




Optimising portfolio diversification and dimensionality

M. Barkhagen¹ · S. García¹ · J. Gondzio¹ · J. Kalcsics¹ · J. Kroeske² · S. Sabanis^{1,3}  · A. Staal⁴

Received: 22 May 2020 / Accepted: 19 April 2022 / Published online: 2 July 2022
© The Author(s) 2022

Abstract

A new framework for portfolio diversification is introduced which goes beyond the classical mean-variance approach and portfolio allocation strategies such as risk parity. It is based on a novel concept called portfolio dimensionality that connects diversification to the non-Gaussianity of portfolio returns and can typically be defined in terms of the ratio of risk measures which are homogenous functions of equal degree. The latter arises naturally due to our requirement that diversification measures should be leverage invariant. We introduce this new framework and argue the benefits relative to existing measures of diversification in the literature, before addressing the question of optimizing diversification or, equivalently, dimensionality. Maximising portfolio dimensionality leads to highly non-trivial optimization problems with objective functions which are typically non-convex and potentially have multiple local optima. Two complementary global optimization algorithms are thus presented. For problems of moderate size and more akin to asset allocation problems, a deterministic Branch and Bound algorithm is developed, whereas for problems of larger size a stochastic global optimization algorithm based on Gradient Langevin Dynamics is given. We demonstrate analytically and through numerical experiments that the framework reflects the desired properties often discussed in the literature.

Keywords Portfolio diversification · Global optimization · Branch and Bound · Gradient Langevin dynamics

1 Introduction

Diversification as a concept is as old as investing itself, coming into focus in particular during crisis periods. The global financial crisis of 2008, for example, induced heavy losses for most asset portfolios held by institutional investors, prompting practitioners to question their

✉ S. Sabanis
s.sabanis@ed.ac.uk

¹ School of Mathematics, The University of Edinburgh, Edinburgh, UK

² abrdn, Edinburgh, UK

³ The Alan Turing Institute, London, UK

⁴ FTSE Russell, London Stock Exchange Group, London, UK

portfolio construction methodologies and understanding of the level of diversification that can thus be achieved. This led to increased activity in both academia and the financial industry seeking to develop new portfolio construction techniques with the goal of obtaining a well diversified portfolio. Despite the fact that an overwhelming majority of investors seeks to hold a well diversified portfolio, there is still no agreed upon definition or measure of diversification in the literature or among practitioners. A common understanding is that a diversified portfolio should provide risk dissemination and be protected against large drawdowns. Expressed differently, the risk of the portfolio should not be concentrated to only a few risk factors and the tail risk of the portfolio should be controlled. In the diversification literature, the focus was initially purely on volatility reduction, but this definition cannot lead to a useful measure as hedging reduces volatility but does not lead to a better risk dissemination. Most diversification measures and construction methods in the literature are based on the covariance matrix which can be traced back to the seminal paper on mean-variance optimization by Markowitz [55]. Other approaches, with a growing literature used by many asset managers and institutional investors, include the risk parity approach coined by Qian [59] (see also e.g. Qian [60], Roncalli [63] and Roncalli and Weisang [64]) and the most diversified portfolio of Choueifaty and Coignard [21] (see also Choueifaty et al. [22]). Another portfolio diversification measure based on the covariance matrix is introduced in Meucci [58]. There, the asset universe is orthogonalized via Principal Component Analysis of the covariance matrix leading to a new universe consisting of uncorrelated so called principal portfolios. Based on the squared weighted volatilities of the portfolio in the new universe, a portfolio diversification measure based on the dispersion of the squared weighted volatilities is defined. The interpretation of this measure put forth by Meucci is that it represents the effective number of uncorrelated bets that the portfolio is exposed to.

The primary drawback with existing portfolio diversification approaches is that they are based on the distribution of portfolio volatilities or marginal contributions to volatility; however, as far as we are aware, there is not a direct connection from these measures to the distributional or sampling properties of portfolio returns. This in turn leads to a situation where different metrics are shown to behave intuitively in particular cases, while singular counterexamples raise questions about how broadly a technique can be applied. For example, unintuitive behaviour is observed in risk parity portfolios when highly correlated assets are added to the portfolio, leading to over-allocation to them. Unintuitive behaviour of the measure introduced in Meucci [58] led to the introduction of a new technique in Meucci (2013). For this reason we argue, following Fleming and Kroske [31], that it is helpful to augment the covariance based frameworks and connect diversification to additional properties of the distribution of portfolio returns such as higher order moments.

Bringing in higher order moments allows us to move beyond limiting Gaussian assumptions. As a relatively extreme example, consider a two strategy portfolio combining an equity index exposure and a volatility selling strategy on the same index. The volatility of most volatility selling strategies is lower than that of the equity index, while the negative skewness and the kurtosis are more pronounced. A portfolio construction strategy based on volatility would thus put larger weight to the volatility selling strategy in order to decrease volatility risk at the expense of being more exposed to tail risk. Such a portfolio would have suffered heavy losses during the VIX spike on 5 February 2018. On that day the VIX index experienced its largest one-day jump in its 25 year history, rising 20 points from 17.31 from the previous day's close to 37.32 at the end of the trading day. This example highlights why we require our diversification measure to have a direct link to the tail properties of the distribution of portfolio returns. Secondly, we want the measure to be leverage invariant as being 100% exposed to S&P 500 is as diversified as being 50% exposed to S&P 500 and leaving the

rest in cash. Thus, the diversification measure should not be based on portfolio volatility or Expected Shortfall alone. After all, the best strategy to reduce volatility or Expected Shortfall is to have more capital allocated in cash but that does not increase the diversification, it simply represents a reduced exposure to risky assets.

In this paper we introduce a framework that offers a coherent foundation for understanding portfolio diversification by connecting it to the non-Gaussian properties of portfolio returns. The requirement that the diversification measure is leverage invariant naturally leads to measures based on ratios of homogeneous functions of equal degree, such as kurtosis (degree 4) or the square of skewness (degree 6). Even though, for example, the fourth moment and the square of variance are both convex functions, the ratio that yields kurtosis is not necessarily convex. Optimizing ratios of convex functions is a global optimization problem with potentially several local optima which are not equal to the global optimum or optima. We therefore develop two methods, one deterministic and one stochastic, for the global optimization of ratios of convex functions and use them to conduct initial numerical experiments.

2 Non-Gaussianity as a measure of diversification

In this section we present a framework which first connects non-Gaussianity and diversification before introducing the notion of portfolio dimensionality. The main goal of the portfolio diversification framework is to manage the distribution of the portfolio returns. One observes in fact that, the distribution of portfolio volatilities or marginal contributions to volatility across assets, which make up the portfolio, is irrelevant within a mean-variance framework. This is due to the underlying assumption of normally distributed returns. All that matters is portfolio variance. From this simple observation, we argue that meaningful measures of diversification must be related to additional properties of the distribution of portfolio returns. A natural extension of existing frameworks is to connect the concept of diversification to the non-Gaussian properties of the distribution of portfolio returns. Given that the mean-variance framework assumes Gaussianity, diversification can then be seen as an augmentation which relates to model limitations.

2.1 A novel approach to portfolio diversification: dimensionality

In an ideal world, we propose that one could define *portfolio dimensionality* as the number of equally sized independent return streams in the portfolio. This definition is intuitively related to risk dissemination and, arguments based on the Central Limit Theorem (CLT) imply that adding independent exposures to the portfolio leads to a portfolio whose distribution is closer to the Gaussian distribution and thereby the tail risk is reduced. Obviously, financial markets do not obey the idealized assumptions of independent and identically distributed (i.i.d.) returns of the standard CLT (see [10], for some examples of the CLT with relaxed assumptions). The idea behind our diversification measures is to base them on the degree of non-Gaussianity of the portfolio return distribution. A portfolio with a low degree of non-Gaussianity is a well diversified portfolio, and vice versa. Measuring the degree of non-Gaussianity is directly related to the tail properties of the portfolio and naturally leads to measures which are leverage invariant. Measuring and optimizing non-Gaussianity have been thoroughly studied in the Independent Component Analysis (ICA) literature, see e.g. Hyvärinen and Oja [44]. A common measure of non-Gaussianity in the ICA literature is kurtosis, and other frequently used measures are based on neg-entropy or Kullback–Leibler

divergence. Inspired by the ICA literature, we initially link the notion of a well diversified portfolio to a portfolio with a low kurtosis which implies a low (symmetric) tail risk. Other attractive aspects of using kurtosis are that we see it as a natural extension of a symmetric risk framework and it is also related to the distribution of sample variance. In particular, it is known that the variance of the distribution of sample variance is positively related to kurtosis (see e.g. [75]). Reducing kurtosis therefore increases confidence in estimates of portfolio variance.

That said, asymmetry in the form of skewness is also of interest to investors where empirical results from the risk premia literature (Lempérière et al. [53]) show that maximising the Sharpe ratio of a portfolio is strongly linked to maximizing the negative skewness of portfolio returns. There are several ways to incorporate skewness into a portfolio diversification framework. In Lassance and Vrins [51], a portfolio risk measure based on exponential Rényi entropy is used in order to incorporate higher order moments into the portfolio decision framework. Through a truncated Gram–Charlier expansion of Rényi entropy they demonstrate that their portfolio risk measure can be directly expressed as a function of portfolio skewness and kurtosis. Another approach, see e.g. Jondeau and Rockinger [46], relies on a higher order Taylor expansion of the investors utility function, which leads to an expression in terms of the non-standardized portfolio moments. This latter approach suffers from the drawback of optimizing an objective function which is not invariant to leverage. In the following, we offer a general framework which allows us to look at various measures including skewness and kurtosis but for the purposes of these initial numerical investigations we focus on kurtosis.

2.2 From non-Gaussianity to dimensionality: definition and examples

With all the above in mind, we proceed with defining a diversification framework which is invariant to leverage and directly linked to the tail properties of the distribution of portfolio returns. It is also flexible enough to allow different objective functions, such as excess kurtosis and the square of skewness (along with any suitable linear or polynomial combination), but within a robust setting for measuring, in an appropriate sense, the level of non-Gaussianity of resulting portfolios. Furthermore, in order to have an intuitive interpretation of diversification we link it with the tail risk of an equally weighted reference portfolio of i.i.d. reference assets. This reference portfolio is representative of the given asset universe and we proceed to define the notion of portfolio dimensionality relative to the tail risk of the reference asset.

Definition 1 Let p be a positive integer and \mathcal{L}^p denote the set of all random variables with finite p -th (absolute) moment. Let also \mathcal{X} be a convex subset of \mathcal{L}^p and \mathcal{N} denote the set of all Gaussian random variables. We define a function $\nu : \mathcal{L}^p \mapsto \mathbb{R}_+$ that satisfies the following properties:

- (i) $\nu(tX) = \nu(X)$, for any $t > 0$ and $X \in \mathcal{L}^p$. (*leverage invariance*)
- (ii) Let $Y \in \mathcal{X}$ and, moreover, let $Y_1, Y_2, \dots \in \mathcal{X}$ and independently follow $\text{Law}(Y)$. The function

$$\phi_{Y,\nu}(n) = \nu\left(\sum_{i=1}^n Y_i\right) \quad (1)$$

is strictly decreasing in $n \in \mathbb{N}$ for any $Y \in \mathcal{X}$. (*strict monotonicity for i.i.d. data*)

- (iii) $\nu(X) \geq 0$ for every $X \in \mathcal{X} \cup \mathcal{N}$ with equality holding only if $X \in \mathcal{N}$. (*positivity*)
- (iv) $\nu(X) = \nu(Y)$ for any $X, Y \in \mathcal{L}^p$ such that $\text{Law}(X) = \text{Law}(Y)$. (*law invariance*)

(v) For constant $c \in \mathbb{R}$, $v(X + c) = v(X)$ for every $X \in \mathcal{L}^p$. (cash invariance)

Remark 2.1 One notes that the functional v , which sometimes is referred to as a risk measure without necessarily adhering to the formal definition of a risk measure as in Artzner et al. [4] but rather follow the generic description in asset management, is only required to be nonnegative over $\mathcal{X} \cup \mathcal{N}$. Moreover, v is used henceforth to measure the level of non-Gaussianity of resulting portfolios. See also the kurtosis example below along with the related discussion.

Then, one proceeds by defining a measure of diversification relative to a reference random variable $Z \in \mathcal{X}$. First, let us recall the definition for the $(n - 1)$ -dimensional probability simplex, which is given by $\mathcal{W}_n = \{\mathbf{w} \in \mathbb{R}^n \mid \sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n\}$ where $\mathbf{w} = [w_1, \dots, w_n]^T$.

Definition 2 Let $Z, X_1, \dots, X_n \in \mathcal{X}$ and v satisfy the properties (i)–(v) from Definition 1. Let also $v(\sum_{i=1}^n w_i X_i)$ be a continuous function in $\mathbf{w} \in \mathcal{W}_n$. Then, the function $D_{Z, v} : \mathcal{W}_n \mapsto \mathbb{R}_+$ is defined by

$$\text{(Diversification measure)} : \quad D_{Z, v}(\mathbf{w}) = \frac{v(Z)}{v(\sum_{i=1}^n w_i X_i)}. \tag{2}$$

Remark 2.2 In the context of asset management, the above setting for the diversification measure is understood as $X_i \in \mathcal{X}$ being the return of the i -th asset of the portfolio and w_i is the corresponding portfolio weight.

Definition 3 Let \mathcal{X} be a convex subset of \mathcal{L}^p . Let also $\{Z_i\}_{i \geq 1}$ be a sequence of i.i.d. random variables such that each $Z_i \in \mathcal{X}$ and let $Z \in \mathcal{X}$. Moreover, suppose that $\text{Law}(Z_1) = \text{Law}(Z)$ and that v satisfies the properties (i)–(v) from Definition 1. We define $\hat{\eta} : \mathbb{N} \setminus \{0\} \mapsto \mathbb{R}_+$ by

$$\hat{\eta}(k) = \frac{v(Z)}{v\left(\frac{1}{k} \sum_{i=1}^k Z_i\right)}, \text{ for any } k \in \mathbb{N} \setminus \{0\}.$$

We further define the function $\eta : \mathbb{R}_+ \mapsto \mathbb{R}_+$ as the continuous, monotonic (linear) interpolation of $\hat{\eta}$ on \mathbb{R}_+ .

Recall here that given a collection of i.i.d. random variables $\{Z_i\}_{1 \leq i \leq n}$ and Z , which belong to \mathcal{X} , such that $\text{Law}(Z_1) = \text{Law}(Z)$, $\hat{\eta}(k)$ represents the diversification measure of an equally weighted portfolio in those Z_i 's. Moreover, due to the leverage invariance of v and the strict monotonicity of the function $\phi_{Z, v}$, see (1), one guarantees the existence of the function $\eta : \mathbb{N} \setminus \{0\} \mapsto \mathbb{R}_+$, which is also a strictly increasing function. Hence, the definition of portfolio dimensionality follows naturally by considering a suitable transformation, which involves the inverse of the function η .

Definition 4 Let $D_{Z, v}(\mathbf{w})$ be described by Definition 2 and the function η by Definition 3. We define the function $d_{Z, v} : \mathcal{W}_n \mapsto \mathbb{R}_+$ by

$$\text{(Portfolio dimensionality)} : \quad d_{Z, v}(\mathbf{w}) = \eta^{-1}(D_{Z, v}(\mathbf{w})). \tag{3}$$

Remark 2.3 One observes that, for any $k \leq n$,

$$D_{Z, v}(\mathbf{w}) = \frac{v(Z)}{v\left(\frac{1}{k} \sum_{i=1}^k Z_i\right)} \frac{v\left(\sum_{i=1}^k Z_i\right)}{v\left(\sum_{i=1}^n w_i X_i\right)} = \eta(k) \frac{v\left(\sum_{i=1}^k Z_i\right)}{v\left(\sum_{i=1}^n w_i X_i\right)}. \tag{4}$$

Thus,

$$v \left(\sum_{i=1}^n w_i X_i \right) = v \left(\frac{1}{k} \sum_{i=1}^k Z_i \right) \Rightarrow d_{Z, v}(\mathbf{w}) = k,$$

since the denominator in (4) is equal to $\phi_{Z, v}(k)$, which implies that $D_{Z, v}(\mathbf{w}) = \eta(k)$. Thus, we see that the portfolio dimensionality is exactly the number of independent return streams.

Remark 2.4 In general, a judicious selection of the reference asset as representative of the investment universe will produce values of $D_{Z, v} > 1$ as we achieve some relative diversification benefit; however, we note that $D_{Z, v} < 1$ is also possible if we worsen the relative non-Gaussianity.

Let us concentrate now on the case where v is either excess kurtosis or the square of skewness (as this allows us to consider other suitable linear or polynomial combinations of interest using these two risk measures). Using the leverage invariance property of v and taking into account the findings of Fleming and Kroeske [31], where the notion of the distribution of portfolio variance is used and the effective size of its support is related to the spectrum of Rényi entropies, one identifies η with the identity function and $d_{Z, v}(\mathbf{w}) = D_{Z, v}(\mathbf{w})$. For completeness, we present here the relevant derivations, although they can be found in Fleming and Kroeske [31], albeit with different notation.

Proposition 2.5 *Let \mathcal{X} be a convex subset of \mathcal{L}^P . Then*

- (a) *For $p \geq 3$, the square of skewness, i.e. $v(\cdot) = \left(\frac{\mathbb{E}[(\cdot - \mathbb{E}[\cdot])^3]}{(\mathbb{E}[(\cdot - \mathbb{E}[\cdot])^2])^{3/2}} \right)^2$, satisfies properties (i)–(v) from Definition 1, provided that $v(X) > 0$ for every $X \in \mathcal{X}$. Moreover, $\hat{\eta}(k) = k$ for any $k \in \mathbb{N}$.*
- (b) *For $p \geq 4$, the excess kurtosis, i.e. $v(\cdot) = \frac{\mathbb{E}[(\cdot - \mathbb{E}[\cdot])^4]}{(\mathbb{E}[(\cdot - \mathbb{E}[\cdot])^2])^2} - 3$, satisfies properties (i)–(v) from Definition 1, provided that $v(X) > 0$ for every $X \in \mathcal{X}$. Moreover, $\hat{\eta}(k) = k$ for any $k \in \mathbb{N}$.*

Proof We start the proof by considering an arbitrary sequence of i.i.d. random variables $\{Y_i\}_{i \geq 1}$ such that each $Y_i \in \mathcal{X}$.

(a) One observes that properties (i), (iii), (iv) and (v) are satisfied trivially due to the definition of the square of skewness, the linearity of expectation and the assumption that $v(X) > 0$ for every $X \in \mathcal{X}$. It remains to show property (ii). To this end, one considers an arbitrary convex combination of n i.i.d. random variables, say $\sum_{i=1}^n w_i Y_i$, where $w_i \in [0, 1]$ and $\sum_{i=1}^n w_i = 1$. Further and without loss of generality, it is assumed that the Y_i 's have mean zero (as the central second and third moments are considered in the definition of skewness). One further defines $p_i := w_i^2 / \sum_i w_i^2$, the vector $\mathbf{p}_n := [p_1, \dots, p_n]^\top$ in \mathbb{R}^n and the function

$$D_{3/2}(\mathbf{p}_n) := \left(\sum_{i=1}^n p_i^{3/2} \right)^{-2}, \text{ with } D_{3/2}(\mathbf{p}_n) = n \text{ when } w_1 = \dots = w_n = 1/n.$$

One then calculates, due to the i.i.d. nature and zero mean property of the random variables under consideration,

$$\begin{aligned} v\left(\sum_{i=1}^n w_i Y_i\right) &= \left(\frac{\sum_{i=1}^n w_i^3 \mathbb{E}[Y_1^3]}{(\sum_{i=1}^n w_i^2 \mathbb{E}[Y_1^2])^{3/2}}\right)^2 = v(Y_1) \left(\frac{\sum_{i=1}^n w_i^3}{(\sum_{i=1}^n w_i^2)^{3/2}}\right)^2 \\ &= v(Y_1) \left(\sum_{i=1}^n p_i^{3/2}\right)^2 = \frac{v(Y_1)}{D_{3/2}(\mathbf{p}_n)}. \end{aligned}$$

Thus, due to property (i), i.e. leverage invariance and the fact that $D_{3/2}(\mathbf{p}_n) = n$ when $w_1 = \dots = w_n = 1/n$, one obtains

$$v\left(\sum_{i=1}^n Y_i\right) = v\left(\sum_{i=1}^n \frac{1}{n} Y_i\right) = \frac{v(Y_1)}{n},$$

which is strictly decreasing in n . Moreover, for any $Y \in \mathcal{X}$ such that $\text{Law}(Y_1) = \text{Law}(Y)$,

$$\hat{\eta}(k) = \frac{v(Y)}{v\left(\frac{1}{k} \sum_{i=1}^k Y_i\right)} = k \frac{v(Y)}{v(Y_1)} = k,$$

and thus the desired result is obtained.

