Modeling asset allocation strategies and a new portfolio performance score

Apostolos Chalkis[1,2] and Ioannis Z. Emiris[1,2]

[1]Department of Informatics & Telecommunications, National & Kapodistrian University of Athens, Greece
[2]ATHENA Research & Innovation Center, Greece

Abstract

We discuss a powerful, geometric representation of financial portfolios and stock markets, which identifies the space of portfolios with the points lying in a simplex convex polytope. The ambient space has dimension equal to the number of stocks, or assets. Although our statistical tools are quite general, in this paper we focus on the problem of portfolio scoring. Our contribution is to introduce an original computational framework to model portfolio allocation strategies, which is of independent interest for computational finance. To model asset allocation strategies, we employ log-concave distributions centered on portfolio benchmarks. Our approach addresses the crucial question of evaluating portfolio management, and is relevant to the individual private investors as well as financial organizations. We evaluate the performance of an allocation, in a certain time period, by providing a new portfolio score, based on the aforementioned framework and concepts. In particular, it relies on the expected proportion of actually invested portfolios that it outperforms when a certain set of strategies take place in that time period. We also discuss how this set of strategies –and the knowledge one may have about them– could vary in our framework, and we provide additional versions of our score in order to obtain a more complete picture of its performance. In all cases, we show that the score computations can be performed efficiently. Last but not least, we expect this framework to be useful in portfolio optimization and in automatically identifying extreme phenomena in a stock market.

1 Introduction

Modern finance has been pioneered by Markowitz who set a framework to study choice in portfolio allocation under uncertainty, see [14], and which earned him the Nobel Prize in economics, 1990. Within this framework, Markowitz characterized portfolios by their return and their risk; the latter is formally defined as the variance of the portfolios' returns. An investor would build a portfolio that will maximize its expected return for a chosen level of risk; it has since become common for asset managers to optimize their portfolio within this framework. This approach has led a large part of the empirical finance research to focus on the so-called efficient frontier which is defined as the set of portfolios presenting the lowest risk for a given expected return. Figure 1 presents such an efficient frontier. The efficient frontier is associated with a well-known family of convex functions, studied by Markowitz in [15]. However, building a portfolio in that (or any other) framework does not always guarantees superior performance in practice comparing to other allocation choices. Thus, evaluating the performance of a certain allocation is a challenging task of special interest. We discuss previous work on portfolio scoring and then we present our main contributions.

1.1 Previous work

The fast growth of asset management industry during the past few decades has highlighted the analysis of portfolio allocation performance as an important aspect of modern finance. Research in this area is axed on Sharpe-like ratios proposed in the 1960's [11, 19, 21]. In practice, the performance of a portfolio manager, over a given period, is usually measured as the ratio of his "excess" return with respect to a benchmark portfolio over a risk measure [8]. Managers are then ranked according to these ratios, and the one achieving the highest and steadiest returns receives the best score. The major drawback of these techniques is the identification of benchmark portfolios, while the formation of such portfolios remains controversial. Thus, we assume that the best score corresponds to a "good" portfolio allocation, but without having a universal measure of goodness for this allocation. Moreover, they suffer from significant estimation errors [12], which prevent any performance comparison to be significant.

In [16] -and independently in [9, 1]- they use the geometric representation of a stock market, presented also in this paper, to define a cross-sectional score of a portfolio given a vector of assets' returns. In particular the score of a portfolio is defined as the proportion of allocations that the portfolio outperforms. The aim is to measure the relative performance -in terms of return- of an asset allocation with respect to all possible alternative allocations offered to the manager. The term *cross section* is used to underline that the score takes into account portfolios that are diversified over all sections of assets, without studying -separately- the performance on specific sections of stocks. In [16, 17, 10], the relative performance of value-weighted indices with respect to long-only portfolios is assessed in the Dutch, Spanish and German markets; they considered the MSCI Netherlands 24, IBEX 35, and DAX 30 components, respectively. Interestingly, in [1], they follow the same approach by defining what they call *naive investor's strategy*. A naive investor's strategy selects uniformly a portfolio from the set of all portfolios, as it is agnostic about the assets' returns generating process, and hence does not use any such information.

In [16, Thm 4.2.2] they compute the score by sophisticated geometrical algorithms. However, this computation is not valid when some asset returns are equal and it presents

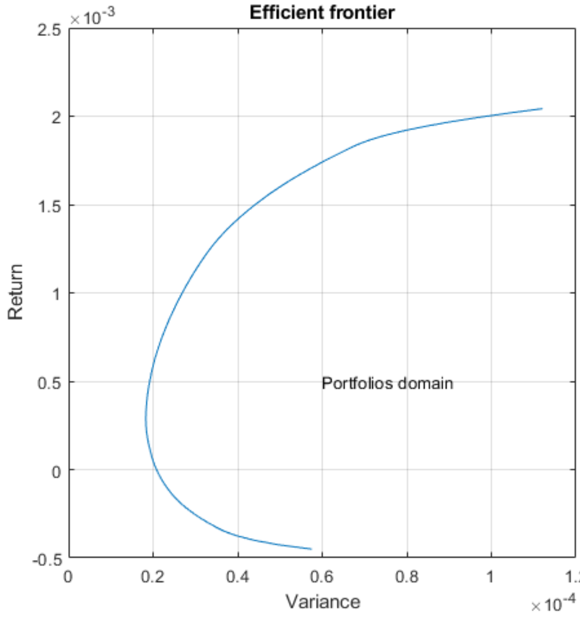

Figure 1: An illustration of the efficient frontier.

floating point errors limiting its use to around 20 assets. As a consequence, in [16] and in related studies [1], the score is estimated by a quasi-Monte Carlo sampling of the portfolios; one may refer to [18] for uniform sampling methods over a simplex of general dimension. Finally, in [2] they show that an algorithm in [22] computes this score very efficiently and robustly (a few milliseconds, in stock markets with thousands of assets). Moreover, in [3] they characterize statistically the distribution of portfolios' returns, where the aforementioned portfolio score corresponds to its Cumulative Density Function (CDF), and they rely on powerful techniques in computational geometry to compute exactly the CDF and Probability Density Function, as well as the moment of portfolios' returns distribution of any order, with several applications [20, 7].

