Noname manuscript No.
(will be inserted by the editor)

Geometric and statistical tools for financial modeling

Apostolos Chalkis · Emmanouil Christoforou · Ioannis Z. Emiris · Theodore Dalamagkas

Received: date / Accepted: date

Abstract We discuss a powerful, geometric representation of stock markets which identifies the space of portfolios with the points lying in a simplex convex polytope. Based on this viewpoint, we survey certain state-of-the-art tools from geometric and statistical computing in order to handle important and difficult problems in digital finance. Although our tools are quite general, in this paper we focus on two specific problems.

The first question concerns crisis detection, which is of prime interest for the public in general and for policy makers in particular because of the significant impact that crises have on the economy, including employment and income. Certain features in stock markets lead to this type of anomaly detection: Given the assets' returns, we describe the relationship between portfolios' return and volatility by means of a copula (a bivariate probability distribution), without making any assumption on investors' strategies. We examine a recent method relying on copulae to construct an appropriate indicator that allows us to

Apostolos Chalkis
Department of Informatics & Telecommunications National & Kapodistrian University of Athens, Greece,
ATHENA Research & Innovation Center, Greece
E-mail: achalkis@di.uoa.gr

Emmanouil Christoforou
Department of Informatics & Telecommunications National & Kapodistrian University of Athens, Greece,
ATHENA Research & Innovation Center, Greece
E-mail: echristo@di.uoa.gr

Ioannis Z. Emiris
Department of Informatics & Telecommunications National & Kapodistrian University of Athens, Greece,
ATHENA Research & Innovation Center, Greece
E-mail: emiris@di.uoa.gr

Theodore Dalamagkas
ATHENA Research & Innovation Center, Greece
E-mail: dalamag@imis.athena-innovation.gr

automate crisis detection. On real data from DJ600 Europe, from 1990 to 2008, the indicator identifies correctly 4 crises and issues one false positive for which we offer an explanation.

Our second contribution is to introduce an original computational framework to model portfolio allocation strategies, which is of independent interest for Digital finance and its applications. Furthermore, we expect this framework to be useful in automatically identifying extreme phenomena in a stock market. Last but not least, we evaluate portfolio performance, by providing a new portfolio score, based on the aforementioned framework and concepts.

1 Introduction

Modern finance has been pioneered by Markowitz who set a framework to study choice in portfolio allocation under uncertainty, see [Markowitz, 1952][1]. 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 (left panel) presents such an efficient frontier. The region to the right of the efficient frontier represents the portfolios domain.

The efficient frontier is associated with a well-known family of convex functions, studied by Markowitz in [Markowitz, 1956]. In particular, in Markwitz's framework the assets' returns are assumed to be normally distributed following $\mathcal{N}(\mu, \Sigma)$. Then, the parameterized function

$$\phi_q(x) = x^T \Sigma x - q\mu^T x, \ x \in K, \quad q \in [0, +\infty], \tag{1}$$

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 and the $\mu^T x$ is called return term. Typically, a manager selects a value q_0 —which determines the level of risk of his allocation— and then 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.

Interestingly, despite the fact that this framework considers the whole set of portfolios, no attention has been given to the distribution of portfolios. Figure 1 (middle panel) presents such a distribution. When comparing the contour of the empirical portfolios distribution and the portfolio domain bounded by the efficient frontier in Figure 1 (right panel), we observe that the density of portfolios along the efficient frontier is dim and that most of the portfolios are located in a small region of the portfolios domain.

[1] for which he earned the Nobel Prize in economics, 1990.

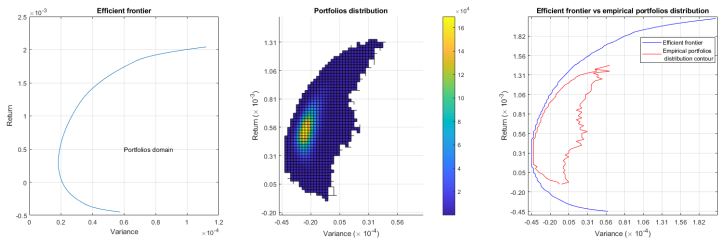

Fig. 1 (left) Efficient frontier, (middle) Empirical portfolio distribution by portfolios' return and variance, (right) Efficient frontier in blue and contour of the empirical portfolio distribution in red. The market considered is made of the 19 sectoral indices of the DJSTOXX 600 Europe. The data is from October 16, 2017 to January 10, 2018.

We also know from the financial literature that financial markets exhibit 3 types of behavior. In normal times, stocks are characterized by slightly positive returns and a moderate volatility, in up-market times (typically bubbles) by high returns and low volatility, and during financial crises by strongly negative returns and high volatility, see [Billio et al., 2012] for details. So, following Markowitz' framework, in normal and up-market times, the stocks and portfolios with the lowest volatility should present the lowest returns, whereas during crises those with the lowest volatility should present the highest returns. These features motivate us to describe the time-varying dependency between portfolios' returns and volatility to detect financial crises in stock markets.

Except from conventional stock markets, these tools can be also used in cryptocurrency markets. For instance, in [Christoforou et al., 2020], they focused on digital assets to offer a neural network expressing a predictor of the asset's "health" based on a variety of parameters ranging from standard financial / economical, to technological (e.g. blockchain), up to software development (e.g. Github) aspects. The tools in this paper can be combined with models that predict asset's return using Machine Learning to improve their results.

