Omega 113 (2022) 102718

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

Multi-Time and Multi-Moment Nonparametric Frontier-Based Fund Rating: Proposal and Buy-and-Hold Backtesting Strategy[☆,☆☆]

Kristiaan Kerstens[a], Paolo Mazza[c], Tiantian Ren[b,c,*], Ignace Van de Woestyne[d]

[a] Univ. Lille, CNRS, IESEG School of Management, UMR 9221 - LEM - Lille Économie Management, Lille F-59000, France
[b] School of Business Administration, Hunan University, Changsha 410081, China
[c] IESEG School of Management, Univ. Lille, CNRS, UMR 9221 – LEM – Lille Economie Management, 3 rue de la Digue, F-59000 Lille, France
[d] KU Leuven, Research Centre for Operations Research and Statistics (ORSTAT), Brussels Campus, Warmoesberg 26, Brussels B-1000, Belgium

ARTICLE INFO

Article history:
Received 8 January 2022
Accepted 20 June 2022
Available online 22 June 2022

JEL classification:
D24
G11

Keywords:
Data Envelopment Analysis
Shortage function
Frontier
Fund rating

ABSTRACT

This contribution introduces new frontier models to rate mutual funds that can simultaneously handle multiple moments and multiple times. These new models are empirically applied to hedge fund data, since this category of funds is known to be subject to non-normal return distributions. We define a simple buy-and-hold backtesting strategy to test for the impact of multiple moments and multiple times separately and jointly. The empirical results demonstrate that the proposed frontier models perform better than most financial performance measures and existing frontier models in selecting promising funds.

1. Introduction

The foundational work of Markowitz [60] in modern portfolio theory has learned every investor that to gauge the performance of portfolio management one must consider risk in addition to return. This mean-variance (MV) dual objective of maximizing returns and minimizing risks turns performance evaluation into a controversial task involving trade-offs related to the risk preferences of the investor. The two-dimensional nature of this nonlinear quadratic optimization problem allows to display the efficient frontier as a Pareto-optimal subset of portfolios whereby the expected return can only increase when also the variance increases.

A large part of modern portfolio theory continues developing variations on these bi-objective MV optimization problems. A wide offer of alternative risk measures is available in the portfolio literature: entropy, expected shortfall, mean absolute deviation, semi-variance and other partial moment measures, Value-at-Risk (VaR) in all its variations, etc. (see, e.g., Bacon [4] and Feibel [35] for surveys).[1]

This focus on the first two moments of a random variable's distribution is only consistent with the von Neumann-Morgenstern axioms of choice underlying expected utility (EU) theory when: (i) asset processes follow normal distributions, or (ii) investors have quadratic utility functions.[2] A substantial empirical literature has documented that normality of asset returns can be rejected for a variety of financial asset classes in both developed and emerging financial markets (e.g., Jondeau and Rockinger [44]). At least since Scott and Horvath [69], investors have been attributed a positive preference for skewness as well as a negative preference for kurtosis to explain financial behavior. Meanwhile, decision-theoretic arguments exist for what has become known as the broad class of mixed risk-aversion utility functions that are characterized by a preference for odd moments and an aversion for even moments

☆ We gratefully acknowledge the help of Mikael Petitjean in collecting the data. We also acknowledge the most constructive comments of Vincent Théroude and two referees of this journal.
☆☆ Area: Data-Driven Analytics. This manuscript was processed by Associate Editor Victor Podinovski.

* Corresponding author.
E-mail addresses: k.kerstens@ieseg.fr (K. Kerstens), p.mazza@ieseg.fr (P. Mazza), ttren@hnu.edu.cn, t.ren@ieseg.fr (T. Ren), ignace.vandewoestyne@kuleuven.be (I. Van de Woestyne).

[1] More rarely alternatives are proposed for the expected return: e.g., Benati [6] focuses on the median as a location parameter of the distribution of returns.

[2] Von Neumann and Morgenstern [77] propose that under some axioms of rational behaviour (i.e, ordering, continuity, and independence axioms), the decisions of agents under uncertainty are based on maximizing their EU. Transposed into a portfolio context, the portfolio selection of investors should maximize the EU associated with the uncertain return of assets.

https://doi.org/10.1016/j.omega.2022.102718
0305-0483/

(see Eeckhoudt and Schlesinger [27]). Furthermore, via surveys and experiments traditional risk preferences like risk aversion, but also higher order risk preferences like prudence and temperance are nowadays better understood (see Trautmann and van de Kuilen [75] for a review).