1.2 Contributions.

At first we employ a geometric representation of the set of portfolios in a stock market (Section 2) also appeared in [2, 3]. In particular, we focus on the long-only strategies and thus, we represent the set of portfolios with the canonical simplex, which is a convex polytope. However, the computational framework we provide can be generalized for any convex set. In the sequel, our aim is twofold:
(a) to introduce original models of portfolio allocation strategies, which should be critical for other problems in finance.
(b) to employ the latter framework in evaluating portfolio's performance in a certain time period by introducing a new score.

We introduce a new mathematical model of portfolio allocation strategies in a stock market. It is of independent interest and may be used to address several questions in fintech besides those in Section 5. We consider the concept where portfolio managers compute and propose asset allocations, which we call *formal allocation proposals*. Then, an investor first decides which allocation proposal to select and second how much to modify this proposal to create his final investment / portfolio. Thus, we expect that the portfolios of the investors, that choose the proposal of a certain portfolio manager, will be "concentrated around" that proposal. To model this procedure we employ multivariate distributions. The support of the Probability Density Function (i.e. the subset of $\mathbb{R}^n$ which are not mapped to zero) of each distribution is the set of all portfolios. In particular, we say that a *portfolio allocation strategy* F_π is induced from a distribution π as follows: to create a portfolio with strategy F_π sample a point/portfolio from π. According to the previous observations, the most intuitive choice for π is a unimodal distribution. Then, we call the mode of π *formal allocation proposal* of the allocation strategy F_π.

We focus on Markowitz's framework to leverage log-concave distributions induced by the family of convex functions of Equation (3) in Section 3.1. We discuss how we parameterize the allocation strategies by the level of risk that a certain group of investors select. Similarly, for a given level of risk, we use the variance to parameterize how stick around the formal allocation proposal a subgroup of investors may decide to be. In other words, when we say "the investors that create their portfolio according to strategy F_π" we denote the proportion of the investors, in a certain stock market and time period, that select risk according to the mode of π and they stick around the formal allocation proposal of F_π according to the variance of π. Finally, as in a stock market appear plenty of strategies followed by group of investors, we define the *mixed strategy* induced by a convex combination of distributions, i.e. a mixture distribution.

We evaluate the performance of a portfolio for a given time period and compare the portfolio against a mixed strategy F_π, when a certain set of strategies take place in that time period. Thus, we define the score of a portfolio as the expected number of actually invested portfolios that the first outperforms, when the portfolios have been invested according to the mixed strategy F_π. We provide an efficient algorithm, based on Markov Chain Monte Carlo integration, to estimate the new score within arbitrarily small error ϵ (Section 3). Furthermore, in extreme cases our new score becomes equal to that of [16, 9, 1]. Thus, it can also be seen as a generalization of the latter score.

Lastly, one may have limited knowledge about a certain stock market and how the investors behave in it, or her/his knowledge may vary from a time period to another. We extend our framework to handle these issues (Section 4). We also provide different versions of our score. Each version provides a different information about the portfolio allocation we would like to evaluate.

We expect that the frameworks and the computational tools we present in the sequel can be generalized and used to handle further problems in fintech. For example, they could be combined with various asset-pricing models and methods to predict assets' returns by Machine Learning and AI methods [6]. Additionally, we believe that the new score can be used to define new performance measures and optimal portfolios according to these measures. Finally, despite the fact that in this paper we focus on the long-only strategies, the tools we present can be easily extend to any set of portfolios.

The work presented here is also appears in [4].

Paper structure. The next section presents the geometric representation of portfolios we use. Section 3 introduces our new framework for modeling allocation strategies, and evaluating portfolio performance by defining a new score of a portfolio. In Section 4 we discuss how one can parameterize our framework to have further in-depth study of portfolio performance. In Section 5 we briefly discuss conclusions and future work.

2 Geometric representation of the set of portfolios

In this section we formalize the geometric representation of sets of portfolios with an arbitrary large number of assets n. We handle the case of long-only strategies. Thus, the set of all portfolios becomes a specific convex set.

In particular, let a portfolio x investing in n assets, whose weights are $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. The portfolios in which a long-only asset manager can invest are subject to $\sum_{i=1}^{n} x_i = 1$ and $x_i \geq 0, \forall i$. Thus, the set of portfolios available to this asset manager is the unit $(n-1)-$dimensional canonical simplex, denoted by Δ^{n-1} and defined as

$$\Delta^{n-1} := \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^{n} x_i = 1, \text{ and } x_i \geq 0, \forall i \in \{1, \ldots, n\} \right\} \subset \mathbb{R}^n. \quad (1)$$

The simplex Δ^{n-1} is the smallest convex polytope with nonzero volume in a given dimension. For instance, in the plane any triangle is a simplex, while a triangular pyramid, or tetrahedron, is the simplex in 3d space.

Here the space dimension n represents the number of assets. Each point in the interior of the simplex represents a portfolio since its coordinate vector is a convex combination of the vertex coordinates: if we use all vertices, this combination is unique and is known as barycentric coordinates of the point. The vertices represent portfolios composed entirely of a single asset. This is the most common investment set —of long-only strategies— in practice today, as portfolio managers are typically forbidden from short-selling or leveraging.

3 Modeling allocation strategies

We now discuss an original method for modeling allocation choices and for evaluating portfolio performance by a new portfolio score. We define the new score of a portfolio as the expected value of the proportion of actually invested portfolios that it outperforms, when the portfolios have been built according to, what we call, a *mixed strategy*.

Here, we assume that in a stock market the portfolio managers make allocation proposals and then the investors choose which proposal to follow and how much to modify it before they create their final portfolio. We model allocation strategies in Markowitz' framework using multivariate log-concave distributions with Δ^{n-1} being the support of each Probability Density Function (PDF). A proper choice of log-concave distributions allows us to parameterize a strategy by the level of risk and the level of dispersion around the formal allocation proposal of the strategy. However, the framework presented in this Section allow us to use any unimodal distribution centered at any benchmark portfolio.

Definition 1. *Let π be a unimodal distribution truncated in Δ^{n-1} with PDF $\pi(x)$. Then, a portfolio allocation strategy $F : \pi \rightarrow \Delta^{n-1}$ is said to be induced by the distribution π, and we write F_π. More precisely, F_π is induced by the following state:*

"To build a portfolio with strategy F_π sample a point/portfolio from π".