Score of a portfolio. Now, let use briefly present existing work on the problem of portfolio scoring. 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 [Jensen, 1967,Sharpe, 1966, Treynor, 2015]. 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 [Grinblatt and Titman, 1994]. 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 [Lo, 2002], which prevent any performance comparison to be significant.

In [Pouchkarev, 2005] -and independently in [Guegan et al., 2011, Banerjee and Hung, 2011]- 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. Interestingly, in [Banerjee and Hung, 2011], 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 particular, they introduce a score as the comparison of the return of an allocation versus the return distribution of naive investors.

In previous works [Pouchkarev, 2005, Banerjee and Hung, 2011, Calès et al., 2018] that discuss efficient computation of that score, the set of all portfolios is usually taken to be the set of long-only strategies which is the most common type of investment, namely portfolios whose weights are non-negative and sum up to one. Thus, the set can be represented with the canonical simplex Δ^{n-1} where n is the number of assets (Section 2).

In [Pouchkarev, 2005, Thm 4.2.2] they compute the score by decomposing the intersection of the simplex with a halfspace into smaller simplices. However, this computation is not valid when some asset returns are equal and it presents floating point errors limiting its use to around 20 assets. As a consequence, in [Pouchkarev, 2005] and in related studies [Banerjee and Hung, 2011], the score is estimated by a quasi-Monte Carlo sampling of the portfolios; one may refer to [Rubinstein and Melamed, 1998] for uniform sampling methods over a simplex of general dimension. Finally, in [Calès et al., 2018] they show that an algorithm in [Varsi, 1973] computes this score very efficiently and robustly (a few milliseconds, in stock markets with thousands of assets). Moreover, in [Calès et al., 2019] 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 (PDF), as well as the moment of portfolios' returns distribution of any order. Overall, the computation of CDF and PDF has found certain applications in asset management, portfolio performance measurement and the study of financial stability, while moments may be useful in return dispersion and noise trading [Stivers and Licheng, 2010, De Long et al., 1989].

Notice that the score in [Pouchkarev, 2005, Guegan et al., 2011, Banerjee and Hung, 2011] does not require any further assumptions on the portfolio allocation strategies that take place in a certain stock market. This observation motivate us to

introduce a new cross-sectional score of a portfolio based on a new framework to model portfolio allocation strategies. This new score would take into account how the truly invested portfolios are distributed in a stock market in a given time period. Then, one could be interested in the number of truly invested portfolios that a certain portfolio has outperformed (Section 4).

Contributions. At first we employ a geometric representation of the set of portfolios in a stock market (Section 2). 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. In the sequel, our aim is (a) to contribute to the problem of crises detection in stock markets. Then, (b) we contribute to the problem of modeling portfolio allocation strategies, which we expect to be useful for (a) and for a few other problems in fintech. Last but not least, (c) we employ the latter framework to evaluate portfolio's performance by introducing a new score.

- For (a) we rely on copula representation to capture the dependency between portfolios' return and volatility. In Section 3.1 we briefly survey the results from [Calès et al., 2018] and we further strengthens them by employing clustering methods on copulae. We call *copula* a discrete approximation of a joint distribution of two continuous random variables, for which the marginal probability distribution of each variable is uniform. Then, for a vector of assets' returns, of a given time interval, we compute the corresponding copula to obtain how the portfolios behave during that period of time. We also build an indicator, which is evaluated on a copula. Following this computational framework on real data from DJ600 we detect all the past crises from 1990 to 2008.
- To address (b), we introduce a new mathematical framework to model portfolio allocation strategies in a stock market. This framework is of independent interest and may be used to address a few other problems in fintech except those presented in this paper (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. Moreover, we use the variance of π to parameterize how dispersed are the portfolios created according to the strategy F_π around the formal allocation proposal (or mode of π).

To be more precise, we focus on Markowitz's framework to leverage log-concave distributions induced by the family of convex functions of Equation (1). In particular, we consider the family of log-concave distributions, with probability density function

$$\pi_{\alpha,q} \propto e^{-\alpha\phi_q},\ \phi_q(x) = x^T\Sigma x - q\mu^T x,$$

where $\sigma^2 = 1/\alpha$ is the variance. Then, we induce the corresponding allocation strategies $F_{\pi_{\alpha,q}}$. We discuss how we use q, which determines the mode of π, to parameterize the strategies by the level of risk that a certain group of investors select. Similarly, for a given q, we use the variance $1/\alpha$ of $\pi_{\alpha,q}$ to parameterize how stick around the formal allocation proposal of $F_{\pi_{\alpha,q}}$ a subgroup of investors may decide to be. In other words, when we say "the investors that create their portfolio according to strategy $F_{\pi_{\alpha,q}}$" we denote the proportion of the investors, in a certain stock market and time period, that select risk according to q and they stick around the formal allocation proposal of $F_{\pi_{\alpha,q}}$ according to α. 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.

– For (c) we evaluate the performance of a portfolio for a given time period we compare the portfolio against a mixed strategy F_π. In particular, we define the score of a portfolio as the expected number of truly 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 cross-sectional score within arbitrarily small error ϵ (Section 4). Furthermore, in extreme cases our new score becomes equal to that of [Pouchkarev, 2005, Guegan et al., 2011,Banerjee and Hung, 2011]. Thus, it can also be seen as a generalization of the latter cross-sectional score.
Last, 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. Thus, we extend our framework to handle these issues (Section 4.3). 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 aforementioned mathematical framework of modeling allocation strategies could be useful to obtain more sophisticated copulae for the problem of detecting financial crises. Thus, it will allow us to try to handle cryptocurrency markets. Additionally, we believe that the new score can be used to define new performance measures and optimal portfolios according to these measures. We leave both interesting directions as a future work. In any case, 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. 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.