(b) Following the approach from (a), one observes that again the properties (i), (iii), (iv) and (v) are satisfied trivially due to the definition of excess kurtosis, the linearity of expectation and the assumption that $v(X) > 0$ for every $X \in \mathcal{X}$. Also, in a similar approach one considers an arbitrary convex combination of n i.i.d. random variables, which is denoted again by $\sum_{i=1}^n w_i Y_i$. One, however, now considers the function

$$D_2(\mathbf{p}_n) := \left(\sum_{i=1}^n p_i^2\right)^{-1}, \text{ with } D_2(\mathbf{p}_n) = n \text{ when } w_1 = \dots = w_n = 1/n.$$

Consequently, one calculates by taking into consideration the i.i.d. nature and zero mean property of the random variables under consideration

$$\begin{aligned} v\left(\sum_{i=1}^n w_i Y_i\right) &= \frac{\sum_{i=1}^n w_i^4 \mathbb{E}[Y_1^4] + 3 \sum_{i \neq j} w_i^2 w_j^2 (\mathbb{E}[Y_1^2])^2}{(\sum_{i=1}^n w_i^2 \mathbb{E}[Y_1^2])^2} - 3 \\ &= \left(\frac{\mathbb{E}[Y_1^4]}{(\mathbb{E}[Y_1^2])^2} \frac{\sum_{i=1}^n w_i^4}{(\sum_{i=1}^n w_i^2)^2} + 3 \sum_{i \neq j} \frac{w_i^2}{\sum_{i=1}^n w_i^2} \frac{w_j^2}{\sum_{i=1}^n w_i^2}\right) - 3 \\ &= \left((v(Y_1) + 3) \sum_{i=1}^n p_i^2 + 3 \sum_{i \neq j} p_i p_j\right) - 3 \\ &= \left((v(Y_1) + 3) \sum_{i=1}^n p_i^2 + 3(1 - \sum_{i=1}^n p_i^2)\right) - 3. \end{aligned}$$

The last equality is due to the fact that $\sum_{i=1}^n p_i = 1$. Thus,

$$v\left(\sum_{i=1}^n Y_i\right) = v\left(\sum_{i=1}^n \frac{1}{n} Y_i\right) = \frac{v(Y_1)}{D_2(\mathbf{p}_n)} = \frac{v(Y_1)}{n},$$

which is strictly decreasing in n and hence the desired result is also obtained in this case as in part (a). □

Corollary 2.6 *Let \mathcal{X} be a convex subset of \mathcal{L}^4 and v_1 and v_2 are used to denote excess kurtosis and squared skewness respectively, i.e.*

$$v_1(\cdot) = \frac{\mathbb{E}[(\cdot - \mathbb{E}[\cdot])^4]}{(\mathbb{E}[(\cdot - \mathbb{E}[\cdot])^2])^2} - 3 \quad \text{and} \quad v_2(\cdot) = \left(\frac{\mathbb{E}[(\cdot - \mathbb{E}[\cdot])^3]}{(\mathbb{E}[(\cdot - \mathbb{E}[\cdot])^2])^{3/2}} \right)^2.$$

Then

- (a) *Any linear combination with positive coefficients of excess kurtosis and squared skewness, i.e.*

$$v(\cdot) = \lambda_1 v_1(\cdot) + \lambda_2 v_2(\cdot), \text{ where } \lambda_1, \lambda_2 \geq 0,$$

satisfies properties (i)–(v) from Definition 1, provided that $v(X) > 0$ for every $X \in \mathcal{X}$.

- (b) *The following polynomial combinations with positive coefficients of excess kurtosis and squared skewness (which includes the Jarque–Bera goodness-of-fit test), namely*

$$v(\cdot) = \lambda_1 v_1^2(\cdot) + \lambda_2 v_2(\cdot), \text{ where } \lambda_1, \lambda_2 > 0,$$

satisfy properties (i)–(v) from Definition 1, provided that $v(X) > 0$ for every $X \in \mathcal{X}$.

Thus, we see that our newly proposed diversification measure can be built by using excess kurtosis and/or squared skewness and that our definition of portfolio dimensionality is satisfied for large and rich sets of random variables. Moreover, since $\phi_{Z,v}$ is a monotonically decreasing function, if

$$\phi_{Z,v}(k + 1) < v \left(\sum_{i=1}^n w_i X_i \right) < \phi_{Z,v}(k), \text{ then } k < D_{Z,v}(\mathbf{w}) < k + 1,$$

with $D_{Z,v}(\mathbf{w})$ taking a non-integer value according to a monotonic interpolation of $\phi_{Z,v}$. As a result, one observes that the higher the values for $D_{Z,v}(\mathbf{w})$, the closer we are to a tail risk similar to the one given by a standard Gaussian. To see this, consider a large enough k and $D_{Z,v}(\mathbf{w}) \geq k$. Consequently, one obtains due to (4), that the number of independent assets is increased accordingly,

$$v \left(\sum_{i=1}^n w_i X_i \right) \leq \phi_{Z,v}(k) = v \left(\frac{\sum_{i=1}^k Z_i}{\sqrt{k\mathbb{E}[Z^2]}} \right), \quad (\text{due to leverage invariance})$$

and thus due to the CLT and property (ii) from Definition 1, one observes the desired result.

Remark 2.7 We focus henceforth on the case of random variables with skewed and leptokurtic distributions (the latter implies positive excess kurtosis). Such random variables represent the asset returns in a given asset universe under consideration. This is chosen in agreement with the relevant literature in quantitative finance and assumes a well-documented phenomenon of asset returns, namely stylized facts which include negative skewness and excess kurtosis due to fat tails, see e.g. Cont [25] and references therein. Thus, \mathcal{X} represents, henceforth, the convex hull of a fixed number of such random variables which are used to form a portfolio under consideration. One notes however that \mathcal{X} cannot always be a subset of random variables with leptokurtic distributions, which implies that excess kurtosis may become negative. For example, in the theoretical case, where two perfectly negatively correlated assets, say $X \in \mathcal{X}$

and $-X \in \mathcal{X}$, are used to form a portfolio, then the convex combination $1/2X + 1/2(-X)$ has such property. Similarly, there may be (very unlikely to happen in real-world) situations where the assets, which form a portfolio under consideration, exhibit very strong negative correlation, which may result in convex combinations of them having (even marginally) negative excess kurtosis. In all such cases, if the measure of non-Gaussianity is simply excess kurtosis (as in our examples below), then one proceeds with the minimization of excess kurtosis, knowing very well that a negative sign implies tails lighter than those of the Gaussian distribution (which is desirable from the point of view of asset management) but no diversification or portfolio dimensionality number can be produced.

2.3 Desirable properties of the diversification measure: Toy example

Despite the fact that there is no agreed upon definition of diversification in the literature, a number of desirable properties of a diversification methodology have been proposed. In Choueifaty et al. [22], the notion of *polico invariance* is introduced. Extending an asset universe by adding a positive linear combination of assets already belonging to the universe should not affect the weights to the original assets when applying the diversification methodology. A special case of polico invariance, denoted duplication invariance, considers the duplication of one of the assets in the universe. This case naturally arises in applications when one of the assets is listed on multiple exchanges. Applying the diversification methodology should produce the same portfolio irrespective of any asset in the universe being duplicated or not. In Koumou [49] further desirable properties of diversification measures are introduced. However, some of the properties presented in Koumou [49] are not consistent with the requirements that we have on a diversification measure. In Sect. 1, we introduced the requirement that the portfolio diversification measure should be leverage invariant. This contrasts one of the desired properties presented in Koumou [49]. Furthermore, in Koumou [49], the portfolio diversification measure is required to be concave or quasi-concave. As we have argued, a leverage invariant diversification measure typically leads to a ratio of two convex functions which in general is neither concave nor quasi-concave.

In the following, a numerical example is used to demonstrate that important desirable properties are satisfied by the newly introduced portfolio diversification measure. The demonstration is based on a toy example with a universe consisting of three assets with the following covariance matrix

$$\mathbf{C} = \begin{bmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (5)$$

As the correlation ρ between asset one and asset two approaches one, these two assets behave as one asset and hence this corresponds to the case when one of the assets in the universe is duplicated. For this case, the weight of asset three should approach $\frac{1}{2}$ as $\rho \rightarrow 1$. When $\rho \rightarrow -1$, this corresponds to the case when either asset one or asset two is a perfect hedge of the other. In this case, assuming that \mathbf{C} is positive definite, the volatility of a portfolio given by the weight vector $\mathbf{w} = [0.5, 0.5, 0]^T$ tends to a small value $c > 0$ as $\rho \rightarrow -1$. In Choueifaty et al. [22], it is demonstrated that risk parity suffers from duplication invariance. It is well known in the literature that the global minimum variance portfolio tends to be highly concentrated to assets with low volatility, see e.g. Clarke et al. [24]. Thus, for an asset universe where the exposure to some assets to a large extent has been hedged away, the global minimum variance portfolio tends to be highly concentrated to the hedged assets. We denote

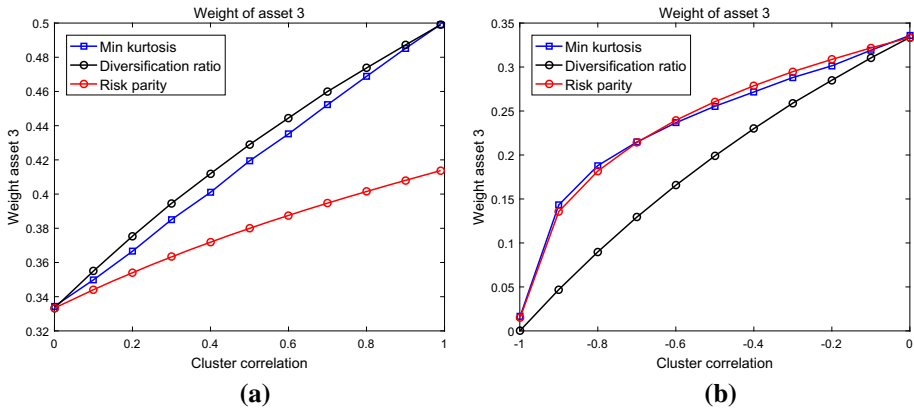


Fig. 1 Weight of asset three for the minimum kurtosis, risk parity and maximized diversification ratio portfolios for the cases when: **(a)** $\rho \in [0, 1]$ and **(b)** $\rho \in (-1, 0]$

this undesirable property of the global minimum variance portfolio the hedging invariance problem.

Consistency with the duplication invariance and hedging invariance properties for the introduced diversification framework is illustrated in Fig. 1 for the case when the marginal distributions of the assets can be assumed to be approximately symmetric. In this case, we assume that non-Gaussianity is adequately captured by portfolio kurtosis. The consistency with the desired properties is monitored through the weight of asset three for the cases when $\rho \rightarrow 1$ and $\rho \rightarrow -1$. The weight of asset three obtained when minimizing portfolio kurtosis is compared to the corresponding weights obtained with risk parity and from maximizing the diversification ratio introduced in Choueifaty and Coignard [21]. Since the volatilities of the three assets are equal, the portfolio obtained from maximizing the diversification ratio coincides with the global minimum variance portfolio, see Choueifaty and Coignard [21]. For risk parity and the most diversified portfolio, the weight of asset three can be solved analytically and is given by

$$w_3^{RP} = \frac{2\sqrt{1+\rho} - (1+\rho)}{3-\rho}, \tag{6}$$

for the risk parity portfolio, and

$$w_3^{DR} = \frac{1+\rho}{3+\rho}, \tag{7}$$

for the maximized diversification ratio and the global minimum variance portfolios. Thus, when $\rho \rightarrow 1$, the weight of the third asset approaches $\sqrt{2} - 1$ for the risk parity portfolio, whereas $w_3 \rightarrow \frac{1}{2}$ for the maximized diversification ratio. From Fig. 1a, one observes that the minimum kurtosis portfolio and the maximized diversification ratio satisfy the duplication invariance property, whereas risk parity does not.

When $\rho \rightarrow -1$, the volatility of a portfolio with weight vector $w = [0.5, 0.5, 0]^T$ approaches the small value $c > 0$. All portfolio construction methodologies that are based on only the covariance matrix will approach this solution when $\rho \rightarrow -1$. The question is at which rate. From Fig. 1b, one observes that w_3^{DR} approaches zero at a rate which is close to linear when ρ varies between 0 and -1 . Since this corresponds to the behaviour of the global minimum variance portfolio, which suffers from the hedging invariance problem, this rate is

too large when ρ is not close to -1 . Fig. 1b reveals that the weight of asset three for both the minimum kurtosis and risk parity portfolios approaches zero at a slower rate compared to the diversification ratio portfolio when ρ is not close to -1 . These portfolios are thus not too heavily concentrated to the partially hedged exposure represented by asset one and asset two in our example. We conclude that the minimum kurtosis and risk parity portfolios satisfy the hedging invariance property. Hence, only the minimum kurtosis portfolio satisfies the two desired properties when the asset distributions are symmetric.

We finally stress that in this paper we do not attempt to accurately estimate higher order moments or joint distributions of assets returns. The multivariate distribution of the asset returns is modelled with a Gaussian copula and marginal distributions which allow for differing skewness and kurtosis parameters for the individual assets. By modelling the dependence structure with a Gaussian copula, we avoid the notoriously difficult task of estimating a nonlinear dependence structure between the assets. The cost of using a model with less uncertainty in the estimated parameters is that we only take linear dependence between the assets returns into account in this paper. In order to obtain a robust implementation of the framework we take the approach of assigning representative tail risk parameters for different asset groups. Based on the assigned tail risk parameters, the diversification framework then lets us measure and optimize the portfolio dimensionality for a given asset universe.

3 Deterministic global optimization of ratios of convex functions

There are numerous applications in finance that involve the optimization of ratios, see, e.g., Stoyanov et al. [73]. In the previous two sections we argued that formulating an appropriately defined portfolio diversification measure naturally leads to functions that are ratios of convex functions. In this section we develop a deterministic algorithm for solving such problems to global optimality.

Let $\mathcal{A} \subseteq \mathbb{R}^n$ be a nonempty compact convex set and consider the maximization problem

$$\max_{\mathbf{w} \in \mathcal{A}} h(\mathbf{w}), \quad \text{where} \quad h(\mathbf{w}) = \frac{f(\mathbf{w})}{g(\mathbf{w})} \tag{8}$$

and $f, g : \mathcal{A} \rightarrow \mathbb{R}$ are positive and continuous functions. In Avriel et al. [7] it is shown that when f is concave and g is convex, then $h(\mathbf{w})$ is a semi-strictly quasi-concave function. Many theoretical results, as well as algorithms of convex programming, apply to the problem of maximizing a quasi-concave function over a convex set (see [27, 66, 67]). In particular, each local maximum is again a global maximum. For the case when f and g are either both convex or both concave, $h(\mathbf{w})$ is in general neither a quasi-concave nor quasi-convex function and the function may have multiple local optima that are different from the global optimum.

3.1 Formulation of the portfolio kurtosis minimization problem

Using the notation for higher order portfolio moments introduced in Appendix A, the portfolio kurtosis as a function of the portfolio weights can be expressed as

$$\kappa_p(\mathbf{w}) = \frac{\mathbb{E}((\mathbf{w}^\top(\mathbf{r} - \boldsymbol{\mu}))^4)}{(\mathbb{E}((\mathbf{w}^\top(\mathbf{r} - \boldsymbol{\mu}))^2))^2} = \frac{\mathbf{w}^\top \mathbf{M}_4(\mathbf{w} \otimes \mathbf{w} \otimes \mathbf{w})}{(\mathbf{w}^\top \mathbf{M}_2 \mathbf{w})^2}, \tag{9}$$

where $\mathbf{w} \in \mathbb{R}^{n+1}$ is the vector of relative portfolio weights, $\mathbf{r} \in \mathbb{R}^{n+1}$ denotes the vector of asset returns, $\boldsymbol{\mu} = \mathbb{E}(\mathbf{r})$, and $\mathbf{M}_2 \in \mathbb{R}^{(n+1) \times (n+1)}$ and $\mathbf{M}_4 \in \mathbb{R}^{(n+1) \times (n+1)^3}$ denote the

covariance and fourth co-moment matrices of the asset returns, respectively. We assume that \mathbf{M}_2 is positive definite and hence that $\mathbf{w}^\top \mathbf{M}_2 \mathbf{w} > 0$ for all non-zero \mathbf{w} . Therefore, the ratio (9) is well defined. By application of Jensen’s inequality we also have that

$$\mathbb{E} \left((\mathbf{w}^\top (\mathbf{r} - \boldsymbol{\mu}))^4 \right) \geq \left(\mathbb{E} \left((\mathbf{w}^\top (\mathbf{r} - \boldsymbol{\mu}))^2 \right) \right)^2 \quad \text{and thus} \quad \mathbf{w}^\top \mathbf{M}_4 (\mathbf{w} \otimes \mathbf{w} \otimes \mathbf{w}) > 0$$

for all non-zero \mathbf{w} . (10)

The convention for the majority of papers in the fractional programming literature is to formulate the fractional program as a maximization problem. Since $\operatorname{argmax}_{\mathbf{w}} f(\mathbf{w})/g(\mathbf{w}) = \operatorname{argmin}_{\mathbf{w}} g(\mathbf{w})/f(\mathbf{w})$, for $f(\mathbf{w}) > 0$ and $g(\mathbf{w}) > 0$, we formulate the portfolio kurtosis optimization problem as the following maximization problem, which we denote by (P)

$$(P) \quad \max_{\mathbf{w} \in \mathcal{W}} h(\mathbf{w}), \quad \text{where} \quad h(\mathbf{w}) = \frac{f(\mathbf{w})}{g(\mathbf{w})} = \frac{(\mathbf{w}^\top \mathbf{M}_2 \mathbf{w})^2}{\mathbf{w}^\top \mathbf{M}_4 (\mathbf{w} \otimes \mathbf{w} \otimes \mathbf{w})} \quad (11)$$

and \mathcal{W} denotes the feasible set for the weights. Since we assume no short selling and a fully invested portfolio, the feasible set is given by $\mathcal{W} = \{\mathbf{w} \in \mathbb{R}^{n+1} \mid \sum_{i=0}^n w_i = 1, w_i \geq 0, i = 0, \dots, n\}$. Letting \mathbf{w}^* denote the optimal weights, the minimum kurtosis over the feasible set is then given by $\kappa_p(\mathbf{w}^*) = 1/h(\mathbf{w}^*)$. Since \mathbf{M}_2 is positive definite, the numerator in (11) is a convex function. In Athayde and Flores [6] it is shown that the fourth moment of the portfolio return is a convex function and hence (11) is a ratio of two convex functions.

3.2 Branch and Bound algorithm for global minimization of portfolio kurtosis

Global optimization of ratios of convex functions is a very difficult optimization problem and has attracted attention in the optimization research community. In this section we present a Branch and Bound (BB) algorithm for global minimization of portfolio kurtosis. The basic idea of BB is to recursively subdivide the solution space geometrically into smaller and smaller subsets, until we can either compute the optimal solution over a subset or rule out that a subset contains the global optimum. A crucial component of the algorithm and key to its efficiency, is the derivation of tight upper and lower bounds on the objective function value, both globally and locally for each subset. Examples of papers in the literature which develop BB algorithms for the special case of ratios of convex quadratic functions are Gotoh and Konno [39], Benson [13] and Yamamoto and Konno [78]. The first, and to the best of our knowledge only, paper which develops a BB algorithm for global optimization of a single ratio of general convex functions is Benson [14]. The generalized problem of optimizing a sum of ratios of convex functions has also attracted considerable attention in the literature. In Shen et al. [70] a BB algorithm for global optimization for the sum of ratios of convex functions over a convex set is developed, while Shen et al. [69] develop a BB algorithm for the case of optimizing the sum of ratios of convex functions when the feasible set is non-convex. Comprehensive treatments of BB algorithms for global optimization can be found in Horst and Tuy [43] and Floudas [32].

We apply the BB algorithm developed by Benson [14] to the problem of portfolio kurtosis minimization and improve the convergence rate by constructing considerably tighter bounds. In the following we first give an overview of the BB algorithm before we describe the steps of the procedure in more detail. As input to the algorithm, one chooses an error tolerance ρ which determines the maximum allowed relative distance between the output value of the algorithm and the global optimum. The output of the algorithm is a ρ -globally optimal solution:

Definition 5 (ρ -globally optimal solution) A solution $\mathbf{w}^k \in \mathcal{W}$ for problem (P) is called ρ -globally optimal, if $h(\mathbf{w}^k) \geq (1 - \rho)h(\mathbf{w}^*)$, where $\rho \in [0, 1)$ and \mathbf{w}^* is an optimal solution for (P).

The basic idea of the BB algorithm is rather simple and consists of the following elements. *Branching process* Consists of choosing a subset $S \subseteq \mathcal{W}$ that is to be subdivided, and then applying a partitioning method for splitting this subset into two smaller subsets.

Upper bounding process Consists of solving a subproblem to obtain an upper bound $UB(S)$ for the maximum of $h(\mathbf{w})$ over each subset $S \subseteq \mathcal{W}$ created by the branching process. Moreover, the upper bound for each subset is used to update a global upper bound UB for the maximum of $h(\mathbf{w})$ over \mathcal{W} .

Lower bounding process Consists of calculating a lower bound $LB(S)$ for the maximum of $h(\mathbf{w})$ over each subset $S \subseteq \mathcal{W}$ created by the branching process. Moreover, the lower bound for each subset is used to update the global lower bound LB for the maximum of $h(\mathbf{w})$ over \mathcal{W} .