The mode of π can be seen as the allocation proposal that a portfolio manager has been made. Then, we expect that the invested portfolios of the investors who have chosen that proposal will be concentrated around that proposal/mode as the mass of π implies.

Definition 2. *Let strategy F_π induced by the unimodal distribution π. We call the mode of π formal allocation proposal or formal proposal of the portfolio allocation strategy F_π.*

In the sequel, we assume that in a stock market the set of actually invested portfolios are created by a combination of different strategies used by the investors (mixed strategy). First, we consider a sequence of log-concave distributions $\pi_1, \ldots, \pi_M$ truncated in Δ^{n-1}. Then, each distribution induces a portfolio allocation strategy, i.e. $F_{\pi_1}, \ldots, F_{\pi_M}$. Then, the mixed strategy is induced by a convex combination of π_i, i.e. by a mixture distribution, as the following definition states.

Definition 3. *Let $\pi_1, \ldots, \pi_M$ be a sequence of unimodal distributions, and let the mixture density be $\pi(x) = \sum_{i=1}^M w_i \pi_i(x)$, where $w_i \geq 0$, $\sum_{i=1}^M w_i = 1$. We call F_π the mixed strategy induced by the mixture density π.*

In Definition 3 each weight w_i corresponds to the proportion of investors that build their portfolios according to the allocation strategy F_{π_i}. Thus the vector of weights $w \in \mathbb{R}^M$ implies how the investors in a certain stock market and time period tends to behave. Now we are ready to define the cross-sectional score of an allocation versus a mixed strategy.

Definition 4. *Let a stock market with n assets and F_π a mixed strategy induced by the mixture density π. For given asset returns $R \in \mathbb{R}^n$ over a single period of time, the score of a portfolio, providing a value of return R^*, is*

$$s = \int_{\Delta^{n-1}} g(x)\pi(x)dx, \quad g(x) = \begin{cases} 1. & \text{if } R^T x \leq R^*, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Notice that the Definition 4 can be generalized for any set of portfolios. The value of the integral in Equation (2) corresponds to the expected proportion of portfolios that an allocation outperforms when the portfolios are invested according to the mixed strategy F_π.

3.1 Log-concave distributions in Markowitz' framework

In this Section, we consider the Markowitz' framework and we discuss the selection of a proper log-concave distribution so that we could fix a sequence $\pi_1, \ldots, \pi_M$. In this framework the assets' returns are random variables distributed normally, with mean μ and covariance matrix Σ.

In general, using Markowitz' framework one can define, under certain assumptions, the optimal portfolio $\bar{x}$ as the maximum of a concave function $h(x)$, $x \in \Delta^{n-1}$. Then the log-concave distribution with PDF $\pi(x) \propto e^{\alpha h(x)}$ has its mode equal to $\bar{x}$ and its variance $\sigma^2 = 1/\alpha$. We again call the mode of π *formal allocation proposal* of the induced strategy F_π as we do in Section 3.

Notice that as the variance grows, π converges to the uniform distribution and as the variance diminishes, the mass of π concentrates around the mode of $\pi(x)$. Thus, we use the variance to parameterize the sequence $\pi_i \propto e^{\alpha_i h(x)}$. Small variances correspond to allocation strategies that are used by investors who stick around the formal proposal. Thus, the created portfolios with such a strategy F_π would be highly concentrated around the formal allocation proposal of F_π (or mode of π) as the mass of π implies. Large variances correspond to allocation strategies that are used by investors who may modify the formal proposal a lot. The portfolios created with such a strategy F_π, would be highly

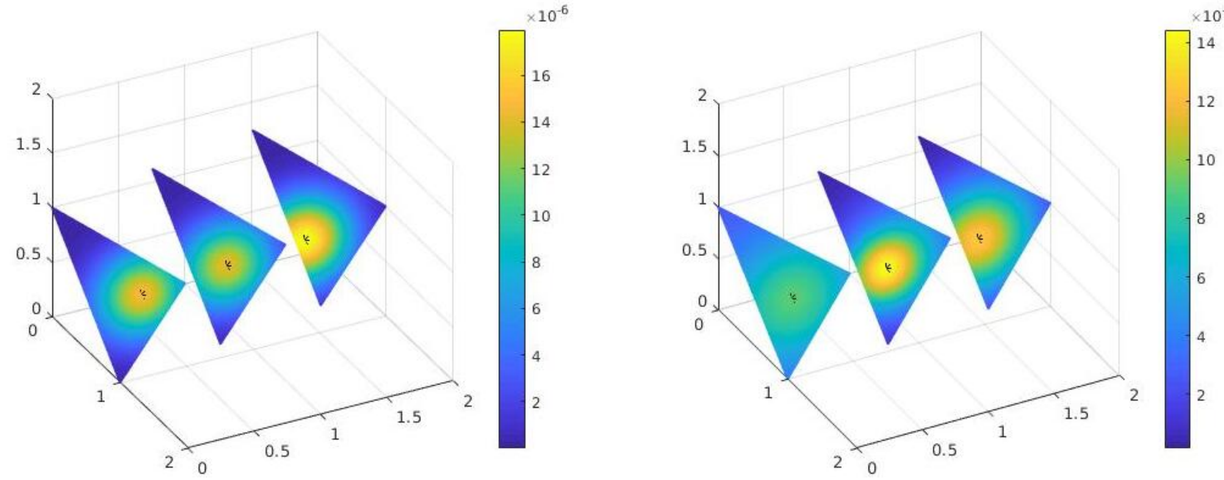

Figure 2: Left: illustration of PDFs $\pi_q \propto e^{-\alpha\phi_q(x)}$, where $\alpha = 1$ and from left to right $q_1 = 0.3,\ q_2 = 1,\ q_3 = 1.5$. Right: 3 illustrations of the mixture density of Equation (5), where $M_1 = 3,\ M_2 = 2$. In both plots the black point corresponds to the formal allocation proposal of each strategy. From yellow to blue: high to low density regions.

dispersed around the mode of π. In the extreme case of very large variance, π is close to the uniform distribution and the induced allocation strategy becomes the naive strategy as defined in [1]. We employ the distance between π_i and the uniform distribution to characterize how dispersed the portfolios created with F_π are, around the formal allocation proposal.