Software. The copulae and indicator computation for crises detection (Section 3) as well as the portfolio cross-sectional score computation -by Varsi's algorithm- of [Calès et al., 2019] are provided by the open source package `volesti` [Chalkis and Fisikopoulos, 2020][2]. `volesti` is a `C++` open source library for high dimensional sampling and volume computation with an `R` interface provided through `CRAN`[3]. The implementations scale up to hundreds or thousands dimensions depending on the application, thus radically increasing the number of assets that had been previously studied (e.g. around 20 in [Pouchkarev, 2005, Guegan et al., 2011]), and eventually allow us to capture any stock market today.

Paper structure. The next section presents the geometric representation of portfolios we use. Section 3 surveys our work on copulae and the ensuing crisis indicator; our approach is corroborated with an application on real data. Most results in this section are presented in [Calès et al., 2018], but here we survey a broader class of techniques and frameworks. Section 4 introduces our new framework for modeling allocation strategies, and evaluating portfolio performance by defining a new score of a portfolio. Finally, 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

$$\Delta^{n-1} := \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^{n} x_i = 1, \text{ and } x_i \geq 0, \forall i \in \{1, \ldots, n\} \right\} \subset \mathbb{R}^n. \tag{2}$$

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

[2] `https://github.com/GeomScale/volume_approximation`

[3] `https://CRAN.R-project.org/package=volesti`

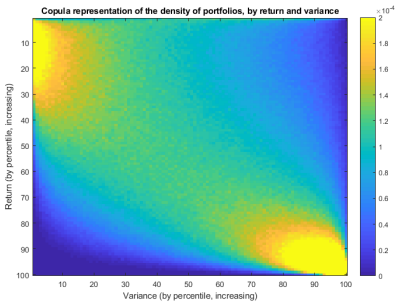

Fig. 2 Copula representation of the portfolios distribution, by return and variance. The market considered is made of the 19 sectoral indices of DJSTOXX 600 Europe. The data is from Oct. 16, 2017 to Jan. 10, 2018. Each line and column sum to 1% of the portfolios.

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 Crises detection

In this section we present our computational methods to address the problem of crises detection in stock markets. The dependency between return and volatility is difficult to capture from the usual mean-variance representation, as in Figure 1 (middle panel), so we will rely on the copula representation of the portfolios distribution. As we follow Markowitz' framework, the variables considered are the portfolios' return and variance. Figure 2 illustrates such a copula and shows a positive dependency between portfolios returns and variances. In the sequel we give definitions for the functions of return f_{ret} and volatility f_{vol} for a given portfolio $x \in \Delta^{n-1}$.

Definition 1 *Given a vector of assets' returns $R \in \mathbb{R}^n$ and the variance-covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ of the distribution that the assets' returns follow, we say that any portfolio $x \in \Delta^{n-1}$ has return $f_{ret}(x, R) = R^T x$ and volatility $f_{vol}(x, \Sigma) = x^T \Sigma x$.*

To capture the relationship between return and volatility we present it in the form of a *copula*. In particular, we discretize the joint distribution between return and volatility to obtain an estimation. Thus, given a vector of assets' returns $R \in \mathbb{R}^n$ and the variance-covariance $\Sigma \in \mathbb{R}^{n\times n}$, we fix two sequences $s_0 < \cdots < s_m$ and $u_0 < \cdots < u_m$ such that

$$\frac{\text{vol}(S_i)}{\text{vol}(\Delta^{n-1})} \approx p \quad \text{and} \quad \frac{\text{vol}(U_i)}{\text{vol}(\Delta^{n-1})} \approx p,\ i = 0, \ldots, m-1, \tag{3}$$

where $S_i := \{x \in \mathbb{R}^n \mid s_i \leq f_{ret}(x, R) \leq s_{i+1}\}$ and $U_i := \{x \in \mathbb{R}^n \mid u_i \leq f_{vol}(x, \Sigma) \leq u_{i+1}\}$ and $p < 1$ a small constant (e.g. $p = 0.01$). Equation (3) implies that a constant percentage p of the portfolios have return less than s_{i+1} and larger than s_i. The same occurs for bodies U_i, which contain portfolios with bounded volatility.

Furthermore, S_i, U_i define a grid of convex bodies, obtained by a family of parallel hyperplanes and a family of concentric ellipsoids intersecting Δ^{n-1}. Precisely, for given integers $i, j \leq m-1$ the body

$$Q_{ij} := \{x \in \Delta^{n-1} \mid s_i \leq f_{ret}(x, R) \leq s_{i+1} \text{ and } u_j \leq f_{vol}(x, \Sigma) \leq u_{j+1}\}, \quad (4)$$

contains the portfolios with return less than s_{i+1} and larger than s_i and volatility less than u_{j+1} and larger than u_j. Now, to obtain the aforementioned copula one has to estimate the ratios $\frac{\text{vol}(Q_{ij})}{\text{vol}(\Delta^{n-1})}$ for $i, j = 0, \ldots, m-1$.

We leverage direct, efficient uniform sampling from Δ^{n-1} following [Rubinstein and Melamed, 1998] to sample N points and then count the number of points in each body in the grid. This is a quasi Monte Carlo method to estimate the volume of an enclosed body in Δ^{n-1}. In Subsection 3.2 this leads to an indicator in order to decide to which state of a market a copula corresponds. Then we use this indicator to detect all past financial crises.

