empirically, for instance using machine learning methods.

Another discrepancy between theory and application, which has not been addressed in the literature so far, is the fact that the times and price levels at which an investor can execute trades (i.e. rebalance their portfolio) in a real-world market are *discrete*. The continuous-time framework of Stochastic Portfolio Theory is only an approximation of this, and the only related research that discretizes time is that by

Pal and Wong (2013, 2016). Therefore, it would be interesting to investigate to what degree the continuous-time results, such as the minimal time required to outperformance of the market, hold in a discrete-time setting, and how they are modified; in a way, one would require a result that is a counterpart to our Theorem 4.1.2. Furthermore, it is crucial for the real-world implementability of SPT portfolios that more empirical studies be executed using data sets pertaining to markets other than the S&P 500 index. For instance, emerging markets have very different characteristics from this index; for one, they often exhibit a very high concentration of wealth in a single stock, as is the case with Samsung in the Korean KOSPI 200 index.

Many opportunities for future work also exist on the theoretical side. First of all, one might object to imposing almost-sure assumptions on the market (such as diversity), on the grounds that this is too restrictive and that one cannot argue from a single observed history that an event has zero probability of happening in the future. If one assumes that these market conditions only hold on a set of probability $\mathfrak{p}$ less than one, then outperformance of the market by the corresponding portfolios on this set will be guaranteed by Fernholz's master equation. However, there will be no lower bound on the market-relative performance outside this set; the relative wealth is unbounded below with probability $1 - \mathfrak{p}$. Thus, this calls for a new approach; one could, for instance, select the (diversity-weighted) portfolio which maximizes log-utility on the complement of the set on which the assumptions hold, or one could change the objective of market-relative outperformance altogether.

Another possibility for extending functional generation follows from the comparison between 'classical' and additive generation in Section 3.3.1. Namely, recalling the multiplicative and additive expressions involving the number of shares $\eta_i(\cdot)$ and the wealth process $V(\cdot)$ of Equations (3.22) and (3.23), respectively, one could consider other ways of equating an expression involving the quantities $\eta_i(\cdot)$ and $V(\cdot)$ to a deterministic function of the current market weights. This could possibly lead to other extensions of functional generation, beyond that of additive functional generation.

On a similar note, all portfolios which have been considered thus far in SPT are long-only. There is no particular reason for this; Fernholz's master equation (0.11) also holds for nonconcave generating functions, and hence allows for portfolios that short-sell. Since practitioners often hedge their positions, and therefore take short positions in many stocks, it would be of interest to construct, for instance, a relative arbitrage using a portfolio that is not long-only.

It should also be noted that many theoretical results relating to outperformance of the market index in SPT models hinge on the non-degeneracy condition (ND$^{\varepsilon}$), which assumes a uniform lower bound on the eigenvalues of the covariance matrix — recall the proof of Theorem 1.1.1, for instance. It is doubtful to what extent this assumption in continuous time can be said to be 'descriptive', as opposed to 'normative', since it cannot be directly observed in real markets. Hence, it is of interest to either find proof of its validity, or to replace the condition with an observable one (or to remove it altogether). The entropy-weighted portfolio of Fernholz and Karatzas (2005), being one of very few explicit relative arbitrages that does not require the assumption of non-degeneracy of the covariance matrix, circumvents this problem altogether.

Finally, portfolios in rank-based models (that is, Itô models whose drift and volatility processes depend on the stocks' ranks by company capitalization — see Section 0.2) have thus far been left unstudied. In particular, no relative arbitrages have been constructed in any of these models; research has been focused on the theoretical properties of these models, and their usability in modeling the observed capital distribution curve. It is of interest to pursue this open question, since it might lead to novel portfolios with desirable properties, and give further insight into the dynamics of these market models. One could ask oneself similar questions for the large-market limit of SPT models, such as those of Shkolnikov (2012, 2013), letting the number of firms $n \to \infty$.

Appendix A

Monte Carlo inference

We give here a very brief introduction to Monte Carlo inference, which is applied in Chapter 2 to sample from the posterior distribution of portfolio function values on a Cartesian grid.