Definition 5. *Let $\pi \propto e^{\alpha h(x)}$ be any log-concave distribution and let F_π be the induced portfolio allocation strategy. We say that F_π is $100(1-D)\%$-dispersed, where D is the distance between π and the uniform distribution, in terms of total variation distance.*

Our main approach is to leverage the family of convex functions which is widely used by investors to compute the efficient frontier (EF). In particular, in Markowitz's framework the assets' returns are assumed to be normally distributed following $\mathcal{N}(\mu, \Sigma)$. Then, the parameterized function

$$\phi_q(x) = x^T\Sigma x - q\mu^T x,\ x \in K, \quad q \in [0, +\infty], \tag{3}$$

where K is the set of portfolios, is used to compute the efficient frontier and optimal portfolios. The $x^T\Sigma x$ is called risk term, the $\mu^T x$ is called return term and the parameter q controls the trade-off between return and risk. To make an efficient portfolio allocation, in modern finance, a portfolio manager typically compute the EF. In particular, the manager selects a value q_0 —which determines the level of risk of his allocation— and then, according to [15] he solves the following optimization problem:

$$\min\ \phi_{q_0}(x) = x^T\Sigma x - q_0\mu^T x,\ \text{ subject to } x \in \Delta^{n-1}.$$

We call the portfolio $\bar{x} = \min_{x\in K} \phi_{q_0}(x)$ as the *optimal mean-variance* portfolio for the risk implied by q_0. Thus, the efficient frontier can be seen as a parametric curve on q (see Figure 1).

Let the log-concave distribution,

$$\pi_{\alpha,q} \propto e^{-\alpha\phi_q(x)}. \tag{4}$$

The left plot in Figure 2 illustrates some examples of the density function $\pi_{\alpha,q}$ where μ and Σ are randomly sampled once. Notice that for different q, the mode (or the formal allocation proposal of the strategy $F_{\pi_{\alpha,q}}$) is shifted.

We can use parameter q to denote the level of risk of a portfolio allocation strategy $F_{\pi_{\alpha,q}}$. Small values of q correspond to low risk strategies whereas large values of q to high risk strategies. Thus a sequence of such densities can be parameterized by both q (risk) and α (dispersion). In particular, a mixed strategy F_π can be induced by the following mixture density:

$$\pi(x) = \sum_{i=1}^{M_1}\sum_{j=1}^{M_2} w_{ij} e^{-a_{ij}\phi_i(x)}, \text{ where } \phi_i = x^T\Sigma x - q_i\mu^T x, \tag{5}$$

where each q_i denotes the level of risk and for each q_i the parameters α_{ij} imply the level of dispersion of $F_{\pi_{ij}}$. Notice that for each level of risk q_i there are M_2 different levels of dispersion that different groups of investors' portfolios may appear around the same formal allocation proposal. The right plot of Figure 2 illustrates some examples of this mixture density.

A definitely important question is how one could set the risk and dispersion parameters q_i, α_{ij} and the weight w_{ij} of each allocation strategy $F_{\pi_{q_i,\alpha_{ij}}}$ in a certain stock market. The issue is that our knowledge about the stock market and the behavior of the investors in it might be weak or vary from a time period to another. In Section 4 we extend our framework to address these issues. We also provide different versions of the score than those given in Section 3. Each version provides a different information about the portfolio allocation we would like to evaluate for given assets returns.

3.2 Computation of the score

This section discusses Markov Chain Monte Carlo (MCMC) integration to guarantee fast and robust approximation within arbitrarily small error for the computation of the score in Section 3. Let the density $\pi(x) = \sum_{i=1}^M w_i\pi_i(x)$ in Equation (2) to be the probability density function of a mixture of log-concave distributions. Furthermore, let the vector of assets' returns $R \in \mathbb{R}^n$, the halfspace $H(R^*) := \{x \in \mathbb{R}^n \mid R^T x \leq R^*\}$ and the indicator function $g(x) = \begin{cases} 1. & \text{if } x \in H(R^*), \\ 0, & \text{otherwise.} \end{cases}$. Then the score of Equation (2) can be written,

$$\begin{aligned} s &= \int_{\Delta^{n-1}} g(x) \sum_{i=1}^M w_i\pi_i(x)dx = \sum_{i=1}^M w_i \int_{\Delta^{n-1}} g(x)\pi_i(x)dx \\ &= \sum_{i=1}^M w_i \int_{\Delta^{n-1}\cap H(R^*)} \pi_i(x)dx = \sum_{i=1}^M w_i \int_S \pi_i(x)dx, \end{aligned} \tag{6}$$

where $S := \Delta^{n-1} \cap H(R^*)$ is the intersection of the canonical simplex with a halfspace.

It is clear that the computation of the score s is reduced to integrate M log-concave functions over a convex set S, i.e. to compute each $\int_S \pi_i(x)dx,\ i = [M]$. For each one of these M integrals we use the algorithm presented in [13] to approximate it within an arbitrarily small error after a polynomial in dimension (number of assets) n number of operations. First, we use an alternative representation of the volume of S, employing a log-concave density $\pi(x)$,

$$\begin{aligned}\mathrm{vol}(S) &= \int_S \pi(x)dx \; \frac{\int_K \pi^{\beta_1}(x)dx}{\int_S \pi(x)dx} \; \frac{\int_S \pi^{\beta_2}(x)dx}{\int_S \pi(x)^{\beta_1}dx} \cdots \frac{\int_S 1dx}{\int_S \pi(x)^{\beta_k}dx} \\ \Rightarrow \int_S \pi(x)dx &= \mathrm{vol}(S) \; \frac{\int_S \pi(x)^{\beta_k}dx}{\int_S 1dx} \cdots \frac{\int_S \pi(x)dx}{\int_S \pi(x)^{\beta_1}dx},\end{aligned} \tag{7}$$

where the sequence $\beta_j,\ j=[k]$ are factors applied on the variance of $\pi(x)$.