Alternative data-driven formulations of the indicator, in particular by unsupervized learning in the space of copulas, may corroborate these results and offer further insight. In [Calès et al., 2018], these methods (discussed in Section 3) are used to study other dependencies, such as the momentum effect [Jegadeesh and Titman, 1993], which is implied by the dependencies of asset returns with their past returns. The momentum effect is a quite usual market phenomenon by which asset prices follow a trend for a rather long time; it is considered as a market anomaly, which finance theory struggles to explain.

3.1 Computing copulae

In financial applications, one considers compound returns over periods of k observations, where typically $k = 20$ or $k = 60$; the latter corresponds to roughly 3 months when observations are daily. Compound returns are obtained using k observations starting at the i-th one where the j-th coordinate corresponds to asset j. These are data in real space of dimension n, where n is the number of assets with returns $r_i = (r_{i,1}, \ldots, r_{i,n}) \in \mathbb{R}^n$, $i \geq 1$. Therefore, component j of the new vector equals:

$$R_j = (1 + r_{i,j})(1 + r_{i+1,j}) \cdots (1 + r_{i+k-1,j}) - 1, \quad j = 1, \ldots, n.$$

This defines vector $R \in \mathbb{R}^n$ normal to a family of parallel hyperplanes, whose equations are fully defined by selecting appropriate constants.

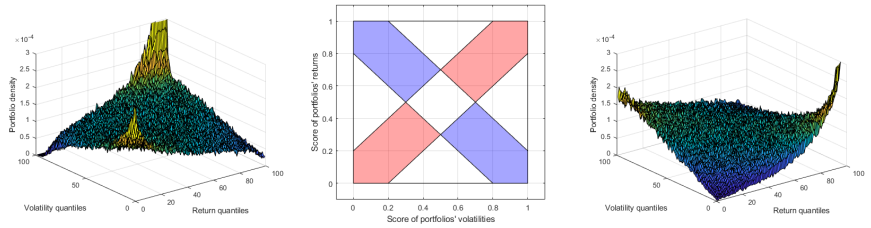

Fig. 3 Returns/variance relationship on 1^{st} Sep. 1999 (left), during the dot-com bubble, and on 1^{st} Sep. 2000 (right), at the beginning of the bubble burst. Blue and yellow indicate low and high density of portfolios, resp. The diagonal bands are considered to specify the indicator (middle).

The covariance matrix Σ of the stock returns is computed using the shrinkage estimator of [Ledoit and Wolf, 2004],[4] as it provides a robust estimate even when the sample size is short with respect to the number of assets.

To compute the copulae, we determine constants defining hyperplanes and ellipsoids so that the volume between two consecutive such objects is $p = 1\%$ of the simplex volume. Let us refer to the method outlined at Equation (3) using notation introduced just before this equation. The sequence of $s_0 < \cdots < s_m$ are determined by bisection using Varsi's algorithm. For ellipsoids, we look for $u_0 < \cdots < u_m$ by sampling the simplex and we set $m = 100$, to compute copulae as an approximation of the joint distribution between return and volatility.

We thus get 100×100 copulae representing the distribution of the portfolios with respect to the portfolio returns and volatilities. Figure 3 illustrates such copulae, and shows the different relationship between returns and volatility in good (left, dot-com bubble) and bad (right, bubble burst) times. In particular, we take into account the 100 components of DJ 600 with longest history, over 60 days ending at the given date. The main computational issue is to compute all the volumes that arise from the intersection of the hyperplane family and of the ellipsoid family, each with simplex Δ^{n-1}. We have implemented sampling uniformly distributed points from Δ^{n-1} using the algorithm in [Rubinstein and Melamed, 1998], which seems to be the method of choice for the problem at hand. Moreover, in [Calès et al., 2018] they have juxtaposed two more algorithms for exact and approximate volume computation for each body in the grid.

We analyze real data consisting of regular interval (e.g. daily) returns of stocks such as the constituents of the Dow Jones Stoxx 600 Europe$^{\text{TM}}$(DJ600). These are points in real space of dimension $n = 600$, respectively: $r_i = (r_{i,1}, \ldots, r_{i,n}) \in \mathbb{R}^n$, $i \geq 1$. We apply the methodology to a subset of assets drawn from the DJ 600 constituents using daily data covering the period from 01/01/1990 to 31/11/2017[5]. Since not all stocks are tracked for the full period

[4] Matlab code at `http://www.econ.uzh.ch/en/people/faculty/wolf/publications.html`.
[5] Our data is from Bloomberg$^{\text{TM}}$.

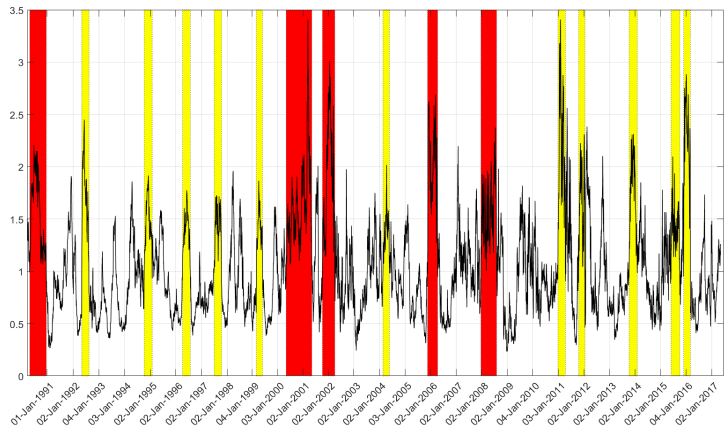

Fig. 4 Representation of the periods over which the indicator is greater than one for 61-100 days (yellow) and over 100 days (red).

of time, we select the 100 assets with the longest history in the index, and juxtapose stock returns and stock returns covariance matrix over the same period to detect crises. Of course, using assets with the longest history implies a survivor bias, but this is used to assess the effectiveness of the methodology. It should soon be computationally feasible to keep all 600 constituents, replacing the exiting stocks with the entering ones along the sample, since our current methods are expected to be efficient for problems of such complexity.

