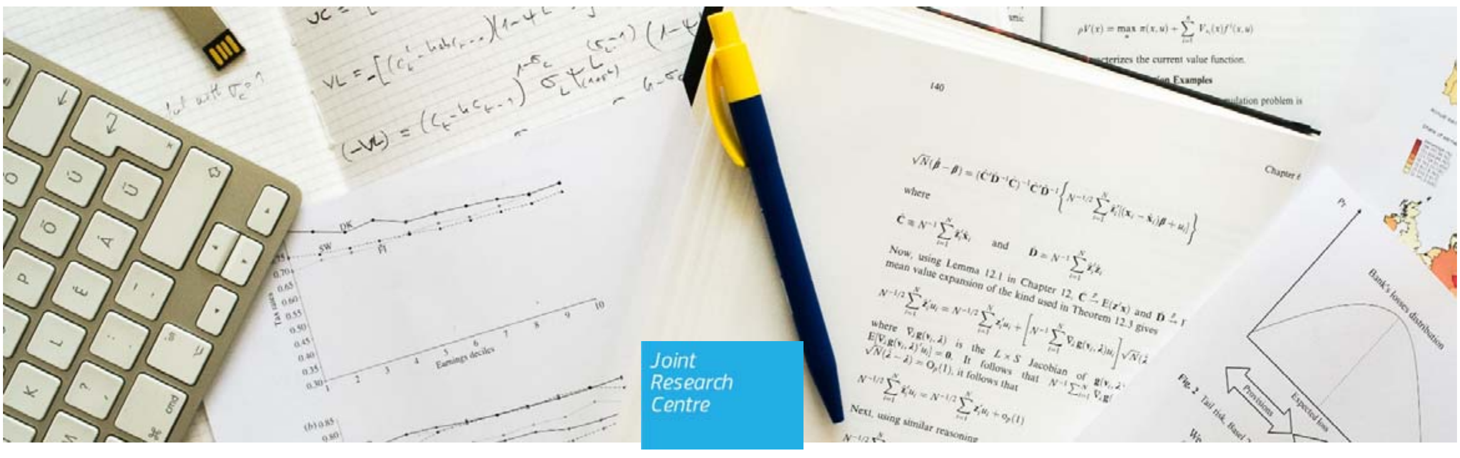
JRC TECHNICAL REPORTS

On the cross-sectional distribution of portfolios' returns

Calès, Ludovic
Chalkis, Apostolos
Emiris, Ioannis Z.

2019

JRC Working Papers in Economics and Finance, 2019/11



Contact information
Name: Ludovic Calès
Address: JRC, via Fermi 2749, 21027 Ispra, Italy
Email: ludovic.cales@ec.europa.eu
Tel.: +39 0332-78-9315

EU Science Hub
https://ec.europa.eu/jrc

JRC116554

PDF ISBN 978-92-76-03963-1 ISSN 2467-2203 doi:10.2760/12907

Luxembourg: Publications Office of the European Union, 2019

How to cite this report: Calès L., Chalkis A., Emiris I.Z., *On the cross-sectional distribution of portfolios' returns*, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-03963-1, doi:10.2760/12907, JRC116554

On the cross-sectional distribution of portfolio returns*

Ludovic Calès[1], Apostolos Chalkis[2], and Ioannis Z. Emiris[2,3]

[1]European Commission, Joint Research Centre, Ispra, Italy
[2]Department of Informatics & Telecommunications
National & Kapodistrian University of Athens, Greece
[3]ATHENA Research & Innovation Center, Greece

May 14, 2019

Abstract

The aim of this paper is to study the distribution of portfolio returns across portfolios and for given asset returns. We focus on the most common type of investment considering portfolios whose weights are non-negative and sum up to 1. We provide algorithms and formulas from computational geometry and the literature on splines to compute the exact values of the probability density function, and of the cumulative distribution function at any point. We also provide closed-form solutions for the computation of its first four moments, and an algorithm to compute the higher moments. All algorithms and formulas allow for equal asset returns.

Keywords: Cross-section of portfolios, Finance, Geometry, B-spline

1 Introduction

The study of the distribution of portfolio returns, across portfolios and for given asset returns, has attracted less attention than it deserves in the finance

*The views expressed are those of the authors and do not necessarily reflect official positions of the European Commission.

literature. However it is a natural tool to understand the relative performance of portfolios, as well as the behavior of asset cross-section and market dynamics in general. Indeed, consider an investment set defined as the set of portfolios in which a manager can invest. The most common is such that the portfolio weights are non-negative and sum up to 1.[1] Let us define the score of a portfolio as the percentage of portfolios, within the investment set, that this portfolio outperforms. For instance, in a market of 3 assets whose returns are 0%, 1% and 1.5%, the score of a portfolio as a function of its return is as given in Figure 1. Here, a portfolio whose return is 0.866% outperforms 50% of the portfolios.

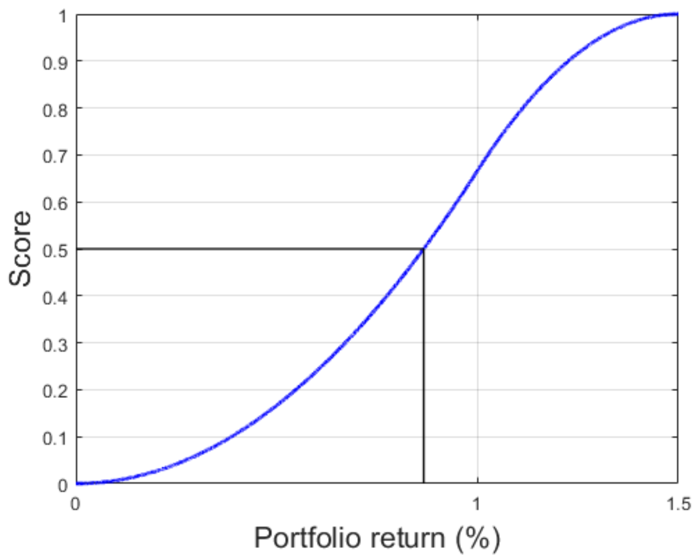


Figure 1: Score of a long-only portfolio in a market of 3 assets whose returns are 0%, 1% and 1.5%.

This portfolio score has been introduced in [Pouchkarev, 2005] and has been used in related studies by the same author: In [Pouchkarev, 2005, Pouchkarev et al., 2004, Hallerbach et al., 2002], the relative performance of value-weighted indices with respect to long-only portfolios is assessed in the Dutch, Spanish and German markets[2]. It leads the authors to question the representativeness of these indices. In [Hallerbach and Pouchkarev, 2005, Hallerbach and Pouchkarev, 2016], the dispersion of the cross-sectional portfolio returns is used to assess the performance of asset managers whose mandate implies

[1]It corresponds to the long-only strategy.

[2]through the MSCI Netherlands 24, IBEX 35, and DAX 30 components, respectively.

tracking error volatility constraints. This score has also been proposed independently in [Billio et al., 2011], and in [Banerjee and Hung, 2011]. In [Billio et al., 2011], this score is used as a portfolio performance measure whose informativeness is shown to be higher than the Sharpe and Sortino ratios in an asset allocation exercise. In [Banerjee and Hung, 2011], a simplified score assigning a reward ranging from -2 to 2 according to the quintile of the score is used to assess the performance of the momentum strategy. This investment strategy is shown to not be outperforming an uninformed naive investor. Recently, in [Calès et al., 2018], the score is used to study the time-varying dependency of portfolios' return and volatility, and relates this dependency to periods of financial turmoils.