Over time, several alternative portfolio selection criteria based on preferences for higher-order moments have been developed. But, so far not a single widely accepted criterion seems to have emerged. It is possible to distinguish between primal and dual approaches to determine such higher-order moments portfolio frontiers. One example of the primal approach is found in Lai [53] who determines mean-variance-skewness (MVS) optimal portfolios via a Polynomial Goal Programming procedure. The dual approach necessitates a specification of some indirect higher-moment utility function and yields optimal portfolios via its parameters reflecting higher-moment preferences (e.g., Harvey et al. [42]).

To our knowledge, Sengupta [71] is the first to introduce an efficiency measure -borrowed from production theory- into a diversified MV portfolio model. This efficiency measure relates to the distance function that for a long time has been employed in consumer theory and especially in production theory (e.g., Cornes [24]). In consumer theory the distance function is dual to the expenditure function: it serves to characterize multiple commodity and single utility choice sets.[3] In production theory the input distance function is dual to the cost function: it basically serves to characterize multiple input multiple output production possibility sets (e.g., Hackman [41]). This has opened up a booming research field where parametric but particularly nonparametric specifications of production and dual (e.g., cost) frontiers are specified based on minimal maintained axioms (e.g., constant or variable returns to scale, convexity or not, etc.). Applied to a plethora of private and public sectors, these frontier methodologies analyse technical, scale or cost efficiency, economies of specialization, mergers, etc. (e.g., Färe et al. [33]).[4]

The introduction of an efficiency measure into portfolio theory allows to gauge performance over multiple dimensions and it opens up new perspectives. On the one hand, following Briec et al. [19] who establish duality between a distance function and MV utility functions, Briec et al. [18] use a general distance function (named shortage function) to look for improvements in efficiency in MVS space by looking for simultaneous expansions in mean return and positive skewness and reductions in risk. Furthermore, these authors provide a duality result with a MVS utility function.[5] Even more general, for the class of mixed risk-aversion utility functions, Briec and Kerstens [17] assess portfolio performance for the general moments case by simultaneously looking for improvements in odd moments and reductions in even moments. In addition, these authors establish duality with general moment utility functions.

Empirical applications of this diversified multi-moment approach are found in Adam and Branda [1], Branda [10], Branda and Kopa [12], Branda [11], Joro and Na [45], Jurczenko et al. [46], Khemchandani and Chandra [50], Krüger [51], Massol and Banal-Estañol [61], among others. Furthermore, Bacmann and Benedetti [3], Boudt et al. [9], and Jurczenko and Yanou [47], among others, are empirical diversified multi-moment contributions focusing on hedge funds (HF).

On the other hand, within a standard MV framework, Morey and Morey [63] develop a multiple time horizon assessment: in particular, these authors use either a risk contraction or a return expansion efficiency measure to evaluate MV performance over three time horizons simultaneously (in particular, a 3, 5 and a 10-year time period). This contribution is slightly generalized in Briec and Kerstens [16].[6] An empirical application is available in Ren et al. [67].

To the best of our knowledge, Murthi et al. [64] is the seminal article that has been rating mutual funds (MF) by simultaneously trying to maximize the return and minimizing standard deviation, expense ratio, load, and turnover using a nonparametric production frontier specification that maintains convexity and constant returns to scale. Following Farrell [34] and Charnes et al. [23], nonparametric production frontiers are transposed into the financial literature in an effort to provide alternative fund ratings. Intuitively, nonparametric production frontiers can envelop the observations of any multi-dimensional choice set and position each of the observations relative to the boundary of the choice set using some efficiency measure. This has led to a growing literature that has been applied to a large variety of financial assets (e.g., exchange traded funds, hedge funds, pension funds, etc.). Furthermore, a wide variety of model specifications are available in terms of some combination of ordinary moments, lower and/and upper partial moments, as well as in terms of production frontier specifications (constant or variable returns to scale, etc.), and the choice of efficiency measure (e.g., reducing variables for which less is better (like inputs), or expanding variables for which more is better (like outputs), or some combination of both). This frontier-based MF rating literature has been rather recently surveyed in Basso and Funari [5].

Following Heffernan [43] and Blake [7], among others, Kerstens et al. [48] interpret this funds rating literature as a transposition of the characteristics approach in consumer theory into finance: MF are seen as fee-based financial products characterized by distributional characteristics of the asset price distribution as summarized by some combination of moments. Compared to the diversified portfolio models that require nonlinear programming, these nonparametric production frontier MF rating models can normally be solved using simple linear programming.