Since S is the intersection of a halfspace with the canonical simplex Δ^{n-1} we use Varsi's algorithm to compute the exact value of $\mathrm{vol}(S)$ after n^2 operations at most. Thus, the computation of $\int_S \pi(x)dx$ is reduced to compute k ratios of integrals. This problem seems intractable at first glance. However, for each ratio we have,

$$\begin{aligned}r_j = \frac{\int_S \pi(x)^{\beta_{j-1}}dx}{\int_S \pi(x)^{\beta_j}dx} &= \frac{1}{\int_S \pi(x)^{\beta_j}dx} \int_S \frac{\pi(x)^{\beta_{j-1}}}{\pi(x)^{\beta_j}(x)} \pi(x)^{\beta_j}(x)dx \\ &= \int_S \frac{\pi(x)^{\beta_{j-1}}}{\pi(x)^{\beta_j}} \frac{\pi(x)^{\beta_j}}{\int_S \pi(x)^{\beta_j}dx} dx.\end{aligned} \tag{8}$$

Thus, to estimate r_j we just have to sample N points from the distribution proportional to $\pi(x)^{\beta_j}$ and truncated to S. Then,

$$r_j \approx \frac{1}{N} \sum_{i=1}^{N} \frac{\pi(x_i)^{\beta_{j-1}}}{\pi(x_i)^{\beta_j}} \tag{9}$$

as N grows. The key for an efficient approximation of r_j using Monte Carlo integration is to set $\beta_j,\ \beta_{j+1}$ such that the variance of r_j is as small as possible (ideally a constant) for N as small as possible. To estimate the score in Equation (6) suffices to estimate each $\int_S \pi_i(x)dx,\ i=1,\dots,M$ as the Equation (7) implies. Then the score $s=\sum_{i=1}^M w_i \int_S \pi_i(x)dx$ can be easily derived. The following Lemma provides the total number of operations required to approximate the score s in Equation (2) within arbitrarily small error, employing MCMC integration and the algorithm in [13].

Lemma 6. *Let the density $\pi(x)$ in the Definition 4 be a mixture of M log-concave densities. Then the portfolio score in Equation (2) can be estimated after $O^*(Mn^5)$ operations, where $O^*(\cdot)$ suppresses polylogarithmic factors and dependence on error e.*

Proof. In [13], they prove that the sequence of $\beta_1,\dots,\beta_k$ can be fixed such that the variance of each $r_j,\ j=[k]$ is bounded by a constant. Moreover, $N=O^*(\sqrt{n})$ points per integral ratio r_j and $k=O^*(\sqrt{n})$ ratios in total suffices to approximate each $\int_S \pi_i(x)dx, i=[M]$ within error e. Thus, $O^*(n)$ points suffices to estimate each $\int_S \pi_i(x)dx$.

To sample from each target distribution proportional to $\pi(x)^{\beta_j}$ and truncated to S in [23] they use the Hit-and-Run random walk [23]. This implies a total number of $O^*(n^4)$ arithmetic operations per generated point. Thus the total number of arithmetic operations to estimate the score s is $O^*(Mn^5)$.

□

Considering practical computations, a plenty of random walks for sampling from log-concave densities in high dimensions are implemented in the software package `volesti` [5]. For an extended introduction to geometric random walks we suggest [23].

4 Determine a mixed strategy

In this Section, we discuss how we set the parameters of a sequence of log-concave distributions

$$\pi_{ij} = e^{-a_{ij}\phi_i(x)},\ \text{where}\ \phi_i = x^T\Sigma x - q_i\mu^T x,\ i = [M_1] \text{ and } j = [M_2]$$

which induce a mixed strategy as in Equation (5). Let $q_i \in [0, Q_U],\ Q_U < \infty\ , i = [M_1]$. When $q_i = Q_U$ the term of risk $x^T\Sigma x$ is negligible in $\phi_i(x)$ with respect to the term of return $\mu^T x$. Thus, $q = Q_U$ corresponds to the optimal mean-variance portfolio with highest expected return. We recall that $q = 0$ corresponds to the allocation strategy of zero risk. Let for each q_i, the parameters $\alpha_{L_i} < \alpha_{ij} < \alpha_{U_i},\ j = [M_2]$. The variance $1/\alpha_{L_i}$ corresponds to a $100(1-e)$%-dispersed allocation strategy and the variance $1/\alpha_{U_i}$ corresponds to the log-concave density $\pi_{\alpha_{U_i},q_i}(x)$, whose mass is almost entirely concentrated around the formal allocation proposal of the induced strategy. The bounds on the parameters α_{ij} and q_i can be easily extracted from the observations in [13].

Now we select equidistant values in both intervals above to set the sequences of q_i and α_{ij}. The aim is to represent allocation strategies with various levels of risk and dispersion in a certain stock market. It is clear that as both M_1, M_2 grow, the representativeness of strategies improves.

Set the sequence of q_i and α_{ij}

1. Select M_1 equidistant values $q_1 < \cdots < q_{M_1}$ from $[0, Q_U]$.
2. For each q_i, select M_2 equidistant values $\alpha_{i1} < \cdots < \alpha_{iM_2}$ from $[\alpha_{L_i}, \alpha_{U_i}]$.

The construction of both sequences of q_i and α_{ij} allow to specify the sequence of log-concave distributions $\pi_{ij} = e^{-\alpha_{ij}\phi_{q_i}(x)}$. However, to determine a mixed strategy one has to determine the weights w_{ij} in the corresponding mixture distribution. We recall that each w_{ij} implies the proportion of investors that create their portfolios according the allocation strategy induced by π_{ij}. Setting w_{ij} forms the mixed strategy F_π while the score of Section 3 becomes,

$$s = \sum_{i=1}^{M_1}\sum_{j-1}^{M_2} w_{ij}\int_S \pi_{ij}(x)dx,\ S := \Delta^{n-1}\cap H(R^*), \tag{10}$$

as also denoted by Equation (6) in Section 3.2. However, one may have a weak knowledge on how the investors behave in a certain stock market, in order to determine explicitly the weights w_{ij}. First we allow to set further bounds on w_{ij}. For example, one would provide an upper bound on the proportion of the investors who chose a specific allocation strategy. We allow these degrees of freedom as follows and we additionally provide three different versions of our score.