In terms of computation, [Pouchkarev, 2005, Theorem 4.2.2] proposes a geometry-based closed form expression of the score. It consists in representing the long only investment set as a simplex. The score is then the volume of the intersection of the simplex and a halfspace. It is computed by decomposing this intersection in smaller simplices. However, this approach is not valid when some asset returns are equal. As a consequence, in [Pouchkarev, 2005] and in related studies, the score is estimated by a quasi-Monte Carlo sampling of the portfolios as described in [Rubinstein and Melamed, 1998]. In [Banerjee and Hung, 2011, Theorem A2], the same approach is considered,[3] for illustration purposes only, and the authors also rely on portfolio sampling for their application. In [Billio et al., 2011], the set of portfolios considered is the specific set of long/short equally weighted zero-dollar portfolios. The estimation of the score relies on combinatorics and order statistics, and it is computationally limited to around 20 assets. Finally, in [Calès et al., 2018], the score is also computed as the volume of the intersection of the simplex and a linear half-space. The authors noticed that an algorithm by [Varsi, 1973] can be used to compute this volume exactly and efficiently for any number of assets, even when some asset returns are equal.

In this paper, we intend to characterize statistically the distribution of the portfolios' returns, the score being its cumulative distribution function (CDF). We provide algorithms and formulas to compute exactly its CDF, its probability density function (PDF) and its moments. We consider the most common investment set, i.e. the set of portfolios whose weights are non-negative and sum up to 1. In Section 2, we formalize the representation

[3]Note that the formula proposed contains a couple of mistakes: the sum is over the number of assets whose returns are lower than the return of the portfolio considered, and the term within parenthesis in the numerator should be the opposite.

of this set as a unit simplex. The portfolios considered are then uniformly distributed over this simplex, allowing us to integrate over it, later on.

In Section 3, we first recall the computation of the CDF as proposed in [Calès et al., 2018]. The set portfolios having same total return is a hyperplane, hence the question consists in computing the volume of the intersection of a simplex and a linear half-space. Based on equivalent results by [Varsi, 1973] using a geometric approach, and [Ali, 1973] with a divided differences approach, the algorithm consists in a recurrence formula. We also propose its computation with a closed-form formula by [Lasserre, 2015] and [Calès, 2019] which considers the case of equal asset returns, as opposed to [Pouchkarev, 2005] and [Banerjee and Hung, 2011]. Next, we compute exactly the PDF. The first approach is based on the geometric interpretation of univariate B-splines by [Curry and Schoenberg, 1966], and it uses the de Boor-Cox recursive formula, see [de Boor, 1972] and [Cox, 1972]. The second approach is a direct derivation of the CDF obtained using the closed-form formula by [Lasserre, 2015]. Finally, since these two approaches suffer of numerical instability in high dimensions, we propose to derive it numerically from Varsi's results.

In Section 4, we derive the moments of the distribution. Our method is based on a result by [Lasserre and Avrachenkov, 2001] which provides an elegant way to integrate symmetric q-linear forms on a simplex. It allows us to propose closed-form solutions for the first four moments, and an algorithm to compute higher moments.

2 Geometric representation of the set of portfolios

In this section we formalize the geometric representation of sets of portfolios with an arbitrary number of assets.

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

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

where $v_1, \ldots, v_n \in \mathbb{R}^{n-1}$ are a set of n affinely independent points in a Euclidean space of dimension $n-1$. The vertices $(v_i)_{i=1,\ldots,n}$ represent the n portfolios made of a single asset and the simplex is the convex hull of these vertices.

For instance, we can define $v_1, \ldots, v_n$ such that:

1. the center of the simplex is set to the origin,
2. the distances of the simplex vertices to its center are equal,
3. the angle subtended by any two vertices through its center is $\arccos(\frac{-1}{n-1})$.

The weights $(x_i)_{i=1,\ldots,n}$ of portfolio x are called its barycentric coordinates, whereas in $\mathbb{R}^{n-1}$ they are called its Cartesian coordinates and are denoted by $\breve{x} = (\breve{x}_1, \ldots, \breve{x}_{n-1})$. We use the Cartesian coordinates in Section 4 to compute the moments of the portfolios' returns distribution.

3 PDF and CDF of the portfolios' returns distribution

In this section we focus on the exact computation of the cumulative distribution function (CDF) and of the probability density function (PDF) of the portfolio returns, given the asset returns.

3.1 The CDF of the portfolios' returns distribution

Let us consider the set of long-only portfolios providing a return lower than a given return R^* over a period of time for which the asset returns were $R = (R_1, \ldots, R_n)$. It corresponds to a linear half-space defined as

$$H(R^*) = \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n \mid \sum_{i=1}^{n} R_i x_i \leq R^* \right\}. \tag{2}$$

Denoting by $V(A)$ the volume of a geometric object A, the allocation score of a portfolio providing a return R^* can be obtained by computing the ratio of the volume of the intersection of the simplex with this half-space over the volume of the simplex, i.e.

$$S(R^*) = \frac{V(H(R^*) \bigcap \Delta^{n-1})}{V(\Delta^{n-1})} \tag{3}$$

We illustrate such a volume in Figure 2. Consider a market of 4 assets whose returns are observed, and a portfolio providing a given return R^*. The pyramid is the simplex representing the set of long-only portfolios. The surface highlighted in the left figure represents the set of portfolios returning R^*. The volume highlighted in the right figure represents the set of portfolios providing a return lower or equal to R^*.

3.1.1 Varsi's algorithm

As noticed in [Calès et al., 2018], there exists an exact, iterative formula for the volume defined by intersecting a simplex with a half-space. It is provided in Algorithm 1. A geometric proof is given in [Varsi, 1973], by subdividing the polytope into pyramids and, recursively, to simplices. For a comparison between alternative proofs and algorithms, the reader may refer to [Calès et al., 2018] and the references thereof.

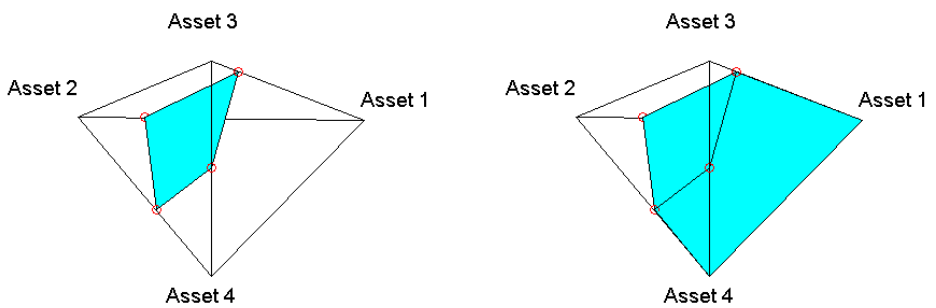


Figure 2: (left) Surface of portfolios providing this given return. (right) Volume of portfolios outperformed by this return.

Algorithm 1. *Let* $H = \{(\omega_1, \ldots, \omega_n) \,|\, \sum_{i=1}^{n} R_i \omega_i \leq R^*\}$ *be a linear half-space.*

1. *Compute* $u_i = R_i - R^*$, $i = 1, 2, \ldots, n$.
2. *Label the non-negative* u_j *as* $Y_1, \ldots, Y_K$ *and the negative ones as* $X_1, \ldots, X_J$.
3. *Initialize* $A_0 = 1$, $A_1 = A_2 = \cdots = A_K = 0$.
4. *For* $h = 1, 2, \ldots, J$ *repeat:* $A_k = \frac{Y_k A_k - X_h A_{k-1}}{Y_k - X_h}$, *for* $k = 1, 2, \ldots, K$.

Then, for $h = J$, $A_K = \frac{V(H(R^*)\bigcap\Delta_{n-1})}{V(\Delta_{n-1})}$.

This algorithm requires $O(n^2)$ operations, and thus can be computed very quickly. As an illustration, we compute the score of a portfolio whose return is 0 in markets of 100, 1000 and $10,000$ assets whose returns are randomly drawn from a centered normal distribution. The computation is repeated 1000 times. We report in Table 1 the average computation time and its standard deviation.

Number of Assets	100	1000	10,000
Mean computation time	5.89e-5	3.63e-3	0.4734
Standard deviation	1.99e-4	1.79e-4	0.0416

Table 1: Mean computation time in seconds and standard deviation of the computation of the CDF at a point for markets of 100, 1000 and $10,000$ assets. The computations have been performed using Matlab© on a bi-xeon E2620 v3 under Windows©.