An open question is how the diversified portfolio models relate to the nonparametric production frontier specifications? Recently, Liu et al. [58] state that a convex variable returns to scale nonparametric production frontier specification provides an inner approximation to the traditional MV diversified portfolio model. This is certainly correct. One basic idea implicit in their contribution is that nonparametric production frontier specifications should ideally underestimate the eventual performance of a diversified portfolio model. In the more general case where we want to explore a nonconvex diversified MV (e.g., with some integer constraints) or a nonconvex higher moment portfolio model, then one can argue that the nonconvex nonparametric production frontier specification with variable returns to scale already advocated by Kerstens et al. [48] provides a conservative underestimation of the corresponding nonconvex diversified portfolio model within some common subspace of moments (see also Germain et al. [37]). By contrast, the more widely used convex nonparametric production frontier specification may overestimate the corresponding nonconvex diversi-

[3] This distance function has sometimes been employed to make welfare comparisons (e.g., Slesnick [72]). More recently, Briec et al. [15] stress that the directional distance function is dually linked to the weighted and indirect Rawlsian social welfare functions.

[4] This nonparametric approach to production is sometimes labeled Data Envelopment Analysis (DEA) because observations are enveloped subject to some minimal set of axioms.

[5] Briec et al. [20] establish a relation between MVS portfolio optimisation using the shortage function and the far more popular Polynomial Goal Programming method proposed by Lai [53].

[6] Note that the use of multiple time horizons within a MV framework is not particularly computationally challenging, but moving from a quadratic convex MV problem to a cubic nonconvex MVS portfolio optimization problem is computationally harder. Evidently, the same remark applies when one moves from a cubic nonconvex MVS to a quartic nonconvex mean-variance-skewness-kurtosis portfolio optimization problem, or beyond by including even higher order moments.

fied portfolio model within the common subspace of moments. The latter argument seems to have escaped attention so far: this explains why most nonparametric production frontier MF rating models with higher moments do impose convexity (for instance, Gregoriou et al. [40]).

The use of distance functions or efficiency measures in both the diversified portfolio models and the nonparametric production frontier specifications leads to the question how these gauges relate to traditional financial performance measures (see, e.g., the surveys in Bacon [4], Feibel [35] and Caporin et al. [22]). While relative performance measures that are variations on returns per unit of risk (e.g., Sharpe ratio) are useful to handle bi-objective (e.g., MV) optimization problems, they are of little use beyond two dimensional problems. E.g., adding a skewness constraint to a MV diversified model weakly decreases return and weakly increases variance inevitably yielding a weakly worse Sharpe ratio: hence, the Sharpe ratio cannot assess higher moment portfolios. If finance wants to handle mixed risk-aversion preferences of investors, then it must consider a multidimensional performance measure. Some performance measures try to assess the tail risk, like VaR or the Conditional Value-at-Risk (CVaR), but they most of the time focus on the risk component and do not include the first moment of the return distribution.

One exception is the Omega ratio that we include in our analysis. Caporin et al. [22] classify the distance (shortage) function approach correctly among the absolute performance measures: these performance measures are based on rewards when compared to those of a reference portfolio on a portfolio frontier. The choice for distance (shortage) function brings finance and portfolio analysis in line with consumer and production analysis where these micro-economic tools have a proven track record in representing multidimensional choice sets.[7]

The first major objective of this contribution is to define new distance functions or efficiency measures that can simultaneously handle both multiple moments and multiple times (instead of either multiple moments or multiple times separately) compatible with general mixed risk-aversion investor preferences. To the best of our knowledge, the existing literature on traditional financial performance indicators as well as the literature on nonparametric frontiers to gauge MF performance have so far produced less general efficiency measures. The application of the shortage function also guarantees the possibility of dealing with negative data for the output-like variables in MF performance assessment. This performance measure offers a simple and powerful tool for assessing MF performance based on the moment characteristics over all time periods. To the best of our knowledge, this basic idea is new and unavailable in the literature. This performance measure thus aims not only to evaluate to which extent a MF performs well in the several moments following mixed risk-aversion preferences, but it simultaneously is assessing to which extent a MF performs well in all these moments over different times. This is important given the concern in the financial literature that traditional performance measures may exhibit limited stability over time (e.g., Bodson et al. [8], Menardi and Lisi [62] and Grau-Carles et al. [38], among others).

As a second major objective, by positioning ourselves into a nondiversified nonparametric frontier-based approach, our contribution avoids computational limitations. The presence of multiple moments besides mean and variance leads to diversified portfolio models that are nonconvex and nonsmooth. On the one hand, diversified portfolio models with higher-order moments suffer from enormous computational costs even for moderate MF universes.[8] On the other hand, it is difficult to guarantee global optima when solving nonconvex and nonsmooth diversified portfolio models. These computational drawbacks are exacerbated when accounting for the performance of MF over different times in a diversified portfolio model. By contrast, the nondiversified nonparametric frontier models are simply solved by a linear programming (LP) problem or an implicit enumeration algorithm for the binary mixed integer linear programming (BMILP). Thus, the nondiversified models are computationally superior over the diversified models when handling multiple moments and multiple times separately and jointly.