Fathoming process Deletes each subset $S \subset \mathcal{W}$ in the partition which satisfies $(1 - \rho)UB(S) \leq LB$. The algorithm stops when all subsets have been fathomed, i.e., the partition is empty.

Unlike heuristic methods, BB algorithms terminate with the guarantee that the value of the best found solution is ρ -globally optimal. BB algorithms are however often slow, and in many cases they require computational effort that grows exponentially with the problem size. This is due to the fact that the size of the partition will grow from iteration to iteration, unless we can fathom subsets. Fathoming subsets, however, depends on the quality of the lower and, especially, the upper bound for a subset. If the upper bound is loose, then a good feasible solution found early in the search may be detected as good only much later in the partitioning process. In other words, the main computational burden of the BB algorithm typically comes from proving global optimality of a feasible point found at an early stage. Thus, in order for the BB algorithm to be efficient, it is crucial to carefully model the functions used for producing upper bounds for each subset generated by the branching process, to be able to fathom them as quickly as possible. Compared to the BB algorithm in Benson [14], we develop two extensions which provide much tighter upper bounds and, thereby, considerably speed up the convergence of the algorithm. Next, we will give a more detailed description of the BB algorithm applied to the problem of minimizing portfolio kurtosis.

3.2.1 Branching process

The branching process splits the feasible set into successively finer partitions. We denote by $\mathcal{Q}_0 = \{\mathcal{W}\}$ the initial partition and by $\mathcal{Q}_k = \{S_i\}_{i \in \mathcal{I}_k}$ the partition in iteration k of the BB algorithm, where \mathcal{I}_k is a finite index set, $\mathcal{W} = \bigcup_{i \in \mathcal{I}_k} S_i$, and $\text{int}(S_i) \cap \text{int}(S_j) = \emptyset$, for $i \neq j$. Note that, strictly speaking, once we start fathoming subsets, \mathcal{Q}_k will no longer form a partition of \mathcal{W} . However, for the ease of exposition, we will still call \mathcal{Q}_k a partition. At the beginning of step $k \geq 1$, the partition \mathcal{Q}_{k-1} consists of subsets not yet deleted by the algorithm. To determine the subset of \mathcal{Q}_{k-1} to be partitioned, we follow the classical *best-first* rule, which selects the subset $S^k \in \mathcal{Q}_{k-1}$ with the largest upper bound. The rationale for this rule is to pick a subset which is likely to contain a good feasible solution, which will, hopefully, allow for a quick increase in the global lower bound and thereby speed up the fathoming process. See Locatelli and Shoen [54] for other common rules.

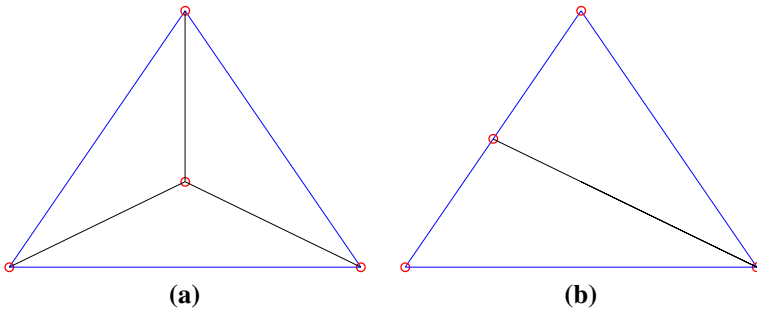


Fig. 2 Examples of subdivision of a 2-simplex: radial subdivision (a) and simplicial bisection (b)

First, we observe that our feasible set \mathcal{W} is identical to the standard n -simplex. In order to refine a partition \mathcal{Q}_{k-1} , we follow Benson [14] and split the chosen subset S^k into two halves by simplicial bisection, which is a special case of radial subdivision introduced in Horst [42]:

Definition 6 (*Radial subdivision*) Let M be an n -simplex with vertex set $\mathcal{V}(M) = \{v^0, v^1, \dots, v^n\}$. Choose a point $m \in M, m \notin \mathcal{V}(M)$ which is uniquely represented by

$$m = \sum_{i=0}^n \lambda_i v^i, \lambda_i \geq 0 \ (i = 0, \dots, n), \sum_{i=0}^n \lambda_i = 1,$$

and for each i such that $\lambda_i > 0$ form the simplex $M(i, m)$ obtained from M by replacing the vertex v^i by m , i.e., $M(i, m) = \{v^0, \dots, v^{i-1}, m, v^{i+1}, \dots, v^n\}$.

A simplicial bisection is obtained by choosing m as the midpoint of a longest edge of the simplex M , see Fig. 2 for an example. Horst and Tuy [43] prove that the set of subsets $M(i, m)$ that can be constructed from an n -simplex M by an arbitrary radial subdivision forms a partition of M into n -simplices. Hence, our subsets S_i are again n -simplices. Let \hat{v} denote the midpoint of one of the longest edges of S^k and v^d, v^e the corresponding endpoints of this edge. In the branching process, we replace S^k by the two n -simplices with vertex sets $S_1^k = M(d, \hat{v})$ and $S_2^k = M(e, \hat{v})$ using simplicial bisection to obtain a refined partition $\mathcal{Q}_k = (\mathcal{Q}_{k-1} \setminus \{S^k\}) \cup \{S_1^k, S_2^k\}$.

3.2.2 Upper bounding process

Let $S \in \mathcal{Q}_k$ be an n -simplex of the partition with vertices $\{v^0, v^1, \dots, v^n\}$. Initially, we follow Benson [14] and overestimate the objective function $h(w) = f(w)/g(w)$ by the ratio of two affine functions: one that overestimates f and one that underestimates g . We will improve these bounding functions in Sect. 3.2.5 in order to obtain tighter upper bounds and thereby increase the speed of convergence. The function g in the denominator is underestimated by a first order Taylor expansion around the barycenter $\hat{v} = 1/(n + 1) \sum_{i=0}^n v^i$ of the simplex S according to

$$g_S(w) = g(\hat{v}) + \nabla_w g(\hat{v})(w - \hat{v})^T. \tag{12}$$

As g is a convex function, $g_S(w) \leq g(w), w \in \mathbb{R}^{n+1}$, and, hence, g_S is an underestimator of g . The gradient of the fourth central moment of the portfolio return is given by (see

Appendix A)

$$\nabla_w g(w) = 4M_4(w \otimes w \otimes w). \tag{13}$$

In order to ensure that the approximation is positive, let

$$z(w) = \max\{\alpha, g_S(w)\}, \quad \text{where } \alpha = \min_{w \in \mathcal{W}} g(w). \tag{14}$$

With g being convex, the minimization problem on the right-hand side can be solved efficiently.

In order to construct a linear overestimator of the function f in the numerator we need the following definition given in Horst and Tuy [43]:

Definition 7 (*Concave envelope*) The concave envelope of a function p taken over a nonempty subset M of its domain is the function p^M that satisfies:

- (i) p^M is a concave function defined over the convex hull of M ,
- (ii) $p^M(x) \geq p(x)$, for all $x \in M$, and
- (iii) if q is a concave function defined over the convex hull of M that satisfies $q(x) \geq p(x)$ for all $x \in M$, then $q(x) \geq p^M(x)$ for all x in the convex hull of M .

Horst [42] shows that when M is an n -simplex and p is a convex function on M , then p^M is the unique affine function that coincides with p at the vertices of M . Denoting by $f^S(w)$ the concave envelope of f over S , we construct the following upper bound for the maximum of h over S

$$UB(S) = \max_{w \in S} \frac{f^S(w)}{z(w)}. \tag{15}$$

Since $z(w) \geq 0$, $w \in \mathbb{R}^{n+1}$, and $f^S(w) \geq f(w) > 0$, $w \in S$, $UB(S)$ is equal to the optimal value of the following problem:

$$(P1(S)) \quad \max_{t, w \in S} \frac{f^S(w)}{t} \tag{16}$$

$$\text{s.t. } t \geq \alpha, \tag{16}$$

$$t - g_S(w) \geq 0. \tag{17}$$

As $S \subseteq \mathcal{W}$ is compact and the objective function is continuous, (P1(S)) has an optimal solution. Moreover, as the ratio of two linear functions is quasi-concave, every local optimum over the closed convex set is also a global optimum. Thus, the fractional program can be solved to global optimality with any local solver. However, as P1(S) has to be solved many times during the BB algorithm, we follow Benson [14] and reformulate the problem as follows. Each $w \in S$ can be written as

$$w = \sum_{i=0}^n \lambda_i v^i, \quad \text{where } \lambda_i \geq 0, i = 0, 1, \dots, n, \text{ and } \sum_{i=0}^n \lambda_i = 1, \tag{18}$$

(see [43]). As $f^S(\mathbf{w})$ is an affine function, we then get $f^S(\mathbf{w}) = \sum_{i=0}^n \lambda_i f(\mathbf{v}^i)$. Substituting $f^S(\mathbf{w})$ and adding the conditions for \mathbf{w} gives the equivalent fractional program

$$(P2(S)) \quad \max_{t, \lambda, \mathbf{w}} \quad \frac{1}{t} \sum_{i=0}^n \lambda_i f(\mathbf{v}^i)$$

$$\text{s.t.} \quad (16), (17),$$

$$\mathbf{w} = \sum_{i=0}^n \lambda_i \mathbf{v}^i, \tag{19}$$

$$\sum_{i=0}^n \lambda_i = 1, \tag{20}$$

$$\lambda_i \geq 0, \quad i = 0, 1, \dots, n. \tag{21}$$

To linearize the objective function, we apply the Charnes–Cooper transformation [17], performing the following change of variables

$$u = \frac{1}{t}, \quad b_i = \frac{\lambda_i}{t}, \quad \mathbf{y} = \frac{\mathbf{w}}{t}, \tag{22}$$

where $\mathbf{y} = [y_0, \dots, y_n]^\top$ and $\mathbf{b} = [b_0, \dots, b_n]^\top$, which results in the equivalent problem

$$(P3'(S)) \quad \max_{u, \mathbf{b}, \mathbf{y}} \quad \sum_{i=0}^n b_i f(\mathbf{v}^i)$$

$$\text{s.t.} \quad u \leq 1/\alpha, \tag{23}$$

$$u \cdot g_S(\mathbf{y}/u) \leq 1, \tag{24}$$

$$\mathbf{y} = \sum_{i=0}^n b_i \mathbf{v}^i, \tag{25}$$

$$\sum_{i=0}^n b_i - u = 0, \tag{26}$$

$$b_i \geq 0, \quad i = 0, \dots, n,$$

$$u > 0. \tag{27}$$

Since $u \cdot g_S(\mathbf{y}/u)$ is an affine function, $(P3'(S))$ is a linear program, except for the domain constraint on u . However, Avriel et al. [7] showed that when a solution to $(P2(S))$ exists, then the strict inequality can be replaced by $u \geq 0$, and we obtain the linear program

$$(P3(S)) \quad \max_{u, \mathbf{b}, \mathbf{y}} \quad \sum_{i=0}^n b_i f(\mathbf{v}^i)$$

$$\text{s.t.} \quad (23) - (27),$$

$$u \geq 0. \tag{28}$$

This formulation can now be solved very efficiently using any linear programming solver.

Finally, the upper bound for $S_l^k, l = 1, 2$, is now computed as $\min\{UB(S_l^k), UB(S^k)\}$. Moreover, for each iteration $k \geq 0$, the upper bounding process also computes an upper bound UB_k for the global optimal value $h(\mathbf{w}^*)$ of the original problem (P) based on the

partition \mathcal{Q}_k :

$$UB_k = \max_{S \in \mathcal{Q}_k} UB(S). \tag{29}$$

By construction, the upper bound is monotonically decreasing in k , i.e., $UB_{k+1} \leq UB_k$, $k \geq 0$.

3.2.3 Lower bounding process

Denoting by w^k the best solution of the problems (P1(S)) encountered up to iteration k , the lower bound LB_k for the global optimal value $h(w^*)$ in iteration k is given by $LB_k = h(w^k)$. The bounds are monotonically increasing in k : $LB_{k+1} \geq LB_k$, $k \geq 0$.

3.2.4 Fathoming process

Based on the lower and upper bounds produced by the algorithm, the fathoming process deletes all subsets $S \in \mathcal{Q}_{k-1}$ from \mathcal{Q}_{k-1} that are guaranteed not to contain the global optimal solution. At the beginning of each iteration, i.e., all $S \in \mathcal{Q}_{k-1}$ are removed for which $(1 - \rho)UB(S) \leq LB_{k-1}$. If this results in \mathcal{Q}_{k-1} being empty, then

$$h(w^{k-1}) \geq (1 - \rho)UB_{k-1} \geq (1 - \rho) \max_{w \in \mathcal{W}} h(w) = (1 - \rho)h(w^*), \tag{30}$$

which means that w^k is a ρ -globally optimal solution to problem (P). Benson [14] shows that when the number of iterations for the BB algorithm is infinite, it generates two sequences of points whose accumulation points are the global optimal solution w^* for (P), and

$$\lim_{k \rightarrow \infty} LB_k = \lim_{k \rightarrow \infty} UB_k = h(w^*). \tag{31}$$

This result implies that whenever $\rho > 0$, the BB algorithm is finite. The complete BB algorithm is summarized below.

Remark For the case with additional constraints, such as position limits, the feasible set is no longer given by the standard n -simplex. The extension of the algorithm to a more general case is however straightforward (see [14], for details).

3.2.5 Improving the upper bound

Preliminary computational tests showed that the BB algorithm spends the vast majority of the computing time calculating the upper bound $UB(S)$ over the n -simplex S . Moreover, while it often took only a few iterations to obtain a very good lower bound LB_k on the optimal value $h(w^*)$, the upper bound was improving only very slowly. In order to achieve faster convergence for the BB algorithm, we present in the following two extensions of the algorithm presented in Benson [14] that lead to a much faster reduction of the global upper bound UB_k . In the first, the lower bound of the function g in the denominator is improved by adding affine functions to the approximation. In the second, the upper bound of the function f in the numerator is enhanced by using a generalization of the concave envelope. This generalization requires the introduction of binary variables, which means that the improved upper bound comes with the cost of having to solve a more difficult combinatorial optimization problem.

BB algorithm

Input: $\rho \in [0, 1)$, n -simplex \mathcal{W} , functions $f(\cdot)$ and $g(\cdot)$.

Output: ρ -globally optimal solution \tilde{w} .

Initialization Set $S^0 = \mathcal{W}$ and $\mathcal{Q}^0 = \{S^0\}$. Calculate $UB_0 = UB(S^0)$ and an optimal solution (w^0, t_0) for $PI(S^0)$. Set $LB_0 = h(w^0)$. If $(1 - \rho)UB_0 \leq LB_0$, then stop; $\tilde{w} = w^0$ is ρ -globally optimal for (P).

Step k ($k = 1, 2, \dots$)

k.1 Delete each n -simplex $S \in \mathcal{Q}_{k-1}$ from \mathcal{Q}_{k-1} for which $(1 - \rho)UB(S) \leq LB_{k-1}$.

k.2 If $\mathcal{Q}_{k-1} = \emptyset$, then stop; $\tilde{w} = w^{k-1}$ is ρ -globally optimal for problem (P).

k.3 Let $UB_k = \max\{UB(S) \mid S \in \mathcal{Q}_{k-1}\}$ and choose an n -simplex $S^k \in \mathcal{Q}_{k-1}$ such that $UB_k = UB(S^k)$. Subdivide S^k into two n -simplices S^k_1, S^k_2 via simplicial bisection.

k.4 For $S = S^k_1, S^k_2$, find the optimal value $UB(S)$ and an optimal solution (w^S, t_S) for $PI(S)$, and set $UB(S) = \min\{UB(S), UB(S^k)\}$.

k.5 Set $LB_k = \max\{LB_{k-1}, h(w^{S^k_1}), h(w^{S^k_2})\}$ and let w^k satisfy $LB_k = h(w^k)$. Set $\mathcal{Q}_k = \mathcal{Q}_{k-1} \setminus \{S^k\} \cup \{S^k_1, S^k_2\}$ and $k = k + 1$.

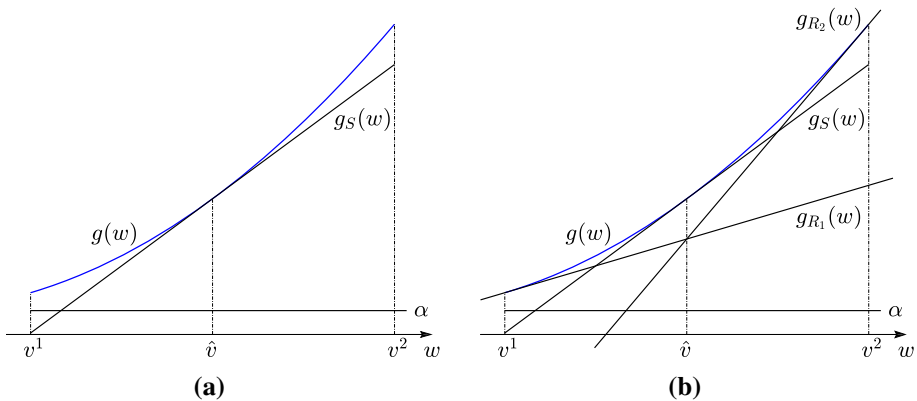


Fig. 3 Improvement of the lower bound for g when S is an 1-simplex: original lower bound (a) and improved lower bound for $p = 2$ (b)

To tighten the lower bound for the function g , we extend the linearization technique in (14) by adding first order Taylor expansions of g around p additional points R_j in S . We then define

$$\tilde{z}(w) = \max(\alpha, g_S(w), g_{R_1}(w), \dots, g_{R_p}(w)), \tag{32}$$

where $g_{R_j}(w) = g(R_j) + \nabla_w g(R_j)(w - R_j)^\top, j = 1, \dots, p$. The idea of the improvement is illustrated in Fig. 3 for the case when S is a 1-simplex and $p = 2$.

For the general case of an n -simplex S , the locations of the points $\{R_j\}_{j=1}^p$ are chosen so that they are evenly distributed in S , see Sect. 3.3 for more details. The resulting problem is then given as

$$\begin{aligned} \text{(P4)(S)} \quad & \max_{u, b, y} \sum_{i=0}^n b_i f(v^i) \\ & \text{s.t. (23) – (27),} \\ & u \geq 0, \\ & u \cdot g_{R_j}(y/u) \leq 1, \quad j = 0, \dots, p. \end{aligned} \tag{33}$$

Obviously, the accuracy of the approximation increases with p , at the expense of adding more linear constraints to the optimization problem.

Next, we turn our attention to improving the accuracy of the approximation of the numerator of the objective function. We start by subdividing the n -simplex S by radial subdivision according to Definition 6. Let the set of n -simplices created by the radial subdivision be given by $\mathcal{T} = \{S_j\}_{j \in \mathcal{J}}$ and the corresponding set of all vertices by $\mathcal{V}(\mathcal{T}) = \{v^i\}_{i \in \mathcal{I}}$. The improved upper bound is then constructed by the combination of the concave envelopes over the n -simplices in \mathcal{T} . The construction is more easily illustrated by the simplest possible example in one dimension given in Fig. 4b. For this example, the set of n -simplices and corresponding vertices are after the radial subdivision given by $\mathcal{T} = \{S_1, S_2\}$ and $\mathcal{V}(\mathcal{T}) = \{v^1, v^2, v^3\}$, respectively. The generalized concave envelope over S is constructed from the concave envelopes over S_1 and S_2 .