To illustrate the CDF obtained using Algorithm 1, let us consider a market of 10 assets whose returns are as in Table 2 and let R^* denote a portfolio return. The CDF of the portfolios returns for any given return is reported in Figure 3. With these asset returns, 10% of the portfolios have a negative return and a bit more than 20% of the portfolios have a return greater than 1%.

R_1	R_2	R_3	R_4	R_5
0.5377%	1.8339%	−2.2588%	0.8622%	0.3188%
R_6	R_7	R_8	R_9	R_{10}
−1.3077%	−0.4336%	0.3426%	3.5784%	2.7694%

Table 2: Some asset returns.

3.1.2 Closed form expression

In [Lasserre, 2015], a closed form formula is proposed to compute this volume taking into account the case of equal asset returns. However, it omitted some extra terms and has been corrected in [Calès, 2019]. We report it here, adapted to our notation: Let $\boldsymbol{R} = (R_i)_{i=1}^n$ be the asset returns, $(S_i)_{i=1}^d$ the d distinct returns, where $d \leq n$, and $(m_i)_{i=1}^d$ their multiplicities (i.e. number of occurrences). We denote by $(J_i)_{i=1}^d$ the subsets of indices in $\{1, \dots, n\}$ associated to each S_i and,

$$\text{for } j = 1, \dots, d, \text{ we let } \boldsymbol{b}_j = \left(\frac{1}{R_i - S_j}\right)_{i\in\{1,\dots,n\}\setminus J_j}.$$

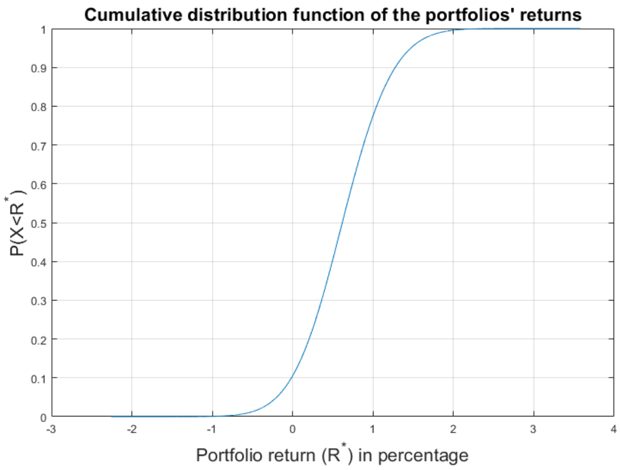


Figure 3: Illustration of the CDF of the portfolios returns.

Among the distinct returns we distinguish between those whose multiplicities are 1, and those whose multiplicities are greater than 1. The indices of the first group form a set denoted by I, while those of the second group form set K, where $S = I \cup K$. Finally, $(x)_+$ stands for $\max\{0, x\}$. The CDF computed in R^* is given by:

$$\begin{aligned} S(R^*, \boldsymbol{R}) = & \sum_{i \in I} \frac{(R^* - S_i)_+^{n-1}}{\prod\limits_{j=1, j\neq i}^{n} (S_j - S_i)} + \\ & + \sum_{i \in K} \left(\sum_{j=0}^{m_i - 1} (-1)^{j+m_i+1} \binom{n-1}{j} \frac{(R^* - S_i)_+^{n-j-1}}{\prod\limits_{k \in S \setminus \{S_i\}} (S_k - S_i)} \Phi_{m_i - 1 - j}(\boldsymbol{b}_i) \right), \end{aligned} \tag{4}$$

with

$$\Phi_k(x) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{i_1} \cdots \sum_{i_k=1}^{i_{k-1}} x_{i_1} \ldots x_{i_k}, \boldsymbol{x} \in \mathbb{R}^n.$$

This formula is similar to the one proposed in [Pouchkarev, 2005, Theorem 4.2.2] and [Banerjee and Hung, 2011, Theorem A2], but with extra terms correcting for the equal asset returns. Unfortunately, when it comes to calculations, the formula becomes numerically unstable for $n \geq 20$ at the usual machine precision.

3.1.3 Properties of the score

As noticed in [Banerjee and Hung, 2011], the score is invariant under some linear transformation of the asset returns. To see this, let us consider a market of n assets providing the returns $\boldsymbol{R} = (R_i)_{i=1}^n$. We are interested in the score of a portfolio $\boldsymbol{\omega} = (\omega_i)_{i=1}^n$ providing a return $R^* = \boldsymbol{\omega}'\boldsymbol{R}$, where $\boldsymbol{\omega}'$ stands for the transpose vector. As explained before, this score is the volume of simplex Δ^{n-1} intersected with half-space $H(R^*)$ as in Equation (2). So, it is

$$S(R^*|\boldsymbol{R}) = \left\{\boldsymbol{\theta} \in \Delta^{n-1} | \sum_{i=1}^{n} R_i\theta_i \leq R^*\right\} \tag{5}$$

Property 1. *The score is invariant under linear transformations of the asset returns such that $\boldsymbol{R} \to \sigma\boldsymbol{R} + \alpha$ with $\alpha \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$.*

Proof.

$$\begin{aligned} S(\sigma R^* + \alpha \,|\, \sigma\boldsymbol{R} + \alpha) &= \left\{\boldsymbol{\theta} \in \Delta^{n-1} \,|\, \sum_{i=1}^{n} (\sigma R_i + \alpha)\theta_i \leq \sigma R^* + \alpha\right\} \\ &= \left\{\boldsymbol{\theta} \in \Delta^{n-1} \,|\, \sigma \sum_{i=1}^{n} R_i\theta_i + \alpha \leq \sigma R^* + \alpha\right\} \\ &= \left\{\boldsymbol{\theta} \in \Delta^{n-1} \,|\, \sigma \sum_{i=1}^{n} R_i\theta_i \leq \sigma R^*\right\} \\ &= S(R^*|\boldsymbol{R}). \end{aligned}$$

The main implication of this property is that the asset returns can be standardized cross-sectionally without affecting the score. It is interesting to note that such a transformation is common in financial event studies, since the seminal work of [Boehmer et al., 1991]. This approach has the advantages of being robust to event-induced heteroskedasticity and of not requiring data from a pre-event estimation period.

3.2 The PDF of the portfolios' returns distribution

In this section we focus on the exact computation of the probability density function (PDF) of the portfolios' returns distribution.

3.2.1 Geometric interpretation of B-splines

In [Curry and Schoenberg, 1966], a seminal paper on splines, Theorem 2 shows that the univariate B-spline resulting from the orthogonal projection of the volumetric slices of a unit simplex on $\mathbb{R}$ can be interpreted as the PDF of these slices' volume. For instance, in Figure 4, we have the projections of the areas of the intersection of planes with the 3-d simplex on the real line. In our case, the simplex is the set of portfolios, while the planes are the equi-return portfolios. The resulting univariate B-spline is the PDF of the portfolio returns. For any value, it can be computed using the de Boor-Cox recursion formula, see [de Boor, 1972] and [Cox, 1972], as shown in Algorithm 2.

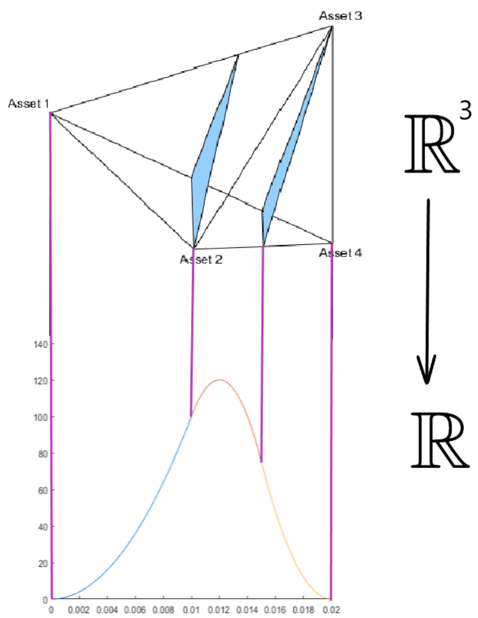


Figure 4: Geometric interpretation of the PDF of the portfolio returns: a univariate B-spline, as the orthogonal projection of the volumetric slices of a unit simplex on $\mathbb{R}$. In this example, we have 4 assets whose returns are 0%, 1%, 1.5% and 2%.