Let us place ourselves in a Bayesian framework, so that we have, as in (2.4), that the posterior distribution of some parameters $\boldsymbol{\theta}$ given data $\mathbf{x}$ is given, by Bayes's Theorem, by

$$p(\boldsymbol{\theta}|\mathbf{x}) \propto p(\mathbf{x}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}).$$

The computational implementation of the above can be problematic, foremost because the normalization of the posterior density requires the computation of the integral

$$\int p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}.$$

This is often analytically intractable, especially if $\boldsymbol{\theta}$ is of high dimensionality. One way around this is by using a Markov chain Monte Carlo (MCMC) method.

A discrete-time Markov chain is a stochastic process $\boldsymbol{\theta}(\cdot)$ that is specified by a transition rule $\mathcal{R}$, which moves the chain from one state $\boldsymbol{\theta}(t)$ to the next, $\boldsymbol{\theta}(t+1)$, without any dependency on earlier states. Hence applying $\mathcal{R}$ a number t times to $\boldsymbol{\theta}(0)$ yields $\boldsymbol{\theta}(t)$, which we write as $\mathcal{R}^t\big(\boldsymbol{\theta}(0)\big) = \boldsymbol{\theta}(t)$. In MCMC, one aims to select a transition rule in such a way that

$$||\mathcal{R}^t\big(\boldsymbol{\theta}\big) - p(\boldsymbol{\theta}|\mathbf{x})|| \to 0 \quad \text{as } t \to \infty,$$

for some suitable norm $||\cdot||$. Then, provided the chain is ergodic and it has $p(\boldsymbol{\theta}|\mathbf{x})$ as a stationary distribution, in the limit, the sequence $\boldsymbol{\theta}(\cdot)$ behaves as a dependent sample from $p(\boldsymbol{\theta}|\mathbf{x})$.

Gibbs sampler

One of the simplest MCMC algorithms is the Gibbs sampler, see for instance Geman and Geman (1984). The underlying idea of this method is to sample from each parameter in the vector $\boldsymbol{\theta}$, conditional on all the other parameters. Namely, writing $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ and denoting by $p(\theta_i|\boldsymbol{\theta}_{(-i)})$ the full conditional density function of $\theta_i \,|\, \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p$, the Gibbs sampler is defined as follows:

A.0.1 Algorithm (Gibbs sampler)**.** Initialize some arbitrary state $\boldsymbol{\theta}(0)$. At each time t, update the state vector $\boldsymbol{\theta}(t)$ one component at a time, by sampling in the following order:

$$\begin{aligned}
\theta_1(t+1) &\sim p\big(\theta_1|\boldsymbol{\theta}_{(-1)}(t)\big)\\
\theta_2(t+1) &\sim p\big(\theta_2|\theta_1(t+1), \theta_3(t), \ldots, \theta_p(t)\big)\\
&\vdots\\
\theta_p(t+1) &\sim p\big(\theta_p|\boldsymbol{\theta}_{(-p)}(t+1)\big).
\end{aligned}$$

Collect a pre-determined number T of samples $\big\{\boldsymbol{\theta}(t)\big\}_{t=1,\ldots,T}$ this way. Discard the first b samples as 'burn-in', allowing the Markov chain to settle into the stationary distribution. □

It is clear by the above that if $\boldsymbol{\theta}(0) \sim p(\boldsymbol{\theta}|\mathbf{x})$, then each step of the Gibbs sampler will leave the distribution of $\boldsymbol{\theta}(t)$ unchanged. It follows that $p(\boldsymbol{\theta}|\mathbf{x})$ is a stationary distribution of the Markov chain.

Finally, a *blocked* Gibbs sampler, which is applied in Chapter 2, groups several variables together and samples from their joint distribution conditional on all other variables, instead of updating each component of $\theta(t)$ individually at every time-step.

Metropolis-Hastings algorithm

Hastings (1970) suggested an alternative algorithm for the one-dimensional case $\boldsymbol{\theta} = \theta$; it implements a form of 'rejection sampling', updating all the components of θ simultaneously.