When calculating the concave envelope, we have to introduce binary variables $q_j, j \in \mathcal{J}$, in order to keep track of which n -simplex in \mathcal{T} is active. The function representing the

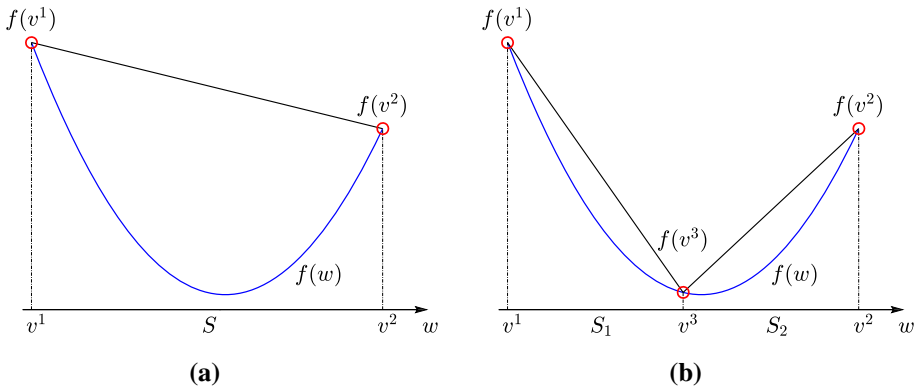


Fig. 4 Improvement of the upper bound for f when S is a 1-simplex: original upper bound (a) and improved upper bound (b)

generalized concave envelope over the n -simplex S can now be formulated as

$$\sum_{i \in \mathcal{I}} \lambda_i f(v^i)$$

(20), (21),

$$\sum_{j \in \mathcal{J}} q_j = 1, \tag{34}$$

$$\lambda_i \leq \sum_{j: v^i \in S_j} q_j, \quad i \in \mathcal{I}, \tag{35}$$

$$q_j \in \{0, 1\}, \quad j \in \mathcal{J}. \tag{36}$$

Condition (35) ensures that only λ_i 's belonging to vertices of the n -simplex that is active, i.e. for which $q_j = 1$, can be non-zero. Using the improved approximation function, we obtain the optimization problem

$$\begin{aligned} \text{(P5}(S)) \quad & \max_{t, \lambda, q, w} \quad \frac{1}{t} \sum_{i \in \mathcal{I}} \lambda_i f(v^i) \\ \text{s.t.} \quad & \text{(16), (17), (19)–(21), (34)–(36)}. \end{aligned}$$

As before, we transform this problem into a mixed-integer linear program (MILP) via Charnes–Cooper transformation through the variable transformations in (22). The last set of constraints is transformed into

$$b_i - \sum_{j: v^i \in S_j} q_j u \leq 0, \quad i \in \mathcal{I}, \tag{37}$$

in the new variables. The product of variables is linearized by introducing the continuous variables $z_j = q_j u$, $j \in \mathcal{J}$, and adding the following constraints for each $j \in \mathcal{J}$:

$$z_j \leq \frac{q_j}{\alpha}, \quad z_j \leq u, \quad z_j \geq u - (1 - q_j)/\alpha, \quad \text{and} \quad z_j \geq 0. \tag{38}$$

If $q_j = 0$, then the first and last constraint ensures that $z_j = 0$, while the third only states that z_j has to be greater than a negative number. If $q_j = 1$, then the first constraint enforces

$z_j \leq 1/\alpha$, and the second and third ensure that $z_j = u$. Summing up, we obtain the following equivalent MILP, which we denote by (P6(S))

$$\begin{aligned}
 \text{(P6(S))} \quad & \max_{u, b, z, q, y} \sum_{i \in \mathcal{I}} b_i f(v^i) \\
 \text{s.t.} \quad & (23)–(27), (34), (36), (37), (38), \\
 & u \geq 0.
 \end{aligned}$$

This problem can easily be enhanced by adding linear terms to the constraints in order to obtain a better approximation of the function g in the denominator. The hope is that the improved upper bound in (P6(S)) will lead to a sufficiently fast decrease of the global upper bound UB_k in order to compensate for the increased computational time induced by solving an MILP instead of an LP for each instance of the upper bounding process.

3.3 Numerical implementation

In this section we will demonstrate the convergence properties of the BB algorithm when applied to the problem of minimizing the portfolio kurtosis for an increasing number of assets. As sample problem we assume that all assets have identical marginal distributions and that all correlations between different assets are assumed to be equal. This problem instance represents a non-convex problem with multiple local optima. When the subproblems for the BB algorithm are given by the MILP (P6(S)), the description in Sect. 3.2 needs to be extended in order to produce an efficient algorithm. Numerical experiments show that radial subdivision does not improve the upper bound of f sufficiently to produce an efficient algorithm. In order to further improve the upper bound of f , the n -simplex S is instead subdivided with barycentric subdivision. Roughly speaking, the barycentric subdivision of an n -simplex S is obtained by radial subdivision of all k -faces of dimension $1 \leq k \leq n$ in decreasing order of dimension. It is also possible with partial barycentric subdivision of S by restricting the radial subdivision to all k -faces of dimension $l \leq k \leq n$, with $l > 1$, in decreasing order of dimension (see [2], for a detailed description of barycentric subdivision). The partial barycentric subdivision of a 2-simplex with $l = 2$ corresponds to radial subdivision as illustrated in Fig. 5 a. The full barycentric subdivision of the 2-simplex is illustrated in Fig. 5 b. Numerical experimentation reveals that full barycentric subdivision is required in order to produce a sufficiently improved upper bound of f for the MILP formulation. Unfortunately, this means that $(n + 1)!$ binary variables need to be introduced when solving the subproblems with the MILP formulation. The formulation of the optimization problem (P6(S)) does not change when subdividing the n -simplex with barycentric subdivision instead of with radial subdivision. The only thing that changes is the set of n -simplices created by the subdivision and the corresponding set of vertices.

In the following we investigate the improvements obtained in terms of iteration count and runtime, when using the enhanced LP formulation and the MILP formulation for solving the subproblems of the algorithm. The BB algorithm was implemented in MATLAB. For all LP formulations of the subproblems we use the solver CPLEX in the implementation, whereas the subproblems arising from the MILP formulation are solved with the built-in solver *intlinprog* in MATLAB. For all comparisons we set the parameter ρ to 10^{-3} . When using the enhanced LP formulation (P4(S)), a choice has to be made regarding how many extra constraints p are added to the problem. The p points defining the added constraints are distributed evenly over the subsimplex S for which the subproblem is solved. Letting n_c denote the number of added constraints per asset, one has that $p = (n + 1)n_c$. We choose to

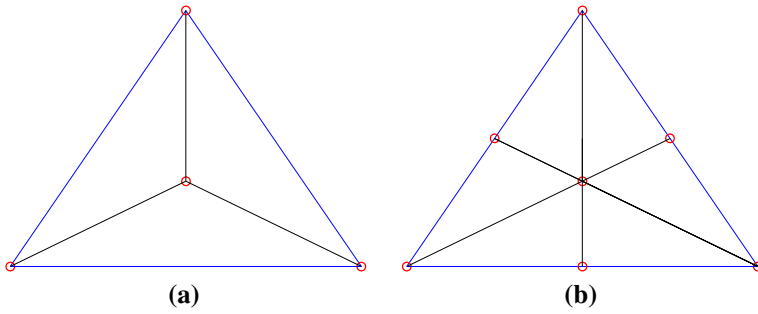


Fig. 5 Examples of subdivision of a 2-simplex: radial subdivision (a) and barycentric subdivision (b)

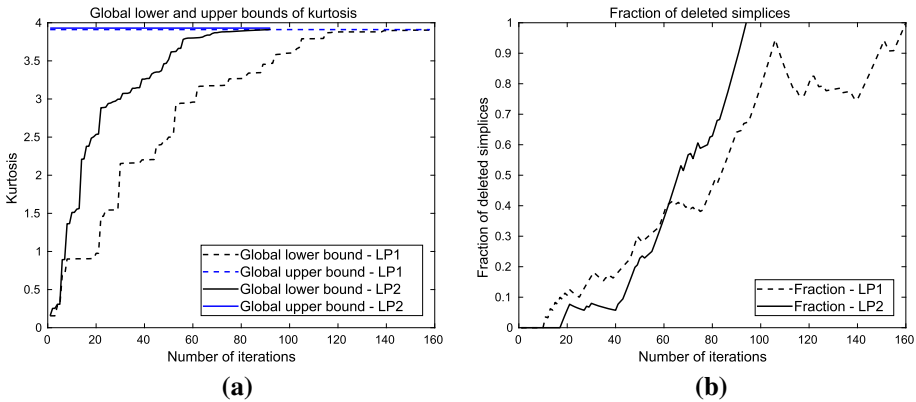


Fig. 6 a Evolution of the global lower and upper bounds of the portfolio kurtosis for the original (LP1) and enhanced LP model (LP2) with $n_c = 2$ for the three asset problem. b The fraction of deleted simplices for the original and enhanced LP model for the three asset problem

distribute the p points evenly between the vertices $\{v^i\}_{i=0}^n$ and the barycenter \hat{v} of S . Thus, for $n_c = 1$ the p points are defined by the vertices. For $n_c \geq 2$, the p points are defined by the vertices and the $(n + 1)(n_c - 1)$ points

$$\frac{j}{n_c} v^i + \left(1 - \frac{j}{n_c}\right) \hat{v}, \quad i = 0, \dots, n; \quad j = 1, \dots, n_c - 1. \tag{39}$$

We also experimented with distributing points evenly between the vertices but that did not bring any noticeable improvement in terms of iteration count. Naturally, adding more constraints in order to obtain a tighter lower bound should decrease the iteration count for the BB algorithm, at the cost of increasing the runtime for each of the subproblems that are solved. This trade-off is now investigated. In the following, we denote the enhanced LP formulation (P4(S)) by LP2 and the LP formulation (P3(S)) as used in Benson [14] by LP1.

Figure 6a displays the evolution of the global lower and upper bounds of the portfolio kurtosis for the iterations of the BB algorithm applied to the three asset problem. Note that these are the inverses of the global lower and upper bounds calculated by the BB algorithm, i.e., for $\kappa_p(w) = g(w)/f(w)$. Simulated return data is used to calculate the moment matrices in the objective function (11). The sample moment matrices \hat{M}_2 and \hat{M}_4 are calculated from 10^7 simulated asset returns with NIG-distributed margins and dependence structure given by the Gaussian copula with a homogeneous correlation matrix. The problem instance is defined

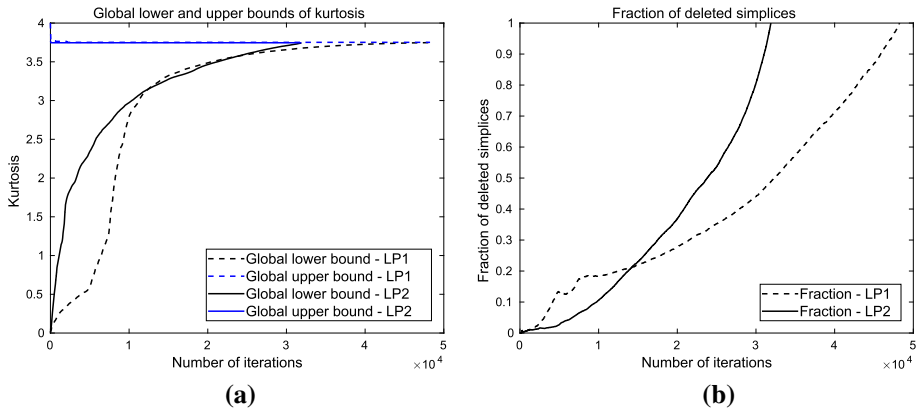


Fig. 7 **a** Evolution of the global lower and upper bounds of the portfolio kurtosis for the original (LP1) and enhanced LP model (LP2) with $n_c = 2$ for the five asset problem. **b** The fraction of deleted simplices for the original and enhanced LP model for the five asset problem

by the homogeneous correlation $\rho = -0.2$ and marginal kurtosis $\kappa_m = 6$ for all the assets. Appendix B contains a description of the simulation procedure. Figure 6b shows the fraction of deleted simplices for the iterations of the algorithm for the three asset problem. Note that the fraction of deleted simplices decreases if the number of deleted simplices is less than the number of simplices that are added by the subdivision procedure. From the graphs it is visible that the enhanced LP formulation LP2 with $n_c = 2$ converges faster to the global optimum in terms of number of iterations compared to the original LP formulation LP1. As can be seen in Table 1, the number of iterations decreases with the number of extra constraints p added to the problem. However, as displayed in Table 2 the decrease in the number of iterations is not significant enough in order to compensate for the increased runtime associated with the larger number of constraints. Thus, the enhanced LP formulation with $p = 1$ has a lower runtime than the formulations with $p > 1$. Compared to the original LP formulation LP1, the runtime of LP2 with $n_c = 1$ is the same for the three asset problem.

Figure 7 a, b display the evolution of the global lower and upper bounds of the portfolio kurtosis and the fraction of deleted simplices for the iterations of the BB algorithm applied to the five asset problem. The solid and dotted lines represent the evolution of the bounds and the fraction of deleted simplices for LP2 with $n_c = 2$ and LP1, respectively. The graphs reveal that there is a significant decrease in iteration count when using the enhanced LP formulation LP2 compared to LP1. As for the three asset case, the iteration count for LP2 decreases with the number of added constraints p . However, as can be seen in Table 2 the decrease in iteration count does not compensate for the added computational cost and hence LP2 with $n_c = 1$ has the lowest runtime. LP2 with $n_c = 1$ also has a lower runtime than LP1 for the five asset case, 175 seconds compared to 193 seconds.

We will now investigate the performance of the MILP formulation against the LP formulation with the lowest runtime for the five asset problem. Figure 8 a, b show the evolution of the global lower and upper bounds of the portfolio kurtosis and the fraction of deleted simplices for the two cases. The solid and dotted lines represent the evolution of the bounds and the fraction of deleted simplices for the MILP formulation and LP2 with $n_c = 1$, respectively. From Fig. 8, one observes that the MILP formulation improves the global lower bound of the kurtosis, corresponding to the global upper bound for the BB algorithm, much faster than the LP formulation up to around iteration count 5000. After that, the global lower bound of the

Table 1 Number of iterations for the BB algorithm for different portfolio sizes and solution methods for the subproblems.

Subproblem method	LP1	LP2 $n_c = 1$	LP2 $n_c = 2$	LP2 $n_c = 4$	LP2 $n_c = 8$	MILP
3 assets	159	119	96	95	92	47
4 assets	1551	1762	1358	1254	1264	1068
5 assets	48254	35943	33374	30824	29041	28179
6 assets	–	620,000	–	–	–	–

The number of iterations for the enhanced LP model, LP2, is given for different numbers (n_c) of extra constraints per asset in the portfolio. In each case, the number of iterations is the median from five runs of the algorithm. The dashes in the table indicate that for the six asset problem, we have only produced results for the best performing algorithm in terms of runtime for the five asset problem

Table 2 Runtime for the BB algorithm for different portfolio sizes and solution methods for the subproblems.

Subproblem method	LP1	LP2 $n_c = 1$	LP2 $n_c = 2$	LP2 $n_c = 4$	LP2 $n_c = 8$	MILP
3 assets	4 s	4 s	4 s	5 s	5 s	6 s
4 assets	8 s	10 s	10 s	12 s	15 s	61 s
5 assets	193 s	175 s	198 s	260 s	425 s	9,390 s
6 assets	–	49,680 s	–	–	–	–

The runtime for the enhanced LP model, LP2, is given for different numbers (n_c) of extra constraints per asset in the portfolio. In each case, the runtime is the median from five runs of the algorithm. The dashes in the table indicate that for the six asset problem, we have only produced results for the best performing algorithm in terms of runtime for the five asset problem

MILP formulation improves slower than for the LP formulation. The overall iteration count is lower for the MILP formulation compared to LP2. However, the improvement in iteration count does not compensate for the increased computational cost associated with solving a MILP instead of an LP as can be seen in Table 2. The runtime for the MILP formulation can however likely be reduced by using a state-of-the art solver instead of the built-in solver in MATLAB.

For the six asset problem with homogeneous correlation $\rho = -0.18$, the number of iterations is 620,000 for the best performing solution method for the subproblems, LP2 with $n_c = 1$. The corresponding runtime for the six asset problem is 49,680 seconds, illustrating the exponential growth in computational effort when the BB algorithm is applied to the portfolio kurtosis minimization problem. The BB algorithm can be enhanced by developing special purpose solvers for the subproblems. Furthermore, the algorithm can be parallelized in order to further reduce the runtime. Moreover, we could combine the MILP formulation and LP2, starting with the former to quickly raise the upper bound, and then switch to LP2 to save runtime. It is, however, unlikely that any of these will admit solving problems with significantly higher number of assets than six.

4 Stochastic global optimization

In Sect. 3.2 we developed a deterministic global optimization algorithm for minimizing the inverse of the introduced portfolio diversification measures. However, as is well known and illustrated by the numerical examples in Sect. 3.3, the BB algorithm suffers from the curse of

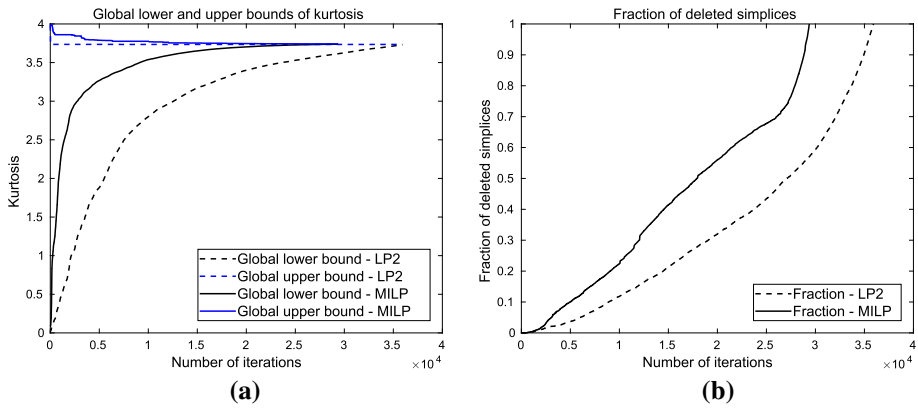


Fig. 8 **a** Evolution of the global lower and upper bounds of the portfolio kurtosis for the enhanced LP model with $n_c = 1$ and the MILP model for the five asset problem. **b** The fraction of deleted simplices for the enhanced LP model and the MILP model for the five asset problem

dimensionality and converges too slowly for problems where the number of assets exceeds six. In this section we develop a stochastic optimization algorithm for global optimization of portfolio kurtosis. The BB algorithm has the desirable property that the objective function value at the obtained solution is guaranteed to be arbitrarily close to the global minimum. For the algorithm developed in this section it is not possible to determine if the solution is a global optimum. However, the algorithm is a special case of stochastic approximation with a rich and well developed theory for convergence analysis. Since the BB algorithm is limited to problems of moderate size, the algorithm developed in this section complements the BB algorithm in the sense that it allows for tackling problems of larger size.

4.1 Stochastic algorithms for global optimization—a very brief overview

There is a huge literature on global optimization algorithms, so called metaheuristic methods, for which it is not possible to guarantee that the obtained solution is a global optimum. These methods iteratively search the feasible set for the global optimum and without prior knowledge there is always the possibility that the optimal point lies in an unexplored region when the algorithm stops. Important examples of metaheuristic methods are genetic algorithms [41], simulated annealing [48] and tabu search [37]. The interested reader may consult Gendreau and Potvin [36] for an overview of metaheuristic methods. Even though, for metaheuristic methods, it is not possible to guarantee that a global optimal point has been found, algorithms that are based on stochastic approximation have a solid theoretical foundation and in many cases non-asymptotic convergence results are available with explicit constants, see Dalalyan [26], Durmus and Moulines [28] and Durmus and Moulines [29]. This can be contrasted to many other popular metaheuristic methods where the theory is often incomplete or even nonexistent, see Spall [72]. A strong aspect of stochastic approximation is the rich convergence theory that has been developed over many years. It has been used to show convergence of many stochastic algorithms such as neural network backpropagation and simulated annealing. For rigorous examples where stochastic approximation methods are applied to problems in finance, see Laurelle and Pages [52] and Sabanis and Zhang [65]. The latter, and more recent result, offers theoretical guarantees for the discovery of near-optimal

solutions of a non-convex optimization problem, namely the optimal allocation of weights for the (unconstrained via a suitable transformation) minimization of CVaR/Expected Shortfall of a portfolio of assets.

In stochastic approximation one is concerned with finding at least one root $\theta^* \in \Theta^* \subseteq \mathbb{R}^d$ to $G(\theta) = 0$, based on noisy measurements of $G(\theta)$. Root finding via stochastic approximation was introduced in Robbins and Monro [62] and important generalizations were made in Kiefer and Wolfowitz [47]. Consider the unconstrained minimization problem

$$\min_{\theta} L(\theta), \tag{40}$$

where L is a smooth function, which has multiple local minima. For the special case when $G(\theta)$ is given by $G(\theta) = \nabla_{\theta} L(\theta)$, the stochastic approximation algorithm is given by the following stochastic gradient descent (SGD) algorithm

$$\theta_{k+1} = \theta_k - a_k H(\theta_k, \mathbf{X}_{k+1}), \tag{41}$$

where $\{\mathbf{X}_k\}_{k \in \mathbb{Z}}$ is a sequence of \mathbb{R}^m -valued i.i.d. data and $H(\theta_k, \mathbf{X}_{k+1})$ is an unbiased estimate of the gradient, i.e. $\nabla_{\theta} L(\theta) = \mathbb{E}(H(\theta, \mathbf{X}_{k+1}))$. In (41), $\{a_k\}$ can either be a decreasing positive sequence satisfying appropriate conditions or a fixed small positive value $a_k = \lambda > 0$, for any $k \geq 0$.

In many estimation problems, a full set of data is collected and G (or L) is chosen by conditioning on the data. This conditioning removes the randomness from the problem and the estimation problem becomes deterministic. In the machine learning literature this is commonly referred to as the batch gradient descent algorithm, which is given by $\theta_{k+1} = \theta_k - a_k \bar{H}(\theta_k)$, where $\{\mathbf{X}_k\}_{k=1}^N$ is the collected data and

$$\bar{H}(\theta) = \frac{1}{N} \sum_{k=1}^N H(\theta, \mathbf{X}_k). \tag{42}$$

Since L has multiple local minima, applying SGD to (40) may yield convergence to a local minimum of L . Under broad conditions, Kushner and Yin [50] show that (41) converges to one of the local minima of L with probability 1. However, the iterates will often be trapped at a local optimum and will miss the global one. Nevertheless, SGD, or one of its various extensions, is commonly used in machine learning for optimization of Deep Neural Networks, see Goodfellow et al. [38]. When L has a unique minimum, Chau et al. [18] provide convergence results for the case with dependent data, discontinuous L , and fixed step size.