3.2 Indicator and crisis detection

When we work with real data in order to build the indicator, we wish to compare the densities of portfolios along the two diagonals. In normal and up-market times, the portfolios with the lowest volatility present the lowest returns and the mass of portfolios should be on the up-diagonal. During crises, the portfolios with the lowest volatility present the highest returns and the mass of portfolios should be on the down-diagonal, see Figure 3 as illustration. Thus, setting up- and down-diagonal bands, we define the indicator as the ratio of the down-diagonal band over the up-diagonal band, discarding the intersection of the two. The construction of the indicator is illustrated in Figure 3 (middle) where the indicator is the ratio of the mass of portfolios in the blue area over the mass of portfolios in the red one.

In the following, the indicator is computed using copulae estimated using the method in [Rubinstein and Melamed, 1998], drawing $500,000$ points. Computing the indicator over a rolling window of $k = 60$ days and with a band of $\pm 10\%$ with respect to the diagonal, we report with yellow color in Figure 4 all the periods over which the indicator is greater than 1 for more than 60 days. The periods should be more than 60 days to avoid the detection of isolated

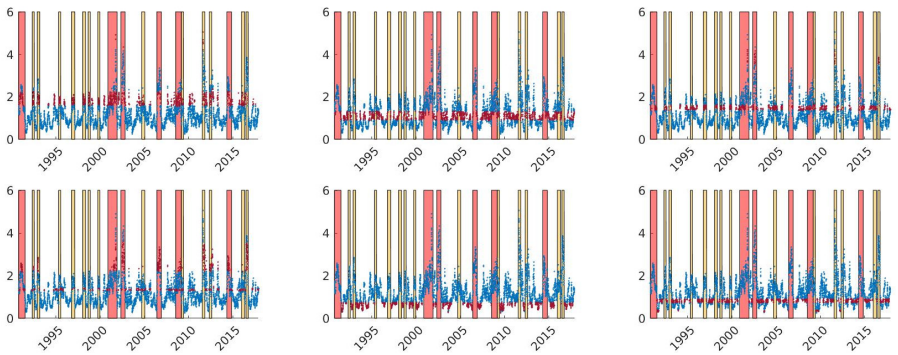

Fig. 5 Spectral clustering of Copulas, with $k = 6$ clusters, on the earth mover's distances (EMD) of the copulas. Results are shown on the values of the indicator for every copula. There are 6 different plots, one for every cluster. Red points indicate the copulas assigned to the specific cluster, while the blue points are the copulas assigned to other clusters. Yellow and red time intervals are the identified by the indicator warning and crises periods respectively.

events whose persistence is only due to the auto-correlation implied by the rolling window. All these periods offer warnings, but only the longest ones correspond to crises. Thus, in Figure 4 we report in red the periods when the indicator was greater than 1 for more than 100 days. For a full discussion and further results we refer the reader to [Calès et al., 2018].

We compare these results with the database for financial crises in European countries proposed in [Duca et al., 2017]. The first crisis (May 1990 to Dec. 1990) corresponds to the early 90's recession, the second one (May 2000 to May 2001) to the dot-com bubble burst, the third one (Oct. 2001 to Apr. 2002) to the stock market downturn of 2002, the fourth one (Nov. 2005 to Apr. 2006) is not listed in the European database and is either a false positive of our method or may be due to a bias in the companies selected in the sample, and the fifth one (Dec. 2007 to Aug. 2008) can be associated with the sub-prime crisis.

3.2.1 Clustering of copulas agrees with indicator

In order to further evaluate our results we applied clustering on the resulting copulas. Aiming to identify whether the copulas are able to distinguish different economic periods (normal, crisis and intermediate), as well as to validate the indicator, we experimented with clustering techniques based on probability distributions distances.

To cluster the probability distributions distances of the copulas, we computed a distance matrix between all the copulas using the earth mover's distance (EMD) [Rubner et al., 2000], a method to evaluate the dissimilarity between two multi-dimensional distributions. The EMD between two distributions is the minimum amount of work required to turn one distribution into the other. Then, we apply spectral clustering [Ng et al., 2001] by converting the distance matrix to affinity. The results of the clustering are shown on the indicators'

values in Figure 5. The clusters appear to contain copulas with similar indicator values. Thus, crisis and normal periods are assigned in clusters with high and low indicator values respectively. Therefore, the clustering of the copulas is proportional to discretising the values of the indicator.

4 Modeling allocation strategies

This Section discusses an original method for modeling allocation choices and for evaluating portfolio performance by a new portfolio score. Notice that the previous analysis is agnostic on allocation strategies by working directly with the set of portfolios Δ^{n-1}. Our current work includes the use of this framework so as to try to obtain more sophisticated copulae by sampling from mixture densities $\pi(x)$ truncated in Δ^{n-1} instead of uniform sampling, in order to identify extreme events in stock markets earlier and with more detail. Moreover, we define a new score of a portfolio, to measure its performance, as the expected value of the proportion of truly 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 here allow us to use any unimodal distribution centered at any benchmark portfolio.

