Information geometry in portfolio theory

Ting-Kam Leonard Wong

1 Introduction

In the first chapter of their influential monograph [1, p.1] Amari and Nagaoka explained the key idea of information geometry:

> Information geometry ... allow[s] us to take problems from a variety of fields: statistics, information theory, and control theory; visualize them geometrically; and from this develop novel tools with which to extend and advance these fields.

In this paper we show that this principle can be fruitfully applied to financial problems. We review some recent development in the field of stochastic portfolio theory (SPT) motivated by information geometry, present illustrative examples and some extensions (announced in [54] which is an early version of this paper), and suggest several directions for further study. It is hoped that this paper will be of interest to researchers in both information geometry and mathematical finance. The topics discussed are heavily influenced by the author's research interests. For clarity we only focus on the main ideas and refer the reader to the references for further details. Other financial applications of information geometry are briefly reviewed in Section 1.2.

1.1 Main ideas: market diversity and volatility

To set the stage, let us consider a universe of stocks represented by a capitalization-weighted index. A typical example is the S&P 500 Index which represents a significant portion of the US stock market. In a capitalization-weighted index, the influence of a stock is proportional to its market capitalization. Following the standard

Ting-Kam Leonard Wong
University of Southern California, e-mail: tingkamw@usc.edu

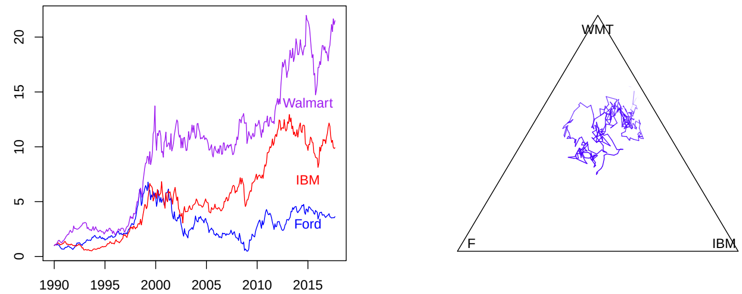

Fig. 1: Left: Prices of the stocks (regarded as the capitalization and normalized to be 1 in January 1990). Right: Path of the corresponding market weights in the simplex Δ_3.

set up of stochastic portfolio theory (see for example [20, 21]), if we let $X_i(t) > 0$ be the market capitalization of stock i at time t, then

$$\mu_i(t) := \frac{X_i(t)}{X_1(t) + \cdots + X_n(t)} \tag{1}$$

is the *market weight* of the stock. Let n be the number of stocks in the market. Then the vector $\mu(t) = (\mu_1(t), \ldots, \mu_n(t))$ takes values in the open unit simplex Δ_n. Throughout this paper we let

$$\Delta_n := \{p = (p_1, \ldots, p_n) \in (0,1)^n : p_1 + \cdots + p_n = 1\} \text{ and}$$
$$\overline{\Delta}_n := \{p = (p_1, \ldots, p_n) \in [0,1]^n : p_1 + \cdots + p_n = 1\}$$

denote respectively the open and closed unit simplices in $\mathbb{R}^n$. The vector $\mu(t)$ may also be regarded as the portfolio weights of the *market portfolio*, and a natural objective is to construct investment strategies that beat the market under suitable conditions. In SPT these portfolios are known as *relative arbitrages*, and a major problem is to construct such strategies under realistic conditions (for precise statements and examples see [20, 21, 22] and their references). As a simple example, Figure 1 plots the path of $\{\mu(t)\}$ for a hypothetical 3-stock market consisting of the US stocks Ford, IBM and Walmart. It should not be surprising that the relative performance of portfolios with respect to the market can be analyzed using appropriate geometries on the simplex[1] which is a fundamental state space in information geometry.

Thinking of the market as a process in the simplex Δ_n, there are two quantities one wants to keep track of: *diversity* and *volatility*. Diversity refers to the degree

[1] In practice the number of stocks changes with time, and the market capitalization may fluctuate due to public offerings and other events. For simplicity these complications are neglected in this paper.

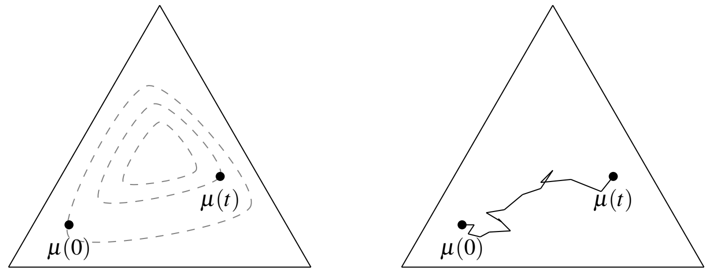

Fig. 2: Left: Change in market diversity measured by φ. The dashed curves represent level sets of φ. Right: Cumulative market volatility measured by the cumulative L-divergence $\mathbf{D}[\cdot \mid \cdot]$.

of capital concentration in the equity market. For example, in Figure 1, the market concentrates towards the vertex representing Walmart. According to [24], many mutual funds tend to overweight small stocks and underweight large stocks (relative to the market), so the change in market diversity is a significant predictor of the performance relative to the market portfolio. To quantify diversity one introduces a positive *concave* function $\Phi : \Delta_n \to (0,\infty)$, and we say that the market is more diverse when $\Phi(\mu(t))$ is large. Typical examples include the Shannon entropy

$$\Phi(p) = -\sum_{j=1}^{n} p_j \log p_j$$

as well as the λ-diversity function

$$\Phi(p) = \left(\sum_{j=1}^{n} (p_j)^{\lambda} \right)^{1/\lambda}, \tag{2}$$

where $0 < \lambda < 1$ is a parameter. More examples can be found in [20, Chapter 3]. Note that we allow Φ to be asymmetric, so it can attain its maximum value at a point other than the barycenter $\overline{e} := \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$. As it turns out, it is more natural to consider its logarithm $\varphi := \log \Phi$. Since $e^{\varphi} = \Phi$ is concave, we say that φ is *exponentially concave*. We remark that φ, being the logarithm of a concave function, is itself a concave function. Given φ, the time series of $\{\varphi(\mu(t))\}$ is an indicator of market diversity (see Figure 2 (left)).

The volatility of the market weight $\mu(t)$ refers to the volatility of the stocks *relative to each other*. Anticipating the use of information geometry, let us note that the Euclidean quadratic variation (in discrete time here)

$$\sum_{s=0}^{t-1} |\mu(s+1) - \mu(s)|^2 \tag{3}$$

may not be appropriate because the Euclidean norm on the simplex may not have a financial meaning (see however Example 4). In particular, the same displacement

$v = \mu(s+1) - \mu(s)$ (approximated by a tangent vector) should have different sizes on different portions of the simplex. Depending on the application, market volatility should be quantified by a sum like

$$\sum_{s=0}^{t-1} \mathbf{D}\left[\mu(s+1) \mid \mu(s)\right],$$

where $\mathbf{D}[\cdot \mid \cdot] : \Delta_n \times \Delta_n \to [0,\infty)$ is possibly asymmetric in its arguments. In information geometry we say that $\mathbf{D}[\cdot \mid \cdot]$ a *divergence*. Intuitively, the asymmetry of $\mathbf{D}[\cdot \mid \cdot]$ reflects the effect of time: the time-reversed path $\tilde{\mu}(t) = \mu(T-t)$ should have different impacts on the portfolio.

The main idea of this paper is the following. A differentiable, exponentially concave function $\varphi : \Delta_n \to \mathbb{R}$ defines an *L-divergence* (L stands for logarithmic)

$$\mathbf{D}^{(1)}\left[q \mid p\right] := \log\left(1 + \nabla\varphi(p) \cdot (q-p)\right) - \left(\varphi(q) - \varphi(p)\right), \quad p, q \in \Delta_n, \tag{4}$$

which can be used to quantify market volatility. (The superscript will become clear in Definition 1.) An important example of L-divergence is the *excess growth rate* (also known as the *diversification return* [6]) defined for a fixed portfolio vector $\pi \in \overline{\Delta_n}$ by

$$\mathbf{T}_\pi\left[q \mid p\right] := \log\left(\sum_{i=1}^n \pi_i \frac{q_i}{p_i}\right) - \sum_{i=1}^n \pi_i \log\frac{q_i}{p_i}. \tag{5}$$

The corresponding exponentially concave function is $\varphi(p) = \sum_{i=1}^n \pi_i \log p_i$. Note that the L-divergence is different from the classical Bregman divergence defined by

$$\mathbf{D}^{(0)}\left[q \mid p\right] := \nabla\varphi(p) \cdot (q-p) - \left(\varphi(q) - \varphi(p)\right). \tag{6}$$

(Note that φ is concave rather than convex.) The most important example of Bregman divergence is the relative entropy

$$\mathbf{H}\left(q \mid p\right) := \sum_{i=1}^n q_i \log\frac{q_i}{p_i}, \tag{7}$$

where the potential function φ is the Shannon entropy. Comparing (5) and (7), we see that the excess growth rate involves a nonlinear transformation of an integral, whereas the relative entropy is itself an integral. In Example 3 we will see that the Rényi entropy generates the Rényi divergence in the sense of L-divergence and is related to the diversity function (2).

The L-divergence determines uniquely an investment strategy, called a *multiplicatively generated portfolio*, whose performance $V(t)$ relative to the market has the pathwise decomposition

$$\log V(t) - \log V(0) = \varphi(\mu(t)) - \varphi(\mu(0)) + \sum_{s=0}^{t-1} \mathbf{D}^{(1)}\left[\mu(s+1) \mid \mu(s)\right]. \tag{8}$$

From this decomposition, we see that as long as $\varphi(\mu(t))$ remains bounded and the cumulative volatility grows at a steady rate, the portfolio will outperform the market in the long run. Furthermore, the dualistic geometry induced by the L-divergence (in the sense of [15, 16]; also see [2]) has interesting financial applications. In a continuous time framework and without using geometric concepts, these portfolios were first introduced by Fernholz [23, 20]. Here we adopt the discrete time, geometric approach established in [43, 44]. Following [54, 53], we will also generalize the portfolio construction using the $L^{(\alpha)}$-divergence:

Definition 1 (**$L^{(\alpha)}$-divergence).** Let φ be α-exponentially concave, i.e, $e^{\alpha\varphi}$ is concave. We define the $L^{(\alpha)}$-divergence of φ is defined for $p, q \in \Delta_n$ by

$$\mathbf{D}^{(\alpha)}\left[q \mid p\right] := \frac{1}{\alpha}\log\left(1 + \alpha\nabla\varphi(p)\cdot(q-p)\right) - \left(\varphi(q) - \varphi(p)\right), \tag{9}$$

This framework covers also the *additively generated portfolio* introduced recently in [33]. Note that the L-divergence is the $L^{(1)}$-divergence, and the Bregman divergence is equal to $\mathbf{D}^{(\alpha)}$ as $\alpha \downarrow 0$, so we will also call it the $L^{(0)}$-divergence. Note that using obvious notations, we have the identity

$$\mathbf{D}_{\varphi}^{(\alpha)}\left[\cdot \mid \cdot\right] \equiv \frac{1}{\alpha}\mathbf{D}_{\alpha\varphi}^{(1)}\left[\cdot \mid \cdot\right]. \tag{10}$$

According to the results obtained recently in [53], it appears that the $L^{(\alpha)}$-divergence is the canonical interpolation between the Bregman divergence and the L-divergence, and plays a fundamental role in information geometry.