The idea behind simulated annealing is that by adding an additional noise term to the iterations one can avoid getting prematurely trapped in a local minimum of L . In Gelfand and Mitter [35], the following modified SGD algorithm is analyzed

$$\theta_{k+1} = \theta_k - a_k H(\theta_k, \mathbf{X}_{k+1}) + b_k \epsilon_{k+1}, \tag{43}$$

where $\{\epsilon_k\}$ is a sequence of standard d -dimensional independent Gaussian random variables, and $\{a_k\}$ and $\{b_k\}$ are decreasing sequences of positive numbers tending to zero. They show that under suitable assumptions, θ_k tends to the global minimizer as $k \rightarrow \infty$ in probability. In the machine learning and Bayesian inference literature, the closely related Stochastic Gradient Langevin Dynamics (SGLD) algorithm has attracted significant interest in the research community in recent years. The SGLD algorithm for global optimization can be formulated as

$$\theta_{k+1} = \theta_k - a_k H(\theta_k, X_{k+1}) + \sqrt{2a_k/\beta} \epsilon_{k+1}, \tag{44}$$

where $\{\epsilon_k\}$ is a sequence of standard d -dimensional independent Gaussian variables and $\beta > 0$ is a temperature parameter. The batch version of this algorithm, Gradient Langevin Dynamics (GLD), is correspondingly given by

$$\theta_{k+1} = \theta_k - a_k \bar{H}(\theta_k) + \sqrt{2a_k/\beta} \epsilon_{k+1}. \tag{45}$$

Assuming that the gradient H is Lipschitz continuous and under further assumptions, Raginsky et al. [61] provide a non-asymptotic analysis of SGLD and GLD applied to non-convex problems for the case when the step size a_k is a positive constant. The analysis provides non-asymptotic guarantees for SGLD and GLD to find an approximate minimizer. The rate of convergence is further improved for both SGLD and GLD in the recent papers by Xu et al. [77] and in Chau et al. [19] even in the presence of dependent data streams.

4.2 A Gradient Langevin Dynamics algorithm for minimization of kurtosis

Motivated by the enormous progress in the aforementioned optimization algorithms, we develop a GLD algorithm for global minimization of portfolio kurtosis. Since portfolio kurtosis is the ratio of two convex functions, the batch gradient is not simply given by the average as in (42). Given a sample of observed return data for a given asset universe, the sample covariance matrix and sample fourth co-moment matrix can be estimated. The batch version of the portfolio kurtosis is then given by

$$\bar{h}(w) = \frac{\bar{f}(w)}{\bar{g}(w)} = \frac{w^\top \hat{M}_4(w \otimes w \otimes w)}{(w^\top \hat{M}_2 w)^2}, \tag{46}$$

where \hat{M}_2 and \hat{M}_4 denote the sample covariance and fourth co-moment matrices, respectively. Given the complicated form of the approximate bias for sample kurtosis, see Bao [9], global minimization of portfolio kurtosis is not easily adapted to the algorithms in Sect. 4.1 which utilize a stochastic unbiased estimate of the gradient. For this reason we only develop a GLD algorithm for the global minimization problem.

The algorithms in Sect. 4.1 are formulated for unconstrained optimization problems and hence need to be adapted to constrained minimization over the standard n -simplex. The GLD algorithm for the constrained problem is given by the following projected iterations

$$w_{k+1} = \Pi_{\mathcal{W}} \left(w_k - \lambda \nabla_w \bar{h}(w_k) + \sqrt{2\lambda/\beta} \epsilon_{k+1} \right), \tag{47}$$

where $\Pi_{\mathcal{W}}$ denotes the Euclidean projection onto the feasible set, $\lambda > 0$ is the fixed step size and $w \in \mathbb{R}^{n+1}$. Euclidean projection of a point onto the standard n -simplex is a quadratic program which can be solved very efficiently. See Chen and Ye [20] for a fast and simple algorithm for computing the projection onto the standard n -simplex. The gradient of the batch version of portfolio kurtosis is given by

$$\nabla_w \bar{h}(w) = \frac{\nabla_w \bar{f}(w)}{\bar{g}(w)} - \frac{\bar{f}(w) \nabla_w \bar{g}(w)}{(\bar{g}(w))^2}, \tag{48}$$

where the explicit form of the gradient $\nabla_w \bar{f}(w)$ is given in Appendix A and

$$\nabla_w \bar{g}(w) = 4 \left(w^\top \hat{M}_2 w \right) \hat{M}_2 w. \tag{49}$$

Most convergence results for SGLD and GLD are only applicable for algorithms without projection. A natural way to avoid the projection step in each iteration would be to extend the objective function with a convex function outside of the feasible set. Naturally, the extended objective function needs to be continuous on the boundary of the feasible set and have a continuous gradient on the boundary. However, in Tawarmalani and Sahinidis [74] it is shown that a sufficient condition for the existence of a convex extension of a function outside of a convex feasible set, is the convexity of the function. Even if the requirement of convexity of the function to be extended is relaxed such that convexity is only required close to the boundary of the feasible set, this does not hold for portfolio kurtosis. It can easily be shown that portfolio kurtosis in general is a non-convex function on the boundary of the feasible set. Hence, it is not possible to find a convex extension of portfolio kurtosis outside of the feasible set \mathcal{W} .

When the objective function is convex, Bubeck et al. [16] provide convergence results for the projected SGLD and GLD algorithms. In the case of a non-convex objective function, no convergence results for projected SGLD and GLD currently exist in the literature to the best of our knowledge, apart from Sabanis and Zhang [65] where the projection is achieved implicitly, via a transformation, under the assumptions of Lipschitz continuity and dissipativity for the gradient of the objective function. These conditions also hold true in our kurtosis minimization problem, see (53) and (57), which provide the theoretical justification for the choice of the proposed projected GLD algorithm as our global optimization approach in higher dimensions. Nevertheless, it is acknowledged here that if the feasible region is made of further constraints, no such guarantees exist. It should however be mentioned that the analysis of SGLD and GLD algorithms is currently a very active research area

In order for the iterations (47) to converge, the gradient $\nabla_w \bar{h}(\mathbf{w})$ needs to be Lipschitz continuous on the domain given by the feasible set \mathcal{W} . The Hessian of h is given by

$$\begin{aligned} \nabla_w^2 \bar{h}(\mathbf{w}) &= \frac{1}{(\bar{g}(\mathbf{w}))^3} \left((\bar{g}(\mathbf{w}))^2 \nabla_w^2 \bar{f}(\mathbf{w}) - \bar{g}(\mathbf{w}) \left(\nabla_w \bar{f}(\mathbf{w}) (\nabla_w \bar{g}(\mathbf{w}))^\top + \nabla_w \bar{g}(\mathbf{w}) (\nabla_w \bar{f}(\mathbf{w}))^\top \right) \right. \\ &\quad \left. + \bar{g}(\mathbf{w}) \bar{f}(\mathbf{w}) \nabla_w^2 \bar{g}(\mathbf{w}) + 2 \bar{f}(\mathbf{w}) \nabla_w \bar{g}(\mathbf{w}) (\nabla_w \bar{g}(\mathbf{w}))^\top \right), \end{aligned} \tag{50}$$

where $\nabla_w \bar{f}(\mathbf{w})$ and $\nabla_w^2 \bar{f}(\mathbf{w})$ are given in Appendix A, $\nabla_w \bar{g}(\mathbf{w})$ is given in (49) and

$$\nabla_w^2 \bar{g}(\mathbf{w}) = 12 \left(\mathbf{w}^\top \hat{\mathbf{M}}_2 \mathbf{w} \right) \hat{\mathbf{M}}_2. \tag{51}$$

In (50), each component of the numerator is a polynomial of degree 10 and the denominator is a polynomial of degree 12. Since it is assumed that $\hat{\mathbf{M}}_2$ is positive definite, the minimum c of $\bar{g}(\mathbf{w})$ is strictly positive over \mathcal{W} , and one has that $|\bar{g}(\mathbf{w})| \geq c > 0$ for $\mathbf{w} \in \mathcal{W}$. As each component of $\nabla_w^2 \bar{h}(\mathbf{w})$ is a continuous function its value is bounded on a closed compact set, and hence

$$\|\nabla_w^2 \bar{h}(\mathbf{w})\|_2 \leq K, \text{ for all } \mathbf{w} \in \mathcal{W}, \tag{52}$$

which, using the mean value theorem, implies

$$\|\nabla_w \bar{h}(\mathbf{u}) - \nabla_w \bar{h}(\mathbf{v})\|_2 \leq K \|\mathbf{u} - \mathbf{v}\|_2, \text{ for all } \mathbf{u}, \mathbf{v} \in \mathcal{W}, \tag{53}$$

where the matrix norm in (52) is defined as the Hilbert–Schmidt norm. Thus, the gradient of the portfolio kurtosis is Lipschitz continuous over the feasible set. In both Raginsky et al. [61] and Xu et al. [77] it is required that the objective function is dissipative in order for the convergence results to hold. The objective function \bar{h} is dissipative on \mathcal{W} if there exists constants $m > 0$ and $b \geq 0$ such that

$$\mathbf{w}^\top \nabla_{\mathbf{w}} \bar{h}(\mathbf{w}) \geq m \|\mathbf{w}\|_2^2 - b, \quad \forall \mathbf{w} \in \mathcal{W}. \tag{54}$$

Since the gradient of $\bar{h}(\mathbf{w})$ is a continuous function it is bounded over \mathcal{W} :

$$\|\nabla_{\mathbf{w}} \bar{h}(\mathbf{w})\|_2 \leq K_2, \quad \forall \mathbf{w} \in \mathcal{W}. \tag{55}$$

Over the n -simplex \mathcal{W} , the Cauchy–Schwartz inequality implies

$$|\mathbf{w}^\top \nabla_{\mathbf{w}} \bar{h}(\mathbf{w})| \leq \|\mathbf{w}\|_2 \|\nabla_{\mathbf{w}} \bar{h}(\mathbf{w})\|_2 \leq \|\nabla_{\mathbf{w}} \bar{h}(\mathbf{w})\|_2 \leq K_2, \tag{56}$$

and hence $\mathbf{w}^\top \nabla_{\mathbf{w}} \bar{h}(\mathbf{w}) \geq -K_2$. Furthermore, $a(\|\mathbf{w}\|_2^2 - 1) \leq 0$, for $a > 0$, implying

$$\mathbf{w}^\top \nabla_{\mathbf{w}} \bar{h}(\mathbf{w}) \geq a(\|\mathbf{w}\|_2^2 - 1) - K_2 = a\|\mathbf{w}\|_2^2 - (K_2 + a) = a\|\mathbf{w}\|_2^2 - b, \tag{57}$$

and hence $\bar{h}(\mathbf{w})$ is dissipative over \mathcal{W} . This means that portfolio kurtosis satisfies the assumptions underlying the convergence results in the non-convex case for GLD and SGLD without projection. Even though we cannot rely on formal convergence results from the literature for GLD with projection, we will in the next section apply the projected GLD algorithm to some example problems with multiple local minima.

4.3 Numerical illustration

In this section we apply the projected GLD algorithm to an artificial problem of kurtosis minimization when all assets are assumed to have identical marginal distributions and where all correlations between different assets are assumed to be negative and identical. Moreover, it is assumed that the weights are positive and sum to one. This problem provides a test bed for testing if the projected GLD algorithm finds an optimal point which is close to the global optimum. The problem has several local optima which, for the non-zero weights, represent equally weighted portfolios with exposure to all or a subset of the assets. To see that this must be the case, consider two assets with non-zero weight at a local optimum. Since the assets are linearly dependent and have the same marginal kurtosis, the optimization will allocate equal weight to both assets when they are non-zero. Hence, all locally optimal portfolios assign equal weight to all weights that are non-zero at the respective local optimum.

The sample covariance matrix $\hat{\mathbf{M}}_2$ and the sample fourth co-moment matrix $\hat{\mathbf{M}}_4$ are calculated from 10^7 simulated asset returns with NIG-distributed margins and dependence structure given by the Gaussian copula with a homogeneous correlation matrix. The simulation procedure is described in Appendix B. Portfolio kurtosis for equally weighted portfolios for the cases with homogeneous correlation matrices with correlation $\rho = -0.2$ and $\rho = -0.05$, respectively, and marginal kurtosis $\kappa_m = 6$, are displayed in Fig. 9. Note that for the described experimental setup, assuming no estimation error, the portfolio kurtosis for an equally weighted portfolio with n assets is equal to the portfolio kurtosis for an $n + 1$ asset portfolio with equal weight in n assets and zero weight in the remaining asset. To see this, consider the definition of portfolio kurtosis

$$\kappa_p(\mathbf{w}) = \frac{\mathbb{E}(\mathbf{w}^\top \mathbf{r})^4}{\left(\mathbb{E}(\mathbf{w}^\top \mathbf{r})^2\right)^2}. \tag{58}$$

From the definition it is apparent that setting one of the weights to zero and the remaining weights to $1/n$ for the kurtosis in the $n + 1$ asset case is identical to the kurtosis for the equally weighted portfolio in the n asset case. Inspecting the graph in Fig. 9 a, one observes that the global optima for the five asset case are located at points with equal weights in four of

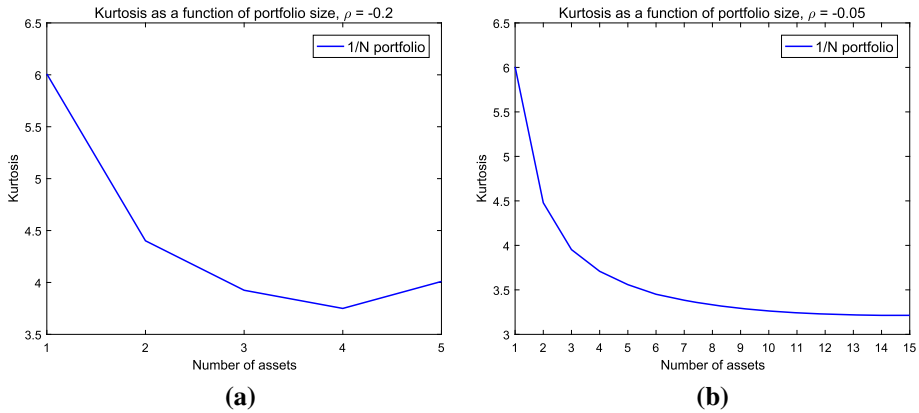


Fig. 9 Portfolio kurtosis for equally weighted portfolios as a function of the number of assets. The asset distributions are generated from a Gaussian copula with homogeneous correlation matrix and NIG-distributed margins with kurtosis $\kappa_m = 6$. (a) Homogeneous correlation $\rho = -0.2$ (b) Homogeneous correlation $\rho = -0.05$

the assets and zero weight in the remaining asset. Thus, assuming no estimation error, there are five global optima for the five asset problem. With 10^7 simulated sets of asset returns, the estimation error is small but nevertheless not zero and hence one of the points represents the unique global optimum with simulated data. Figure 9 b displays the portfolio kurtosis for equally weighted portfolios with up to 15 assets for the case when the homogeneous correlation is -0.05 . Even though not distinguishable from the graph, the portfolio kurtosis for the equally weighted portfolio with 14 assets is slightly lower than the equally weighted portfolio with 15 assets, for every 14 asset sub-portfolio. Thus, for the 15 asset problem the global optimum for the kurtosis minimization problem is given by assigning equal weight to 14 of the assets and zero weight to the remaining asset.

The projected GLD algorithm is applied to the problem of minimizing portfolio kurtosis for the experimental setup described above, with five and 15 assets, respectively. We implement a multistart version of the algorithm, where the iterations (47) are started from points $w_0 \in \mathcal{W}$ uniformly sampled over the feasible set. In order to generate starting points that are evenly distributed over the n -simplex defining the feasible set, the method described in Shaw [68] is used. For each generated path of the projected GLD iterations, the point with the smallest recorded objective function value is stored. The output from the algorithm \tilde{w} is the point with the smallest overall recorded objective function value. Finally, the optimal solution is taken to be

$$w^{GLD} = \operatorname{argmin}\{\kappa_p(\tilde{w}), \kappa_p(w^{Loc})\}, \tag{59}$$

where w^{Loc} denotes the solution from a local solver started at \tilde{w} . The complete multistart projected GLD algorithm is summarized below. The fixed step size λ is chosen to be 0.01 (based on numerical experimentation) for both the five and 15 asset problems. The temperature parameter β is chosen large enough so that the iterations from the projected GLD algorithm can jump between different local optima. Based on initial experimentation, the following formula for the temperature was chosen

$$\beta = \frac{2\lambda(n + 1)^2}{c^2}, \tag{60}$$

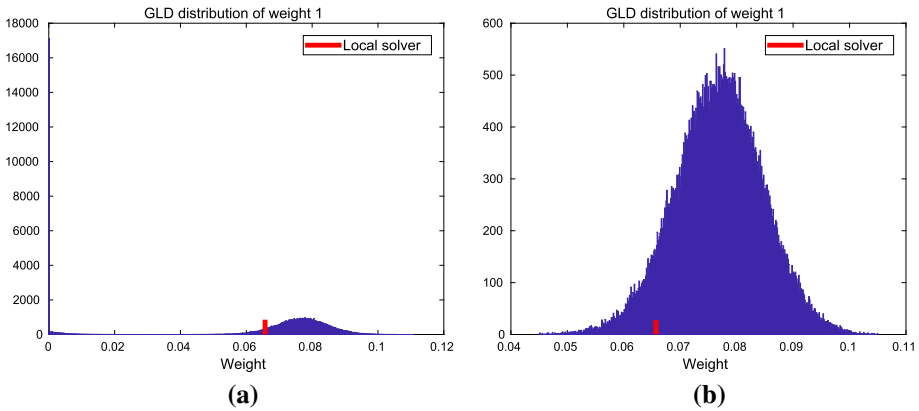


Fig. 10 The distribution of the final iterate for the weight of asset 1 in the 15 asset case: the full distribution (a) and the zoomed in distribution (b)

where $n + 1$ is the number of assets and c was chosen to be 0.06 for the five asset case and 0.1 for the 15 asset case. For the implementation, the number of paths n_{sim} was 10^5 and the number of iterations for each path n_{iter} was 10^4 , implying that 10^9 points in the search space were visited by the algorithm. Given the multistart implementation, the algorithm is very easy to parallelize. The algorithm was parallelized and implemented on a multi-core processor with 24 cores. For the five asset case, the multistart projected GLD algorithm finds a solution with equal weight in four assets and zero weight in one asset. By running the BB algorithm on the same problem it was confirmed that the GLD algorithm finds the global optimum. The runtime for the parallelized algorithm with 24 cores was 2,476 seconds for the five asset case and 8,322 seconds with 15 assets.

Multistart projected GLD algorithm

Input: $\lambda, \beta, n_{sim}, n_{iter}, \hat{M}_2, \hat{M}_4$.
for $i = 1, 2, \dots, n_{sim}$ **do**
 Generate $w_0 \in \mathbb{R}^{n+1}$ uniformly on \mathcal{W} ;
 for $k = 0, 1, \dots, n_{iter}$ **do**
 Generate $\epsilon_{k+1} \sim N(\mathbf{0}, \mathbf{I})$;
 $w_{k+1} = \Pi_{\mathcal{W}}(w_k - \lambda \nabla_w \bar{h}(w_k) + \sqrt{2\lambda/\beta} \epsilon_{k+1})$;
 end
 $w_i^s = \operatorname{argmin}\{\kappa_p(w_0), \kappa_p(w_1), \dots, \kappa_p(w_{n_{iter}})\}$;
end
Output: $\tilde{w} = \operatorname{argmin}\{\kappa_p(w_0^s), \kappa_p(w_1^s), \dots, \kappa_p(w_{n_{sim}}^s)\}$;

For the 15 asset case, the output from the GLD algorithm is a portfolio with equal weights in 14 of the assets and zero weight in one asset. As argued above, this represents a global optimum for the 15 asset problem. As a comparison, a local solver was started from the generated starting points w_0 for each of the n_{sim} outer simulations of the GLD algorithm. The built-in interior point solver in MATLAB was used as local solver. For all of the generated starting points, the output from the local solver is the equally weighted portfolio with non-zero weights in all of the 15 assets. Thus, the multistart projected GLD algorithm jumps

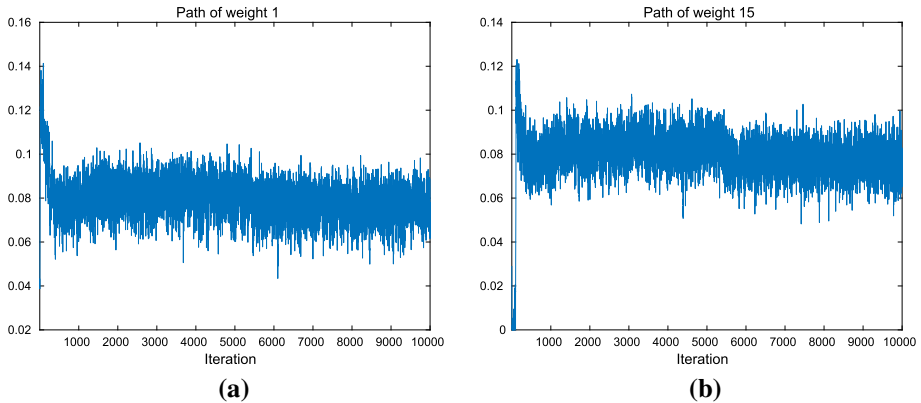


Fig. 11 One of the paths produced by the projected GLD algorithm for the weight of: asset 1 (a) asset 15 (b)

between the local optima and is able to locate the global optimum for the 15 asset problem, whereas a multistart algorithm which uses a local solver finds a local optimum in all cases. The distribution of the final iterate from the GLD algorithm for one of the assets is illustrated in Fig. 10. Figure 10 a displays the full distribution where many of the final iterates are concentrated around zero, whereas Fig. 10 b shows the distribution zoomed in around the non-zero weights. The ability of the GLD algorithm to produce iterates that jump between different local optima is illustrated by the graphs in Fig. 11. In general it is not possible to verify if the output from the GLD algorithm is a global optimum, but the experiments indicate that the algorithm is a useful tool for locating the global optimum for problems for which the number of assets is out of reach for the BB algorithm.