Algorithm 2. *Let $\mathbf{R} = (R_i)_{i=1}^n$ be the returns of the n assets, ordered such that $R_i \le R_j$ for $i < j$, and let $k = n - 1$ be the B-spline order. To evaluate the PDF at x, we set $j = 1$ and call function bspline_pdf() as defined below. The outcome is then normalized such that $y = y\frac{k}{R_n - R_1}$.*

function y = bspline_pdf(j,k,$\mathbf{R}$,x)

1. *$y = 0$*
2. *if $k > 1$ then*

(a) b = *bspline_pdf(j,k-1,*$\mathbf{R}$*,x);*

(b) *if* $R_{j+k} \neq R_{j+1}$*, then* $y = y + b\left(\frac{x-R_{j+1}}{R_{j+k}-R_{j+1}}\right)$

(c) b = *bspline_pdf(j+1,k-1,*$\mathbf{R}$*,x);*

(d) *if* $R_{j+k+1} \neq R_{j+2}$*, then* $y = y + b\left(\frac{R_{j+k+1}-x}{R_{j+k+1}-R_{j+2}}\right)$

3. *elseif* $R_{j+1} \leq x$

   (a) *if* $R_{j+2} < R_n$ *and* $x < R_{j+2}$*, then* $y = 1$*, else* $y = 0$

4. *else*

   (a) *if* $R_{j+1} <= x$ *and* $R_{j+2} > R_n$*, then* $y = 1$*, else* $y = 0$

Number of Assets	10	20
Mean computation time	2.99e-4	0.2870
Standard deviation	2.99e-5	0.0024

Table 3: Mean computation time in seconds and standard deviation of the computation of the PDF at a point for markets of 10 and 20 assets. The computations have been performed using Matlab© on a bi-xeon E2620 v3 under Windows©.

As an illustration, let us consider the previous example with 10 assets. Using Algorithm 2, we compute the PDF of the portfolios' returns and report it in Figure 5. In this example, we observe that the distribution is uni-modal with most portfolios having a return close to 0.6%. In the next section, we shall focus on the computation of the moments of this distribution.

3.2.2 Closed form expression

It is straightforward to get the PDF by deriving Equation (4). Using the same notation as in Section 3.1.2, this leads to the following closed-form formula for the PDF:

$$\begin{aligned} f(R^*, \boldsymbol{R}) &= (n-1) \sum_{i \in I} \frac{(R^*-S_i)_+^{n-2}}{\prod\limits_{j=1, j\neq i}^{n} (S_j - S_i)} + \\ &+ \sum_{i \in K} \left(\sum_{j=0}^{m_i-1} (-1)^{j+m_i+1} (n-j-1) \binom{n-1}{j} \frac{(R^*-S_i)_+^{n-j-2}}{\prod\limits_{k \in S\setminus\{S_i\}} (S_k - S_i)} \Phi_{m_i-1-j}(\boldsymbol{b}_i) \right). \end{aligned} \tag{6}$$

Its computation suffers from the same drawback as the computation of the CDF, providing numerically unstable results for $n \geq 20$ at the usual machine precision.

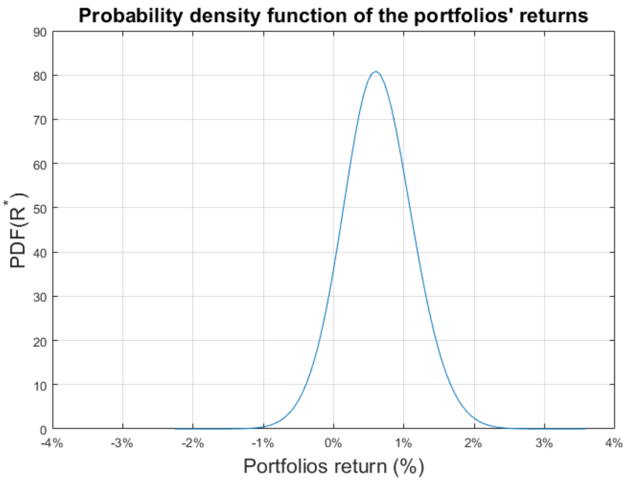


Figure 5: Illustration of the PDF of the portfolios returns.

3.2.3 Numerical derivation

The iterative nature of the de Boor-Cox formula makes the computation of the PDF slow for large number of assets, say ≥ 20. Moreover, its computation using the closed form formula above is numerically unstable for large number of assets, say ≥ 20. So, a practical alternative is to estimate the PDF by deriving numerically the CDF obtained earlier, using Varsi's algorithm.

Let F be the CDF and x_0 the point in which we wish to estimate its derivative. One may employ central differences and the five points' method, thus having

$$F'(x_0) = \frac{-F(x_0+2h) + 8F(x_0+h) - 8F(x_0-h) + F(x_0-2h)}{12h} + \frac{h^4}{30}F^{(5)}(c), \tag{7}$$

with $c \in [x_0 - 2h, x_0 + 2h]$. The truncation error is then $O(h^4)$. Even though it is only an estimate of the PDF, this approach enables us to scale up to thousands of assets with computation times being 5 times higher than those reported in Table 1 for an estimate at a single point with the five points method.

4 Moments of the portfolios' returns distribution

In this section, we compute the moments of the portfolios returns. In Section 4.1, we provide affine maps to pass from barycentric to Cartesian coordinates, and vice-versa, as well as some useful identities. In Section 4.2, we state the moments' definitions. In Section 4.3, we recall a theorem proposed in [Lasserre and Avrachenkov, 2001] which provides an elegant expression for the integral of a symmetric q-linear form on a simplex. In Section 4.4, since the theorem makes use of nested sums, we provide identities for some of these nested sums. In Section 4.5, after having developed the moments expressions, we compute the individual terms that make up the moments up to the fourth one. In Section 4.6, we provide the closed form solution of the first four moments. Finally, in Section 4.7, we propose a general algorithm to compute any moment.

4.1 Barycentric and Cartesian representations

There are affine maps to pass

- from barycentric to Cartesian coordinates:

$$\begin{array}{rccl} m_{bc}: & \mathbb{R}^n & \rightarrow & \mathbb{R}^{n-1}, \\ & x & \mapsto & \breve{x} = Tx + v_n, \end{array} \tag{8}$$

  where vertices v_i correspond to $(n-1)$-dimensional column vectors, and $T = \left[\begin{array}{ccc} v_1 - v_n, & \cdots, & v_{n-1} - v_n \end{array} \right]$ is an $(n-1) \times (n-1)$ matrix.

- from Cartesian to barycentric coordinates:

$$\begin{array}{rccl} m_{cb}: & \mathbb{R}^{n-1} & \rightarrow & \mathbb{R}^n, \\ & \breve{x} & \mapsto & x = \left[\begin{array}{c} I_{n-1} \\ -1'_{n-1} \end{array} \right] T^{-1}(\breve{x} - v_n) + \left[\begin{array}{c} 0_{n-1} \\ 1 \end{array} \right], \end{array} \tag{9}$$

  where 0_{n-1} and 1_{n-1} are the $(n-1)$-dimensional column vectors of 0's and 1's, respectively, and I_{n-1} is the $(n-1) \times (n-1)$ identity matrix.