This new performance measure is applied to HFs, a fund accessible only to institutional investors and high net worth individuals. Among MFs, HFs have a unique compensation structure. The most widespread fee structure is the so-called 2/20, i.e., 2% of assets under management for annual management fees and 20% of any profits made as a performance incentive fee. Consequently, HFs are marked by their heterogeneity and unusual (i.e., non-normal) statistical properties, as compared to more traditional MFs. Indeed, HFs tend to exhibit some more strongly asymmetric and fat tailed return characteristics compared to other MFs (see Gregoriou [39], Darolles and Gourieroux [25], Eling and Faust [29], among others, and especially El Kalak et al. [28] for a survey). Furthermore, Racicot and Théoret [65,66] develop time-varying measures of co-skewness and co-kurtosis: their work reveals that the behavior of HFs tends to trade off return for higher moments when building optimal portfolios, and this behavior is asymmetric in relation to the phase of the economic cycle. They are globally viewed as riskier but are also associated with higher rewards. This is why our empirical study specifically focuses on HFs since these are most likely to be affected by higher order moments.

The traditional financial performance measures (e.g., Sharpe ratio, Sortino ratio, etc.) used for HF rating have been subject to some criticism, because they basically follow the theoretical assumptions of the Capital Asset Pricing Model (CAPM) that the capital market is efficient and financial asset returns are normally, independently and identically distributed, among others. When asset returns do not obey the normal distribution, then the mean and variance no longer suffice to effectively summarize its return distribution. Several studies extend the conventional two-moment CAPM by incorporating the effects of systematic skewness and kurtosis. In the four-moment CAPM, in addition to systematic variance also systematic kurtosis and skewness contribute to the risk premium of an asset (e.g., Fang and Lai [32], Friend and Westerfield [36] and Sears [70], among others).[9] While the four-moment CAPM to some extent refines the conventional CAPM, it still makes stronger assumptions on the return distribution of assets compared to the nonparametric frontier models when applied to HF appraisal. Given the complexities to assess the performance of HFs using traditional performance measures (e.g., see Smith [73]), we think that our new performance measure may provide a suitable framework to evaluate both persistence across moments and across times.

In a HF context, the need for multiple moments is apparent in a multitude of nonparametric production frontier studies: examples include, e.g., Gregoriou et al. [40], Kumar et al. [52], Germain et al. [37], among others. However, to the best of our knowledge none of these studies appeal to the characteristics approach as proposed

[7] Tammer and Zălinescu [74] show that the shortage function is linked to the scalarization function that is used in vector optimization problems, of which multi-objective optimization problem is a special case.

[8] For a given financial universe containing n MFs, the co-variances, co-skewnesses and co-kurtosises of the MFs are $n \times n$, $n \times n \times n$, and $n \times n \times n \times n$ tensors, respectively. Furthermore, a simple case is provided in Appendix to explain the computational costs of the diversified models with multiple moments.

[9] Consistent with the co-variance definition, the measures of co-skewness and co-kurtosis are proposed based on an four-moment CAPM model. Back [2] provides a systematic discussion of the properties of these two measures.

by Kerstens et al. [48]. Furthermore, all these existing nonparametric production frontier studies are single time: this contribution is the first to develop a multi-time evaluation framework. Therefore, as a third major objective, we focus on the impact of multiple moments and multiple times separately and jointly surrounding the application of nonparametric frontiers when assessing the performance of HFs. We employ a Li-test statistic (initially proposed in Li [54]) to empirically test the necessity of multiple moments and multiple times in HF appraisal. Thereafter, by means of a backtesting approach in a buy-and-hold setting, the potential benefits and superiority of the multi-moment and multi-time frontier ratings compared to most existing traditional and frontier MF ratings are empirically illustrated.

The remainder of this contribution is organized as follows. The next Section 2 introduces the nonparametric production frontiers that serve to approximate the diversified portfolio models: we first discuss single-time multi-moment models, then we introduce the new multi-time multi-moment models. In Section 3, we develop the buy-and-hold backtesting strategy in detail. Section 4 describes the hedge fund data in detail and comments upon the empirical results. Finally, Section 5 concludes.