1.2 Financial applications of information geometry

In the literature one can find numerous financial applications of information geometry. Instead of attempting an exhaustive literature review, we contend with giving some examples and we apologize for interesting works that are not mentioned here. There are mainly two (overlapping) directions: (i) geometries on the state space of financial dynamics; (ii) optimization using information-geometric quantities such as entropy and divergence.

Regarding the first direction, the paper [9] identifies a yield curve with a distribution function and studies its corresponding dynamics using the Fisher information metric. In option pricing, [48] applies Tsallis's deformed exponentials and generalized the Black-Scholes model to fat-tailed distributions. Though not directly related to finance, the paper [41] generalizes the concept of multiplicatively generated portfolio map (reviewed in this paper) to a large class of optimal transport problems. The recent work [35] applies the Fisher metric in the study of systemic risk.

On the other hand, divergences are frequently useful as objective/cost functionals. In [37, 38], the authors generalize Markowitz's mean-variance model to a mean-divergence model, and show that the resulting portfolios have superior performance.

Optimization of probability functionals under a divergence constraint is studied [8] and is applied to model risk.

1.3 Outline of the paper

In Section 2 we present the discrete time market model and introduce various ways of representing a trading strategy and the associated value process. Section 3 reviews known results about multiplicatively generated portfolios with an emphasis on ideas and clarity. Motivated by these results and the recently introduced additively generated portfolio, in Section 4 we introduce a general framework of functional portfolio generation where the $L^{(\alpha)}$-divergence arises naturally. Further properties of the L-divergence are discussed in Section 5 and several related problems are suggested.

2 The market model

We work in a discrete time, pathwise framework that is used in our previous papers [42, 51, 43] to which the reader is referred for further details. Let $n \geq 2$, the number of stocks in the market, be fixed. The data of our model is a sequence $\{\mu(t) = (\mu_1(t), \ldots, \mu_n(t))\}_{t=0}^{\infty}$ with values in the open unit simplex Δ_n. We regard $\mu(t)$ as the vector of market weights at time t. At this point we do not impose any condition on the sequence $\{\mu(t)\}_{t=0}^{\infty}$, and in Proposition 1 we will impose path properties that lead to relative arbitrages. Extension to continuous time is discussed briefly in Section 4.5.

In this market we consider various self-financing trading strategies. Let us express a strategy in terms of the number of shares held at each point in time. Furthermore, we use the market portfolio as the numéraire (i.e., unit of price). This means that the (relative) value of stock i is simply the market weight $\mu_i(t)$. We assume that trading is frictionless.

Definition 2 (Trading strategy). A self-financing trading strategy is a sequence $\eta = \{\eta(t)\}_{t=0}^{\infty}$, with values in $\mathbb{R}^n$, such that the self-financing identity

$$\sum_{i=1}^{n} \eta_i(t)\mu_i(t+1) \equiv \sum_{i=1}^{n} \eta_i(t+1)\mu_i(t+1) \tag{11}$$

holds for all time t. We always assume η is adapted in the sense that for each $t \geq 0$, $\eta(t)$ is a deterministic function of $\{\mu(s)\}_{0 \leq s \leq t}$. The (relative) value process of η is defined by

$$V_\eta(t) = V_\eta(0) + \sum_{s=0}^{t-1} \eta(s) \cdot (\mu(s+1) - \mu(s)), \tag{12}$$

where $V_\eta(0) = \eta(0) \cdot \mu(0)$ and $a \cdot b$ is the Euclidean inner product.

In the above definition, the portfolio's initial value is determined implicitly by $V_\eta(0) = \eta(0) \cdot \mu(0)$. In this paper we only study the value of a trading strategy relative to the market portfolio, so for simplicity we may omit the word 'relative'. We may interpret

$$V_\eta(t) = \frac{\text{nominal value of portfolio at time } t}{(X_1(t) + \cdots + X_n(t))/(X_1(0) + \cdots + X_n(0))},$$

where the X_i's are the market capitalizations of the stocks. Note that because we allow both long and short positions in the portfolio, the value $V_\eta(t)$ may take negative values. The self-financing identity (11) means that all changes in the portfolio value are due to price changes (but not addition or withdrawal of capital).

If the portfolio value $V_\eta(t)$ is strictly positive for all t, we may define the *portfolio weight* at time t by

$$\pi(t) = (\pi_1(t), \ldots, \pi_n(t)) = \left(\frac{\eta_1(t)\mu_1(t)}{V_\eta(t)}, \ldots, \frac{\eta_n(t)\mu_n(t)}{V_\eta(t)} \right). \tag{13}$$

The components of $\pi(t)$ represent the percentages of current capital invested in each of the stocks; clearly $\sum_{i=1}^n \pi_i(t) \equiv 1$. We call $\pi(t)$ the portfolio weight vector at time t. In this case, the value $V_\eta(t)$ can be expressed *multiplicatively* in the form

$$V_\eta(t) = V_\eta(0) \prod_{s=0}^{t-1} \left(\pi(s) \cdot \frac{\mu(s+1)}{\mu(s)} \right), \tag{14}$$

where $\frac{\mu(s+1)}{\mu(s)}$ is the vector of componentwise ratios. Compare this with the *additive* representation (12). If $\pi_i(t) \geq 0$ for all i and t, we say that the portfolio is *all-long*.

3 Multiplicatively generated portfolio

In this section we review the definition and main results of multiplicative functional generation, following the approach of [43]. For simplicity of exposition we will assume that the generating functions are smooth.

3.1 Pathwise decomposition and relative arbitrage

Definition 3 (Multiplicatively generated portfolio). Let $\varphi : \Delta_n \to \mathbb{R}$ be smooth and exponentially concave. Given the generating function φ, we define a mapping $\pi : \Delta_n \to \overline{\Delta}_n$, called the portfolio map, by

$$\pi_i(p) = p_i \left(1 + D_{e_i - p}\varphi(p)\right), \quad i = 1, \ldots, n, \tag{15}$$

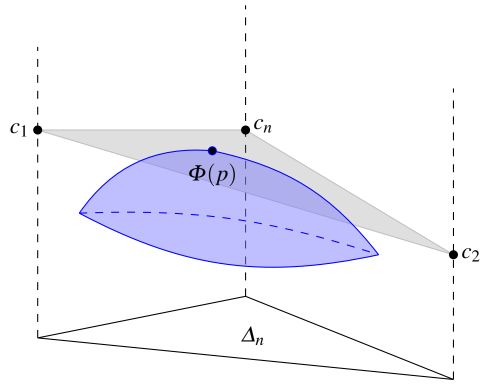

Fig. 3: Geometric interpretation of multiplicatively generated portfolio.

where $(e_1, \ldots, e_n)$ is the standard Euclidean basis and $D_{e_i - p}$ is the directional derivative along the tangent vector $e_i - p$. It defines a self-financing trading strategy η such that the portfolio weight at time t is

$$\pi(t) = \left(\frac{\eta_1(t)\mu_1(t)}{V_\eta(t)}, \ldots, \frac{\eta_n(t)\mu_n(t)}{V_\eta(t)} \right) = \pi(\mu(t)). \tag{16}$$

We say that η (and π) are generated multiplicatively by φ.

Here is a geometric interpretation of the formula (15). Consider the graph of the positive concave function $\Phi = e^\varphi$. Given $p \in \Delta_n$, let the tangent hyperplane to Φ at p be given by $q \mapsto \sum_{i=1}^n c_i q_i$ (see Figure 3). Then, it can be verified that the portfolio vector $\pi(p)$ is given by

$$\pi_i(p) = \frac{c_i p_i}{c_1 p_1 + \cdots + c_n p_n}, \quad i = 1, \ldots, n. \tag{17}$$

In particular, $\pi(p)$ is an element of the closed simplex $\overline{\Delta}_n$ (so the portfolio is all-long), and the weight ratio $\pi_i(p)/p_i$ is proportional to c_i. We say that the trading strategy is generated *multiplicative* by φ because we are specifying the weight ratios in terms of the derivatives of φ.

The following is the main result about multiplicatively generated portfolios.

Theorem 1 (Multiplicative decomposition). *[23, 43] Let η be the trading strategy generated multiplicatively by the exponentially concave function φ as in Definition 3. Then the value process of η is given by*

$$\log V_\eta(t) - \log V_\eta(0) = \varphi(\mu(t)) - \varphi(\mu(0)) + \sum_{s=0}^{t-1} \mathbf{D}^{(1)}\left[\mu(s+1) \mid \mu(s)\right], \tag{18}$$

where $\mathbf{D}^{(1)}[\cdot \mid \cdot]$ is the $L^{(1)}$-divergence of φ defined by (4).

Proof. As this result is fundamental let us give a complete proof (following [43]) which also motivates our later development. We also note that this proof is more transparent than the original proof (see [20, Theorem 3.1.5]). Consider the relative value $V_\eta(t)$ of the strategy. Using the multiplicative representation (14), we have

$$\frac{V_\eta(s+1)}{V_\eta(s)} = \sum_{i=1}^n \pi_i(\mu(s)) \frac{\mu_i(s+1)}{\mu_i(s)}.$$

From (15), we have

$$\frac{\pi_i(\mu(s))}{\mu_i(s)} = 1 + D_{e_i - \mu(s)} \varphi(\mu(s)),$$

so we get the useful identity

$$\begin{aligned} \frac{V_\eta(s+1)}{V_\eta(s)} &= 1 + \sum_{i=1}^n \mu_i(s+1) D_{e_i - \mu(s)} \varphi(\mu(s)) \\ &= 1 + D_{\mu(s+1)-\mu(s)} \varphi(\mu(s)) \\ &= 1 + \nabla \varphi(\mu(s)) \cdot (\mu(s+1) - \mu(s)). \end{aligned} \tag{19}$$

Here we think of the gradient $\nabla \varphi(\mu(s))$ as operating on tangent vectors of Δ_n. Financially, (19) says that the relative return $(V_\eta(s+1) - V_\eta(s))/V_\eta(s)$ of the portfolio is nothing but the directional derivative of φ.