The return of portfolio x is then given by

$$R'x = A\breve{x} - Av_n + R_n, \tag{10}$$

where $A = R' \left[\begin{array}{c} I_{n-1} \\ -1'_{n-1} \end{array} \right] T^{-1}$. By construction, we also have these useful identities:

Lemma 2. *For $n \in \mathbf{N}$, vertices $(v_i)_{i=1}^n$, and matrix A as defined previously, it holds*

$$\sum_{i=1}^{n} Av_i = A \sum_{i=1}^{n} v_i = 0. \tag{11}$$

Lemma 3. *For $n \in \mathbf{N}$, $(R_i)_{i=1}^n$, $(v_i)_{i=1}^n$, A and M_1 as defined previously, it holds*

$$R_i = Av_i - Av_n + R_n,\ i \in \{1, \dots, n\}. \tag{12}$$

Lemma 4. *For $n \in \mathbf{N}$, $(R_i)_{i=1}^n$, $(v_i)_{i=1}^n$, A and M_1 as defined previously, it holds*

$$M_1 = R_n - Av_n. \tag{13}$$

Lemma 5. *For $n \in \mathbf{N}$, $(R_i)_{i=1}^n$, $(v_i)_{i=1}^n$, A and M_1 as defined previously, it holds*

$$Av_i = R_i - M_1. \tag{14}$$

4.2 Moments

We are interested in the distribution of the portfolio returns given the observed individual asset returns $\boldsymbol{R} = (R_1, \dots, R_n)$.

By definition, the moments of the portfolio returns distribution are given as follows, where $V(\cdot)$ stands for Euclidean volume:

$$M_1 = \frac{1}{V(\Delta^{n-1})} \int\limits_{\Delta^{n-1}} A\breve{x} - Av_n + R_n \, d\breve{x}, \tag{15}$$

$$M_2 = \frac{1}{V(\Delta^{n-1})} \int\limits_{\Delta^{n-1}} (A\breve{x} - Av_n + R_n - M_1)^2 \, d\breve{x}, \tag{16}$$

$$M_k = \frac{1}{V(\Delta^{n-1})(\sqrt{M_2})^k} \int\limits_{\Delta^{n-1}} (A\breve{x} - Av_n + R_n - M_1)^k d\breve{x},\ k \geq 3, \tag{17}$$

where the term $\frac{1}{V(\Delta^{n-1})}$ is normalizing the equations. Indeed, the distance between the vertices v_i is arbitrary, and so is the volume of Δ^{n-1}. An alternative is to choose the distance between the vertices v_i such that $V(\Delta^{n-1}) = 1$.

By employing Lemma 4, M_2 and M_k simplify to

$$M_2 = \frac{1}{V(\Delta^{n-1})} \int\limits_{\Delta^{n-1}} (A\breve{x})^2 \, d\breve{x}, \tag{18}$$

$$M_k = \frac{1}{V(\Delta^{n-1})(\sqrt{M_2})^k} \int\limits_{\Delta^{n-1}} (A\breve{x})^k d\breve{x},\ k \geq 3. \tag{19}$$

4.3 Integrating over Δ^{n-1}

From [Lasserre and Avrachenkov, 2001], and slightly adapted to our notations, we have the following

Theorem 6. *Let $v_1, \ldots, v_n$ be the vertices of an $(n-1)-$dimensional simplex Δ^{n-1}. Then, for a symmetric q-linear form $H : (\mathbb{R}^{n-1})^q \to \mathbb{R}$, we have*

$$\int_{\Delta^{n-1}} H(X, \ldots, X)\, d\breve{x} = \frac{V\left(\Delta^{n-1}\right)}{\begin{pmatrix} n-1+q \\ q \end{pmatrix}} \sum_{1 \le i_1 \le i_2 \le \cdots \le i_q \le n} H(v_{i_1}, v_{i_2}, \ldots, v_{i_q}), \tag{20}$$

where $V\left(\Delta^{n-1}\right) = \int_{\Delta^{n-1}} 1\, d\breve{x}$ stands for the volume of the simplex Δ^{n-1}.

4.4 Nested sum identities

In the following, we employ these identities:

Lemma 7. *For $n \in \mathbf{N}$, it holds*

$$2 \sum_{i=1}^{n} \sum_{j=i}^{n} x_i x_j = \left(\sum_{i=1}^{n} x_i \right)^2 + \sum_{i=1}^{n} x_i^2. \tag{21}$$

Proof. Trivial.

Lemma 8. *For $n \in \mathbf{N}$, it holds*

$$(3!) \sum_{i=1}^{n} \sum_{j=i}^{n} \sum_{k=j}^{n} x_i x_j x_k = \left(\sum_{i=1}^{n} x_i \right)^3 + 3 \left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} x_i^2 \right) + 2 \sum_{i=1}^{n} x_i^3. \tag{22}$$

See proof in Annex A.

Lemma 9. *For $n \in \mathbf{N}$, it holds*

$$(4!) \sum_{i=1}^{n} \sum_{j=i}^{n} \sum_{k=j}^{n} \sum_{t=k}^{n} x_i x_j x_k x_t =$$

$$= \left(\sum_{i=1}^{n} x_i \right)^4 + 6 \left(\sum_{i=1}^{n} x_i \right)^2 \left(\sum_{i=1}^{n} x_i^2 \right) + 8 \left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} x_i^3 \right) + 6 \left(\sum_{i=1}^{n} x_i^4 \right) + 3 \left(\sum_{i=1}^{n} x_i^2 \right)^2.$$

See proof in Annex B.

4.5 Expression of moments' individual terms

We now need to compute $\int_{\Delta^{n-1}}(A\breve{x})^p \, d\breve{x}$, where p corresponds to a single term of the development.

Lemma 10. *For $n \in \mathbf{N}$, it holds*

$$\int_{\Delta^{n-1}} (A\breve{x})^2 \, d\breve{x} = \frac{V\left(\Delta^{n-1}\right)}{n(n+1)} \sum_{i=1}^{n} (Av_i)^2. \tag{23}$$

Proof. Apply Theorem 6, with $q = 2$, and $H(\breve{x}_1, \breve{y}_2) = (A\breve{x}_1)(A\breve{x}_2)$, by replacing the nested sum in the theorem as in Lemma 7, then recalling Lemma 2.

Lemma 11. *For $n \in \mathbf{N}$, it holds*

$$\int\limits_{\Delta^{n-1}} (A\breve{x})^3 \, d\breve{x} = \frac{2V\left(\Delta^{n-1}\right)}{n(n+1)(n+2)} \sum_{i=1}^{n} (Av_i)^3. \tag{24}$$

Proof. Apply Theorem 6, with $q = 3$, and $H(\breve{x}_1, \breve{y}_2, \breve{y}_3) = (A\breve{x}_1)(A\breve{x}_2)(A\breve{x}_3)$, by replacing the nested sum in the theorem as in Lemma 8, then recalling Lemma 2.

Lemma 12. *For $n \in \mathbf{N}$, it holds*

$$\int\limits_{\Delta^{n-1}} (A\breve{x})^4 \, d\breve{x} = \frac{V\left(\Delta^{n-1}\right)}{n(n+1)(n+2)(n+3)} \left(6 \left(\sum_{i=1}^{n} (Av_i)^4 \right) + 3 \left(\sum_{i=1}^{n} (Av_i)^2 \right)^2 \right).$$

Proof. Apply Theorem 6, with $q = 4$, and $H(\breve{x}_1, \breve{y}_2, \breve{y}_3, \breve{y}_4) = (A\breve{x}_1)(A\breve{x}_2)(A\breve{x}_3)(A\breve{x}_4)$, replacing the nested sum in the theorem as in Lemma 9, then recalling Lemma 2.

4.6 Closed form expression of the first four moments

In this section, we derive the following closed form expressions for the first four moments, reported in Theorems 13 to 16, respectively.

Theorem 13. *In a market of n assets, $n \in \mathbf{N}$, whose returns are $\boldsymbol{R} = (R_i)_{i=1}^n$, the first moment of the portfolios' returns is*

$$M_1 = \frac{1}{n} \sum_{i=1}^{n} R_i. \tag{25}$$

Proof. See proof in Annex C.