In particular, let us assume that we estimate the $M = M_1M_2$ integrals of Equation (10) as described in Section 3.2, where M is the number of allocation strategies in a certain stock market. Then, let the M values to form a vector $c \in \mathbb{R}^M$ and also let the corresponding weights w_{ij} in Equation (10) to be given as a vector $w \in \mathbb{R}^M$. Then the score,

$$s = \langle c, w\rangle, \tag{11}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product between two vectors. Given a matrix $A \in \mathbb{R}^{N \times M}$ and a vector $b \in \mathbb{R}^N$ which express N further constraints on the weights (e.g. specify lower, upper bounds or any linear constraint on w_{ij}), let $Q \subset \mathbb{R}^M$ the following feasible region of weights,

$$\begin{aligned} Aw &\leq b \\ w_i &\geq 0 \\ \sum_i^M w_i &= 1 \end{aligned} \tag{12}$$

Notice that if no further constraints are given on the weights, then the feasible region Q is the canonical simplex Δ^{M-1}. Now let us define three new versions of score s. Each new score provides a different information about the allocation we evaluate.

Let the weights $w \in Q$, where $Q \subset \mathbb{R}^M$ the feasible region in Equation (12).

1. **min score**, $s_1 := \min\langle c, w \rangle,$ subject to Q.
2. **max score**, $s_2 := \max\langle c, w \rangle,$ subject to Q.
3. **mean score**, $s_3 := \frac{1}{\mathrm{vol}(Q)} \int_Q \langle c, w \rangle \; dw$.

For the scores s_1 and s_2 one has to solve a linear program for each one of them. The score s_3 requires the computation of an integral which can be computed with MCMC integration employing uniform sampling from Q; otherwise it can be reduced to the computation of the volume of a convex polytope $P \subseteq \mathbb{R}^M$ since $\langle c, w \rangle$ is a linear function of w with the domain being the set Q. For the latter computation there are several randomized approximations algorithms and efficient `C++` software provided by package `volesti` [5].

Let $w_1 \in Q$ such that the min score $s_1 = \langle c, w_1 \rangle$. The weights denoted by the vector w_1 imply the proportions of the investors that follow each allocation strategy such that the portfolio score s takes its minimum value. Similarly, the vector of weights $w_2 \in Q$ such that the max score $s_2 = \langle c, w_2 \rangle$, implies the proportions of the investors that follow each allocation strategy such that the portfolio score s takes its maximum value. Moreover, it is easy to prove that the mean score $s_3 = \langle c, \bar{w} \rangle$, where the vector of weights $\bar{w}$ is the center of mass of Q. For example, if $Q = \Delta^{M-1}$ (i.e. the case where no further constraints are given on the weights) the vector $\bar{w}$ is the equally weighted vector.

However, one may have an additional knowledge on how the investors tend to behave in a certain stock market, i.e. which allocation strategies they tend to prefer. We also allow for these degrees of freedom by providing the notion of *behavioural functions*.

4.1 Behavioural functions

In this Section we assume that we are given a set of functions which represents the knowledge, that one may have, related to which allocation strategies the investors tend to prefer in a certain stock market and time period. We assume that we are given $M_1 + 1$ functions f_q, f_{α_i} with the domain being $[0, Q_U]$ and $[\alpha_{L_i}, \alpha_{U_i}],\ i = [M_1]$ respectively. We call these functions behavioural functions and we use them to create a vector of weights

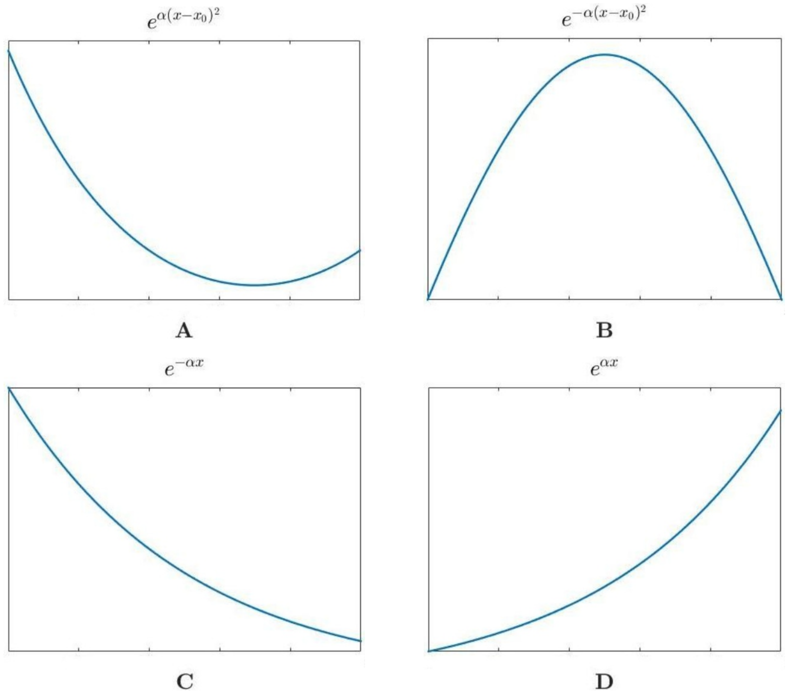

Figure 3: Examples of behavioral functions.

$w \in \mathbb{R}^M$, that emphasizes specific strategies, where $M = M_1 M_2$ the number of allocation strategies that take place in the stock market.

The plots in Figure 3 demonstrate 4 possible choices of such functions. For example, if plot C is f_q then the investors tend to prefer low risk investments; the value of f_q is high for small values of q (low risk) and low for high values of q (high risk). If in addition the plot D is f_{α_i} then the investor tends to be highly sticked around the formal allocation proposal that corresponds to q_i; the value of f_{α_i} is large for large values of α (low dispersion) and small for small values of α (high dispersion). The following pseudo-code describes how we compute such a weight vector w when $M_1 + 1$ behavioural functions are given.

Construct the vector weight w
Input: risk and dispersion parameters q_i and α_{ij}, $i = [M_1],\ j = [M_2]$ computed as in Section 4 and $M_1 + 1$ behavioural functions $f_q,\ f_{\alpha_i}$.