5 Diversification of a typical multi-asset universe–U.S. investor

In this section we apply the introduced novel portfolio diversification framework to an asset universe which is representative of the constituents of a typical U.S. institutional investor portfolio. The asset universe is multi-asset in nature and consists of exposures to U.S. and international equities in developed and emerging markets as well as exposures to property, corporate bonds across both investment grade and high yield, government bonds in developed and emerging markets and also inflation protected securities. For the analysis, each of the 12 asset classes is represented by a suitable index which accurately captures the respective asset class characteristics. The chosen indices are listed in Table 3, which also includes a short description of each index. The asset universe is divided into three broader categories: equities including REITS (EQ), higher yielding credit including emerging markets debt (HY) and government and investment grade bonds (BD). The categorisation can be justified by the correlation structure as well as common practices in the industry. It is likely that the relative weights of these three categories are in practise decided by strategic decisions, constraints and risk appetite. It is therefore not relevant or realistic to compare portfolio construction methodologies that can allocate freely across these categories. In particular, we believe that empirical studies which allow large overweights in bonds (such as a naïve application of risk parity) are unlikely to be useful for practitioners when making forward-looking decisions given the unprecedented bull-run over the last 30 years in bonds and the historically low level

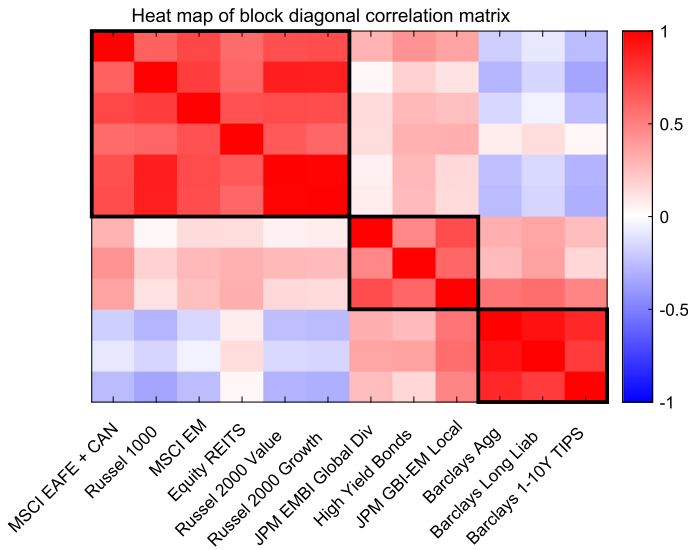


Fig. 12 Heat map of the approximately block diagonal correlation matrix for the 12 indices as of the start of the backtesting period in February 2005. The three blocks represent the equity (EQ), high yield (HY) and bond (BD) portfolios, respectively. Positive and negative correlations are indicated in red and blue, respectively, whereas the magnitudes of the correlations are indicated by the intensity of the colour

of yields available at this point in time. We therefore chose a hierarchical approach, which assumes fixed weights across the three categories in line with the median asset allocation of large U.S. institutional investors (55% in equities, 20% in higher yielding credit and 25% in bonds). The weights within each category are determined by minimizing portfolio kurtosis and through various other portfolio construction methodologies for comparison.

Due to the large estimation errors associated with estimates of higher order moments, each index is assigned a representative kurtosis parameter. The parameter values, given in Table 3, were determined from a combination of estimation from historical data and consistency with estimates reported in the literature, see e.g. Xiong and Idzorek [76]. Historical data reveals, in many cases strong, positive correlations between the indices within the equity, higher yielding credit and bond sub-portfolios, whereas the correlations between indices belonging to different sub-portfolios are comparatively small and time varying. The heat map of the correlation matrix for the asset universe as of February 2005 in Fig. 12 illustrates the approximate block diagonal structure. In the heat map, a red colour represents a positive correlation, whereas a negative correlation is given in blue. The intensity of the colour indicates the magnitude of the positive and negative correlations. From the heat map it is evident that the intra-block correlations are strongly positive, whereas the inter-block correlations are much weaker with both positive and negative values.

The portfolio diversification framework developed in this paper was applied to the multi-asset universe in an out-of-sample analysis. Historical returns expressed in U.S. dollar terms for the 12 indices over the period from September 2001 to September 2020, obtained from Bloomberg and Thomson Reuters DataStream, were used as input to the analysis. We are considering a realistic real world multi-asset allocation problem with semi-annual rebalancing of the portfolio. At each rebalancing date the covariance matrix was estimated from the previous 180 observations of weekly returns using an EWMA estimator with a half-life of one year. Since the first 180 weeks of the return history were used for estimating the

Table 3 The indices that constitute the asset universe of a typical U.S. institutional investor.

Index	Description	Asset number	Portfolio	Kurtosis
MSCI EAFE + Canada	The MSCI EAFE equity index covers developing market countries in Europe, Australasia, Israel, and East Asia.	1	EQ	6
Russel 1000	The large-cap segment of the U.S. equity universe.	2	EQ	6
MSCI EM	Large- and mid-cap segments across emerging markets equities.	3	EQ	6
Equity REITS	Includes all tax-qualified real estate investment trusts (REITs) that are listed on the major U.S. stock exchanges.	4	EQ	6
Russel 2000 Value	The small-cap segment of U.S. value stocks.	5	EQ	6
Russel 2000 Growth	The small-cap segment of U.S. growth stocks.	6	EQ	6
JPM EMBI Global Diversified	Tracks USD denominated bonds issued by EM governments.	7	HY	12
High Yield Bonds	U.S. corporate high yield bond index.	8	HY	12
JPM GBI-EM Local	Tracks local currency bonds issued by EM governments.	9	HY	12
Barclays Aggregated	Includes most of the U.S. traded investment grade bonds.	10	BD	4
Barclays Long Liabilities	Long duration U.S. corporate bonds (10y+).	11	BD	4
Barclays 1-10Y TIPS	Measures the performance of the U.S. TIPS market.	12	BD	8

The multi-asset universe includes broad equity indices, REITS, U.S. and emerging markets high yield bonds, U.S. bonds, liability-driven investments and inflation protected bonds. The data was obtained from Bloomberg and Thomson Reuters DataStream. The table also includes the asset number, sub-portfolio and the chosen representative kurtosis parameters of each asset

covariance matrix for the first rebalancing date in February 2005, there are 32 semi-annual rebalancing dates over the backtesting period.

In Sect. 2.2 it was shown that when the measure of non-Gaussianity ν is given by excess kurtosis, portfolio dimensionality is defined as

$$d_{Z,\nu}(\mathbf{w}) = \frac{\nu(Z)}{\nu(\mathbf{w}^\top \mathbf{r})}, \tag{61}$$

where \mathbf{r} is the return vector, \mathbf{w} is the corresponding weight vector and Z represents the reference asset. For all problem instances encountered over the backtesting period, the minimum portfolio excess kurtosis is always positive. Thus, in this case maximizing portfolio dimensionality is equivalent to minimizing portfolio kurtosis. In the remainder of this section we will therefore denote the minimum kurtosis portfolio as the optimised dimensionality portfolio. As explained in the beginning of this section, the relative weights across the three sub-portfolios were held constant for each rebalancing date and hence the following minimum kurtosis problem was solved for each sub-portfolio and rebalancing date

$$\min_{\mathbf{w} \in \mathcal{W}} \frac{\mathbf{w}^\top \mathbf{M}_4 (\mathbf{w} \otimes \mathbf{w} \otimes \mathbf{w})}{(\mathbf{w}^\top \mathbf{M}_2 \mathbf{w})^2}, \tag{62}$$

where the feasible set is given by $\mathcal{W} = \{\mathbf{w} \in \mathbb{R}^{n_s+1} \mid \sum_{i=0}^{n_s} w_i = 1, w_i \geq 0, i = 0, \dots, n_s\}$, and $n_s + 1$ is the number of assets in sub-portfolio s . As input to the optimization, the moment matrix \mathbf{M}_4 was estimated by simulating 10^7 realisations of the asset returns whose multivariate distribution was modeled by a Gaussian copula and NIG-distributed margins, see Appendix B. Even though the estimated covariance matrix is an input to the portfolio construction at each rebalancing date, in order to be consistent, also the moment matrix \mathbf{M}_2 was estimated from the simulated returns. We chose to model the dependence structure with a Gaussian copula as it allows practitioners to base the model on readily available model data, since existing portfolio diversification models are based on estimated covariance matrices. The model can hence be seen as the simplest possible enhancement to existing portfolio diversification models, where the only further model input that is needed are the marginals for the assets. The minimum kurtosis problem was solved for each sub-portfolio and each rebalancing date over the backtesting period. At each rebalancing date the total portfolio was then constructed by allocating 55% of the capital to the equity sub-portfolio, 20% to the high yield sub-portfolio and 25% to the bond sub-portfolio,

$$\mathbf{w}_{OD}^\top = \left[0.55 \mathbf{w}_{EQ}^\top, 0.2 \mathbf{w}_{HY}^\top, 0.25 \mathbf{w}_{BD}^\top \right], \tag{63}$$

where \mathbf{w}_{OD} , \mathbf{w}_{EQ} , \mathbf{w}_{HY} and \mathbf{w}_{BD} denote the weight vectors for the optimised dimensionality portfolio and the equity-, high yield- and bond sub-portfolios, respectively. Obviously, given that the weights across the sub-portfolios are fixed, the total optimised dimensionality portfolio is likely not optimal for the full 12 asset universe.

As exemplified by the numerical example in Sect. 3.3, the computational time to solve the six asset problem with the BB algorithm is around 14 hours. Since the minimum kurtosis problems were solved for all 32 rebalancing dates, instead of solving the six asset problem with the BB algorithm an alternative strategy was used. The strategy exploits an empirical observation of the minimum kurtosis problem: in the case of non-negative correlations and positive weights that sum to one, there is just one global minimum, no other local minima, and the global minimum can always be found by local optimiser. For such problem instances, extensive empirical runs have failed to generate even a single counter example where the solution from a local solver is different from the solution returned by the BB algorithm. Since

the correlation matrices for each sub-portfolio and rebalancing date only contain positive values, the minimum kurtosis problem instances all have this empirically observed property. This observation was also verified for the problem instances at hand by solving a reduced version of the minimum kurtosis problem where one of the assets from the equity sub-portfolio was removed, resulting in a five asset problem. In a first step, the reduced minimum kurtosis problem was solved with a local solver for each sub-portfolio and rebalancing date, using the same seed for the random number generator when generating the scenarios for calculation of the moment matrices. In the next step, the same sequence of minimum kurtosis problems was solved using the BB algorithm and the same seed for the random number generation. This allowed us to confirm that the local solver found the global optimum for all rebalancing dates and all sub-portfolios. This is a strong indication that the local solver finds the global optimum also for the full problem for all problem instances. When solving the minimum kurtosis problems we therefore used the built-in Matlab interior point solver for the six asset problem.

In the analysis, the optimised dimensionality portfolio is compared to a portfolio which is representative of the median U.S. institutional investor, as well as the portfolios obtained from applying four commonly used portfolio construction methodologies to the multi-asset universe: risk parity, diversification ratio, minimum variance and equal weights. The portfolio construction methodologies are applied to the three sub-portfolios separately, constraining the weights to be non-negative and summing to one. The portfolio weights for the median U.S. investor are based on internal estimates from Aberdeen Standard Investments. Naturally, the allocation changes over time but the weights are a good representation of the typical U.S. institutional investor portfolio. As for the optimised dimensionality portfolio, the total portfolio weights for each methodology are finally determined by assigning 55% of the capital to the equity sub-portfolio, 20% to the high yield sub-portfolio and 25% to the bond sub-portfolio. In Bai et al. [8] it is shown that the risk parity portfolio subject to non-negative weights summing to one can be found by solving a convex optimization problem. The risk parity weights for the sub-portfolios are thus determined as the solution to a convex optimization problem which uses the estimated covariance matrix at each rebalancing date as input. As shown in Choueifaty et al. [22], also the problem of maximising the diversification ratio

$$\frac{\mathbf{w}^T \boldsymbol{\sigma}}{\sqrt{\mathbf{w}^T \mathbf{M}_2 \mathbf{w}}}, \quad (64)$$

where $\boldsymbol{\sigma}$ is the vector of asset volatilities, when the weights are constrained to be non-negative and summing to one, is a convex optimisation problem. Since the problem of minimizing portfolio variance $\mathbf{w}^T \mathbf{M}_2 \mathbf{w}$ for a fully invested long-only portfolio is a quadratic program, the portfolios used for comparison with the optimised dimensionality portfolio were either determined by a convex optimisation problem using the estimated covariance matrix as input, or were using fixed weights for all rebalancing dates. The fixed weights of the U.S. institutional portfolio and the equally weighted portfolio are displayed in Table 4.

We now analyse the portfolios obtained from applying the minimum kurtosis as well as the described alternative methodologies to the multi-asset universe for the described out-of-sample time series study. We have chosen a realistic setup with semi-annual rebalancing and fixed weights between the sub-portfolios, reflecting a portfolio allocation process of a typical large U.S. institutional investor. The average weights over time for the six portfolio construction methodologies are given in Table 4 and illustrated in Fig. 13. From the table, one observes that the risk parity portfolio has the most evenly distributed weights out of the dynamic portfolios, whereas the minimum variance methodology produces the most

Table 4 Average weights for the six portfolios obtained from the out-of-sample analysis over the backtesting period from 16 February 2005 to 16 September 2020.

Asset	1 [%]	2 [%]	3 [%]	4 [%]	5 [%]	6 [%]	7 [%]	8 [%]	9 [%]	10 [%]	11 [%]	12 [%]
σ_{mean}	20.7	15.2	17.4	17.8	20.4	21.1	3.3	6.5	6.8	3.4	9.0	4.1
w_{OD}	15.1	2.8	7.3	12.1	7.8	9.9	11.8	6.7	1.5	18.0	7.0	0.0
w_{Inst}	16.5	22.0	5.5	5.5	2.8	2.8	5.0	10.0	5.0	15.0	5.0	5.0
w_{RP}	8.5	10.6	9.7	10.5	8.0	7.7	9.9	5.6	4.5	10.8	4.3	9.9
w_{DR}	15.7	2.7	6.6	12.1	7.9	9.9	12.2	6.8	1.0	1.5	7.1	16.4
w_{MV}	1.2	33.3	5.6	15.0	0.0	0.0	17.4	2.6	0.0	19.3	0.0	5.7
w_{EW}	9.2	9.2	9.2	9.2	9.2	9.2	6.7	6.7	6.7	8.3	8.3	8.3

The average weight vectors for the optimised dimensionality, institutional, risk parity, diversification ratio, minimum variance and equally weighted portfolios are denoted w_{OD} , w_{Inst} , w_{RP} , w_{DR} , w_{MV} and w_{EW} , respectively. The table also includes, for each asset, the average of the annualized volatilities estimated at each rebalancing date. The vector of averaged volatilities is denoted σ_{mean} in the table

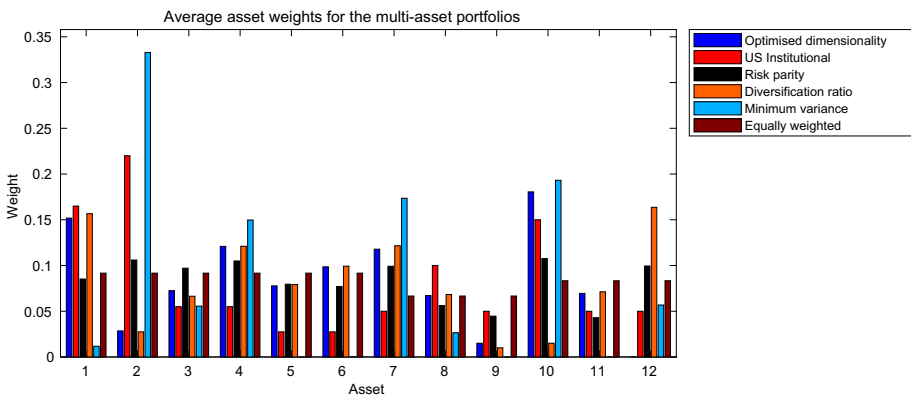


Fig. 13 Average weights for the six portfolios obtained from the out-of-sample analysis over the backtesting period from 16 February 2005 to 16 September 2020

concentrated portfolio. The average weights of the optimised dimensionality portfolio and the diversification ratio portfolio are very similar for the equity and high yield sub-portfolios, whereas the weights differ substantially for the bond portfolio. The weights for these two portfolios are more unevenly distributed than for the risk parity but much less concentrated than the minimum variance portfolio.

Out of the dynamic portfolios, the risk parity has the most stable weights over time, whereas the minimum variance portfolio has the highest turnover. The optimised dimensionality portfolio and the diversification ratio portfolio lie in between the risk parity and minimum variance portfolios in terms of weight stability over time, with the optimised dimensionality portfolio showing the lowest turnover of the two.

The realized performances of the six portfolios over the backtesting period from 16 February 2005 to 16 September 2020 are shown in Fig. 14. The sample period incorporates the periods of the Global Financial Crisis of 2007–2008, the subsequent European debt crisis and the recent sharp decline in asset prices caused by the economic impact of the Corona virus. From the graph it can be observed that the performances of the six portfolios are very similar. Since we are considering a realistic multi-asset portfolio allocation example,

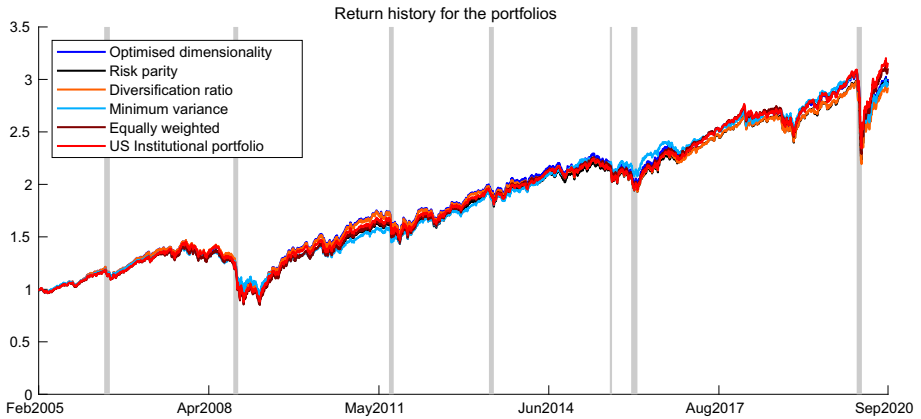


Fig. 14 Realized performance of the six portfolios obtained from the out-of-sample analysis over the backtesting period from 16 February 2005 to 16 September 2020. The grey shaded areas represent the time intervals for the bear market scenarios defined in Table 5

where the weights across sub-portfolios are held constant, we are not expecting dramatically different behaviours across portfolios which is confirmed in the graph. In addition to the portfolio performances, the graph also displays the time intervals for seven historical bear market scenarios as indicated by the grey shaded areas. The definitions of the bear market scenarios are given in Table 5 together with a description of each scenario. As can be seen from the graph in Fig. 14, the most dramatic drop in portfolio value for all of the portfolios was caused by the recent market decline due to the Corona virus, which caused sharp falls in asset prices and liquidity to evaporate.