Definition 2 *Let π be a unimodal distribution truncated in Δ^{n-1} with PDF $\pi(x)$. Then, a portfolio allocation strategy $F : \pi \to \Delta^{n-1}$ 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 3 *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 truly 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 4 *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 4 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 5 *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*

$$s = \int_{\Delta^{n-1}} g(x)\pi(x)dx, \quad g(x) = \begin{cases} 1. \ \textit{if } R^T x \leq R^*, \\ 0, \textit{ otherwise.} \end{cases} \tag{5}$$

Notice that the Definition 5 can be generalized for any set of portfolios. The value of the integral in Equation (5) corresponds to the expected proportion of portfolios that an allocation outperforms when the portfolios are invested according to the mixed strategy F_π.

4.1 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 4.

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 dispersed around the mode of π. In the extreme case of very large variance, π is

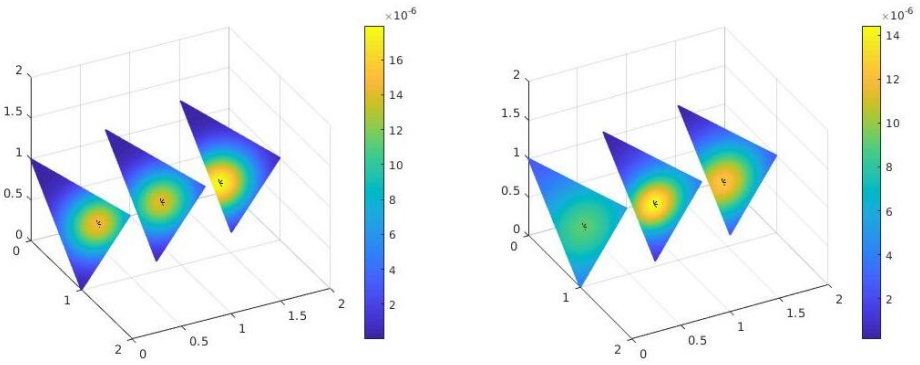

Fig. 6 Left: illustration of PDFs $\pi_q \propto e^{-\alpha\phi_q(x)}$, where $\alpha = 1$ and from left to right $q_1 = 0.3,\ q_2 = 1,\ q_3 = 1.5$. Right: 3 illustrations of the mixture density of Equation (7), where $M_1 = 3,\ M_2 = 2$. In both plots the black point corresponds to the optimal choice of each strategy. From yellow to blue: high to low density regions.

close to the uniform distribution and the induced allocation strategy becomes the naive strategy as defined in [Banerjee and Hung, 2011]. 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 proposal.

Definition 6 *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 convex function which is widely used by investors to compute the efficient frontier (EF). To make an efficient portfolio allocation, in modern finance, a portfolio manager typically compute the EF. In particular, according to [Markowitz, 1956], they solve the following optimization problem:

$$\min\ x^T\Sigma x - q\mu^T x, \quad \text{subject to } x \in \Delta^{n-1},$$

where $q \in [0, +\infty)$. The $x^T\Sigma x$ is called risk term and the $\mu^T x$ is called return term. Parameter q controls the trade-off between risk and return. Thus, the EF is a parametric curve on q (see Figure 1).

Let the log-concave distribution,

$$\pi_{\alpha,q} \propto e^{-\alpha\phi_q(x)}, \text{ where } \phi_q(x) = x^T\Sigma x - q\mu^T x \tag{6}$$

The left plot in Figure 6 illustrates some examples of the density function π_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 an investor's 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:

$$\pi(x) = \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} w_{ij} e^{-a_{ij}\phi_i(x)}, \text{ where } \phi_i = x^T \Sigma x - q_i \mu^T x, \tag{7}$$

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 6 illustrates some examples of this mixture density.

A definitely important question is how one could set the risk and dispersion parameters $q_i,\ \alpha_{ij}$ 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 4.3 we extend our framework to address these issues. We also provide different versions of the score given in Section 4. Each version provides a different information about the portfolio allocation we would like to evaluate for given assets returns.

4.2 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 4. Let the density $\pi(x) = \sum_{i=1}^{M} w_i \pi_i(x)$ in Equation (5) 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 $H(R^*) := \{x \in \mathbb{R}^d \mid R^T x \leq R^*\}$ 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 (5) can be written,

$$\begin{aligned} s &= \int_{\Delta^{n-1}} g(x) \sum_{i=1}^{M} w_i \pi_i(x) dx = \sum_{i=1}^{M} w_i \int_{\Delta^{n-1}} g(x) \pi_i(x) dx \\ &= \sum_{i=1}^{M} w_i \int_{\Delta^{n-1} \cap H(R^*)} \pi_i(x) dx = \sum_{i=1}^{M} w_i \int_S \pi_i(x) dx, \end{aligned} \tag{8}$$