2. Nonparametric Frontier Rating Models: Methodology

2.1. Single-Time and Multi-Moment Rating Framework

The nonparametric frontier rating methods gauge the financial performance of MF, and these evaluations are done mostly using frontier-based models which originate from production theory. In this section, we only introduce the basic definitions and properties needed for applications within finance. Assume that there are n MFs under evaluation over a given time horizon. At time t in this time horizon, the j-th MF ($j \in \{1,\dots,n\}$) is characterized by m input-like values x^t_{ij} ($i \in \{1,\dots,m\}$) and s output-like values y^t_{rj} ($r \in \{1,\dots,s\}$). Input-like variables need to be minimized and output-like variables need to be maximized.

We introduce one widely used production frontier-based model with variable returns to scale (VRS).[10] Following Briec et al. [21], a unified algebraic representation of convex and nonconvex production possibility sets (PPS) under the VRS assumption for a sample of n MFs at time t is:

$$P^t_\Lambda = \Bigg\{ (x^t, y^t) \in \mathbb{R}^m \times \mathbb{R}^s \mid \forall i \in \{1,\cdots,m\} : x^t_i \geq \sum_{j=1}^n \lambda_j x^t_{ij}, \forall r \in \{1,\cdots,s\} : y^t_r \leq \sum_{j=1}^n \lambda_j y^t_{rj}, \lambda \in \Lambda \Bigg\}, \tag{1}$$

where:

$\Lambda \equiv \Lambda^C = \{\lambda \in \mathbb{R}^n \mid \sum_{j=1}^n \lambda_j = 1 \text{ and } \forall j \in \{1,\dots,n\} : \lambda_j \geq 0\}$ if convexity is assumed, and

$\Lambda \equiv \Lambda^{NC} = \{\lambda \in \mathbb{R}^n \mid \sum_{j=1}^n \lambda_j = 1 \text{ and } \forall j \in \{1,\dots,n\} : \lambda_j \in \{0,1\}\}$ if nonconvexity is assumed.

At time t, if there exists an input-output combination $(\sum_{j=1}^n \lambda_j x^t_{ij}, \sum_{j=1}^n \lambda_j y^t_{ij})$ in the convex or nonconvex PPS using less inputs and producing more outputs than the observed MF, then this MF is considered inefficient since it can improve its inputs and/or outputs. MFs are efficient if no improved input-output combinations can be found. The input-output combinations of these efficient MFs are all located at the boundary of P^t_Λ which is called the convex or nonconvex VRS (VRSc and VRSnc for short hereafter) nonparametric frontier.

Using the nonparametric PPS defined in (1), the shortage function of any observed MF at time t is now defined as follows:

Definition 2.1. At time t, let $g^t = (-g^t_x, g^t_y) \in \mathbb{R}^m_- \times \mathbb{R}^s_+$ and $g^t \neq 0$. For any observation $z^t = (x^t, y^t) \in \mathbb{R}^m \times \mathbb{R}^s$, the shortage function S^t_Λ at time t in the direction of vector g^t is defined as:

$$S^t_\Lambda(z^t; g^t) = \sup\{\beta \in \mathbb{R} \mid z^t + \beta g^t \in P^t_\Lambda\}.$$

This shortage function simultaneously permits the enhancement of output-like variables and the reduction of input-like variables. If the shortage function value $S^t_\Lambda(z^t; g^t) > 0$ for the input-output combination $z^t = (x^t, y^t)$ of a specific MF at time t, then z^t is not located on the frontier of P^t_Λ. Hence, its inputs and/or outputs can be improved to catch up with the VRS nonparametric frontier. By contrast, if the shortage function value $S^t_\Lambda(z^t; g^t) = 0$, then z^t is located on the frontier.

Consider a MF with index $o \in \{1,\dots,n\}$ in need of assessment at time t by means of the shortage function with direction vector $g^t_o = (-g^t_{xo}, g^t_{yo}) \in \mathbb{R}^m_- \times \mathbb{R}^s_+$. Combining (1) and Definition 2.1, the efficiency status of this MF at time t can be determined by solving the following model:

$$\begin{aligned} \max\ & \beta \\ \text{s.t.}\ & \sum_{j=1}^n \lambda_j x^t_{ij} \leq x^t_{io} - \beta g^t_{io}, \quad i = 1,\cdots,m, \\ & \sum_{j=1}^n \lambda_j y^t_{rj} \geq y^t_{ro} + \beta g^t_{ro}, \quad r = 1,\cdots,s, \\ & \sum_{j=1}^n \lambda_j = 1, \quad \beta \geq 0, \\ & \forall j = 1,\cdots,n : \begin{cases} \lambda_j \geq 0, & \text{under convexity}, \\ \lambda_j \in \{0,1\}, & \text{under nonconvexity}. \end{cases} \end{aligned} \tag{2}$$

Note that model (2) results in a LP problem under convexity and a BMILP problem under nonconvexity. In the empirical application, the direction vector is in general set as $g^t_o = (-x^t_{1o},\dots,-x^t_{mo}, |y^t_{1o}|,\dots,|y^t_{so}|)$ to accommodate eventual negative values of return and skewness, whereby all input-like values x^t_{io}, $(i = 1,\dots,m)$ and output-like values y^t_{ro}, $(r = 1,\dots,s)$ are simultaneously increased and decreased in proportion to their initial values, respectively. The optimal value β^* measures the resulting proportional amount of inefficiency representing the shortage function.