A.0.2 Algorithm (Metropolis-Hastings algorithm)**.** Initialize some arbitrary state $\theta(0)$. Then at each time t:

1. Simulate a proposal value ϑ from some *proposal density* $q(\vartheta|\theta(t))$.

2. Generate an *acceptance indicator* $A \sim \text{Bernoulli}\big(\alpha(\theta(t), \vartheta)\big)$.

3. If $A = 1$, set $\theta(t+1) = \vartheta$, else $\theta(t+1) = \theta(t)$.

Collect a pre-determined number T of samples $\{\theta(t)\}_{t=1,\ldots,T}$ this way, and discard the first b samples as burn-in. $\square$

Hence, one must specify the proposal density (which can be fairly arbitrary), and the *acceptance probability* α. The latter needs to be set to

$$\alpha(\theta, \vartheta) = \min\left\{1\,,\ \frac{f(\vartheta)}{f(\theta)}\right\}; \tag{A.1}$$

with $f(\theta) \propto p(\theta|\mathbf{x})$.

Bibliography

Francesco Audrino, Robert Fernholz, and Roberto G. Feretti. A forecasting model for stock market diversity. *Annals of Finance*, 3:213–240, 2007.

Adrian D. Banner and Daniel Fernholz. Short-term relative arbitrage in volatility-stabilized markets. *Annals of Finance*, 4:445–454, 2008.

Adrian D. Banner and Raouf Ghomrasni. Local times of ranked continuous semi-martingales. *Stochastic Processes and their Applications*, 118(7):1244–1253, 2008.

Adrian D. Banner, Daniel Fernholz, and Ioannis Karatzas. Atlas models of equity markets. *The Annals of Applied Probability*, 15(4):2296–2330, 2005.

Erhan Bayraktar, Yu-Jui Huang, and Qingshuo Song. Outperforming the market portfolio with a given probability. *The Annals of Applied Probability*, 22(4):1465–1494, 2012a.

Erhan Bayraktar, Constantinos Kardaras, and Hao Xing. Strict local martingale deflators and valuing American call-type options. *Finance and Stochastics*, 16(2): 275–291, 2012b.

Mathias Beiglböck, Alexander MG Cox, and Martin Hüsmann. Optimal transport and Skorokhod embedding. *Inventiones mathematicae*, pages 1–74, 2016.

Huy N Chau and Peter Tankov. Market models with optimal arbitrage. *SIAM Journal on Financial Mathematics*, 6(1):66–85, 2015.

Samuel N. Cohen and Robert J. Elliott. *Stochastic Calculus and Applications*. Springer, 2015.

Thomas M Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.

Alexander M.G. Cox and David G. Hobson. Local martingales, bubbles and option prices. *Finance and Stochastics*, 9:477–492, 2005.

Christa Cuchiero, Walter Schachermayer, and Ting-Kam Leonard Wong. Cover's universal portfolio, stochastic portfolio theory and the numeraire portfolio. *arXiv preprint arXiv:1611.09631*, 2016.

Freddy Delbaen and Walter Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300(3):463–520, 1994.

Freddy Delbaen and Walter Schachermayer. Arbitrage possibilities in Bessel processes and their relations to local martingales. *Probability Theory and Related Fields*, 102 (3):357–366, 1995.

Yan Dolinsky and Halil Mete Soner. Martingale optimal transport and robust hedging in continuous time. *Probability Theory and Related Fields*, 160(1-2):391–427, 2014.

Daniel Fernholz and Ioannis Karatzas. On optimal arbitrage. *The Annals of Applied Probability*, 20(4):1179–1204, 2010.

Daniel Fernholz and Ioannis Karatzas. Optimal arbitrage under model uncertainty. *The Annals of Applied Probability*, 21(6):2191–2225, 2011.

Robert Fernholz. Portfolio generating functions. In M. Avellaneda, editor, *Quantitative Analysis in Financial Markets*. World Scientific, River Edge, NJ, 1999a.

Robert Fernholz. On the diversity of equity markets. *Journal of Mathematical Economics*, 31:393–417, 1999b.

Robert Fernholz. Equity portfolios generated by functions of ranked market weights. *Finance and Stochastics*, 5:469–486, 2001.

Robert Fernholz. *Stochastic Portfolio Theory*. Springer, 2002.

Robert Fernholz. A new decomposition of portfolio return. *arXiv preprint arXiv:1606.05877*, 2016.

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

Robert Fernholz and Ioannis Karatzas. Stochastic Portfolio Theory: A survey. In Alain Bensoussan and Qiang Zhang, editors, *Handbook of Numerical Analysis. Vol. XV. Special volume: mathematical modeling and numerical methods in finance*, volume 15 of *Handbook of Numerical Analysis*. Elsevier/North-Holland, Amsterdam, 2009.