Theorem 14. *In a market of n assets, $n \in \mathbf{N}$, whose returns are $\boldsymbol{R} = (R_i)_{i=1}^n$, the second moment of the portfolios' returns is*

$$M_2 = \frac{1}{n(n+1)} \sum_{i=1}^{n} (R_i - M_1)^2 = \frac{1}{n+1} Var(\boldsymbol{R}), \tag{26}$$

where Var is the biased sample variance.

Proof. Replace $\int\limits_{\Delta^{n-1}} (A\breve{x})^2 \, d\breve{x}$ in Equation (18) by the expression from Lemma 10, then apply Lemma 5.

Theorem 15. *In a market of n assets, $n \in \mathbf{N}$, whose returns are $\boldsymbol{R} = (R_i)_{i=1}^n$, the third moment of the portfolios' returns is*

$$M_3 = \frac{1}{M_2^{3/2}} \frac{2}{n(n+1)(n+2)} \sum_{i=1}^{n} (R_i - M_1)^3. \tag{27}$$

Proof. By replacing $\int\limits_{\Delta^{n-1}} (A\breve{x})^3 \, d\breve{x}$ in Equation (19) with the expression from Lemma 11, and applying Lemma 5, one obtains the result.

Theorem 16. *In a market of n assets, $n \in \mathbf{N}$, whose returns are $\boldsymbol{R} = (R_i)_{i=1}^n$, the fourth moment of the portfolios' returns is*

$$M_4 = \frac{1}{M_2^2} \frac{1}{n(n+1)(n+2)(n+3)} \left(6 \sum_{i=1}^{n} (R_i - M_1)^4 + 3 \left(\sum_{i=1}^{n} (R_i - M_1)^2 \right)^2 \right). \tag{28}$$

Proof. By replacing $\int\limits_{\Delta^{n-1}} (A\breve{x})^4 \, d\breve{x}$ in Equation (19) with the expression from Lemma 12, and applying Lemma 5, the claim is established.

4.7 General algorithm to compute the moments

We now wish to compute the k^{th} moment of the portfolios' returns, i.e., restating Equation (19), we compute

$$M_k = \frac{1}{Vol\left(\Delta^{n-1}\right)\left(\sqrt{M_2}\right)^k} \int\limits_{\Delta^{n-1}} (A\breve{x})^k d\breve{x} \quad , k \geq 3.$$

Let us set $H(X_1, \ldots, X_k) = \prod_{i=1}^{k} X_i$. It is a symmetric k-linear form, and we have

$$H(\underbrace{A\breve{x}, \ldots, A\breve{x}}_{k \text{ times}}) = (A\breve{x})^k.$$

Thus from Theorem 6 we have

$$\int_{\Delta^{n-1}} H(\underbrace{A\breve{x}, \ldots, A\breve{x}}_{k \text{ times}})\, d\breve{x} = \frac{V\left(\Delta^{n-1}\right)}{\begin{pmatrix} n-1+k \\ k \end{pmatrix}} \sum_{1 \le m_1 \le m_2 \le \cdots \le m_k \le n} H(Av_{m_1}, Av_{m_2}, \ldots, Av_{m_k}).$$

Let q_i be the number of occurrences of value i, $1 \le i \le n$. Then,

$$\int_{\Delta^{n-1}} (A\breve{x})^k \; d\breve{x} = \frac{V\left(\Delta^{n-1}\right)}{\begin{pmatrix} n-1+k \\ k \end{pmatrix}} \sum_{\sum_{i=1}^{n} q_i = k} \prod_{i=1}^{n} (Av_i)^{q_i}.$$

Now, we change the notations. Let $\lambda = (\lambda_1, \ldots, \lambda_k)$ be a partition of k, and Λ_k the set of partitions of k. We denote by $l = (l_i)_{i=1}^{d}$ the d unique non-zero values in λ, $d \le k$, and by $(p_i)_{i=1}^{d}$ the multiplicities of $(l_i)_{i=1}^{d}$. For instance, $\lambda = (2, 1, 1, 0)$ is a partition of $k = 4$, with $d = 2$, $l = (1, 2)$ and the associated multiplicities $p = (2, 1)$.

From [Macdonald, 1995, Eq. (2.14')], we have

$$\sum_{\sum_{i=1}^{n} q_i = k} \prod_{i=1}^{n} (Av_i)^{q_i} = \sum_{\lambda \in \Lambda_k} \frac{\prod_{i=1}^{d} \left(\sum_{j=1}^{n} (Av_j)^{l_i} \right)^{p_i}}{\prod_{i=1}^{d} p_i! l_i^{p_i}}.$$

Set Λ_k can be obtained with Algorithm $ZS1$ in [Zoghbi and Stojmenovic, 1994], and is still tractable for large moments, as shown on Table 4:

k	1	5	10	20	30	40
$\lvert\Lambda_k\rvert$	1	7	42	627	5604	37338

Table 4: Number of partitions $|\Lambda_k|$ for different values of k.

The computation is formally presented in Algorithm 3 below. As an illustration of the computation times, we compute the k^{th} order moments, $k = 5, 10, 15, 20$, for markets of 100, 1000 and $10,000$ assets whose returns are randomly drawn. The computation is repeated 1000 times. We report in Table 5 the average computation time in seconds and its standard deviation.

Algorithm 3. *Let $\boldsymbol{R}$ be the asset returns, N the number of assets, and k the moment order.*

1. *Compute M_2 by Theorem 14*
2. *Compute Av as in Lemma 5*
3. *Compute Λ using Algorithm $ZS1$ in [Zoghbi and Stojmenovic, 1994]*
4. *Set $S = 0$*
5. *For each $\lambda \in \Lambda$:*
   (a) *decompose λ in its d non-zero elements $(l_i)_{i=1}^{d}$ with multiplicities $(p_i)_{i=1}^{d}$*
   (b) $a = \prod_{i=1}^{d} p_i!\, l_i^{p_i}$
   (c) $b = \prod_{i=1}^{d} \left(\sum_{j=1}^{n} (Av_j)^{l_i} \right)^{p_i}$
   (d) $S = S + a/b$
6. *Set $M_k = S \,/ \left(\sqrt{M_2}^{k} \cdot \binom{n-1+k}{k} \right)$*

Moment order: Nb of Assets	5	10	15	20
100	0.0006 (0.0021)	0.0024 (0.0003)	0.0100 (0.0005)	0.0398 (0.0005)
1000	0.0008 (0.0000)	0.0034 (0.0003)	0.0117 (0.0002)	0.0420 (0.0007)
10000	0.0020 (0.0001)	0.0053 (0.0000)	0.0145 (0.0011)	0.0506 (0.0028)

Table 5: Mean runtime in seconds, and standard deviation in parenthesis, of computing the moment of order k in markets of 100, 1000 and $10,000$ assets. The assets returns are drawn randomly before each of the 1000 computations. The experiments were performed with Matlab© on a bi-xeon E2620 v3 under Windows©.

5 Concluding remarks and future work

In this paper, we reviewed different approaches to compute the PDF, CDF and moments of the distribution of portfolio returns, across portfolios and for the long-only strategy.

For the CDF, the computations improve upon existing work by providing exact results, allowing for equal asset returns, and handle a large number of assets, thus removing the need of Monte Carlo sampling for its estimation. These computations can be based on:

- the volume algorithm by [Varsi, 1973] which is fast and exact even for a large number of assets,
- the closed form expression of CDF, which is exact but numerically unstable for a large number of assets.

For the PDF, our methods are new, based on what follows:

- the algorithm by de Boor and Cox [de Boor, 1972, Cox, 1972], which is exact but slow for a large number of assets,
- the closed form expression of PDF, which is exact but numerically unstable for a large number of assets,
- the numerical derivation of the CDF using [Varsi, 1973], which only provides an estimate but is fast and applies to a large number of assets.