Note that the value of inefficiency β^* determined by model (2) in the nonconvex case is always less than that determined in the convex case. Naturally, the number of efficient MFs obtained using the nonconvex case is also larger than the ones obtained using the convex case when assessing a set of MFs to be evaluated. Mathematically, these properties are a direct consequence of the restrictions of weights λ_j $(j = 1,\dots,n)$ in model (2) under convexity and nonconvexity. Clearly, the nonconvex model is more restrictive on the weights λ_j than the convex model. In terms of efficient frontiers, the VRSnc nonparametric frontier is always located below the VRSc one (see Kerstens et al. [48]), since the latter imposes the convexity axiom that allows to linearly combine MFs.[11]

The setting defined in the previous section is general and flexible and can thus handle a large choice of inputs and outputs. We now particularize the above formulation to characterize the

[10] Remark that a VRS frontier model is the most general representation of a technology allowing for increasing, constant, or decreasing returns to scale at different points on the production frontier.

[11] Convexity is not always useful for guiding investors in terms of selecting MFs whenever there are nonconvexities at stake (e.g., higher moments of the asset returns, cardinality constraints on the number of assets, etc.).

efficient frontier in the MVS and the mean-variance-skewness-kurtosis (MVSK) spaces. Suppose that there are n MFs under evaluation. At time t, let $R_1^t, \dots, R_n^t$ denote the raw returns of the n funds, which are characterized by their expected return $E(R_j^t)$, variance $V(R_j^t)$, skewness $S(R_j^t)$ and kurtosis $K(R_j^t)$ for $j \in \{1, \dots, n\}$. Here, the calculations of variance, skewness and kurtosis are expressed as follows: $V(R_j^t) = E[(R_j^t - E(R_j^t))^2]$, $S(R_j^t) = E[(R_j^t - E(R_j^t))^3]$, and $K(R_j^t) = E[(R_j^t - E(R_j^t))^4]$.[12] To obtain a detailed specification of the PPS, as defined in (1), we need to classify the different goals of the investor in terms of either inputs (i.e., objectives to minimize), or outputs (i.e., objectives to maximize). As discussed in the previous section, the need for multiple moments is apparent to assess MFs (and most particularly HFs) whose return distributions may exhibit strong asymmetry and fat tails. Given mixed risk-aversion utility functions, investors express a preference for odd moments and a dislike for even moments of the distribution of asset returns. Therefore, when the MVSK framework is considered, we can define the first and second inputs of MFs as $x_{1j}^t = V(R_j^t)$ and $x_{2j}^t = K(R_j^t)$, and the first and second outputs as $y_{1j}^t = E(R_j^t)$ and $y_{2j}^t = S(R_j^t)$ for $j \in \{1, \dots, n\}$. Obviously, for the MVS case only the first input is considered.

For a MF o under evaluation at time t, denote $E_o = E(R_o^t)$, $V_o = V(R_o^t)$, $S_o = S(R_o^t)$ and $K_o = K(R_o^t)$. Then both models, either with convexity or nonconvexity, allow to project the input-output combination (V_o, K_o, E_o, S_o) of this MF in such a way that inputs (i.e., variance and kurtosis) are decreased and outputs (i.e., expected return and skewness) are increased in the direction g_o^t. The optimal solution β^* of model (2) measures how many times the direction vector g_o^t fits in the line segment from the input-output combination of the MF o to the efficient frontier in the direction of g_o^t.

In model (2) under convexity, the left-hand sides of the constraints are all linear. All possible linear combinations of inputs and outputs of the observed MFs are used to construct a convex VRS frontier for evaluation. For the MF o, if $\beta^* = 0$, the corresponding input-output combination is on the convex frontier and efficient at time t. If $\beta^* > 0$, there exist input-output combinations yielding a higher or equal return and skewness together with a lower or equal variance and kurtosis. When nonconvexity is assumed in model (2), evaluation is done with respect to a nonconvex VRS frontier determined by all efficient MFs (excluding the convex input-output combinations of these).