Robert Fernholz, Robert Garvy, and John Hannon. Diversity-weighted indexing. *The Journal of Portfolio Management*, 24(2):74–82, 1998.

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

Robert Fernholz, Tomoyuki Ichiba, and Ioannis Karatzas. A second-order stock market model. *Annals of Finance*, 9:439–454, 2013a.

Robert Fernholz, Tomoyuki Ichiba, and Ioannis Karatzas. Two Brownian particles with rank-based characteristics and skew-elastic collisions. *Stochastic Processes and their Applications*, 123:2999–3026, 2013b.

Robert Fernholz, Tomoyuki Ichiba, Ioannis Karatzas, and Vilmos Prokaj. Planar diffusions with rank-based characteristics and perturbed Tanaka equations. *Annals of Finance*, 156:343–374, 2013c.

Robert Fernholz, Ioannis Karatzas, and Johannes Ruf. Volatility and arbitrage. *arXiv preprint arXiv:1608.06121*, 2016.

Hans Föllmer. The exit measure of a supermartingale. *Probability Theory and Related Fields*, 21(2):154–166, 1972.

Claudio Fontana. Weak and strong no-arbitrage conditions for continuous financial markets. *International Journal of Theoretical and Applied Finance*, 18:1550005, 2015.

Claudio Fontana, Monique Jeanblanc, and Shiqi Song. On arbitrages arising with honest times. *Finance and Stochastics*, 18:515–543, 2014.

Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 24:97–109, 1970.

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

Tomoyuki Ichiba, Ioannis Karatzas, and Mykhaylo Shkolnikov. Strong solutions of stochastic equations with rank-based coefficients. *Probability Theory and Related Fields*, 156:229–248, 2013a.

Tomoyuki Ichiba, Soumik Pal, and Mykhaylo Shkolnikov. Convergence rates for rank-based models with applications to portfolio theory. *Probability Theory and Related Fields*, 156:415–448, 2013b.

Benjamin Jourdain and Julien Reygner. Capital distribution and portfolio performance in the mean-field Atlas model. *Annals of Finance*, 11(2):151–198, 2015.

Ioannis Karatzas and Constantinos Kardaras. The numéraire portfolio in semimartingale financial models. *Finance and Stochastics*, 11:447–493, 2007.

Ioannis Karatzas and Constantinos Kardaras. Optional decomposition for continuous semimartingales under arbitrary filtrations. *Electronic Communications in Probability*, 20, 2015.

Ioannis Karatzas and Johannes Ruf. Trading strategies generated by Lyapunov functions. *arXiv preprint arXiv:1603.08245, to appear in Finance and Stochastics*, 2016.

Ioannis Karatzas and Andrey Sarantsev. Diverse market models of competing Brownian particles with splits and mergers. *arXiv preprint arXiv:1404.0748*, 2014.

Constantinos Kardaras. Market viability via absence of arbitrage of the first kind. *Finance and Stochastics*, 16:651–667, 2012.

Constantinos Kardaras. Valuation and parities for exchange options. *SIAM Journal of Financial Mathematics*, 6:140–157, 2015.

Yves-Laurent Kom Samo and Stephen Roberts. Generalized spectral kernels. *arXiv preprint arXiv:1506.02236*, 2015a.

Yves-Laurent Kom Samo and Stephen Roberts. String Gaussian processes. *arXiv preprint arXiv:1507.06977*, 2015b.

Yves-Laurent Kom Samo and Alexander Vervuurt. Stochastic Portfolio Theory: a machine learning perspective. In *Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence*, 2016.

Dmitrij O. Kramkov. Optional decomposition of supermartingales and hedging contingent claims in incomplete security markets. *Probability Theory and Related Fields*, 105(4):459–479, 1996.

Rémi Lassalle. Causal transference plans and their Monge-Kantorovich problems. *arXiv preprint arXiv:1303.6925*, 2013.

David J. C. MacKay. Introduction to Gaussian processes. In C. M. Bishop, editor, *Neural Networks and Machine Learning*, pages 133–165. Springer, Berlin, 1998.

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

Aleksandar Mijatović and Mikhail Urusov. Deterministic criteria for the absence of arbitrage in diffusion models. *Finance and Stochastics*, 16(2):225–247, 2012.