By the concavity of $\Phi = e^\varphi$, for any $p, q \in \Delta_n$ we have

$$\Phi(p) + \nabla \Phi(p) \cdot (q - p) \geq \Phi(q).$$

Rewriting the inequality in terms of φ and taking logarithm on both sides, we have

$$\mathbf{D}^{(1)}[q \mid p] = \log(1 + \nabla \varphi(p) \cdot (q - p)) - (\varphi(q) - \varphi(p)) \geq 0.$$

Taking logarithm on both sides of (19), we have

$$\log V_\eta(s+1) - \log V_\eta(s) = \varphi(\mu(s+1)) - \varphi(\mu(s)) + \mathbf{D}^{(1)}[\mu(s+1)|\mu(s)].$$

Finally, summing over time gives the desired pathwise decomposition (18).

Here is the financial intuition behind the decomposition (18). From (18), the performance of the portfolio relative to the market can be attributed to two quantities. The first is the change in market diversity $\varphi(\mu)$. It depends only on the beginning location $\mu(0)$ and the current location $\mu(t)$ of the market. Note that change in $\varphi(\mu(t))$ is only caused by the component of market movement along the direction of $\nabla \varphi(\mu(t))$ which is perpendicular to the level set of φ. In particular, displacement along the same level set is not visible in this first term. The second term in (18) measures the volatility of the market, as it travels from $\mu(0)$ to $\mu(t)$, measured by the sum of $\mathbf{D}^{(1)}[\mu(s+1) \mid \mu(s)]$ over time. Intuitively, the functionally generated trading strategy η outperforms the market if and only if the volatility is greater than the change in market diversity. In SPT, this decomposition allows us to formulate con-

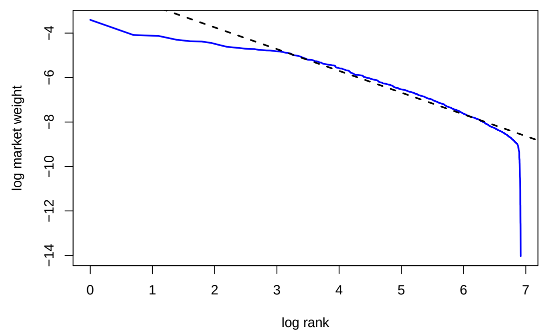

Fig. 4: Capital distribution of the Russel 1000 Index in June 2015 (taken from [40]).

ditions under which relative arbitrage (with respect to the market portfolio) exists. Here is a simple version of this idea:

Proposition 1 (Relative arbitrage). *Fix a smooth, exponentially concave function $\varphi : \Delta_n \to \mathbb{R}$ and let $M, T > 0$. Then there exists a self-financing trading strategy η such that $V_\eta(T)/V_\eta(0) > 1$ (i.e., the portfolio outperforms the market portfolio over the horizon $[0,T]$) for all market weight sequences $\{\mu(t)\}_{t=0}^{\infty}$ such that $\varphi(\mu(t)) > -M$ for all t and $\sum_{s=0}^{T-1} \mathbf{D}^{(1)}[\mu(s+1)|\mu(s)] > M$, where $\mathbf{D}^{(1)}$ is the $L^{(1)}$-divergence of φ.*

Proof. Let η be the trading strategy generated multiplicatively by φ. The statement is then an immediate consequence of Theorem 1.

We note that the proof of Proposition 1 looks almost trivial because we already have the concept of multiplicatively generated portfolio. Without knowing this construction, it is not immediate why a relative arbitrage η exists and how it is constructed. The usefulness of this result comes from the following observations (see [20]). Consider the *capital distribution* of the market defined by the reversed order statistics of the components of $\mu(t)$:

$$\mu_{(1)}(t) \geq \mu_{(2)}(t) \geq \cdots \geq \mu_{(n)}(t).$$

Empirically, it is found that if one plots $\log \mu_{(k)}(t)$ against $\log k$ (log of the rank), one gets an approximately linear curve (except the tail) which is relatively stable over time. This means that the capital distribution has an approximate Pareto distribution, and the market diversity $\varphi(\mu(t))$ is mean-reverting for approximately chosen φ. On the other hand, for typical φ the market volatility $\sum_{s=0}^{T-1} \mathbf{D}[\mu(s+1)|\mu(s)]$ grows roughly linearly in time [26]. Thus, it appears that the market satisfies the conditions of (1).

The stability of the capital distribution has inspired many works on the construction and analysis of market models that exhibit such behaviors. Mathematically, these are systems of Brownian particles (representing the market capitalizations) where the drift and volatility coefficients depend on their relative rankings, and so are called *rank-based models*. For more details we refer the reader to the papers [5, 31, 39, 30, 25, 32, 14] and their references.

Proposition 1 only addresses long term relative arbitrages. In practice, short term relative arbitrages are much more relevant and interesting. Naturally their constructions require more work and conditions (see for example [27, 4, 19, 40]). The paper [22] proves that market volatility alone does not imply the existence of short term relative arbitrage (this problem had been open in SPT for more than 10 years). It is interesting to note that the counterexample in [22] involves constructing a martingale that lives on an evolving submanifold of the simplex.

3.2 Examples

In this subsection we give some examples of exponentially concave functions on Δ_n, the portfolios they generate as well as the corresponding $L^{(1)}$-divergences.

Example 1 (Market portfolio). Let $\varphi(p) \equiv c$ be a constant function. It generates the identity portfolio map $\pi(p) \equiv p$ which represents the market portfolio. Its $L^{(1)}$-divergence vanishes identically, i.e., $\mathbf{D}^{(1)}[q \mid p] \equiv 0$ for $p, q \in \Delta_n$, so it is not a divergence in the technical sense of Definition 5 below.

Example 2 (Constant-weighted portfolio). Fix a probability vector $\pi \in \overline{\Delta}_n$ and consider the function

$$\varphi(p) = \sum_{j=1}^{n} \pi_j \log p_j.$$

It is exponentially concave since $\Phi = e^{\varphi} = (p_1)^{\pi_1} \cdots (p_n)^{\pi_n}$ is the geometric mean which is concave on Δ_n. This function generates the constant-weighted portfolio $\pi(p) \equiv \pi$, and the $L^{(1)}$-divergence is the excess growth rate $\mathbf{T}_\pi [q \mid p]$ given by (5). We also observe that

$$\begin{aligned} \varphi(\mu(t)) - \varphi(\mu(0)) &= \sum_{j=1}^{n} \pi_j \log \frac{\pi_j}{\mu_j(0)} - \sum_{j=1}^{n} \pi_j \log \frac{\pi_j}{\mu_j(t)} \\ &= H(\pi \mid \mu(0)) - H(\pi \mid \mu(t)) \end{aligned}$$

is the negative of the change in the relative entropy $H(\pi \mid \cdot)$. In [42], we call the decomposition (18) for this portfolio the *energy-entropy decomposition*.

Example 3 (Diversity-weighted portfolio). For $\lambda \in (0,1)$ fixed, let φ be the function

$$\varphi(p) = \frac{1}{\lambda} \log \sum_{j=1}^{n} (p_j)^{\lambda}.$$

Then φ is exponentially concave and generates the diversity-weighted portfolio where

$$\pi_i(p) = \frac{(p_i)^\lambda}{\sum_{j=1}^n (p_j)^\lambda}, \quad i = 1, \ldots, n. \tag{20}$$

Note that this portfolio interpolates between the equal-weighted portfolio $\pi(p) \equiv \overline{e}$ (when $\lambda \downarrow 0$) and the market portfolio $\pi(p) \equiv p$ (when $\lambda \uparrow 1$). See [50] where the portfolio is studied for negative values of λ. We remark that these cover, except the log case, portfolios constructed using *Tukey's transformation ladder*; see [18] for a detailed empirical study.

For $\lambda \in (0,1)$ fixed, let $p^{(\lambda)} \in \Delta_n$ be given by $\pi(p)$ as in (20). In information geometry, $p^{(\lambda)}$ is called the λ-*escort distribution* corresponding to the distribution p (see [2, Section 4.3]). Using this we may interpret the portfolio using the Rényi entropy and divergence.

Proposition 2. *For $0 < \lambda < 1$ and $p \in \Delta_n$, we have*

$$\varphi(p) = \frac{1}{\lambda} \log \left(\sum_{j=1}^n (p_j)^\lambda \right) = (\alpha - 1) \mathbf{H}_\alpha (p^{(\lambda)}),$$

where $\alpha =: \frac{1}{\lambda} \in (1, \infty)$, and

$$\mathbf{H}_\alpha(r) = \frac{1}{1-\alpha} \log \left(\sum_{j=1}^n (r_j)^\alpha \right)$$

is the Rényi entropy of order α.

Moreover, the $L^{(1)}$-divergence of φ is given by

$$\mathbf{D}^{(1)}[q \mid p] = (\alpha - 1) \mathbf{D}_\alpha (q^{(\lambda)} || p^{(\lambda)}), \tag{21}$$

where $\mathbf{D}_\alpha(\cdot||\cdot)$ is the Rényi divergence of order α defined by

$$\mathbf{D}_\alpha(p||q) = \frac{1}{\alpha - 1} \log \left(\sum_{j=1}^n (p_j)^\alpha (q_j)^{1-\alpha} \right). \tag{22}$$

Proof. This is a direct computation and we only give the proof of the first statement. Using the fact that $p = \left(p^{(\lambda)}\right)^{(1/\lambda)}$, we have

$$\begin{aligned}
\frac{1}{\lambda}\log\left(\sum_{i=1}^{n} p_j^{\lambda}\right) &= \frac{1}{\lambda}\log\left(\sum_{i=1}^{n}\left(\frac{\left(p_i^{(\lambda)}\right)^{1/\lambda}}{\sum_{j=1}^{n}\left(p_j^{(\lambda)}\right)^{1/\lambda}}\right)^{\lambda}\right)\\
&= -\log\left(\sum_{j=1}^{n}\left(p_j^{(\lambda)}\right)^{\frac{1}{\lambda}}\right)\\
&= (\alpha-1)\frac{1}{1-\alpha}\log\left(\sum_{j=1}^{n}\left(p_j^{(\lambda)}\right)^{\alpha}\right),
\end{aligned}$$

which is $\alpha - 1$ times the Rényi entropy.

Corollary 1. *Let $\lambda \in (0,1)$ and $\alpha := \frac{1}{\lambda} \in (1,\infty)$. The relative value of the diversity-weighted portfolio with parameter λ is given by*

$$\log V_{\eta}(t) = (\alpha-1)\left[\mathbf{H}_{\alpha}\left(\mu^{(\lambda)}(t)\right) - \mathbf{H}_{\alpha}\left(\mu^{(\lambda)}(0)\right) + \sum_{s=0}^{t-1}\mathbf{D}_{\alpha}\left(\mu^{(\lambda)}(t+1)\middle|\middle|\mu^{(\lambda)}(t)\right)\right].$$

Consider a portfolio manager who tries to optimize over the parameter λ for a diversity-weighted portfolio. By Corollary 1, this means comparing the dynamics of the market weight $\mu(t)$ using different escort geometries of the simplex. In particular, it is well-known that the Rényi divergence satisfies

$$\mathbf{D}_{\alpha}(r+tv||r) = \frac{\alpha}{2}t^2\|v\|_p^2 + o(t^3), \tag{23}$$

where $\|v\|_p^2 := \sum_{i=1}^{n} v_i^2/p_i$ is the Fisher information metric of the tangent vector v. This geometric viewpoint may lead to new statistical methods and algorithms. In this regard, let us mention the recent work [36] which studies the dynamics of market diversity in the context of large rank-based models, as well as the paper [3] which suggests a model for predicting change in market diversity. More generally, optimization of multiplicatively generated portfolio amounts to finding the geometry in which $\mu(t)$ has the least change in diversity and has the greatest cumulated volatility.