For the moments, the computations are new and can based on

- closed form expressions up to the fourth-order moment,
- a new algorithm using Algorithm $ZS1$ by [Zoghbi and Stojmenovic, 1994], for higher moments, which is fast and exact even for a large number of assets.

It should be noted that most of these computations can easily be vectorized, thus further extending their realm of applications.

These results have several statistical and econometric implications.

- The asset returns can be standardized cross-sectionally without altering the relative performance of portfolios. The series obtained are then robust to systemic heteroskedasticity.

- The closed form expressions of the first three moments show a direct mapping between the moments of the cross-sectional asset returns distribution and those of the distribution of portfolio returns.

  In particular, the first moments are identical for both distributions. The second and third moments are proportional to each other, the factor depending on the number of assets. For the second moment, the factor is $\frac{1}{N+1}$ where N is the number of assets, implying that it might be more difficult to find a portfolio which performs significantly better than the equally weighted portfolio, when the number of assets increases.

  The closed form expression of the fourth moment behaves differently with an extra positive term implying fatter tails in the distribution of portfolio returns than in the cross-sectional distribution of the asset returns. Its implications have to be further analyzed.

The relevance of computing high moments can be discussed. We believe that these moments should be useful in an alternative method recovering the PDF. Indeed, since the distribution is bounded, this problem is known as the Hausdorff moment problem, which has been addressed, see for instance [Mnatsakanov, 2008]. This approach has been inconclusive for the authors.

Future work can take several directions. On the theoretical side, it would be interesting to study the distribution of portfolio volatilities in order to better assess the dependency between portfolios returns and volatilities, which is done by sampling in [Calès et al., 2018] It would also be interesting to consider different investment sets, e.g. including short selling[4] and leverage[5].

In terms of applications, it should be noticed that the paper focuses on portfolio returns but the methodology can be applied to any linear combination of asset characteristics, e.g. by defining the portfolio dividend yield as the weighted sum of the asset dividend yields. The use of the score, i.e., the CDF, has already found applications in portfolio performance measures. These may be improved, for instance by considering the random nature of the asset returns. The PDF can be used to finely assess the distribution of portfolios with possible applications in portfolio diversification and turn-over analysis. The moments can find applications in the literature on return dispersion, see e.g. [Yu and Sharaiha, 2007], [Stivers and Sun, 2010], [Gorman et al., 2010], [Bhootra, 2011] and [Verousis and Voukelatos, 2018]. They can also find applications in the literature on noise trading, see e.g. [De Long et al., 1989].

[4] i.e. negative portfolio weights.

[5] i.e. sum of portfolio weights greater than one.

References

[Ali, 1973] Ali, M. (1973). Content of the frustum oa a simplex. *Pacific J. Math.*, 48(2):313–322.

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

[Bhootra, 2011] Bhootra, A. (2011). Are momentum profits driven by the cross-sectional dispersion in expected stock returns? *J. Financial Markets*, 14(3):494–513.

[Billio et al., 2011] Billio, M., Calès, L., and Guégan, D. (2011). A cross-sectional score for the relative performance of an allocation. *Intern. Review Appl. Financial Issues & Economics*, 3(4):700–710.

[Boehmer et al., 1991] Boehmer, E., Masumeci, J., and Poulsen, A. B. (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics*, 30(2):253 – 272.

[Calès, 2019] Calès, L. (2019). Erratum on: Volume of slices and sections of the simplex in closed form. Technical report, European Commission, Joint Research Centre.

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

[Cox, 1972] Cox, M. (1972). The numerical evaluation of b-splines. *IMA J. Applied Mathematics*, 10(2):134–149.

[Curry and Schoenberg, 1966] Curry, H. B. and Schoenberg, I. J. (1966). On pólya frequency functions IV: The fundamental spline functions and their limits. *Journal d'Analyse Mathématique*, 7:71–107.

[de Boor, 1972] de Boor, C. (1972). On calculating with b-splines. *Journal of Approximation Theory*, 6:50–62.

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

[Gorman et al., 2010] Gorman, L. R., Sapra, S. G., and Weigand, R. A. (2010). The cross-sectional dispersion of stock returns, alpha and the information ratio. *J. Investing*.

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

[Hallerbach and Pouchkarev, 2005] Hallerbach, W. and Pouchkarev, I. (2005). A relative view on tracking error. Technical Report ERS-2005-063-F&A, Erasmus University Rotterdam.

[Hallerbach and Pouchkarev, 2016] Hallerbach, W. and Pouchkarev, I. (2016). Active portfolio management with conditional tracking error. *Bankers, market and investors*, 143:18–25.

[Lasserre and Avrachenkov, 2001] Lasserre, J. and Avrachenkov, K. (2001). The multi-dimensional version of $\int_a^b x^p dx$. *The American Mathematical Monthly*, 108(2):151–154.

[Lasserre, 2015] Lasserre, J. B. (2015). Volume of slices and sections of the simplex in closed form. *Optimization Letters*, 9:1263–1269.

[Macdonald, 1995] Macdonald, I. (1995). *Symmetric Functions and Hall Polynomials*. Oxford University Press, 1 edition.

[Mnatsakanov, 2008] Mnatsakanov, R. M. (2008). Hausdorff moment problem: Reconstruction of probability density functions. *Statistics & Probability Letters*, 78(13):1869 – 1877.

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

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

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

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

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

[Verousis and Voukelatos, 2018] Verousis, T. and Voukelatos, N. (2018). Cross-sectional dispersion and expected returns. *Quantitative Finance*, 18(5):813–826.

[Yu and Sharaiha, 2007] Yu, W. and Sharaiha, Y. M. (2007). Alpha budgetingcross-sectional dispersion decomposed. *Journal of Asset management*, 8(1):58–72.

[Zoghbi and Stojmenovic, 1994] Zoghbi, A. and Stojmenovic, I. (1994). Fast algorithms for generating integer partitions. *Intern. J. Computer Mathematics*, 70:319–332.

A Proof of Lemma 8

Lemma 8 For $n \in \mathbf{N}$, it holds

$$6\sum_{i=1}^{n}\sum_{j=i}^{n}\sum_{k=j}^{n} x_i x_j x_k = \left(\sum_{i=1}^{n} x_i\right)^3 + 3\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i^2\right) + 2\sum_{i=1}^{n} x_i^3.$$

Proof. Let $S(n) = \sum_{i=1}^{n}\sum_{j=i}^{n}\sum_{k=j}^{n} x_i x_j x_k$. The cases $n = 0$ and $n = 1$ are easily verifiable. Let us assume that the theorem holds for $S(n), n \geq 1$. We shall prove it for $S(n+1)$. Clearly,

$$S(n+1) = S(n) + \sum_{i=1}^{n}\sum_{j=i}^{n} x_i x_j x_{n+1} + \sum_{i=1}^{n} x_i x_{n+1}^2 + x_{n+1}^3,$$

which, by inductive hypothesis, yields

$$6S(n+1) =$$
$$\left(\sum_{i=1}^{n} x_i\right)^3 + 3\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i^2\right) + 2\sum_{i=1}^{n} x_i^3 + \underline{6x_{n+1}\sum_{i=1}^{n}\sum_{j=i}^{n} x_i x_j} + 6x_{n+1}^2\sum_{i=1}^{n} x_i + 6x_{n+1}^3.$$

The underlined term becomes

$$3x_{n+1}\left(2\sum_{i=1}^{n} x_i^2 + 2\sum_{i=1}^{n}\sum_{j>i}^{n} x_i x_j\right), \tag{29}$$

where the last index j is strictly larger than i. The overall sum is re-written as the sum of the following three terms, where we have underlined terms corresponding to sum (29):

$$\left(\sum_{i=1}^{n} x_i\right)^3 + \underline{3x_{n+1}\left(\sum_{i=1}^{n} x_i\right)^2} + 3x_{n+1}^2\sum_{i=1}^{n} x_i + x_{n+1}^3 = \left(\sum_{i=1}^{n} x_i + x_{n+1}\right)^3,$$

$$\begin{aligned}&3\left[\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i^2\right) + \underline{x_{n+1}\sum_{i=1}^{n} x_i^2} + x_{n+1}^2\sum_{i=1}^{n} x_i + x_{n+1}^3\right] = \\ =&3\left(\sum_{i=1}^{n} x_i + x_{n+1}\right)\left(\sum_{i=1}^{n} x_i^2 + x_{n+1}^2\right),\end{aligned}$$

$$2\sum_{i=1}^{n} x_i^3 + 2x_{n+1}^3,$$

which correspond to the three sums of the original claim.