In this paper we have argued strongly that measures of diversification should be related to the tail properties of portfolio returns and thus introduced the notion of dimensionality. In practise, tail risk is perceived different across investor types depending on institutional mandate, regulatory restrictions or simply risk appetite. We therefore compare portfolio construction techniques not by Sharpe ratio as is commonly done, but across a variety of commonly used measures of tail risk; statistical measures, such as skewness and kurtosis in addition to measures used more in practise, such as maximum drawdown and Expected Shortfall at different confidence levels scaled by the volatility. Instead of using Expected Shortfall directly as a tail risk measure, we consider Expected Shortfall divided by the portfolio volatility

$$\frac{ES_{\alpha}(w)}{\sqrt{w^T M_2 w}}, \tag{65}$$

for three different confidence levels α . The ratio in (65) satisfies our requirement of being a leverage invariant tail risk measure, since Expected Shortfall and portfolio volatility are both homogeneous functions of degree one. For a chosen time interval, the maximum drawdown (MDD) of a portfolio is defined as the maximum observed loss from a peak to a trough, before a new peak is attained. The drawdown (DD) of a portfolio with value $W(t)$ at time $t \geq 0$ is defined as

$$DD(t) = \frac{W(t) - W_{peak}(t)}{W_{peak}(t)}, \quad \text{where } W_{peak}(t) = \max_{\tau \in [0,t]} W(\tau). \tag{66}$$

Table 5 Definitions and descriptions of the bear market scenarios

Scenario	Start date	End date	Description
EM selloff	2006-05-01	2006-06-08	Emerging market crisis with sharp losses in equity and FX markets
Bank meltdown	2008-09-12	2008-10-15	Great financial crisis of 2008 triggered by a house price crisis in the U.S.
Euro crisis	2011-07-22	2011-08-23	Sovereign debt crisis in the Eurozone with fears of defaults and break-up of the currency union
QE jitters	2013-05-22	2013-06-24	The mentioning of an eventual tapering of QE in the U.S. by Fed president Bernanke causes markets to sell off aggressively
China devaluation	2015-08-10	2015-08-25	Surprise devaluation of the Chinese Renminbi causes steep market falls and a commodity price collapse
U.S. arrest	2015-12-31	2016-02-11	Expectations of rate hikes in the U.S. are rapidly scaled back as the macro data deteriorates
Corona virus	2020-02-19	2020-03-25	Covid-19 pandemic causes markets to fall and liquidity to evaporate

The MDD over the time interval $[0, T]$ is then formally defined as

$$MDD_{[0, T]} = \min_{t \in [0, T]} DD(t). \quad (67)$$

The realized tail risk measures, together with the realized mean return, volatility and Sharpe ratio, of the six portfolios over the backtesting period are given in Table 6. For each measure, the value of the best performing portfolio is displayed in bold font. Even though the portfolio weights of most of the dynamic construction methodologies differ substantially, the realized volatility of the aggregated portfolios are of approximately the same magnitude for all of the six portfolios, which can be attributed to the fixed weights across the sub-portfolios. Since the realized mean return of the portfolios are of the same magnitude, also the Sharpe ratio is very similar across methodologies. As we have argued, the diversification properties of the portfolio construction methodologies should be evaluated through realisations of commonly used tail risk measures over the sample period.

In a realistic setting, where rebalancing cannot happen very often (we assume semi-annual rebalancing, as is often the practise) and using fixed weights for the different sub-portfolios, it is unlikely to observe dramatically different behaviour across portfolios and it is also not to be expected that one technique would dominate others on all measures of tail risk. We therefore rank portfolio construction methodologies for each measure of tail risk as demonstrated in Table 7. From the table we conclude that our proposed portfolio construction methodology through dimensionality achieved the best tail properties on average. The table also reveals that the optimised dimensionality portfolio has the fourth highest out-of-sample kurtosis. This can be explained by two main factors. First, the optimised dimensionality portfolio is found by minimizing portfolio kurtosis, given the chosen kurtosis parameters for the individual assets in the portfolio. The realized values of kurtosis for the individual assets over the backtesting period were in many cases far from the chosen parameters. Second, even if the portfolio kurtosis for the optimised dimensionality portfolio is lower for the sub-portfolios, the fact that we use fixed weights between the sub-portfolios can create a sub-optimal overall portfolio

Table 6 Out-of-sample realized mean return, volatility, Sharpe ratio and six measures of tail risk for the six portfolios over the backtesting period from 16 February 2005 to 16 September 2020.

	Optimised dimensionality	Risk parity	Diversification ratio	Minimum variance	Equally weighted	U.S. institutional
Mean return	7.00%	6.95%	6.82%	6.92%	7.22%	7.33%
Volatility	10.07%	10.02%	10.15%	9.81%	10.20%	9.70%
Cash rate	1.70%	1.70%	1.70%	1.70%	1.70%	1.70%
Sharpe ratio	0.53	0.52	0.50	0.53	0.54	0.58
Skewness	−0.87	−0.85	−0.86	−0.60	−0.90	−0.83
Kurtosis	17.14	16.95	16.79	20.47	17.06	17.41
$ES_{0,90}/\sigma$	1.78	1.79	1.78	1.78	1.78	1.79
$ES_{0,95}/\sigma$	2.43	2.45	2.44	2.47	2.44	2.45
$ES_{0,99}/\sigma$	4.69	4.75	4.70	4.75	4.75	4.80
Maximum drawdown	−38.96%	−38.80%	−39.53%	−32.62%	−40.01%	−39.74%

The realized value of the best performing portfolio for each measure is shown in bold font. Even though it doesn't show in the table due to rounding, the optimised dimensionality portfolio obtained the lowest realized value of $ES_{0,90}/\sigma$

Table 7 Tail risk rankings of the six portfolios for each of the six tail risk measures

	Optimised dimensionality	Risk parity	Diversification ratio	Minimum variance	Equally weighted	U.S. institutional
Skewness	2	4	3	6	1	5
Kurtosis	3	5	6	1	4	2
$ES_{0,90}/\sigma$	6	2	4	5	3	1
$ES_{0,95}/\sigma$	6	2	5	1	4	3
$ES_{0,99}/\sigma$	6	2	5	4	3	1
Maximum drawdown	4	5	3	6	1	2
Average ranking	4.50	3.33	4.33	3.83	2.67	2.33

The portfolios are ranked from 1 to 6 for each tail risk measure, where 6 corresponds to the best value and 1 to the worst value. The best value for each tail risk measure and for the aggregate is shown in bold font

in terms of kurtosis. The small difference in average tail risk ranking between the optimised dimensionality and the diversification ratio portfolios can be attributed to the fact that the two methodologies produce very similar weights for the equity- and high yield sub-portfolios. If more precise estimates of tail properties are known and used as input, the results of the optimised dimensionality portfolio can be refined.

In Table 8 we complement the analysis by investigating the performance of the portfolio construction methodologies over the bear market scenarios defined in Table 5. As can be observed in the table, the minimum variance portfolio realized losses of the smallest magnitude for all but one of the bear market scenarios. This is an indication that with perfect foresight, one should concentrate the allocation to the assets with the lowest volatility during periods of extreme market stress. However, in practise it is of course not possible to perfectly forecast the timing of these periods. As demonstrated through the realized tail risk measures, the best overall tail risk properties over the full backtesting period, which includes several periods of extreme market stress, is obtained via the optimised dimensionality methodology.

Table 8 Returns for the six portfolios over the bear market scenarios with start and end dates as defined in Table 5

%Scenario	%Optimised dimensionality	%Risk parity	%Diversification ratio	%Minimum variance	%Equally weighted	%U.S. institutional
EM selloff	-4.45	-4.24	-4.51	-2.61	-4.39	-4.56
Bank meltdown	-19.83	-19.61	-20.36	-15.78	-20.61	-20.59
Euro crisis	-8.42	-7.89	-8.58	-6.42	-8.12	-7.76
QE jitters	-6.77	-6.62	-7.08	-5.37	-6.82	-7.07
China devaluation	-4.82	-4.86	-4.94	-3.97	-5.00	-5.23
U.S. arrest	-6.77	-6.49	-6.88	-4.52	-6.77	-5.92
Corona virus	-18.33	-17.10	-18.24	-16.89	-17.68	-16.36

The return of the best performing portfolio in each scenario is shown in bold font

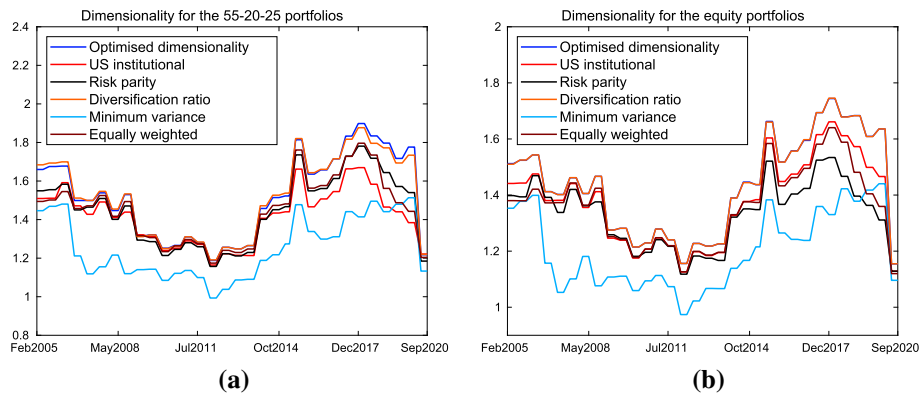


Fig. 15 **a** Dimensionality of the six aggregated 55-20-25 portfolios over the backtesting period from 16 February 2005 to 16 September 2020. **b** Dimensionality of the six equity portfolios over the same period

The dimensionalities of the aggregated portfolio, as well as of the three sub-portfolios, were measured for all of the portfolio construction methodologies for each rebalancing date over the backtesting period. The excess kurtosis $\nu(Z)$ of the reference asset in the portfolio dimensionality expression (61) was chosen to be representative of the respective sub-portfolio. Hence, the excess kurtosis for the reference asset when measuring portfolio dimensionality for the equity-, high yield- and bond portfolios were chosen to be 3, 9, and 1, respectively. For the aggregated portfolio we chose an excess kurtosis of 3 for the reference asset, allowing us to interpret the dimensionalities of the aggregated portfolios as the equivalent number of independent equity exposures. In Figs. 15 and 16, the measured dimensionalities of the aggregated, as well as the three sub-portfolios, are displayed for all six portfolio construction methodologies. From the graphs in Figs. 15b and 16a, one can observe that, for the equity- and high yield sub-portfolios, the dimensionalities of the optimised dimensionality and the diversification ratio portfolios are very close, and higher than for the other portfolios, for all dates over the backtesting period. For the bond sub-portfolio, the optimised dimensionality portfolio dominates the other portfolios over the full sample period in terms of measured dimensionality.

Since the weights across sub-portfolios are fixed, the dimensionality obtained with the optimised dimensionality methodology is not guaranteed to be optimal for the aggregate

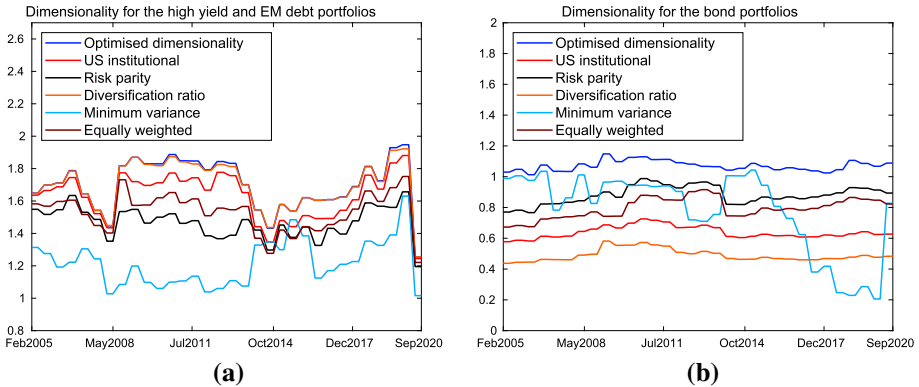


Fig. 16 **a** Dimensionality of the six high yield portfolios over the backtesting period from 16 February 2005 to 16 September 2020. **b** Dimensionality of the six bond portfolios over the same period

portfolio. This can be observed in Fig. 15a, where the dimensionality of the diversification ratio portfolio dominates all other portfolios for some dates at the beginning of the backtesting period. The graph in Fig. 15a reveals that the dimensionalities of five of the aggregate portfolios are very close during bear market periods when correlations between asset classes increase, indicating that opportunities for diversification decrease in such circumstances.

To conclude, we have, using a realistic setup with fixed weights across sub-portfolios and semi-annual rebalancing, demonstrated the usefulness of our proposed methodology based on dimensionality. Compared to four commonly used portfolio construction methodologies and the U.S. institutional median portfolio, the optimised dimensionality portfolio showed the best overall tail risk properties over the sample period. Finally we observe that an easily interpretable statistic, such as portfolio dimensionality, has the additional advantage of being readily explainable and being a relevant statistic in itself. As a final remark, we mention that it can be very beneficial to extend the investment universe if dimensionality optimization or tail risk mitigation is the objective. The investment universe could be extended with strategies that are not driven by traditional risk premia, but utilise structural or behavioral effects in the markets. Examples of such strategies are momentum and low beta strategies which show low correlations to traditional asset classes, see e.g. Jegadeesh and Titman [45], Asness et al. [5], and Frazzini and Pedersen [34].

6 Conclusions

In this paper we have introduced a portfolio diversification framework based on a novel measure called portfolio dimensionality. This measure is directly related to the tail risk of the portfolio and it is leverage invariant, which means that it can typically be expressed as the ratio of convex functions. In order to solve the global optimization problem that arises when minimizing portfolio kurtosis, two complementary global optimization algorithms have been formulated, one deterministic BB algorithm and one stochastic GLD algorithm. Solving the problem with the BB algorithm, one can guarantee that the global optimum has been found. However, it suffers from the curse of dimensionality which limits the size of the problem when the BB algorithm is used for the optimization. A complementary stochastic optimization algorithm for the global optimization problem has therefore been formulated. As illustrated in Sect. 4.3, the multistart projected GLD algorithm can find the global optimum for cases

when a multistart local solver algorithm does not. The projected GLD algorithm therefore complements the BB algorithm and allows for solving problems in higher dimensions, albeit without the guarantee that the global optimum will be found. An alternative solution method could be to run the BB algorithm for a fixed number of iterations when solving larger problems. Empirically we observed that the vast majority of the solution time for the BB algorithm is spent on proving optimality for a point found early on. This is a heuristic method that may be used instead of the projected GLD algorithm when solving larger problems. Furthermore, we observed empirically that for problem instances where all correlations are positive, a local solver finds the global optimum as verified by the BB algorithm.

Our introduced framework extends the diversification frameworks in the literature that are based on only the covariance matrix. Through numerical experiments we have illustrated that our framework possess desirable properties as introduced in the portfolio diversification literature. This can be contrasted to commonly used diversification frameworks such as risk parity and the most diversified portfolio. In order to avoid the problem of obtaining robust estimates of asymmetric tail dependencies between asset returns (see [33]), we have in this paper chosen to model the dependence structure with a Gaussian copula. It is possible to extend the framework to also taking dynamic volatilities and correlations as well as non-linear dependence into account. The model can be extended by using a dynamic GARCH model with skewed and leptokurtic innovations for the marginal distributions, as well as a dynamic conditional correlation model for the copula correlations, see Engle [30]. Furthermore, in order to capture the asymmetric tail dependence observed in the financial markets, a skewed t copula can be used as in Christoffersen et al. [23]. Alternatively, a non-linear dependence structure can be modelled with regime shifts as in Ang and Bekaert [3].

Acknowledgements This work was supported by two EPSRC Impact Acceleration Account grants [EPSRC IAA PIII011 and EPSRC P049].

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Useful notation for higher order portfolio moments

Given the weight vector $\mathbf{w} \in \mathbb{R}^{n \times 1}$ of relative portfolio weights and the random vector $\mathbf{r} \in \mathbb{R}^{n \times 1}$ containing the returns of the assets in the portfolio, the third and fourth central moments of the portfolio return r_p are given by

$$\mu_3 = \mathbb{E}((r_p - \mu_p)^3) = \mathbb{E}((\mathbf{w}^\top (\mathbf{r} - \boldsymbol{\mu}))^3) = \mathbf{w}^\top \mathbf{M}_3 (\mathbf{w} \otimes \mathbf{w}), \text{ and} \tag{68}$$

$$\mu_4 = \mathbb{E}((r_p - \mu_p)^4) = \mathbb{E}((\mathbf{w}^\top (\mathbf{r} - \boldsymbol{\mu}))^4) = \mathbf{w}^\top \mathbf{M}_4 (\mathbf{w} \otimes \mathbf{w} \otimes \mathbf{w}), \tag{69}$$

where \mathbf{M}_3 and \mathbf{M}_4 denote the third and fourth co-moment matrices, respectively, \otimes denotes the Kronecker product, $\mu_p = \mathbb{E}(r_p)$ and $\boldsymbol{\mu} = \mathbb{E}(\mathbf{r})$. In reality, the third and fourth co-moments of the asset returns are three and four dimensional tensors with dimensions $n \times n \times n$ and $n \times n \times n \times n$, respectively. In order to make these tensors mathematically tractable we follow Athayde and Flores [6] and Jondeau and Rockinger [46], among others, and convert these

tensors into two-dimensional matrices. The $n \times n^2$ third co-moment matrix \mathbf{M}_3 in (68) is defined as

$$\mathbf{M}_3 = \mathbb{E} \left((\mathbf{r} - \boldsymbol{\mu})(\mathbf{r} - \boldsymbol{\mu})^\top \otimes (\mathbf{r} - \boldsymbol{\mu})^\top \right) = \{s_{ijk}\}, \tag{70}$$

where

$$s_{ijk} = \mathbb{E} \left((r_i - \mu_i)(r_j - \mu_j)(r_k - \mu_k) \right), \quad i, j, k = 1, \dots, n. \tag{71}$$

The third co-moment matrix can be written in the following block matrix form

$$\mathbf{M}_3 = [\mathbf{S}_{1jk} \ \mathbf{S}_{2jk} \ \dots \ \mathbf{S}_{njk}], \tag{72}$$

where the $n \times n$ matrices \mathbf{S}_{ijk} are given by

$$\mathbf{S}_{ijk} = \begin{bmatrix} s_{i11} & s_{i12} & \dots & s_{i1n} \\ s_{i21} & s_{i22} & \dots & s_{i2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{in1} & s_{in2} & \dots & s_{inn} \end{bmatrix}, \quad i = 1, \dots, n. \tag{73}$$

Similarly, the $n \times n^3$ fourth co-moment matrix \mathbf{M}_4 in (69) is defined as

$$\mathbf{M}_4 = \mathbb{E} \left((\mathbf{r} - \boldsymbol{\mu})(\mathbf{r} - \boldsymbol{\mu})^\top \otimes (\mathbf{r} - \boldsymbol{\mu})^\top \otimes (\mathbf{r} - \boldsymbol{\mu})^\top \right) = \{k_{ijkl}\},$$

where

$$k_{ijkl} = \mathbb{E} \left((r_i - \mu_i)(r_j - \mu_j)(r_k - \mu_k)(r_l - \mu_l) \right), \quad i, j, k, l = 1, \dots, n, \tag{74}$$

and the block matrix form is given by

$$\mathbf{M}_4 = [\mathbf{K}_{11kl} \ \mathbf{K}_{12kl} \ \dots \ \mathbf{K}_{1nkl} \mid \mathbf{K}_{21kl} \ \dots \ \mathbf{K}_{2nkl} \mid \dots \mid \mathbf{K}_{n1kl} \ \dots \ \mathbf{K}_{nnkl}], \tag{75}$$

where the $n \times n$ matrices \mathbf{K}_{ijkl} are given by

$$\mathbf{K}_{ijkl} = \begin{bmatrix} k_{ij11} & k_{ij12} & \dots & k_{ij1n} \\ k_{ij21} & k_{ij22} & \dots & k_{ij2n} \\ \vdots & \vdots & \ddots & \vdots \\ k_{ijn1} & k_{ijn2} & \dots & k_{ijnn} \end{bmatrix}, \quad i, j = 1, \dots, n. \tag{76}$$

The dimension of each block in (75) is given by

$$[\mathbf{K}_{i1kl} \ \mathbf{K}_{i2kl} \ \dots \ \mathbf{K}_{inkl}] \in \mathbb{R}^{n \times n^2}, \quad i = 1, \dots, n, \tag{77}$$

and hence $\mathbf{M}_4 \in \mathbb{R}^{n \times n^3}$. As is apparent from Eqs. (71) and (74), the matrices \mathbf{M}_3 and \mathbf{M}_4 contain certain symmetries, which means that not all elements need to be explicitly computed. The number of unique elements in \mathbf{M}_3 is $n(n + 1)(n + 2)/6$ and \mathbf{M}_4 contains $n(n + 1)(n + 2)(n + 3)/24$ unique elements (see e.g. [46]). The number of unique elements in \mathbf{M}_3 and \mathbf{M}_4 , respectively, for different portfolio sizes are summarized in Table 9. As is apparent from the table the number of unique elements grows dramatically with portfolio size. This leads to large estimation errors when attempting to estimate the co-moments from historical return data. In the literature this curse of dimensionality is typically handled by assuming that the asset returns are generated by a factor model through which the number of parameters to be estimated is considerably reduced. Examples from the literature are Martinelli and Ziemann [56], who use a single factor model, and Boudt et al. [15], who use a multi-factor model.