3.3 Multiplicative cyclical monotonicity

In this subsection we provide a financial argument (given in [43]) which motivates the definition of the multiplicatively generated portfolio.

Let us restrict to all-long trading strategies defined by portfolio maps, i.e., the portfolio weights satisfies $\pi(t) = \pi(\mu(t))$ where $\pi : \Delta_n \to \overline{\Delta}_n$ is a fixed deterministic function. When is π able to profit from market volatility? Intuitively it should satisfy the following property. Let O be a (small) neighborhood in the simplex Δ_n, and suppose $\mu(t) \in O$ for all t. From the discussion in Section 1.1 market diversity is stable. Then, we expect that the portfolio will outperform the market asymptotically as long as there is enough volatility. Specifically, it should outperform the market whenever it is periodic. This idea leads to the following definition.

Definition 4 (multiplicative cyclical monotonicity (MCM)). A portfolio map $\pi : \Delta_n \to \overline{\Delta}_n$ is multiplicatively cyclical monotone if for any cycle $\{\mu(t)\}_{t=0}^m$ with $\mu(0) = \mu(m)$ we have $V_\eta(m) \geq 1$, i.e.,

$$\prod_{t=0}^{m-1} \left(\pi(\mu(t)) \cdot \frac{\mu(t+1)}{\mu(t)} \right) \geq 1. \tag{24}$$

In [43] we observed that this property characterizes multiplicatively generated portfolio. The following result is the multiplicative analogue of Rockafellar's theorem which characterizes the subdifferentials of convex functions in terms of cyclical monotonicity [46, Section 24].

Theorem 2. *Suppose the portfolio map $\pi : \Delta_n \to \overline{\Delta}_n$ is continuous. Then it is multiplicatively cyclical monotone if and only if there exists a differentiable, exponentially concave function $\varphi : \Delta_n \to \mathbb{R}$ which generates π in the sense of* (15)*.*

Proof. Let us provide a sketch of proof. Continuity of π is included here only to simplify the statement (in the general case φ is not necessarily differentiable and we need to use supergradients). Suppose π is generated by φ. Consider a market weight sequence with $\mu(m) = \mu(0)$. By the decomposition (18), we have

$$\log V_\eta(m) - \log V_\eta(0) = \sum_{t=0}^{m-1} \mathbf{D}^{(1)} \left[\mu(t+1) \mid \mu(t) \right] \geq 0.$$

Thus $V_\eta(m) \geq 1$ and π is MCM.

Conversely, suppose that π is MCM. Consider the function φ defined by

$$\begin{aligned} \varphi(p) &= \varphi(p_0) + \inf \left\{ \log V_\eta(t) - \log V_\eta(0) \right\} \\ &= \varphi(p_0) + \inf \left\{ \sum_{s=0}^{t-1} \log \left(\sum_{i=1}^n \pi_i(\mu(s)) \frac{\mu_i(s+1)}{\mu_i(s)} \right) \right\}, \end{aligned}$$

where $p_0 \in \Delta_n$ is fixed, $\varphi(p_0) \in \mathbb{R}$ is arbitrary, and the infimum is taken over $t \geq 0$ and all market weight sequences $\{\mu(s)\}_{s=0}^t$ for which $\mu(0) = p_0$ and $\mu(t) = p$. Then it can be shown that φ is differentiable, exponentially concave, and generates the given portfolio map π. It can be shown that the function φ is unique up to an additive constant.

Using this characterization, in [43] we introduced a Monge-Kantorovich optimal transport problem and showed that the optimal coupling can be represented using exponentially concave functions and the portfolios they generate.

4 Generalized functional portfolio generation

4.1 Motivation

As it turns out, Theorem 1 is not the only way to generate a portfolio such that a pathwise decomposition holds. In the recent paper [33] the authors introduce a new form of *additive* generation and use it to construct relative arbitrages (see [47] for another extension which involves an additional finite variation process). The following result uses the terminology of [49, Section 3.3] and adapts their construction to our discrete time setting. We omit the proof as it is contained (in the limit) in Theorem 4 below.

Theorem 3 (Additively generated portfolio). *Let $\varphi : \Delta_n \to (0,\infty)$ be a smooth concave function. The trading strategy generated additively by φ is defined by*

$$\eta_i(t) = D_{e_i - \mu(t)} \varphi(\mu(t)) + V_\eta(t), \quad i = 1, \ldots, n. \tag{25}$$

This is well-defined for all t once the initial value $V_\eta(0)$ is fixed. Then the trading strategy η defined by (25) *is self-financed, and the relative value satisfies the decomposition*

$$V_\eta(t) - V_\eta(0) = \varphi(\mu(t)) - \varphi(\mu(0)) + \sum_{s=0}^{t-1} \mathbf{D}^{(0)} \left[\mu(t+1) \mid \mu(t)\right], \tag{26}$$

where $\mathbf{D}^{(0)}[\cdot|\cdot]$ is the Bregman (or $L^{(0)}$) divergence of φ as in (6).

Note that here it is the number of shares that is given in terms of the derivatives of φ.

Example 4. Consider the function

$$\varphi(p) = \frac{-1}{2}|p|^2 = \frac{-1}{2}\left(p_1^2 + \cdots + p_n^2\right).$$

It generates the trading strategy $\eta(t)$ given by $\eta_i(t) = |p|^2 - p_i + V_\eta(t)$. It is interesting to note that the Bregman divergence of φ is half of the squared Euclidean distance:

$$\mathbf{D}^{(0)}\left[q \mid p\right] = \frac{1}{2}\|p - q\|^2.$$

Thus the squared Euclidean distance indeed has a financial meaning for this specific trading strategy.

Observe that both decompositions (18) and (26) can be written in the form

$$g(V_\eta(t)) - g(V_\eta(0)) = \varphi(\mu(t)) - \varphi(\mu(0)) + \mathbf{D}\left[\mu(t+1) \mid \mu(t)\right], \tag{27}$$

where g, φ and $\mathbf{D}[\cdot \mid \cdot]$ are suitable functions:

- (Multiplicative generation) $g(x) = \log x$ and $\mathbf{D}[\cdot \mid \cdot]$ is the $L^{(1)}$-divergence of the exponentially concave function φ.
- (Additive generation) $g(x) = x$ and $\mathbf{D}[\cdot \mid \cdot]$ is the $L^{(0)}$-divergence of the concave function φ.

It is natural to ask if there exists other portfolio constructions that admit pathwise decompositions of the form (27). To formulate this question we introduce the general concept of divergence.

Definition 5 (Divergence on Δ_n). A divergence on Δ_n is a non-negative functional $\mathbf{D}[\cdot \mid \cdot] : \Delta_n \times \Delta_n \to [0, \infty)$ satisfying the following conditions:

(i) $\mathbf{D}[q \mid p] = 0$ if and only if $p = q$.
(ii)It admits a quadratic approximation of the form

$$\mathbf{D}[p + \Delta p \mid p] = \frac{1}{2} \sum_{i,j=1}^{n} g_{ij}(p) \Delta p_i \Delta p_j + O(|\Delta p|^3) \tag{28}$$

as $|\Delta p| \to 0$, and the matrix $G(p) = (g_{ij}(p))$ varies smoothly in p and is strictly positive definite in the sense that

$$\sum_{i,j=1}^{n} g_{ij}(p) v_i v_j > 0 \tag{29}$$

for all vectors $v \in \mathbb{R}^n$ that are tangent to Δ_n, i.e., $v_1 + \cdots + v_n = 0$.

If condition (i) is dropped and in (29) we do not strict inequality, we call $\mathbf{D}[\cdot \mid \cdot]$ a pseudo-divergence.

Example 5. We let $\operatorname{Hess}\varphi$ denote the Euclidean Hessian of φ. If $\alpha > 0$ and φ is α-exponentially concave, then its $L^{(\alpha)}$-divergence satisfies

$$\mathbf{D}^{(\alpha)}[p + \Delta p \mid p] = \frac{-1}{2} (\Delta p)^\top \left(\operatorname{Hess}\varphi(p) + \alpha (\nabla \varphi(p))(\nabla \varphi(p))^\top \right) (\Delta p) + O(|\Delta p|^3). \tag{30}$$

If φ is concave, then its $L^{(0)}$-divergence satisfies

$$\mathbf{D}^{(0)}[p + \Delta p \mid p] = \frac{-1}{2} (\Delta p)^\top \operatorname{Hess}\varphi(p)(\Delta p) + O(|\Delta p|^3).$$

It is easy to verify that the corresponding matrix $G(p)$ is semi-positive definite. They become true divergences if $\operatorname{Hess} e^{\alpha\varphi}$ and $\operatorname{Hess}\varphi$ respectively are strictly positive definite.

Definition 6 (General functional portfolio construction). Let $\eta = \{\eta(t)\}_{t=0}^{\infty}$ be a self-financing trading strategy whose relative value process is $\{V_\eta(t)\}$, and let $\varphi, g : \Delta_n \to \mathbb{R}$ be functions on Δ_n where g is strictly increasing. We say that η is generated by φ with scale function g if there exists a pseudo-divergence $\mathbf{D}[\cdot : \cdot]$ on Δ_n such that (27) holds for all market sequences $\{\mu(t)\}_{t=0}^{\infty}$.

In this section we will introduce a new (α, C)-generation, and, after giving an empirical example, show that this characterizes all forms of functional portfolio generation in the sense of Definition 6.

4.2 A new functional generation

In what follows we always assume that φ is smooth.

Definition 7 (**(α, C)-generation).** Let $\alpha > 0$ and $C \geq 0$ be fixed parameters. Let $\varphi : \Delta_n \to \mathbb{R}$ be smooth and α-exponentially concave. Also let the initial value $V_\eta(0) > 0$ be fixed. The trading strategy (α, C)-generated by φ is defined by

$$\eta_i(t) = \alpha(C + V_\eta(t)) D_{e_i - \mu(t)} \varphi(\mu(t)) + V_\eta(t), \quad i = 1, \ldots, n. \tag{31}$$

In Theorem 4 we show that this trading strategy corresponds to the scale function given by

$$g(x) = \frac{1}{\alpha} \log(C + x). \tag{32}$$

Moreover, in Section 4.4 we show that up to an additive constant this (together with $g(x) = x$) is the most general scale function. It is easy to verify that (31) defines a self-financing trading strategy. Comparing (31) with (15) and (25), we see that multiplicative generation corresponds to the case $C = 0$ and $\alpha = 1$, and additive generation corresponds to the limit when $\alpha = \frac{1}{C} \to 0$.