1. For each pair of (i, j) set $r_{(i-1)M_1+j} \leftarrow f_q(q_i) f_{\alpha_i}(\alpha_{ij})$
2. Normalize the vector $r_j \leftarrow r_j / \sum_{i=1}^{M} r_i,\ j = [M]$ and $M = M_1 M_2$
3. Set the weight vector $w \leftarrow r$.

Note that for each q_i we request a behavioural function f_{α_i} to emphasize strategies with level of risk q_i and level of dispersion denoted by f_{α_i}. Given the behavioral functions, one could use the vector of weights —determined as in the above pseudo-code— to compute the portfolio score $s = \langle c, w \rangle$, while c is again the vector that contains the values of the integrals of Equation (10) in Section 4.

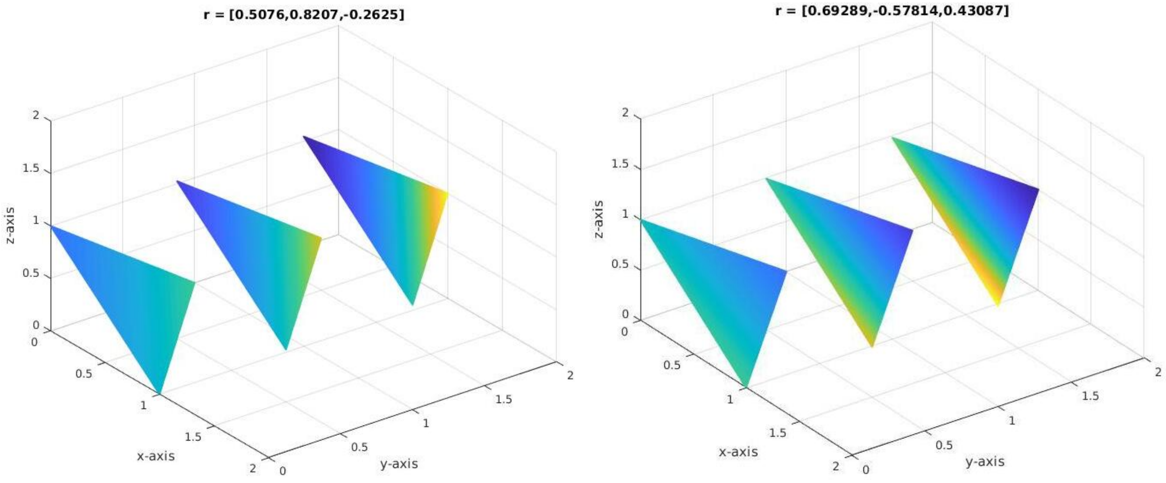

Figure 4: In both plots: Probability density functions $p_{T_i}(w) \propto e^{rw/T_i}$ where (from left to right) $T_1 = 2,\ T_2 = 1,\ T_3 = 2/3$. The bias vector $r \in \mathbb{R}^3$ is given in each title.

4.2 Parametric score

In this Section we allow a weaker knowledge than in Section 4.1 that we might have about how the investors tends to behave. Thus, we do not explicitly determine the vector of weights $w \in \mathbb{R}^M$ —M is the number of allocation strategies in a certain stock market— as in Section 4.1. In particular, let the coordinates of the vector $r \in \mathbb{R}^M$ as in Section 4.1,

$$r_{(i-1)M_1+j} \leftarrow f_q(q_i) f_{\alpha_i}(\alpha_{ij}),\ i = [M_1],\ j = [M_2]$$

where $f_q,\ f_{\alpha_i}$ the $M_1 + 1$ behavioral functions. Then, we use the vector r to denote a *bias* on the behavior of the investors. First, we again allow further bounds and linear constraints on the weights. Thus, we assume —as in Section 4— that the feasible region of the weights is the set Q of Equation (12). To denote the bias on the behavior of the investors we employ the exponential distribution

$$p_T(w) \propto e^{rw/T},\ T > 0,$$

with the support of $p_T(w)$ being the set Q. The distribution $p_T(w) \propto e^{rw/T}$ is usually called Boltzmann distribution and the vector r *bias vector*. In general, Boltzmann distribution gives the probability that a system will be in a certain state as a function of that state's energy and the temperature of the system. The bias vector r determines how the mass tends to distribute in Q and the (temperature) parameter T how strong the bias denoted by r is. The plots in Figure 4 illustrate some examples of the density function of p_T in the simple case of $Q = \Delta^2$ and two different choices of the bias vector r. Notice that the mass tends to concentrate around the vertices which correspond to the coordinates of r with larger values than the other coordinates. Moreover, as the temperature $T \to 0$ this tendency becomes stronger until almost all the mass concentrates around the vertex which corresponds to the coordinate of the largest value of r. As $T \to \infty$, p_T converges to the uniform distribution and the bias denoted by r disappears.

It is clear that our intention is to use the temperature T to parameterize how strong the tendency on the investors' behavior, that the behavioral functions and the bias vector

r imply, is. Then the parametric score is given as,

$$s(T) := \int_S \langle c, w\rangle \; p_T(w)dw, \text{ where } p_T(w) \propto e^{rw/T}, \; T > 0$$
$$\text{and each coordinate} \quad r_{(i-1)M_1+j} = f_q(q_i) f_{\alpha_i}(\alpha_{ij}), \; i = [M_1], \; j = [M_2] \tag{13}$$

Let the center of mass $\bar{w_T}$ in Q when the mass is distributed according to $p_T(w)$. Notice that $\bar{w}_T$ can be seen as a parametric curve on T. Furthermore, it is easy to prove that, for fixed T, the parametric score $s(T) = \langle c, \bar{w}_T\rangle$. Thus, the score $s(T)$ is evaluated on that parametric curve. Following these observations we are ready to state the following Lemma.