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 s is reduced to integrate M log-concave functions over a convex set S, i.e. to compute each $\int_S \pi_i(x) dx,\ i = 1, \ldots, M$. For each one of these M integrals we use the algorithm presented in [Lovasz and Vempala, 2006] to approximate it within an arbitrarily small error e 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 $\pi(x)$,

$$\begin{aligned} \mathrm{vol}(S) &= \int_S \pi(x)dx \ \frac{\int_K \pi^{\beta_1}(x)dx}{\int_S \pi(x)dx} \ \frac{\int_S \pi^{\beta_2}(x)dx}{\int_S \pi(x)^{\beta_1}dx} \ \cdots \ \frac{\int_S 1dx}{\int_S \pi(x)^{\beta_k}dx} \\ \Rightarrow \int_S \pi(x)dx &= \mathrm{vol}(S) \ \frac{\int_S \pi(x)^{\beta_k}dx}{\int_S 1dx} \ \cdots \ \frac{\int_S \pi(x)dx}{\int_S \pi(x)^{\beta_1}dx}, \end{aligned} \tag{9}$$

where the sequence $\beta_j,\ j = 1, \ldots, k$ 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,

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

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,

$$r_j \approx \frac{1}{N}\sum_{i=1}^{N}\frac{\pi(x_i)^{\beta_{j-1}}}{\pi(x_i)^{\beta_j}} \tag{10}$$

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 (8) suffices to estimate each $\int_S \pi_i(x)dx,\ i = 1, \ldots, M$ as the Equation (9) 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 (5) within error ϵ, employing MCMC integration and the algorithm in [Lovasz and Vempala, 2006].

Lemma 7 *Let the density $\pi(x)$ in the Definition 5 be a mixture of M log-concave densities. Then the portfolio score in Equation (5) can be estimated within error e after $O^*(Mn^5)$ operations, where $O^*(\cdot)$ suppresses polylogarithmic factors and dependence on e.*

Proof In [Lovasz and Vempala, 2006], they prove that the sequence of $\beta_1, \ldots, \beta_k$ can be fixed such that the variance of all $r_j,\ j = 1, \ldots, k$ 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 $\int_S \pi_i(x)dx, i = 1, \ldots, M$ within error e, where $O^*(\cdot)$ suppresses polylogarithmic factors and dependence on 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 [Vempala, 2005] they use Hit-and-Run random walk [Vempala, 2005]. This implies a total number of $O^*(n^4)$ arithmetic operations per generated point. Thus the total number of arithmetic operations to estimate 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` [Chalkis and Fisikopoulos, 2020]. For an extended introduction to geometric random walks we suggest [Vempala, 2005].

4.3 Determine a mixed strategy

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

$$\pi_{ij} = e^{-a_{ij}\phi_i(x)}, \text{ where } \phi_i = x^T\Sigma x - q_i\mu^T x, \ i = 1, \ldots, M_1 \text{ and } j = 1, \ldots, M_2$$

which induce a mixed strategy as in Equation (7). Let $q_i \in [0, Q_U]$, $Q_U < \infty$, $i = 1, \ldots, M_1$. 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 allocation strategy with highest expected return. We recall that $q = 0$ corresponds to the allocation strategy of zero risk. Let for each q_i, the parameters $\alpha_{L_i} < \alpha_{ij} < \alpha_{U_i}$, $j = 1, \ldots, M_2$. 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 [Lovasz and Vempala, 2006].

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 4 becomes,

$$s = \sum_{i=1}^{M_1}\sum_{j-1}^{M_2} w_{ij}\int_S \pi_{ij}(x)dx, \ S := \Delta^{n-1} \cap H(R^*), \tag{11}$$

as also denoted by Equation (8) in Section 4.2. However, one may have weak knowledge on how the investors behave in a certain stock market, to determine explicitly the weights w_{ij}. First we allow to set further bounds on w_{ij}. For example, one would provide a 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 $M = M_1 M_2$ integrals of Equation (11) as described in Section 4.2, 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 (11) to be given as a vector $w \in \mathbb{R}^M$. Then the score,

$$s = \langle c, w \rangle,$$

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 many polynomial in M, randomized approximations algorithms and efficient `C++` software provided by package `volesti` [Chalkis and Fisikopoulos, 2020].

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 implies 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

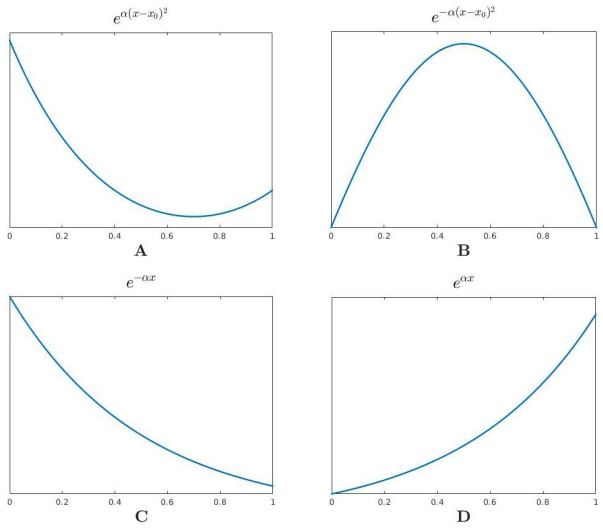

Fig. 7 Examples of behavioral functions.

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 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*.

4.3.1 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 $w \in \mathbb{R}^M$, 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 7 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 values of f_q are 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 values of f_{α_i} are 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 vector weight w
Input: risk and dispersion parameters q_i and α_{ij}, $i = [M_1],\ j = [M_2]$ computed as in Section 4.3 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 (11) in Section 4.3.