Table 9 Number of unique elements in \mathbf{M}_3 and \mathbf{M}_4 for different portfolio sizes

Portfolio size	2	3	4	10	50	100
Number of unique elements in \mathbf{M}_3	4	10	20	220	22,100	171,700
Number of unique elements in \mathbf{M}_4	5	15	35	715	292,825	4,421,275

The third and fourth central moments of the portfolio return are homogeneous functions of degree three and four, respectively. Their gradients can be found by applying Euler’s theorem for positively homogeneous functions

$$\mathbf{w}^\top \nabla_{\mathbf{w}} \mu_3 = 3\mu_3 = \mathbf{w}^\top 3\mathbf{M}_3(\mathbf{w} \otimes \mathbf{w}), \text{ and} \tag{78}$$

$$\mathbf{w}^\top \nabla_{\mathbf{w}} \mu_4 = 4\mu_4 = \mathbf{w}^\top 4\mathbf{M}_4(\mathbf{w} \otimes \mathbf{w} \otimes \mathbf{w}), \tag{79}$$

and hence

$$\nabla_{\mathbf{w}} \mu_3 = 3\mathbf{M}_3(\mathbf{w} \otimes \mathbf{w}), \text{ and} \tag{80}$$

$$\nabla_{\mathbf{w}} \mu_4 = 4\mathbf{M}_4(\mathbf{w} \otimes \mathbf{w} \otimes \mathbf{w}), \tag{81}$$

where each component of the gradients are homogeneous functions of degree two and three, respectively. Letting f_i denote the i th component of $\nabla_{\mathbf{w}} \mu_3$ and applying Euler’s theorem for homogeneous functions yields

$$\nabla_{\mathbf{w}} f_i \mathbf{w}^\top = 2f_i. \tag{82}$$

Letting J denote the Jacobian, the Hessian of μ_3 is given by

$$\nabla_{\mathbf{w}}^2 \mu_3 = (J(\nabla_{\mathbf{w}} \mu_3))^\top = \left[\frac{\partial \nabla_{\mathbf{w}} \mu_3}{\partial w_1} \quad \frac{\partial \nabla_{\mathbf{w}} \mu_3}{\partial w_2} \quad \dots \quad \frac{\partial \nabla_{\mathbf{w}} \mu_3}{\partial w_n} \right]^\top = [\nabla_{\mathbf{w}} f_1 \quad \nabla_{\mathbf{w}} f_2 \quad \dots \quad \nabla_{\mathbf{w}} f_n]. \tag{83}$$

Since the Hessian is symmetric one obtains

$$\nabla_{\mathbf{w}}^2 \mu_3 \mathbf{w} = \begin{bmatrix} (\nabla_{\mathbf{w}} f_1)^\top \\ (\nabla_{\mathbf{w}} f_2)^\top \\ \vdots \\ (\nabla_{\mathbf{w}} f_n)^\top \end{bmatrix} \mathbf{w} = 2 \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} = 2\nabla_{\mathbf{w}} \mu_3, \tag{84}$$

and thus

$$\nabla_{\mathbf{w}}^2 \mu_3 = 6\mathbf{M}_3(\mathbf{w} \otimes \mathbf{I}_n), \tag{85}$$

where \mathbf{I}_n denotes the $n \times n$ identity matrix. Similarly, the Hessian of μ_4 is given by

$$\nabla_{\mathbf{w}}^2 \mu_4 = 12\mathbf{M}_4(\mathbf{w} \otimes \mathbf{w} \otimes \mathbf{I}_n). \tag{86}$$

B Generating the return distribution

In this section we describe a meta-Gaussian distribution (see [57]) defined as the combination of the Gaussian copula and arbitrary distributions for the margins. The meta-Gaussian distribution is in this paper used to generate sample returns from a multivariate distribution with a desired linear dependence structure and marginal distributions with arbitrary skewness and

kurtosis parameters. The Normal Inverse Gaussian (NIG) distribution is used for generating marginal distributions with possibly different skewness and kurtosis parameters.

The NIG distribution

The NIG distribution is a special case of the Generalized Hyperbolic distribution introduced by Barndorff–Nielsen [11]. It was introduced by Barndorff–Nielsen [12] and is commonly used in financial applications to model skewed and leptokurtic distributions. The univariate probability density function (PDF) of the NIG distribution can be expressed as

$$f_X(x) = \frac{\delta\alpha \exp\left(\delta\sqrt{\alpha^2 - \beta^2}\right) K_1\left(\alpha\sqrt{\delta^2 + (x - \mu)^2}\right)}{\pi\sqrt{\delta^2 + (x - \mu)^2}}, \tag{87}$$

where $\delta > 0$, $0 \leq |\beta| < \alpha$, and K_1 is the modified Bessel function of the third kind of order 1, see Abramowitz and Stegun [1]. The parameters μ and δ determine the location and scale, respectively, while α and β control the shape of the density. In particular, $\beta = 0$ corresponds to a symmetric distribution. The mean, variance, skewness and kurtosis of the NIG distribution are given by

$$\begin{aligned} \mathbb{E}(X) &= \mu + \frac{\delta\beta}{\sqrt{\alpha^2 - \beta^2}}, \quad \text{Var}(X) = \frac{\delta\alpha^2}{(\alpha^2 - \beta^2)^{3/2}}, \quad \gamma = 3\frac{\beta}{\alpha} \frac{1}{\delta^{1/2}(\alpha^2 - \beta^2)^{1/4}}, \quad \text{and} \\ \kappa &= 3 + 3\left(1 + 4\left(\frac{\beta}{\alpha}\right)^2\right) \frac{1}{\delta(\alpha^2 - \beta^2)^{1/2}}. \end{aligned} \tag{88}$$

From the condition $0 \leq |\beta| < \alpha$, it follows that the skewness-kurtosis bound $\gamma^2 < 3(\kappa - 3)/5$, must hold for the NIG distribution.

The Gaussian copula

Sklar’s Theorem [71] shows that all multivariate cumulative distribution functions (CDFs) contain copulas and that copulas may be used together with univariate CDFs in order to construct multivariate CDFs. A formulation of Sklar’s theorem, taken from McNeil et al. [57], is given below.

Theorem B.1 (Sklar 1959) *Let F be a joint CDF with marginal CDFs F_1, \dots, F_d . Then there exists a copula $C : [0, 1]^d \rightarrow [0, 1]$ such that, for all $x_1, \dots, x_d \in (-\infty, \infty)$,*

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)). \tag{89}$$

If the margins are continuous, then C is unique. Conversely, if C is a copula and F_1, \dots, F_d are univariate CDFs, then the function F defined in (89) is a joint CDF with margins F_1, \dots, F_d .

Thus, modelling the multivariate return distribution with copulas allows for separating the modelling of the dependence structure and the marginal distributions. In particular, it allows for modelling of the marginal asset return distributions with differing skewness and kurtosis. In this paper, the Gaussian copula is used for modelling of the dependence structure between the asset returns. Let $\Phi_{\mathbf{R}}$ denote the joint CDF of the d -dimensional normally distributed

$X \sim N_d(\mathbf{0}, \mathbf{R})$, where $\mathbf{R} \in \mathbb{R}^{d \times d}$ denotes the correlation matrix. The Gaussian copula is then given by

$$C_{\mathbf{R}}^{\text{Ga}}(u_1, \dots, u_d) = \Phi_{\mathbf{R}}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)), \tag{90}$$

where Φ denotes the standard univariate Gaussian CDF.

Adjusting the input correlation matrix

When generating simulated returns from the meta-Gaussian distribution, one needs to take into account that the realized correlation matrix of the generated returns depends on the copula as well as the marginal return distributions. From Hoeffding’s covariance identity [40], the covariance between two random variables X and Y can be expressed as

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F_{X,Y}(x, y) - F_X(x)F_Y(y)) dx dy, \tag{91}$$

where $F_{X,Y}(\cdot, \cdot)$ denotes the joint CDF and $F_X(\cdot), F_Y(\cdot)$ denote the two marginal CDFs. Using Sklar’s theorem, Eq. (91) can be expressed as

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (C(F_X(x), F_Y(y)) - F_X(x)F_Y(y)) dx dy, \tag{92}$$

where C for the meta-Gaussian distribution is given by the Gaussian copula. Let $C_{\rho_{\text{in}}}^{\text{Ga}}$ denote the two-dimensional Gaussian copula with linear correlation parameter ρ_{in} . The correlation between the two variables X and Y as a function of the correlation parameter ρ_{in} is then given by

$$\rho_{\text{out}}(\rho_{\text{in}}) = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (C_{\rho_{\text{in}}}^{\text{Ga}}(F_X(x), F_Y(y)) - F_X(x)F_Y(y)) dx dy}{\sqrt{\text{Var}(X)\text{Var}(Y)}}. \tag{93}$$

For the generation of samples with the desired realized linear correlations between returns, the input correlation matrix for the Gaussian copula is used by inverting (93). Its entries are determined by numerical integration as, in general, no analytic solution for the inverse exists.

Generating returns from the meta-Gaussian distribution

The procedure for generating returns from the meta-Gaussian distribution is given below.

- (1) Given the marginal distributions and the Gaussian copula, find the input correlation matrix \mathbf{R}_{in} by numerical inversion of Eq. (93).
- (2) Generate $\mathbf{Z} \in N_d(\mathbf{0}, \mathbf{R}_{\text{in}})$.
- (3) Calculate $\mathbf{U} = [\Phi(Z_1), \dots, \Phi(Z_d)]^{\top}$. The CDF of the random vector \mathbf{U} is given by $C_{\mathbf{R}_{\text{in}}}^{\text{Ga}}$.
- (4) Use quantile transformation to obtain $\mathbf{X} = [F_1^{-1}(U_1), \dots, F_d^{-1}(U_d)]^{\top}$ by numerical inversion of the marginal CDFs. The random vector \mathbf{X} has marginal CDFs F_1, \dots, F_d and multivariate CDF $C_{\mathbf{R}_{\text{in}}}^{\text{Ga}}(F_1(x_1), \dots, F_d(x_d))$.

References

1. Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions. Dover, New York (1972)
2. Ahmad, S., Welker, V.: On partial barycentric subdivision. *RM* **73**, 21 (2018)
3. Ang, A., Bekaert, G.: International asset allocation with regime shifts. *Rev. Financ. Stud.* **15**, 1137–1187 (2002)
4. Artzner, P., Delbaen, F., Eber, J., Heath, D.: Coherent Measures of Risk. *Math. Financ.* **9**, 203–228 (1999)
5. Asness, C.S., Moskowitz, T.J., Pedersen, L.H.: Value and momentum everywhere. *J. Finance* **68**, 929–985 (2013)
6. Athayde, G., Flores, R.: Incorporating skewness and kurtosis in portfolio optimization: A multidimensional efficient set. In: *Advances in Portfolio Construction and Implementation*, edited by S. Satchell and A. Scowcroft, Elsevier (2003)
7. Avriel, M., Diewert, W.E., Schaible, S., Zang, I.: Generalized Concavity. Plenum Press, New York (1988)
8. Bai, X., Scheinberg, K., Tutuncu, R.: Least-squares approach to risk parity in portfolio selection. *Quant. Finance* **16**, 357–376 (2016)
9. Bao, Y.: On sample skewness and kurtosis. *Economet. Rev.* **32**, 415–448 (2013)
10. Barbieri, A., Dubikovskiy, V., Gladkevich, A., Goldberg, L.R., Hayes, M.Y.: Central limits and financial risk. *Quant. Finance* **10**, 1091–1097 (2010)
11. Barndorff-Nielsen, O.E.: Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. A. Math. Phys. Sci.* **353**, 401–419 (1977)
12. Barndorff-Nielsen, O.E.: Normal Inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Stat.* **24**, 1–13 (1997)
13. Benson, H.P.: Fractional programming with convex quadratic forms and functions. *Eur. J. Oper. Res.* **173**, 351–369 (2006)
14. Benson, H.P.: Maximizing the ratio of two convex functions over a convex set. *Nav. Res. Logist.* **53**, 309–317 (2006)
15. Boudt, K., Lu, W., Peeters, B.: Higher order comoments of multifactor models and asset allocation. *Financ. Res. Lett.* **13**, 225–233 (2015)
16. Bubeck, S., Eldan, R., Lehec, J.: Sampling from a log-concave distribution with projected Langevin Monte Carlo. *Discrete Comput. Geom.* **59**, 757–783 (2018)
17. Charnes, A., Cooper, W.: Programming with linear fractional functionals. *Nav. Res. Logist.* **9**, 181–186 (1962)
18. Chau, H.N., Kumar, C., Rásonyi, M., Sabanis, S.: On fixed gain recursive estimators with discontinuity in the parameters. *ESAIM: PS* **23**, 217–244 (2019)
19. Chau, H.N., Moulines, E., Rásonyi, M., Sabanis, S., Zhang, Y.: On stochastic gradient Langevin dynamics with dependent data streams: the fully non-convex case. In: *press, SIAM Journal on Mathematics of Data Science* (2021)
20. Chen, Y., Ye, X.: Projection onto a simplex. *arXiv preprint*, (2011)
21. Choueifaty, Y., Coignard, Y.: Toward maximum diversification. *J. Portf. Manag.* **35**, 40–51 (2008)
22. Choueifaty, Y., Froidure, T., Reynier, J.: Properties of the most diversified portfolio. *J. Invest. Strategies* **2**, 49–70 (2013)
23. Christoffersen, P., Errunza, V., Langlois, H.: Is the potential for international diversification disappearing? A dynamic copula approach. *Rev. Financ. Stud.* **25**, 3711–3751 (2012)
24. Clarke, R., De Silva, H., Thorley, S.: Risk parity, maximum diversification, and minimum variance: an analytic perspective. *J. Portf. Manag.* **39**, 39–53 (2013)
25. Cont, R.: Empirical Properties of Asset Returns Stylized Facts and Statistical Issues. *Quant. Finance* **1**, 223–236 (2001)
26. Dalalyan, A.S.: Theoretical guarantees for approximate sampling for smooth and log-concave densities. *J. Roy. Stat. Soc.* **79**, 651–676 (2017)
27. Dinkelbach, W.: On nonlinear fractional programming. *Manage. Sci.* **13**, 492–498 (1967)
28. Durmus, A., Moulines, E.: Nonasymptotic convergence analysis for the Unadjusted Langevin Algorithm. *Ann. Appl. Probab.* **27**, 1551–1587 (2017)
29. Durmus, A., Moulines, E.: High-dimensional Bayesian inference via the Unadjusted Langevin Algorithm. *Bernoulli* **25**, 2854–2882 (2019)
30. Engle, R.: Dynamic conditional correlation: A simple class of multivariate GARCH models. *J. Bus. Econom. Stat.* **20**, 339–350 (2002)
31. Fleming, B., Kroeske, J.: Diversification and the distribution of portfolio variance, part 1: Sums of iid variables and higher-order moments. Available at SSRN, (2017)
32. Floudas, C.A.: *Deterministic Global Optimization - Theory, Methods and Applications*. Springer, US (2000)

33. Frahm, G., Junker, M., Schmidt, R.: Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance Math. Econom.* **37**, 80–100 (2005)
34. Frazzini, A., Pedersen, L.H.: Betting against beta. *J. Financ. Econ.* **111**, 1–25 (2014)
35. Gelfand, S.B., Mitter, S.K.: Recursive stochastic algorithms for global optimization in \mathbb{R}^d . *SIAM J. Control Optim.* **29**, 999–1018 (1991)
36. Gendreau, M., Potvin, J.: *Handbook of Metaheuristics*. Springer, New York (2010)
37. Glover, F.: Future paths for integer programming and links to artificial intelligence. *Comput. Oper. Res.* **13**, 533–549 (1986)
38. Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*. The MIT Press, Cambridge, MA (2016)
39. Gotoh, J.Y., Konno, H.: Maximization of the ratio of two convex quadratic functions over a polytope. *Comput. Optim. Appl.* **20**, 43–60 (2001)
40. Hoeffding, W.: Masstabinvariante Korrelationstheorie. *Schrift. des. Math. Semin. und des Inst. für angew. Math. der Univ. Berlin* **5**, 181–233 (1940)
41. Holland, J.H.: *Adapt. Natur. Artific. Syst.* University of Michigan Press, US (1975)
42. Horst, R.: An algorithm for nonconvex programming problems. *Math. Program.* **10**, 312–321 (1976)
43. Horst, R., Tuy, H.: *Global Optim. - Determin. Approaches*. Springer-Verlag, Berlin Heidelberg (1996)
44. Hyvärinen, A., Oja, E.: Independent component analysis: Algorithms and applications. *Neural Netw.* **13**, 411–430 (2000)
45. Jegadeesh, N., Titman, S.: Returns to buying winners and selling losers: Implications for stock market efficiency. *J. Finance* **48**, 65–91 (1993)
46. Jondeau, E., Rockinger, M.: Optimal portfolio allocation under higher moments. *Eur. Financ. Manag.* **12**, 29–55 (2006)
47. Kiefer, J., Wolfowitz, J.: Stochastic estimation of a regression function. *Ann. Math. Stat.* **223**, 462–466 (1952)
48. Kirkpatrick, S., Gelatt, C.D., Jr., Vecchi, M.P.: Optimization by simulated annealing. *Sci.* **220**, 671–680 (1983)
49. Koumou, N.B.G.: Rao's quadratic entropy, risk management and portfolio theory. PhD thesis, Université Laval, (2017)
50. Kushner, H.J., Yin, G.G.: *Stoch. Approx. Algorithms Appl.* Springer-Verlag, New York (1997)
51. Lassance, N., Vrins, F.: Minimum Rényi entropy portfolios. arXiv preprint, (2018)
52. Laurelle, S., Pages, G.: Stochastic approximation with averaging innovation applied to finance. *Monte Carlo Methods Appl.* **18**, 1–51 (2012)
53. Lempérière, Y., Deremble, C., Nguyen, T.T., Seager, P., Potters, M., Bouchaud, J.P.: Risk premia: Asymmetric tail risks and excess returns. *Quant. Finance* **17**, 1–14 (2017)
54. Locatelli, M., Shoen, F.: *Global Optimization - Theory, Algorithms and Applications*, SIAM (2013)
55. Markowitz, H.M.: Portfolio selection. *J. Finance* **7**, 77–91 (1952)
56. Martinelli, L., Ziemann, V.: Improved estimates of higher-order comoments and implications for portfolio selection. *Rev. Financ. Stud.* **23**, 1467–1502 (2010)
57. McNeil, A.J., Frey, R., Embrechts, P.: *Quantitative Risk Management*. Princeton University Press, Princeton, NJ (2005)
58. Meucci, A.: Managing diversification. *RISK* **22**, 74–79 (2009)
59. Qian, E.: Risk parity portfolios. PanAgora Asset Management Research Paper, (2005)
60. Qian, E.: Risk parity and diversification. *J. Invest.* **20**, 119–127 (2011)
61. Raginsky, M., Rakhlin, A., Telgarsky, M.: Non-convex learning via Stochastic Gradient Langevin Dynamics: A nonasymptotic analysis. *Conference on Learning Theory*, pp. 1674–1703 (2017)
62. Robbins, H., Monro, S.: A stochastic approximation method. *Ann. Math. Stat.* **22**, 400–407 (1951)
63. Roncalli, T.: *Introduction to Risk Parity and Budgeting*. CRC Press, Boca Raton FL (2014)
64. Roncalli, T., Weisang, G.: Risk parity portfolios with risk factors. *Quant. Finance* **16**, 377–388 (2016)
65. Sabanis, S., Zhang, Y.: A fully data-driven approach to minimizing CVaR for portfolio of assets via SGLD with discontinuous updating. arXiv preprint, (2020)
66. Schaible, S.: Parameter-free convex equivalent and dual programs of fractional programs. *Z. Oper. Res.* **18**, 187–196 (1974)
67. Schaible, S.: Minimization of ratios. *J. Optim. Theory Appl.* **19**, 347–352 (1976)
68. Shaw, W.T.: Monte Carlo portfolio optimization for general investor risk-return objectives and arbitrary return distributions: A solution for long-only portfolios. arXiv preprint, (2010)
69. Shen, P., Duan, Y., Pei, Y.: A simplicial branch and duality bound algorithm for the sum of convex-convex ratios problem. *J. Comput. Appl. Math.* **223**, 145–158 (2009)
70. Shen, P., Li, W., Bai, X.: Maximizing for the sum of ratios of two convex functions over a convex set. *Comput. Oper. Res.* **40**, 2301–2307 (2013)

71. Sklar, A.: Fonctions de répartition à n dimensions et leurs marges. *Publ. de l'Inst. de Statist. de l'Univ. de Paris* **8**, 229–231 (1959)
72. Spall, J.C.: *Introd. Stoch. n.a Optim.* John Wiley & Sons, Hoboken, NJ (2003)
73. Stoyanov, S.V., Rachev, S.T., Fabozzi, F.J.: Optimal financial portfolios. *Appl. Math. Finance* **14**, 401–436 (2007)
74. Tawarmalani, M., Sahinidis, N.V.: Convex extensions and envelopes of lower semi-continuous functions. *Math. Program.* **93**, 247–263 (2002)
75. Van der Vaart, A.W.: *Asymptot. Stat.* Cambridge University Press, UK (1998)
76. Xiong, J.X., Idzorek, T.M.: The impact of skewness and fat tails on the asset allocation decision. *Financ. Anal. J.* **67**, 23–35 (2011)
77. Xu, P., Chen, J., Zou, D., Gu, Q.: Global convergence of Langevin Dynamics based algorithms for nonconvex optimization. *arXiv preprint*, (2018)
78. Yamamoto, R., Konno, H.: An efficient algorithm for solving convex-convex quadratic programs. *J. Optim. Theory Appl.* **133**, 241–255 (2007)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.