B Proof of Lemma 9

Lemma 9 For $n \in \mathbf{N}$, it holds

$$(4!) \sum_{i=1}^{n} \sum_{j=i}^{n} \sum_{k=j}^{n} \sum_{t=k}^{n} x_i x_j x_k x_t =$$
$$\left(\sum_{i=1}^{n} x_i\right)^4 + 6\left(\sum_{i=1}^{n} x_i\right)^2 \left(\sum_{i=1}^{n} x_i^2\right) + 8\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i^3\right) + 6\left(\sum_{i=1}^{n} x_i^4\right) + 3\left(\sum_{i=1}^{n} x_i^2\right)^2$$

Proof. Let $S(n) = \sum_{i=1}^{n} \sum_{j=i}^{n} \sum_{k=j}^{n} \sum_{t=k}^{n} x_i x_j x_k x_t$. The cases $n = 0$ and $n = 1$ are easily verifiable. Let us assume that the theorem holds for $S(n), n \geq 1$. We shall prove it for $S(n+1)$. Clearly,

$$S(n+1) = S(n) + \sum_{i=1}^{n} \sum_{j=i}^{n} \sum_{k=j}^{n} x_i x_j x_k x_{n+1} + \sum_{i=1}^{n} \sum_{j=i}^{n} x_i x_j x_{n+1}^2 + \sum_{i=1}^{n} x_i x_{n+1}^3 + x_{n+1}^4,$$

which, by the inductive hypothesis, yields:

$$24S(n+1) =$$
$$\left(\sum_{i=1}^{n} x_i\right)^4 + 6\left(\sum_{i=1}^{n} x_i\right)^2 \left(\sum_{i=1}^{n} x_i^2\right) + 8\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i^3\right) + 6\left(\sum_{i=1}^{n} x_i^4\right) + 3\left(\sum_{i=1}^{n} x_i^2\right)^2$$
$$+ 24x_{n+1} \sum_{i=1}^{n} \sum_{j=i}^{n} \sum_{k=j}^{n} x_i x_j x_k + 24x_{n+1}^2 \sum_{i=1}^{n} \sum_{j=i}^{n} x_i x_j + 24x_{n+1}^3 \sum_{i=1}^{n} x_i + 24x_{n+1}^4.$$

Let $A = x_{n+1}^3 \sum_{i=1}^{n} x_i$ and $B = x_{n+1}^4$. From Lemma 8, we have

$$\begin{aligned}
&24x_{n+1} \sum_{i=1}^{n} \sum_{j=i}^{n} \sum_{k=j}^{n} x_i x_j x_k = \\
=&4\, x_{n+1} \left(\left(\sum_{i=1}^{n} x_i\right)^3 + 3\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i^2\right) + 2\sum_{i=1}^{n} x_i^3 \right) \\
=&\underbrace{4x_{n+1}\left(\sum_{i=1}^{n} x_i\right)^3}_{=C} + \underbrace{12x_{n+1}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i^2\right)}_{=D} + \underbrace{8x_{n+1}\sum_{i=1}^{n} x_i^3}_{=E}
\end{aligned}$$

And from Lemma 8, we have

$$\begin{aligned}
&24x_{n+1}^2 \sum_{i=1}^{n}\sum_{j=i}^{n} x_i x_j = \\
&=12\, x_{n+1}^2 \left(\left(\sum_{i=1}^{n} x_i \right)^2 + \sum_{i=1}^{n} x_i^2 \right) \\
&= \underbrace{12\, x_{n+1}^2 \left(\sum_{i=1}^{n} x_i \right)^2}_{=2F} + \underbrace{12x_{n+1}^2 \left(\sum_{i=1}^{n} x_i^2 \right)}_{=2G}.
\end{aligned}$$

The overall sum is rewritten using the sum of the following four terms:

$$\left(\sum_{i=1}^{n+1} x_i \right)^4 = \left(\sum_{i=1}^{n} x_i \right)^4 + \underbrace{4x_{n+1} \left(\sum_{i=1}^{n} x_i \right)^3}_{=C} + \underbrace{6x_{n+1}^2 \left(\sum_{i=1}^{n} x_i \right)^2}_{=F} + \underbrace{4x_{n+1}^3 \left(\sum_{i=1}^{n} x_i \right)}_{=4A} + \underbrace{x_{n+1}^4}_{=B}.$$

$$\begin{aligned}
6\left(\left(\sum_{i=1}^{n+1} x_i \right)^2 \left(\sum_{i=1}^{n+1} x_i^2 \right) \right) =&6\left(\left(\left(\sum_{i=1}^{n} x_i \right)^2 + 2x_{n+1} \left(\sum_{i=1}^{n} x_i \right) + x_{n+1}^2 \right) \left(x_{n+1}^2 + \sum_{i=1}^{n} x_i^2 \right) \right) \\
=&6\left(\sum_{i=1}^{n} x_i \right)^2 \left(\sum_{i=1}^{n} x_i^2 \right) + \underbrace{6x_{n+1}^2 \left(\sum_{i=1}^{n} x_i \right)^2}_{=F} + \underbrace{12x_{n+1}^3 \left(\sum_{i=1}^{n} x_i \right)}_{=12A} \\
&+ \underbrace{12x_{n+1} \left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} x_i^2 \right)}_{=D} + \underbrace{6x_{n+1}^4}_{=6B} + \underbrace{x_{n+1}^2 \left(\sum_{i=1}^{n} x_i^2 \right)}_{=G}.
\end{aligned}$$

$$8\left(\sum_{i=1}^{n+1} x_i \right) \left(\sum_{i=1}^{n+1} x_i^3 \right) =8\left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} x_i^3 \right) + \underbrace{8x_{n+1} \left(\sum_{i=1}^{n} x_i^3 \right)}_{=E} + \underbrace{8x_{n+1}^3 \left(\sum_{i=1}^{n} x_i \right)}_{=8A} + \underbrace{8x_{n+1}^4}_{=8B}.$$

$$3\left(\sum_{i=1}^{n+1} x_i^2 \right)^2 =3\left(\sum_{i=1}^{n} x_i^2 \right)^2 + \underbrace{6x_{n+1}^2 \left(\sum_{i=1}^{n} x_i^2 \right)}_{=G} + \underbrace{3x_{n+1}^4}_{=3B}.$$

$$6\left(\sum_{i=1}^{n+1} x_i^4\right) = 6\left(\sum_{i=1}^{n} x_i^4\right) + \underbrace{6x_{n+1}^4}_{=6B}.$$

C Proof of Theorem 13

By definition of the second moment in Eq. 15, we have

$$V\left(\Delta^{n-1}\right) M_1 = \int\limits_{\Delta^{n-1}} A\breve{x} - Av_n + R_n \; d\breve{x} \;\; = \int\limits_{\Delta^{n-1}} A\breve{x} \; d\breve{x} + (-Av_n + R_n) V\left(\Delta^{n-1}\right).$$

From Lemma 9, and simplifying by $V\left(\Delta^{n-1}\right)$, we get

$$\begin{aligned} M_1 &= \left(\frac{1}{n}\sum_{i=1}^{n} Av_i\right) + (-Av_n + R_n) \\ &= \frac{1}{n}\sum_{i=1}^{n}(Av_i - Av_n + R_n) \;= \frac{1}{n}\sum_{i=1}^{n} R_i, \end{aligned}$$

which concludes the proof.

KJ-AE-19-011-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/12907

ISBN 978-92-76-03963-1