4.3.2 Parametric score

In this Section we allow a weaker knowledge that we might have about how the investors tends to behave than in Section 4.3.1. 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 4.3.1. In particular, let the vector of Section 4.3.1 $r \in \mathbb{R}^M$ with coordinates

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

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 4.3— 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

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

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 8 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 diminishes this tendency becomes stronger until almost all the mass concentrates around the vertex which corresponds to the coordinate

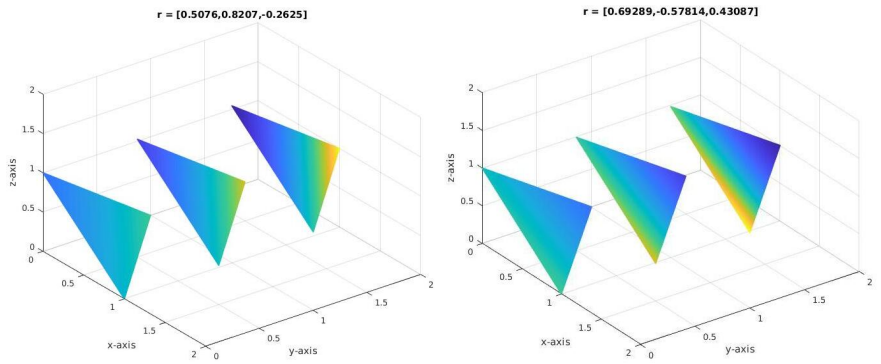

Fig. 8 In both plots: Probability density functions $p_{T_i}(w) \propto e^{rw/T_i}$ where (from left to right) $T_1 = 2,\ T_2 = 1,\ T_3 = 2/3$. The bias vector $r \in \mathbb{R}^3$ is given in each title.

of the largest value of r. As T grows 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 denote, is. Then the parametric score is given as,

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

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 8 *Let a stock market with M allocation strategies. Assume that we are given the parameters $q_i,\ \alpha_{ij}$ of Section 4.3 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 4.3 and the parametric score in Equation (13). Then, the followings hold,*

$$\begin{aligned} s_1 &= \lim_{T\to-\infty} s(T) \\ s_1 &= \lim_{T\to+\infty} s(T) \\ s_3 &= \lim_{T\to 0} s(T) \end{aligned} \tag{14}$$

Notice that the Equation (14) holds for any set of behavioral functions. Thus, the scores $s_1,\ s_2,\ s_3$ can be seen as "extreme" cases of the parametric

score. Therefore, the intuition behind Lemma 8 is that the score s takes the value of $s_1,\ s_2,\ s_3$ when the behavior of the investors is "extreme". These extreme cases are determined by the feasible region of the weights Q. For example, in the simple case where $Q = \Delta^{M-1}$, the extreme cases hold when the investors are equally splitted to all the strategies (when $s = s_3$) or when all of the investors create their portfolios according to a single strategy (when $s = s_1$ or $s = s_2$).

5 Future work

We briefly survey existing work on crises detection and we strengthen its results employing clustering algorithms for bivariate distributions. This problem motivate us to develop a new computational framework to model portfolio allocation strategies in a stock market. A future direction would be to compute copulae as in Section 3.1 but instead of uniform sampling to employ sampling from a mixture distribution as in Equation (7). The latter will allow us to estimate the joint distribution between return and volatility of the truly invested portfolios. Moreover, we could introduce parametric copulae following the notion of parametric score in Section 4.3.2. Detecting crises in Cryptocurrency markets would be a challenging problem.

Furthermore, we 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, we believe that new efficient practical methods based on sampling via state of the art random walks (e.g. Hamiltonian Monte Carlo) will be required. Thus, we leave as future work to develop such practical methods and the corresponding software to handle those applications. Considering software, the package `volesti` is a great starting point.

Acknowledgment

Section 3 of this paper surveys work in collaboration with L. Calès (JRC Ispra, Italy).

References

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

Billio et al., 2012. Billio, M., Getmansky, M., and Pelizzon, L. (2012). Dynamic risk exposures in hedge funds. *Comput. Stat. & Data Analysis*, 56(11):3517–3532.

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

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

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

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

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

Duca et al., 2017. Duca, M. L., Koban, A., Basten, M., Bengtsson, E., Klaus, B., Kusmierczyk, P., Lang, J., Detken, C., and Peltonen, T. (2017). A new database for financial crises in european countries. Technical Report 13, European Central Bank & European Systemic Risk Board, Frankfurt, Germany.

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

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

Jegadeesh and Titman, 1993. Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *J. Finance*, 48:65–91.

Jensen, 1967. Jensen, M. (1967). The Performance of Mutual Funds in the Period 1945-1964. SSRN Scholarly Paper ID 244153, Social Science Research Network, Rochester, NY.

Ledoit and Wolf, 2004. Ledoit, O. and Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *J. Portfolio Management*, 30(4):110–119.

Lo, 2002. Lo, A. (2002). The Statistics of Sharpe Ratios. *Financial Analysts Journal*, 58:36–52.

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

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

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

Ng et al., 2001. Ng, A. Y., Jordan, M. I., and Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, NIPS'01, page 849–856, Cambridge, MA, USA. MIT Press.

Pouchkarev, 2005. Pouchkarev, I. (2005). *Performance evaluation of constrained portfolios*. PhD thesis, Erasmus Research Institute of Management, The Netherlands.

Rubinstein and Melamed, 1998. Rubinstein, R. and Melamed, B. (1998). *Modern simulation and modeling*. Wiley, New York.

Rubner et al., 2000. Rubner, Y., Tomasi, C., and Guibas, L. (2000). The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40:99–121.

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

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

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

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

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