The trading strategy (31) can be interpreted as follows.

Lemma 1 (Portfolio weight of η). *Let $\pi^{(\alpha)}$ be the portfolio process generated multiplicatively by the 1-exponentially concave function $\alpha\varphi$. If $V_\eta(t) > 0$, the portfolio weight vector $\pi(t)$ of the (α, C)-generated trading strategy η is given by*

$$\pi(t) = \left(\frac{\eta_1(t)\mu_1(t)}{V_\eta(t)}, \ldots, \frac{\eta_n(t)\mu_n(t)}{V_\eta(t)} \right) = \frac{C + V_\eta(t)}{V_\eta(t)} \pi^{(\alpha)}(t) - \frac{C}{V_\eta(t)} \mu(t). \tag{33}$$

In particular, $\eta(t)$ longs the multiplicatively generated portfolio $\pi^{(\alpha)}$ and shorts the market portfolio with weights depending on $V_\eta(t)$ and C.

Proof. Direct computation using (31).

By increasing C, we may construct portfolios that are more aggressive than the multiplicatively generated portfolio. Note that we keep the parameter α so that we can generate different portfolios with the same generating function φ (as long as $e^{\alpha\varphi}$ is concave).

Next we show that the new portfolio generation admits a pathwise decomposition for the portfolio value.

Theorem 4 (Pathwise decomposition). *Consider an (α, C)-generated trading strategy η as in Definition 7. If $V_\eta(\cdot) > -C$, the value process satisfies the pathwise decomposition*

$$\frac{1}{\alpha} \log \frac{C + V_\eta(t)}{C + V_\eta(0)} = \varphi(\mu(t)) - \varphi(\mu(0)) + \sum_{s=0}^{t-1} \mathbf{D}^{(\alpha)} \left[\mu(s+1) \mid \mu(s)\right], \tag{34}$$

where $\mathbf{D}^{(\alpha)}$ is the $L^{(\alpha)}$-divergence of φ.

Proof. The proof is similar to that of Theorem 1. By (31), for each time t we have

$$\begin{aligned} &\frac{1}{\alpha} \log(C + V_\eta(t+1)) - \frac{1}{\alpha} \log(C + V_\eta(t)) \\ &= \frac{1}{\alpha} \log \frac{C + V_\eta(t) + \alpha(C + V_\eta(t)) \nabla \varphi(\mu(t)) \cdot (\mu(t+1) - \mu(t))}{C + V_\eta(t)} \\ &= \frac{1}{\alpha} \log\left(1 + \alpha \nabla \varphi(\mu(t)) \cdot (\mu(t+1) - \mu(t))\right) \\ &= \varphi(\mu(t+1)) - \varphi(\mu(t)) + \mathbf{D}^{(\alpha)} \left[\mu(t+1) \mid \mu(t)\right]. \end{aligned} \tag{35}$$

This yields the desired decomposition.

4.3 An empirical example

Consider a smooth and exponentially concave function φ. It is α-exponentially concave for all $0 < \alpha \leq 1$ and is concave (which corresponds to the case $\alpha \downarrow 0$). Thus both the additive and multiplicatively generated portfolios are well-defined. Unfortunately, while the $L^{(\alpha)}$-divergence is a natural interpolation, there does not seem to be a canonical choice for the constant C that connects the two basic cases.

In this example we consider instead the parameterized family $\{\eta^{(\alpha)}\}_{0 \leq \alpha \leq 1}$ where $\eta^{(\alpha)}$ is the trading strategy $(\alpha, \frac{1}{\alpha})$-generated by φ (so when $\alpha = 0$ it is the additively generated portfolio), and compare their empirical performance. Note that $\eta^{(1)}$ is not the multiplicatively generated portfolio as it also shorts the market portfolio.

Consider as in Figure 1 the (beginning) monthly stock prices of the US companies Ford, Walmart and IBM from January 1990 ($t = 0$) to September 2017 ($t = 332$). We normalize the prices so that at $t = 0$ the market weight is at the barycenter $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. The path of the market weight $\mu(t)$ in the simplex Δ_3 is plotted in Figure 1 (right).

We consider the 1-exponentially concave function

$$\varphi(p) = \sum_{i=1}^{3} \frac{1}{3} \log p_i \tag{36}$$

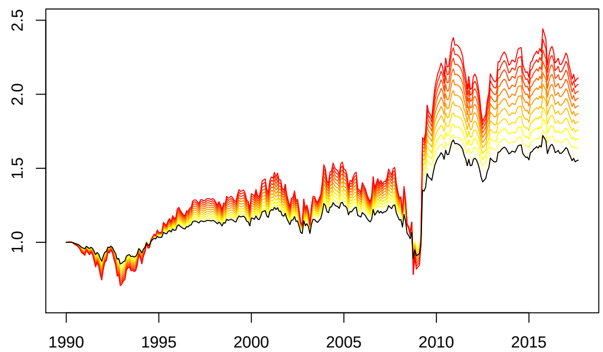

Fig. 5: Time series of the relative portfolio value $V_{\eta^{(\alpha)}}(t)$, from $\alpha = 0$ (yellow) to $\alpha = 1$ (red). The value of the equal-weighted portfolio is shown in black.

which generates multiplicatively the equal-weighted portfolio $\pi(p) \equiv \overline{e} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. By (31), for each $\alpha \in [0,1]$ the trading strategy is given by

$$\eta_i^{(\alpha)}(t) = (1 + \alpha V_\eta(t)) \left(\frac{1}{3\mu_i(t)} - 1 \right) + V_\eta(t).$$

In terms of portfolio weights, we have

$$\pi^{(\alpha)}(t) = \frac{1 + \alpha V_\eta(t)}{V_\eta(t)} \overline{e} - \frac{1 + \alpha V_\eta(t) - V_\eta(t)}{V_\eta(t)} \mu(t).$$

Thus the portfolio longs more and more the equal-weighted portfolio as α increases. We also set $V_\eta(0) = 1$. The corresponding $L^{(\alpha)}$-divergence is given by

$$\mathbf{D}^{(\alpha)}\left[q \mid p\right] = \frac{1}{\alpha} \log \left(1 + \alpha \sum_{i=1}^{n} \frac{1}{np_i} (q_i - p_i) \right) - \sum_{i=1}^{n} \frac{1}{n} \log \frac{q_i}{p_i}.$$

The relative values of the simulated portfolios are plotted in Figure 5. At the end of the period the portfolio value is increasing in α, and the additive portfolio ($\alpha = 0$) has the smallest value. It is interesting to note that the reverse is true at the beginning. Note that the values fluctuate widely in the period 2008–2009 corresponding to the financial crisis. For comparison, we also simulate the multiplicatively generated equal-weighted portfolio (i.e., $(\alpha, C) = (1, 0)$) and plot the result in Figure 5. Interestingly, the additive and multiplicative portfolios have similar behaviors here. In this period, shorting the market by using a positive value for C gives significant advantage over both the additive and multiplicative portfolios. Dynamic optimization over our extended functionally generated portfolios is an interesting problem.

4.4 Characterizing functional portfolio generation

Now we show that our (α, C)-generation is the most general one. Throughout this subsection we let η be a functionally generated trading strategy as in Definition 6. We assume that the scale function g is smooth and $g'(x) > 0$ for all x. We also require that the domain of g contains the positive real line $(0, \infty)$. Furthermore, we assume that φ is smooth, and η is non-trivial in the sense that for all $t \geq 0$ and all market weight paths $\{\mu(s)\}_{s=0}^t$ up to time t, the profit-or-loss

$$V_\eta(t+1) - V_\eta(t) = \eta(t) \cdot (\mu(t+1) - \mu(t))$$

is not identically zero as a function of $\mu(t+1) \in \Delta_n$.

Theorem 5. *Under the above conditions, the scale function has one of the following forms. Either*

$$g(x) = c_1 x + c_2 \tag{37}$$

where $c_1 > 0$ and $c_2 \in \mathbb{R}$, or

$$g(x) = c_2 \log(c_1 + x) + c_3 \tag{38}$$

where $c_1 \geq 0$, $c_2 > 0$ and $c_3 \in \mathbb{R}$. In the first case φ is concave and η is additively generated by φ, whereas in the second case φ is α-exponentially concave with $c_2 = \frac{1}{\alpha}$ and η is (α, c_1)-generated by φ. The corresponding pseudo-divergence is the $L^{(\alpha)}$-divergence of φ.

Note that in (37) and (38) the additive constants are irrelevant and may be discarded. We will prove Theorem 5 with several lemmas. First we observe that the decomposition (27) already implies a formula of the trading strategy.

Lemma 2. *For any t and any tangent vector v of Δ_n (i.e., $v_1 + \cdots + v_n = 0$) we have*

$$\eta(t) \cdot v = \frac{1}{g'(V_\eta(t))} \nabla \varphi(\mu(t)) \cdot v. \tag{39}$$

In particular, for each $i = 1, \ldots, n$ we have

$$\eta_i(t) = \frac{1}{g'(V_\eta(t))} D_{e_i - \mu(t)} \varphi(\mu(t)) + V_\eta(t). \tag{40}$$

Proof. From (27) we have the identity

$$g(V_\eta(t+1)) - g(V_\eta(t)) = \varphi(\mu(t+1)) - \varphi(\mu(t)) + \mathbf{D}[\mu(t+1) \mid \mu(t)] \tag{41}$$

which holds for all values of $\mu(t+1)$. Moreover, we have

$$V_\eta(t+1) = V_\eta(t) + \eta(t) \cdot (\mu(t+1) - \mu(t)).$$

Now let $\mu(t+1) - \mu(t) = \delta v$, $\delta > 0$ sufficiently small, and compute the first order approximation of both sides of (41). Since $\mathbf{D}[\cdot \mid \cdot]$ is a pseudo-divergence, by (28) its first order approximation vanishes. Evaluating the derivatives and dividing by $\delta > 0$, we obtain (39).

Letting $v = e_i - \mu(t)$ in (40) for $i = 1, \ldots, n$, we get the explicit formula (40).