Lemma 7. *Let a stock market with M allocation strategies. Assume that we are given the parameters $q_i, \; \alpha_{ij}$ of Section 4 and any behavioral functions $f_q, \; f_{\alpha_i}, \; i = [M_1], \; j = [M_2]$ and $M = M_1M_2$ the number of allocation strategies that take place in the stock market. Let the feasible set $Q \subset \mathbb{R}^M$ of the weights as in Equation (12), the min score s_1, the max score s_2 and the mean score s_3 of Section 4 and the parametric score in Equation (13). Then, the followings hold,*

$$s_1 \leq s(T) \leq s_2, \; \forall T > 0,$$
$$s_3 = \lim_{T\to\infty} s(T) \tag{14}$$

Notice that the Equation (14) holds for any set of behavioral functions. Thus, the scores $s_1, \; s_2$ always bound the parametric score. In particular, for given M allocation strategies, the parametric score when $T \to 0$ is equal to the score s when all the investors select the allocation strategy denoted by the largest coordinate of the bias vector r. Furthermore, when $T \to \infty$ the distribution $p_T(w)$ converges to the uniform distribution over the feasible region of the weights Q and thus the parametric score is equal to the mean score s_3. On the other hand, let the weights $w_1, \; w_2$ that correspond to scores $s_1, \; s_2$ as in Section 4. The $w_1, \; w_2$ imply how the investors have to be distributed among the allocations strategies such that the score s takes the smallest possible and the largest possible value respectively. Thus, they can be seen as lower and upper bound on the parametric score respectively.

5 Future work

A future direction would be to employ the present computational framework for the problem of detecting financial crisis. In particular, we could compute copulae as in [2] but instead of uniform sampling to employ sampling from a mixture distribution as in Equation (5). Moreover, we could introduce parametric copulae following the notion of parametric score in Section 4.2.

We also believe that it would be of special interest to use the new score to define new performance measures and thus, compute the optimal portfolios with respect to those measures. In particular, for a given portfolio one could estimate its score distribution. Then, the problem reduces to compute a portfolio with a “good” score distribution.

From an implementation point of view, the latter two applications require to sample from various log-concave distributions truncated to convex sets and perform MCMC integration multiple times. Thus, new, practical sampling methods leveraging modern random walks (e.g. Hamiltonian Monte Carlo) will be required. We plan to develop the corresponding methods based on package `volesti` [5].

References

[1] A. Banerjee and C.-H. Hung. Informed momentum trading versus uninformed "naive" investors strategies. *J. Banking & Finance*, 35(11):3077–3089, 2011.

[2] L. Calès, A. Chalkis, I. Z. Emiris, and V. Fisikopoulos. Practical volume computation of structured convex bodies, and an application to modeling portfolio dependencies and financial crises. In B. Speckmann and C.D. Tóth, editors, *Proc. Intern. Symp. Computational Geometry (SoCG)*, volume 99 of *Leibniz Intern. Proc. Informatics (LIPIcs)*, pages 19:1–19:15, Dagstuhl, Germany, 2018.

[3] L. Calès, A. Chalkis, and I.Z. Emiris. On the cross-sectional distribution of portfolio returns. Working Papers 2019-11, Joint Research Centre, European Commission (Ispra site), July 2019.

[4] A. Chalkis, M. Christoforou, I. Z. Emiris, and T. Dalamagkas. Modeling asset allocations and a new portfolio performance score. *Submitted for journal publication*, 2020.

[5] A. Chalkis and V. Fisikopoulos. volesti: Volume Approximation and Sampling for Convex Polytopes in R. *arXiv preprint arXiv:2007.01578*, 2020.

[6] E. Christoforou, I.Z. Emiris, and A. Florakis. Neural networks for cryptocurrency evaluation and price fluctuation forecasting. In P. Pardalos, I. Kotsireas, Y. Guo, and W. Knottenbelt, editors, *Mathematical Research for Blockchain Economy*, pages 133–149, Cham, 2020. Springer.

[7] J.B. De Long, A. Shleifer, L.H. Summers, and R.J. Waldmann. The size and incidence of the losses from noise trading. *J. Finance*, 44(3):681–696, 1989.

[8] M. Grinblatt and S. Titman. A Study of Monthly Mutual Fund Returns and Performance Evaluation Techniques. *J. Financial and Quantitative Analysis*, 29(3):419–444, 1994.

[9] D. Guegan, L. Calès, and M. Billio. A Cross-Sectional Score for the Relative Performance of an Allocation. Université Paris1 Panthéon-Sorbonne (Post-Print and Working Papers) halshs-00646070, HAL, 2011.

[10] W. Hallerbach, C. Hundack, I. Pouchkarev, and J. Spronk. A broadband vision of the development of the dax over time. Technical Report ERS-2002-87-F&A, Erasmus University Rotterdam, 2002.

[11] M.C. Jensen. The Performance of Mutual Funds in the Period 1945-1964. SSRN Scholarly Paper ID 244153, Social Science Research Network, Rochester, NY, May 1967.

[12] A.W. Lo. The Statistics of Sharpe Ratios. *Financial Analysts Journal*, 58:36–52, 2002.

[13] L. Lovasz and S. Vempala. Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. In *Proc. IEEE Symposium Foundations of Computer Science (FOCS'06)*, pages 57–68, 2006.

[14] H. Markowitz. Portfolio selection. *J. Finance*, 7(1):77–91, 1952.

[15] H. Markowitz. The optimization of a quadratic function subject to linear constraints. *Naval Research Logistics Quarterly*, 3(1-2):111–133, 1956.

[16] I. Pouchkarev. *Performance evaluation of constrained portfolios*. PhD thesis, Erasmus Research Institute of Management, The Netherlands, 2005.

[17] I. Pouchkarev, J. Spronk, and J. Trinidad. Dynamics of the spanish stock market through a broadband view of the IBEX 35 index. *Estudios Econom. Aplicada*, 22(1):7–21, 2004.

[18] R.Y. Rubinstein and B. Melamed. *Modern simulation and modeling*. Wiley, New York, 1998.

[19] W.F. Sharpe. Mutual Fund Performance. *J. Business*, 39(1):119–138, 1966.

[20] C. Stivers and S. Licheng. Cross-sectional return dispersion and time variation in value and momentum premiums. *J. Financial and Quantitative Analysis*, 45(4):987–1014, 2010.

[21] J. L. Treynor. How to Rate Management of Investment Funds. In *Treynor on Institutional Investing*, pages 69–87. John Wiley & Sons, Ltd, 2015.

[22] G. Varsi. The multidimensional content of the frustum of the simplex. *Pacific J. Math.*, 46:303–314, 1973.

[23] S.S. Vempala. Geometric random walks: a survey. In *Combinatorial and Computational Geometry*, volume 52 of *MSRI Publications*, Berkeley, 2005. MSRI.