Gianluca Oderda. Stoxx® index-based risk-controlled portable smart beta strategies, 2016. URL `https://www.stoxx.com/document/Research/insights/Risk-controlled%20Portable%20Smart%20Beta%20Strategies.pdf`.

Jörg R. Osterrieder and Thorsten Rheinländer. Arbitrage opportunities in diverse markets via a non-equivalent measure change. *Annals of Finance*, 2:287–301, 2006.

Soumik Pal. Exponentially concave functions and high dimensional Stochastic Portfolio Theory. *arXiv preprint arXiv:1603.01865*, 2016.

Soumik Pal and Ting-Kam Leonard Wong. Energy, entropy, and arbitrage. *arXiv preprint arXiv:1308.5376*, 2013.

Soumik Pal and Ting-Kam Leonard Wong. The geometry of relative arbitrage. *Mathematics and Financial Economics*, 10(3):263–293, 2016.

Radka Picková. Generalized volatility-stabilized processes. *Annals of Finance*, 10(1):101–125, 2014.

Eckhard Platen and David Heath. *A Benchmark Approach to Quantitative Finance*. Springer, 2006.

Philip Protter. Strict local martingales with jumps. *Stochastic Processes and their Applications*, 125(4):1352–1367, 2015.

Philip E. Protter. *Stochastic Integration and Differential Equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, second edition, 2005.

Joaquin Quiñonero Candela and Carl E. Rasmussen. A unifying view of sparse approximate Gaussian process regression. *The Journal of Machine Learning Research*, 6:1939–1959, 2005.

Ralph T. Rockafellar. *Convex analysis*. Princeton University Press, 2015.

L. C. G. Rogers. *Optimal Investment*. Springer-Verlag, New York, 2013.

Johannes Ruf. *Optimal Trading Strategies Under Arbitrage*. PhD thesis, Columbia University, 2011.

Johannes Ruf. Hedging under arbitrage. *Mathematical Finance. An International Journal of Mathematics, Statistics and Financial Economics*, 23(2):297–317, 2013.

Johannes Ruf and Wolfgang Runggaldier. A systematic approach to constructing market models with arbitrage. Arbitrage, Credit and Informational Risks. Proceedings of the Sino-French Research Program in Financial Mathematics Conference, Beijing June 2013.

Andrey Sarantsev. On a class of diverse market models. *Annals of Finance*, 10(2): 291–314, 2014.

Alexander Schied, Leo Speiser, and Iryna Voloshchenko. Model-free portfolio theory and its functional master formula. *arXiv preprint arXiv:1606.03325*, 2016.

Mykhaylo Shkolnikov. Large systems of diffusions interacting through their ranks. *Stochastic Processes and their Applications*, 122(4):1730–1747, 2012.

Mykhaylo Shkolnikov. Large volatility-stabilized markets. *Stochastic Processes and their Applications*, 123:212–228, 2013.

Christophe Stricker and Jia-An Yan. Some remarks on the optional decomposition theorem. In *Séminaire de Probabilités XXXII*, pages 56–66. Springer, 1998.

Winslow Strong. Generalizations of functionally generated portfolios with applications to statistical arbitrage. *SIAM Journal on Financial Mathematics*, 5(1):472–492, 2014a.

Winslow Strong. Fundamental theorems of asset pricing for piecewise semimartingales of stochastic dimension. *Finance and Stochastics*, 18(3):487–514, 2014b.

Winslow Strong and Jean-Pierre Fouque. Diversity and arbitrage in a regulatory breakup model. *Annals of Finance*, 7:349–374, 2011.

Alexander Vervuurt. Topics in Stochastic Portfolio Theory. *arXiv preprint arXiv:1504.02988*, 2015.

Alexander Vervuurt and Ioannis Karatzas. Diversity-weighted portfolios with negative parameter. *Annals of Finance*, 11(3-4):411–432, 2015.

Cédric Villani. *Optimal Transport: Old and New*, volume 338. Springer-Verlag, 2009.

Ting-Kam Leonard Wong. Optimization of relative arbitrage. *Annals of Finance*, 11 (3-4):345–382, 2015a.

Ting-Kam Leonard Wong. Universal portfolios in Stochastic Portfolio Theory. *arXiv preprint arXiv:1510.02808*, 2015b.