Observe that (39) reduces to (25) when $g(x) = x$, and to (15) when $g(x) = \log x$. Also, we note that the trading strategy depends only on $\mu(t)$ and the current value $V_\eta(t)$. Putting $v = \mu(t+1) - \mu(t)$ in (39), we have

$$V_\eta(t+1) - V_\eta(t) = \frac{1}{g'(V_\eta(t))} \nabla\varphi(\mu(t)) \cdot (\mu(t+1) - \mu(t)). \tag{42}$$

Consider the expression

$$\begin{aligned} &g(V_\eta(t+1)) - g(V_\eta(t)) \\ &= g(V_\eta(t) + (V_\eta(t+1) - V_\eta(t))) - g(V_\eta(t)) \\ &= g\left(V_\eta(t) + \frac{1}{g'(V_\eta)} \nabla\varphi(\mu(t)) \cdot (\mu(t+1) - \mu(t)\rangle)\right) - g(V_\eta(t)). \end{aligned} \tag{43}$$

By (27), this equals

$$\varphi(\mu(t+1)) - \varphi(\mu(t)) + \mathbf{D}[\mu(t+1)|\mu(t)],$$

which is a function of $\mu(t)$ and $\mu(t+1)$ only. Thus, the expression in (43) does not depend on the current portfolio value $V_\eta(t)$. From this observation we will derive a differential equation satisfied by g.

Lemma 3. *The scale function g satisfies the third order nonlinear ODE*

$$g' g''' = 2(g'')^2. \tag{44}$$

Proof. Let $x = V_\eta(t)$, and let $\delta = \nabla\varphi(\mu(t)) \cdot (\mu(t+1) - \mu(t))$. From (43), for any δ, the expression

$$g(x + \frac{1}{g'(x)}\delta) - g(v) \tag{45}$$

does not depend on x.

Differentiating (45) with respect to x, we have

$$g'(x + \frac{1}{g'(x)}\delta)\left(1 - \delta\frac{g''(x)}{(g'(x))^2}\right) - g'(x) = 0.$$

Next we differentiate with respect to δ (since η is assumed to be non-trivial, this can be done by varying $\mu(t+1)$):

$$g''(x+\frac{1}{g'(x)}\delta)\frac{1}{g'(x)}\left(1-\delta\frac{g''(x)}{(g'(x))^2}\right)+g'(x+\frac{1}{g'(x)}\delta)\frac{-g''(x)}{(g'(x))^2}=0.$$

Differentiating one more time with respect to δ, we have

$$g'''(x+\frac{1}{g'(x)}\delta)\frac{1}{(g'(x))^2}\left(1-\delta\frac{g''(x)}{(g'(x))^2}\right)$$
$$+g''(x+\frac{1}{g'(x)}\delta)\frac{-g''(x)}{(g'(x))^3}-\frac{g''(x+\frac{1}{g'(x)}\delta)g''(x)}{(g'(x))^3}=0.$$

Setting $\delta=0$, we get $\frac{g'''(x)}{(g'(x))^2}-2\frac{(g''(x))^2}{(g'(x))^3}=0$ which gives the ODE (44).

With the differential equation in hand, it is not difficult to find the general solution:

Lemma 4. *All solutions to the ODE* (44) *can be written in the form*

$$g(x)=c_0+c_1x \quad or \quad g(x)=c_2\log(c_1+x)+c_3, \tag{46}$$

where the c_i's are real constants. The constraints on the constants follow from our assumptions of g.

Now we are ready to complete the proof of Theorem 5. By Lemma 4, the scale function (up to an additive constant which is irrelevant) has the form (37) or (38). Consider the second case (the first case is similar). Then by (42), (43) and the third equality of (35) (which does not depend on α-exponential concavity of φ), for any $p=\mu(s)$ and $q=\mu(s+1)$, we have

$$\mathbf{D}\left[q\mid p\right]=\frac{1}{\alpha}\log\left(1+\alpha\nabla\varphi(p)\cdot(q-p)\right)-(\varphi(q)-\varphi(p))$$

which is exactly the expression of the $L^{(\alpha)}$-divergence. By assumption $\mathbf{D}\left[\cdot\mid\cdot\right]$ is a pseudo-divergence so it is non-negative for all p, q. It is easy to check that this implies that φ is α-exponentially concave, and so η is the (α,c_1)-generated trading strategy.

4.5 Extension to continuous time

We briefly discuss the extension of the (α,C)-portfolio generation to continuous time which is the conventional setting of mathematical finance in general and SPT in particular. Detailed study is left for future research.

In continuous time, we let $(\Omega,\mathscr{F},\{\mathscr{F}_t\}_{t\geq 0},\mathbb{P})$ be a given filtered probability space satisfying the usual conditions (see [34, 45] for introductions to stochastic calculus), and all processes are defined on this space. Now the market weight $\{\mu(t)\}_{t\geq 0}$ is modeled as a continuous semimartingale with values in Δ_n.

A self-financing trading strategy is a progressively measurable process $\eta(t) = (\eta_1(t), \ldots, \eta_n(t))$ such that if $V_\eta(t) := \sum_{i=1}^n \eta_i(t)\mu_i(t)$ is the (relative) value of the portfolio, then V_η satisfies the stochastic differential equation

$$dV_\eta(t) = \sum_{i=1}^n \eta_i(t) d\mu_i(t),$$

which is the infinitesimal analogue of (11). Let φ be smooth and α-exponentially concave, and let $C \geq 0$. Then, under under suitable conditions not mentioned here, there exists a self-financing trading strategy η given by (31), and the relative value process satisfies the decomposition

$$\frac{1}{\alpha} \log \frac{C + V_\eta(t)}{C + V_\eta(0)} = \varphi(\mu(t)) - \varphi(\mu(0)) + \int_0^t \mathbf{G}(d\mu, d\mu), \tag{47}$$

where

$$\mathbf{G}(p)(u,v) := u^\top \left(-\operatorname{Hess} \varphi(p) - \alpha (\nabla \varphi(p))(\nabla \varphi(p))^\top \right) v \tag{48}$$

is the Riemannian metric of the $L^{(\alpha)}$-divergence as in (30), and the finite variation process $\int_0^t \mathbf{G}(d\mu, d\mu)$ is the $\mathbf{G}$-*quadratic variation* of μ in the sense of [17, Chapter 3]. Thus we may regard (47) as different ways of writing down Itô's formula for the process $\{\varphi(\mu(t))\}$. It is an interesting problem to relate this with information geometry (in particular, the dualistic geometry of the $L^{(\alpha)}$-divergence [53]) and stochastic differential geometry [29].

5 Further properties of L-divergence

In this section we gather some further properties of L-divergence and point out some related problems. For simplicity we focus on the $L^{(1)}$-divergence, and refer the reader to [53] for a systematic study of the $L^{(\alpha)}$-divergence. We always assume that the generating functions are smooth. It is clear that some of the problems make sense on domains other than the unit simplex.

5.1 Interpolation and comparison

If $\varphi^{(0)}$ and $\varphi^{(1)}$ are exponentially concave functions on Δ_n, then it is an easy consequence of the AM-GM inequality that

$$\varphi^{(\lambda)} := (1 - \lambda)\varphi^{(0)} + \lambda \varphi^{(1)}$$

is exponentially concave for any $0 < \lambda < 1$. If $\pi^{(0)}$ and $\pi^{(1)}$ are the portfolio maps generated multplicatively by $\varphi^{(0)}$ and $\varphi^{(1)}$ respectively, then $\varphi^{(\lambda)}$ generates the portfolio map

$$\pi^{(\lambda)}(\cdot) \equiv (1-\lambda)\pi^{(0)}(\cdot) + \lambda\pi^{(1)}(\cdot),$$

which is a constant-weighted portfolio of $\pi^{(0)}$ and $\pi^{(1)}$ [51, Lemma 4.4]. Thus, the spaces of exponentially concave functions and MCM portfolio maps (see Definition 3) are convex. Moreover, the L-divergence $\mathbf{D}^{(1)}[\cdot \mid \cdot]$ is concave in the function φ. In [44], this interpolation provides a new displacement interpolation for a logarithmic optimal transport problem.

Let $\mathbf{D}_\varphi$ and $\mathbf{D}_\psi$ be the L-divergences generated respectively by the exponentially concave functions φ and ψ, we say that $\mathbf{D}_\psi$ *dominates* $\mathbf{D}_\varphi$ if

$$\mathbf{D}_\psi^{(1)}[q \mid p] \geq \mathbf{D}_\varphi^{(1)}[q \mid p] \tag{49}$$

for all $p, q \in \Delta_n$. Financially, this means that the portfolio generated by ψ captures more volatility than that generated by φ. An interesting problem is to find the maximal elements in this partial order, and the following result is obtained in [51] using the relative convexity lemma in [11].

Theorem 6. *Suppose φ is symmetric, i.e., $\varphi(p_1, \ldots, p_n) = \varphi(p_{\sigma(1)}, \ldots, p_{\sigma(n)})$ for any permutation σ of the coordinates. If*

$$\int_0^1 e^{-2\varphi((1-t)e_1 + t\bar{e})} dt = \infty,$$

then $\mathbf{D}_\varphi$ is maximal in the partial order (49)*: if $\mathbf{D}_\varphi$ is dominated by $\mathbf{D}_\psi$, then $\varphi - \psi$ is constant on Δ_n and $\mathbf{D}_\psi \equiv \mathbf{D}_\varphi$.*

As an example, the function $\varphi(p) = \frac{1}{n}\sum_{i=1}^n \log p_i$ (which generates the equal-weighted portfolio) is maximal. Another example is $\varphi(p) = \log\left(-\sum_{i=1}^n p_i \log p_i\right)$, the logarithm of the Shannon entropy.

5.2 Universal portfolio

Let η be a given self-financing trading strategy in the market, and we assume that η is all-long, i.e., $\eta_i(t) \geq 0$ for all i and t. Then it can be described by the portfolio process π, and we denote its portfolio value at time t by $V_\pi(t)$. Given an exponentially concave function φ, we let

$$\log V_\varphi(t) = \varphi(\mu(t)) - \varphi(\mu(0)) + \sum_{s=0}^{t-1} \mathbf{D}_\varphi^{(1)}[\mu(s+1) \mid \mu(s)]$$

be its (log) portfolio value. Without loss of generality we assume all portfolios start at 1 at time 0: $V_\pi(0) = V_\varphi(0) = 1$.

Consider the average regret of the strategy η with respect to φ:

$$r(\eta;\varphi) := \frac{1}{t}\log\frac{V_\varphi(t)}{V_\eta(t)}.$$

A natural financial problem is to minimize the regret in a model-independent way. Specifically, we ask for a (non-anticipating) strategy η such that

$$\lim_{t\to 0}\sup_{\varphi} r(\eta;\varphi) = 0 \tag{50}$$

for 'all' sequences $\{\mu(t)\}_{t=0}^{\infty}$ in the simplex, and the supremum is taken over all exponentially concave functions. Phrased this way, it is an online learning problem [10] where the set of experts is nonparametric. This type of problem was first proposed and solved by Cover [12] in the case where the supremum is taken over constant-weighted portfolios (which form a finite-dimensional simplex). Also see [28] which studies Cover's problem in the context of online convex optimization. Extending Cover's construction, in [52] the author proved the following.

Theorem 7. *(Rough statement) Let ν_0 be a (prior) probability measure over the set of MCM portfolios τ with dense support. For each t, define the posterior distribution ν_t by*

$$\frac{d\nu_t}{d\nu_0}(\tau) \propto V_\tau(t), \quad \tau \text{ MCM},$$

and invest in the posterior mean

$$\pi(t) = \int \tau(t)d\nu_t(\tau).$$

Then the limit (50) *holds for any market weight sequences $\{\mu(t)\}_{t=0}^{\infty}$ such that the empirical measure $\frac{1}{t}\sum_{s=0}^{t-1}\delta_{(\mu(s),\mu(s+1))}$ converge weakly to an absolutely continuous measure on $\Delta_n \times \Delta_n$.*

Further results were obtained in [13], but we were not able to obtain explicit rates. A challenge is to construct and analyze appropriate prior distributions on the set of multiplicatively generated portfolios or exponentially concave functions. Thus let us pose the following

Problem 1. Find a portfolio process with an explicit rate of convergence in (50).

5.3 Dualistic geometry and the generalized Pythagorean theorem

Consider the $L^{(1)}$-divergence $\mathbf{D}[\cdot \mid \cdot]$ of an exponentially concave function φ on the simplex Δ_n. It defines a Riemannian metric g and a dual pair of torsion-free affine connections (∇, ∇^*) (see [2, Chapter 6]). In particular, the Riemanniain metric is given by (48). In [44] this geometry is derived and many interesting properties are shown. It generalizes the dually flat geometry of Bregman divergence (a unified framework is established recently in [53]).

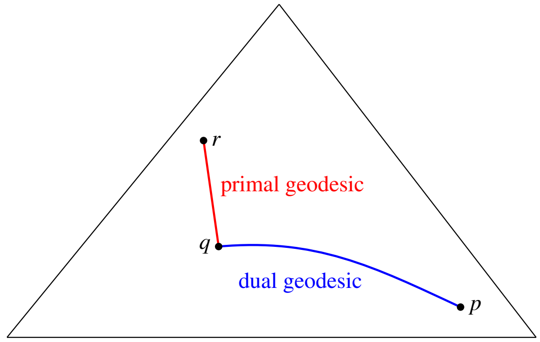

Fig. 6: Generalized Pythagorean theorem for L-divergence.

As before we let p denote a generic element of Δ_n. Now we regard this a global coordinate system of the manifold $M = \Delta_n$, and call it the *primal* coordinate system. Let π be the portfolio map generated by φ. It defines a *dual* coordinate system

$$p^* := \left(\frac{\pi_1(p)/p_1}{\sum_{j=1}^n \pi_j(p)/p_j}, \ldots, \frac{\pi_n(p)/p_n}{\sum_{j=1}^n \pi_j(p)/p_j} \right)$$

which also takes values in the unit simplex. The main properties of the geometry is summarized in the following theorem, and we refer the reader to [44, 53] for further properties related to the geodesic equations, gradient flows and connections with optimal transport.

Theorem 8. *[44] Consider the dualistic geometry induced by* $\mathbf{D}[\cdot \mid \cdot]$.

(i) The trace of a primal geodesic is a straight line under the primal coordinate system.
(ii) The trace of a dual geodesic is a straight line under the dual coordinate system.
(iii) The geometry has constant (primal and dual) sectional curvature -1 *(when* $n \geq 3$*) with respect to the induced Riemannian metric.*

In particular, the induced geometry is dually *projectively* flat but not flat. Furthermore, the $L^{(1)}$-divergence satisfies a generalized Pythagorean theorem.

Theorem 9 (Generalized Pythagorean theorem). *Given* $(p,q,r) \in (\Delta_n)^3$*, consider the dual geodesic joining q and p and the primal geodesic joining q and r. Consider the Riemannian angle between the geodesics at q. Then the difference*

$$\mathbf{D}[q \mid p] + \mathbf{D}[r \mid q] - \mathbf{D}[r \mid p] \tag{51}$$

is positive, zero or negative depending on whether the angle is less than, equal to, or greater than 90 *degrees (see Figure 6).*

Further properties of the Pythagorean theorem can be studied. To give a flavor we present an interesting result for the excess growth rate $\mathbf{T}_\pi[\cdot \mid \cdot]$ defined by (5).

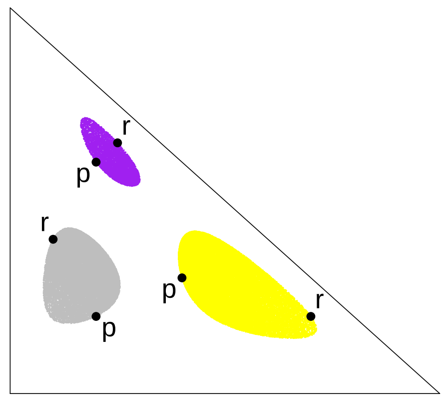

Fig. 7: The sublevel set $\{q : f(q) \leq 0\}$ for several pairs of p and r, for the equal-weighted portfolio $\pi = \bar{e}$. Here f is defined by (52).

Proposition 3. *Consider the excess growth rate $\mathbf{T}_\pi [\cdot \mid \cdot]$ for a fixed portfolio weight vector $\pi \in \overline{\Delta_n}$. For $p, r \in \Delta_n$ fixed, the mapping*

$$q \in \Delta_n \mapsto f(q) := \mathbf{D}[q \mid p] + \mathbf{D}[r \mid q] - \mathbf{D}[r \mid p] \tag{52}$$

is quasiconvex, i.e, the sublevel sets $\{q : f(q) \leq \lambda\}$ are convex (see Figure 7).

Proof. It suffices to show that the map

$$g(x) := \log\left(\pi \cdot \frac{x}{p}\right) + \log\left(\pi \cdot \frac{r}{x}\right)$$

is quasiconvex (Here x/p and r/x are the vectors of componentwise ratios.), and we will use the following characterization (see [7, Section 3.4.3]): g is quasiconvex if and only if

$$g(y) \leq g(x) \Rightarrow \nabla g(x) \cdot (y - x) \leq 0. \tag{53}$$

for any x and y.

Let $x, y \in \Delta_n$ be such that $g(y) \leq g(x)$. We have

$$\partial_i g(x) = \frac{\frac{\pi_i}{p_i}}{\pi \cdot \frac{x}{p}} + \frac{-\frac{\pi_i r_i}{x_i^2}}{\pi \cdot \frac{r}{x}}.$$

After some simplifications, we have

$$\begin{aligned}\nabla g(x)\cdot(y-x) &= \sum_{i=1}^{n}\left(\frac{\frac{\pi_i}{p_i}}{\pi\cdot\frac{x}{p}} + \frac{-\frac{\pi_i r_i}{x_i^2}}{\pi\cdot\frac{r}{x}}\right)(y_i-x_i) \\ &= \frac{\pi\cdot\frac{y}{p}}{\pi\cdot\frac{x}{p}} - \frac{\sum_{i=1}^{n}\pi_i\frac{r_i}{x_i}\frac{y_i}{x_i}}{\pi\cdot\frac{r}{x}}.\end{aligned} \tag{54}$$

Since $g(y)\le g(x)$, we have

$$\frac{\pi\cdot\frac{x}{p}}{\pi\cdot\frac{x}{p}} \le \frac{\pi\cdot\frac{r}{x}}{\pi\cdot\frac{r}{y}}.$$

Substituting this into (54), we get

$$\begin{aligned}\nabla g(x)\cdot(y-x) &\le \frac{\pi\cdot\frac{r}{x}}{\pi\cdot\frac{r}{y}} - \frac{\sum_{i=1}^{n}\pi_i\frac{r_i}{x_i}\frac{y_i}{x_i}}{\pi\cdot\frac{r}{x}} \\ &= \frac{1}{\left(\pi\cdot\frac{r}{x}\right)\left(\pi\cdot\frac{r}{y}\right)}\left[\left(\pi\cdot\frac{r}{x}\right)^2 - \sum_{i,j=1}^{n}\pi_i\pi_j r_i r_j\frac{1}{x_i x_j}\frac{y_j}{x_j}\frac{x_i}{y_i}\right] \\ &= \frac{1}{\left(\pi\cdot\frac{r}{x}\right)\left(\pi\cdot\frac{r}{y}\right)}\sum_{i,j=1}^{n}\frac{\pi_i r_i}{x_i}\frac{\pi_j r_j}{x_j}\left(1-\frac{x_i}{y_i}\frac{y_j}{x_j}\right).\end{aligned} \tag{55}$$

Let $A=\sum_{i=1}^{n}\frac{\pi_i r_i}{x_i}$ and let $\alpha_i=\frac{\pi_i r_i}{x_i}/A$. Note that α is a probability vector. Now we may write (55) in the form

$$C\left(1-\sum_{i,j=1}^{n}\alpha_i\alpha_j\frac{x_i}{y_i}\frac{y_j}{x_j}\right),$$

where $C>0$ is a constant. Let X and Y be independent and identically distributed random variables such that

$$\mathbb{P}\left(X=\frac{x_i}{y_i}\right)=\mathbb{P}\left(Y=\frac{x_i}{y_i}\right)=\alpha_i,\quad i=1,\dots,n.$$

By Jensen's inequality, we have

$$\sum_{i,j=1}^{n}\alpha_i\alpha_j\frac{x_i}{y_i}\frac{y_j}{x_j}=\mathbb{E}\left[X\cdot\frac{1}{Y}\right]=\mathbb{E}[X]\mathbb{E}\left[\frac{1}{Y}\right]\ge\mathbb{E}[X]\frac{1}{\mathbb{E}[X]}=1.$$

Thus $\nabla g(\mathbf{x})\cdot(y-x)\le 0$ and we have proved that g is quasiconvex.

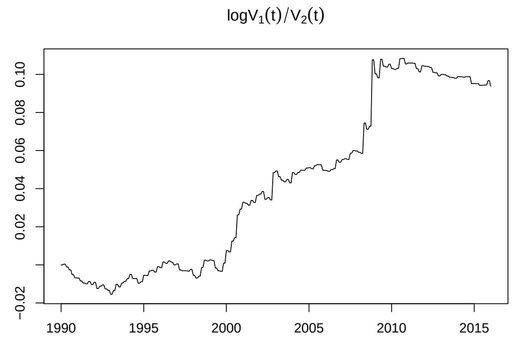

Fig. 8: Relative performance of the portfolio V_1 (rebalanced every month) versus V_2 (rebalanced every two months).

5.4 Optimal rebalancing frequency

Finally we mention a practical problem related to the L-divergence. The generalized Pythagorean theorem gives a geometric way to study the rebalancing frequency of a functionally generated portfolio. To give a simple example, consider the empirical example as in Figure 1 and the equal-weighted portfolio $\pi(p) \equiv \overline{e} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

Consider two ways of implementing this portfolio: (i) rebalance every month; (ii) rebalance every two months. If V_1 and V_2 denote respectively the values of these implementations, Figure 9 plots the time series of the relative performance $\log V_1(t)/V_2(t)$. For this data set, rebalancing every month boost the return by about 10% in log scale. Using the decomposition (18), we see that if T is the terminal time, then

$$\begin{aligned}
&\log \frac{V_1(T)}{V_2(T)} \\
&= \sum_k \left(\mathbf{D}\left[\mu(2k+1) \mid \mu(2k)\right] + \mathbf{D}\left[\mu(2k+1) \mid \mu(2k+2)\right] - \mathbf{D}\left[\mu(2k+2) \mid \mu(k)\right]\right).
\end{aligned}$$

By Theorem 9, the sign of each term is determined by the Riemannian angle of the geodesic triangle. This angle summarizes in a single number the correlation among the stock returns that is relevant to the rebalancing frequency. Further work should study the joint relationship between the angle and the *size* of the geodesic triangle which determines the magnitude of (51).

In practice trading incurs transaction costs which have been neglected so far. Transaction costs create a drag of the portfolio value. A common setting is that the transaction cost is *proportional*, i.e., we pay a fixed percentage of the value exchanged. In our model, the transaction cost may be approximated by a functional $\mathbf{C}[q \mid p] \geq 0$ where p and q are the beginning and ending market weights of the hold-

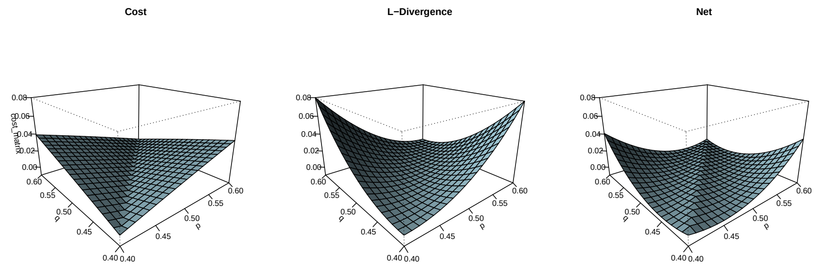

Fig. 9: Left: Transaction cost $\mathbf{C}[q \mid p]$. Middle: The L-divergence $\mathbf{D}[q \mid p]$. Right: The difference $\mathbf{D} - \mathbf{C}$. In these figures $n = 2$ and the x and y-axes are the first coordinates of p and q respectively.

ing period (Figure 9). Since the transaction cost is proportional, the cost is of linear order when $q \approx p$. On the other hand, the L-divergence is approximately quadratic when $q \approx p$. Thus the net difference $\mathbf{D}[q \mid p] - \mathbf{C}[q \mid p]$ is negative when q is sufficiently close to p. Financially, this means that the investor should not rebalance too often – at least when the increments of the market weights are 'small'. We end this paper with the following problem:

Problem 2. Design a robust strategy for rebalancing a given trading strategy.

References

1. S. Amari, H. Nagaoka, and D. Harada. *Methods of Information Geometry*. Translations of mathematical monographs. American Mathematical Society, 2002.
2. S.-I. Amari. *Information Geometry and Its Applications*. Springer, 2016.
3. F. Audrino, R. Fernholz, and R. G. Ferretti. A forecasting model for stock market diversity. *Annals of Finance*, 3(2):213–240, 2007.
4. A. D. Banner and D. Fernholz. Short-term relative arbitrage in volatility-stabilized markets. *Annals of Finance*, 4(4):445–454, 2008.
5. A. D. Banner, R. Fernholz, and I. Karatzas. Atlas models of equity markets. *The Annals of Applied Probability*, 15(4):2296–2330, 2005.
6. D. G. Booth and E. F. Fama. Diversification returns and asset contributions. *Financial Analysts Journal*, 48(3):26–32, 1992.
7. S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
8. T. Breuer and I. Csiszár. Information geometry in mathematical finance: Model risk, worst and almost worst scenarios. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 404–408. IEEE, 2013.
9. D. C. Brody and L. P. Hughston. Interest rates and information geometry. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 457, pages 1343–1363. The Royal Society, 2001.
10. N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

11. M. Chuaqui, P. Duren, and B. Osgood. Schwarzian derivative criteria for valence of analytic and harmonic mappings. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 143, pages 473–486. Cambridge University Press, 2007.
12. T. M. Cover. Universal portfolios. *Math. Finance*, 1(1):1–29, 1991.
13. C. Cuchiero, W. Schachermayer, and T.-K. L. Wong. Cover's universal portfolio, stochastic portfolio theory and the numeraire portfolio. *arXiv preprint arXiv:1611.09631*, 2016.
14. A. Dembo and L.-C. Tsai. Equilibrium fluctuation of the atlas model. *The Annals of Probability*, 45(6b):4529–4560, 2017.
15. S. Eguchi. Second order efficiency of minimum contrast estimators in a curved exponential family. *The Annals of Statistics*, 11(3):793–803, 1983.
16. S. Eguchi et al. Geometry of minimum contrast. *Hiroshima Mathematical Journal*, 22(3):631–647, 1992.
17. M. Émery. *Stochastic Calculus in Manifolds*. Springer, 1989.
18. P. A. Ernst, J. R. Thompson, and Y. Miao. Tukeys transformational ladder for portfolio management. *Financial Markets and Portfolio Management*, 31(3):317–355, 2017.
19. D. Fernholz and I. Karatzas. On optimal arbitrage. *Ann. Appl. Probab.*, 20(4):1179–1204, 2010.
20. E. R. Fernholz. *Stochastic Portfolio Theory*. Springer, 2002.
21. E. R. Fernholz and I. Karatzas. Stochastic portfolio theory: an overview. In P. G. Ciarlet, editor, *Handbook of Numerical Analysis*, volume 15 of *Handbook of Numerical Analysis*, pages 89 – 167. Elsevier, 2009.
22. E. R. Fernholz, I. Karatzas, and J. Ruf. Volatility and arbitrage. *Annals of Probability*, forthcoming.
23. R. Fernholz. Portfolio generating functions. In *Quantitative Analysis in Financial Markets: Collected Papers of the New York University Mathematical Finance Seminar*, volume 1, pages 344–367. World Scientific, 1999.
24. R. Fernholz, R. Garvy, and J. Hannon. Diversity-weighted indexing. *The Journal of Portfolio Management*, 24(2):74–82, 1998.
25. R. Fernholz, T. Ichiba, and I. Karatzas. A second-order stock market model. *Annals of Finance*, 9(3):439–454, 2013.
26. R. Fernholz and I. Karatzas. Relative arbitrage in volatility-stabilized markets. *Annals of Finance*, 1(2):149–177, 2005.
27. R. Fernholz, I. Karatzas, and C. Kardaras. Diversity and relative arbitrage in equity markets. *Finance and Stochastics*, 9(1):1–27, 2005.
28. E. Hazan and S. Kale. An online portfolio selection algorithm with regret logarithmic in price variation. *Mathematical Finance*, 25(2):288–310, 2015.
29. E. P. Hsu. *Stochastic Analysis on Manifolds*, volume 38. American Mathematical Society, 2002.
30. T. Ichiba, S. Pal, and M. Shkolnikov. Convergence rates for rank-based models with applications to portfolio theory. *Probability Theory and Related Fields*, 156(1-2):415–448, 2013.
31. T. Ichiba, V. Papathanakos, A. Banner, I. Karatzas, and R. Fernholz. Hybrid atlas models. *The Annals of Applied Probability*, 21(2):609–644, 2011.
32. B. Jourdain and J. Reygner. Capital distribution and portfolio performance in the mean-field atlas model. *Annals of Finance*, 11(2):151–198, 2015.
33. I. Karatzas and J. Ruf. Trading strategies generated by Lyapunov functions. *Finance and Stochastics*, 21(3):753–787, 2017.
34. I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113. Springer, 2012.
35. K. Khashanah and H. Yang. Evolutionary systemic risk: Fisher information flow metric in financial network dynamics. *Physica A: Statistical Mechanics and its Applications*, 445:318–327, 2016.
36. S. A. A. Monter, M. Shkolnikov, and J. Zhang. Dynamics of observables in rank-based models and performance of functionally generated portfolios. *arXiv preprint arXiv:1802.03593*, 2018.

37. R. Nock, B. Magdalou, E. Briys, and F. Nielsen. On tracking portfolios with certainty equivalents on a generalization of Markowitz model: the fool, the wise and the adaptive. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 73–80, 2011.
38. R. Nock, B. Magdalou, E. Briys, and F. Nielsen. Mining matrix data with bregman matrix divergences for portfolio selection. In *Matrix Information Geometry*, pages 373–402. Springer, 2013.
39. S. Pal. Analysis of market weights under volatility-stabilized market models. *The Annals of Applied Probability*, 21(3):1180–1213, 2011.
40. S. Pal. Exponentially concave functions and high dimensional stochastic portfolio theory. *arXiv preprint arXiv:1603.01865*, 2016.
41. S. Pal. Embedding optimal transports in statistical manifolds. *Indian Journal of Pure and Applied Mathematics*, 48(4):541–550, 2017.
42. S. Pal and T.-K. L. Wong. Energy, entropy, and arbitrage. *arXiv preprint arXiv:1308.5376*, 2013.
43. S. Pal and T.-K. L. Wong. The geometry of relative arbitrage. *Mathematics and Financial Economics*, 10(3):263–293, 2016.
44. S. Pal and T.-K. L. Wong. Exponentially concave functions and a new information geometry. *Annals of Probability*, forthcoming.
45. D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293. Springer, 2013.
46. R. T. Rockafellar. *Convex Analysis*. Princeton Landmarks in Mathematics. Princeton University Press, 1997.
47. J. Ruf and K. Xie. Generalised Lyapunov functions and functionally generated trading strategies. *arXiv preprint arXiv:1801.07817*, 2018.
48. B. Trivellato. Deformed exponentials and applications to finance. *Entropy*, 15(9):3471–3489, 2013.
49. A. Vervuurt. *On Portfolio Construction Through Functional Generation*. PhD thesis, Oxford Unviersity, 2016.
50. A. Vervuurt and I. Karatzas. Diversity-weighted portfolios with negative parameter. *Annals of Finance*, 11(3-4):411–432, 2015.
51. T.-K. L. Wong. Optimization of relative arbitrage. *Annals of Finance*, 11(3-4):345–382, 2015.
52. T.-K. L. Wong. Universal portfolios in stochastic portfolio theory. *preprint arXiv:1510.02808*, 2015.
53. T.-K. L. Wong. Logarithmic divergences from optimal transport and Rényi geometry. *arXiv preprint arXiv:1712.03610*, 2017.
54. T.-K. L. Wong. On portfolios generated by optimal transport. *arXiv preprint arXiv:1709.03169*, 2017.