2.2. Multi-Time and Multi-Moment Rating Framework

Differing from MF ratings in a single-time framework, MF ratings in a multi-time framework consider performance over a time horizon consisting of multiple discrete time periods. In this respect, Morey and Morey [63] and Briec and Kerstens [16] emphasize the importance of the multiple period assessment to find out the maximum improvements possible for a MV portfolio over a multi-period time horizon. Inspired by this multi-time perspective, the objective of our methodological extension is to offer a generalized efficiency measure for evaluating MF performance based on the multiple moment characteristics over all time periods simultaneously. The multi-time and multi-moment ratings thereby developed potentially identify promising MFs that have both persistence across moments and across times. To develop the nonparametric frontier rating models in this multi-time framework, some definitions and properties are presented.

The fundamental idea of multi-time rating is to combine the static evaluation based on nonparametric multi-moment frontiers with the concept of a temporal shortage function proposed in Briec et al. [14]. This temporal efficiency measure is developed based on the assumption of time separability: there are no temporal linkages between each of the estimated technologies in each sub-period. In an investment context, the typical investor attempts to select MFs by their performances over multiple periods (e.g., 1 year, 3 years, and 5 years) starting from a certain initial time in which he decides on an investment. Therefore, an explicit temporal linkage between multiple time periods is not needed in this context, since only a given investment in the initial time period is at stake and not some optimal investment trajectory. Therefore, our multi-time MF rating is based on the idea of the temporal efficiency measurement that explicitly aims to provide an overall weighting scheme for MF performance of a given initial investment over multiple time periods. Thus, it is only dynamic in a limited sense.

By contrast, if one would be interested to analyse the overall performance of MFs based on a series of adjacent periods throughout the whole investment process in which investments are made in some optimal way throughout the whole investment horizon, then this involves a dynamic structure accounting for the intermediate connections between adjacent periods for the dynamic portfolio assessment. Therefore, for cases in which the temporal separability does not hold, one ideally needs truly dynamic portfolio models in either continuous or discrete time: see, e.g., Lin et al. [56] for an example.

Consider n MFs under evaluation. Let T denote the number of consecutive times in a time horizon of interest. In addition, define a multi-time path of inputs and outputs as $Z_j = (x_j^t, y_j^t)_{t=1}^T$ for MF j, $(j = 1, \dots, n)$, where $x_j^t = (x_{1j}^t, \dots, x_{mj}^t)$ and $y_j^t = (y_{1j}^t, \dots, y_{sj}^t)$ represent m inputs and s outputs at time t, respectively. Assuming VRS for all times $t \in \{1, \dots, T\}$ and strong free disposability of all inputs and outputs, the multi-time PPS with convexity and nonconvexity can be defined as:

$$\mathbf{P}_\Lambda^T = P_\Lambda^1 \times \cdots \times P_\Lambda^T \subset (\mathbb{R}^m \times \mathbb{R}^s)^T \cong \mathbb{R}^{m\times T} \times \mathbb{R}^{s\times T}, \tag{3}$$

where P_Λ^t, $(t = 1, \dots, T)$, is the PPS at time t mentioned previously in (1).

The idea is now for each MF to simultaneously expand its multiple outputs and decrease its multiple inputs over all discrete times in a given time horizon by means of the multi-time shortage function. To allow a general definition, we first introduce some abbreviating notations.

The time dependent direction vector denoted by $G = (g^1, \dots, g^T) \in (\mathbb{R}_-^m \times \mathbb{R}_+^s)^T \cong \mathbb{R}_-^{m\times T} \times \mathbb{R}_+^{s\times T}$ represents a given multi-time direction path, where $g^t = (-g_x^t, g_y^t) \in \mathbb{R}_-^m \times \mathbb{R}_+^s$ represents the direction vector at time $t \in \{1, \dots, T\}$. In addition, we denote $\Theta = (\beta_1, \dots, \beta_T) \in \mathbb{R}^T$ and $\Theta \cdot G = (\beta_1 g^1, \dots, \beta_T g^T) \in (\mathbb{R}^m \times \mathbb{R}^s)^T \cong \mathbb{R}^{m\times T} \times \mathbb{R}^{s\times T}$. Considering the time preference of an investor in a portfolio context, we introduce a time discounting factor denoted ξ $(0 < \xi < 1)$ to weight the efficiency measures over the time horizon. Then, the time discounted multi-time shortage function assuming convexity or nonconvexity is defined as follows:

Definition 2.2. With the notations introduced above, for any observation $Z \in (\mathbb{R}^m \times \mathbb{R}^s)^T \cong \mathbb{R}^{m\times T} \times \mathbb{R}^{s\times T}$, the time discounted multi-time shortage function S_Λ^T in the direction of G is defined as:

$$S_\Lambda^T(Z; G) = \sup \left\{ \frac{1}{T} \sum_{t=1}^{T} \xi^{T-t} \beta_t \mid Z + \Theta \cdot G \in \mathbf{P}_\Lambda^T \right\}.$$

For a given time horizon T, this amounts to looking for the largest arithmetic mean of time discounted distances over all times in a given time horizon of the input-output combinations of an observed MF to boundary of $\mathbf{P}_\Lambda^T$. This definition adapts a weighted (discounted) temporal efficiency measure, whereby the weights de-

[12] Note that the four moments of the return distribution are computed based on the historical returns observed in an estimation time window of a given length.

cline as one moves away from the present into the past.[13] If the time discounted multi-time shortage function value $S^T_\Lambda(Z;G) > 0$ for the input-output path Z of the MF being evaluated, then it means that its inputs and outputs can be reduced and improved simultaneously in one or more time periods.

Based on Definition 2.2, we are now in the position to determine the nonparametric frontier rating models in a general formulation. Suppose there are n MFs under evaluation. Let T denote the number of consecutive times in a time horizon under consideration. In particular, the multi-time rating methods used in Section 3 focus on 3 distinct time periods: 1, 3 and 5 years. For a given multi-time direction path $G = (g^t)_{t=1}^T \in \mathbb{R}_-^{m\times T} \times \mathbb{R}_+^{s\times T}$, the efficiency of the MF o under evaluation can be determined by the time discounted multi-time shortage function value resulting from the following program:

$$\begin{aligned}
\max\ & \frac{1}{T}\sum_{t=1}^{T}\xi^{T-t}\beta_t \\
\text{s.t.}\ & \sum_{j=1}^{n}\lambda_j^t x_{ij}^t \le x_{io}^t - \beta_t g_{io}^t, \quad i=1,\cdots,m, \\
& \sum_{j=1}^{n}\lambda_j^t y_{rj}^t \ge y_{ro}^t + \beta_t g_{ro}^t, \quad r=1,\cdots,s, \\
& \sum_{j=1}^{n}\lambda_j^t = 1, \quad \beta_t \ge 0, \quad t=1,\cdots,T, \\
& \forall j=1,\cdots,n: \begin{cases}\lambda_j^t \ge 0,\ t=1,\cdots,T, & \text{under convexity,}\\ \lambda_j^t \in \{0,1\},\ t=1,\cdots,T, & \text{under nonconvexity.}\end{cases}
\end{aligned} \tag{4}$$

In the multi-time framework, we select variance and kurtosis of each time t, $(t = 1,\ldots,T)$, as inputs and expected return and skewness as outputs, whereas for the MVS case only variance for each t is considered as inputs. With the help of the time discounted multi-time shortage function, the observed MF with index o can improve its multiple return and skewness dimensions and reduce its multiple variance and kurtosis dimensions along a given direction path G over all time periods. The value of the objective function of model (4) indicates the amount of (in)efficiency of the MF o representing the multi-time shortage function. A value greater than zero indicates that the inputs and outputs of the evaluated MF can be improved in one or more time periods. The path of input-output combinations is thus situated below the boundary of the multi-time PPS, and thus is inefficient from a multi-time perspective.[14]

Again, it is trivial to prove that the value of the multi-time shortage function computed by model (4) with nonconvexity is always less than the one computed by this model with convexity. Furthermore, the number of efficient MFs determined by the multi-time shortage function in the nonconvex context is larger than that determined by the multi-time shortage function in the convex context for MF assessments.

[13] For retrospective benchmarking based on observed past behavior when assessing performance, the distant past is less valuable than the nearby present (as indicated by Briec and Kerstens [16]). In that sense, the distant past contributes less weight to efficiency gains than the nearby past.

[14] Practically speaking, our frontier rating methods are remarkably flexible in terms of the inclusion of either multiple moments or multiple times. For instance, we can specify only including expect return and variance to measure the performance of MFs in case a normal return distribution is valid. Also, the desired times to be included depend on the actual needs of investors involved in the MFs assessment and selection process. While using frontier-based methods, it is recommended that one employs Li-test statistics to assess these key methodological choices in terms of moments and times to include. In our empirical study, the HFs sample database has been tested in terms of the necessity of multiple moments (skewness and kurtosis) and multiple times separately and jointly.

Due to the time separability assumption mentioned above, the mathematical program (4) is a block-diagonal LP or BMILP, since there are no temporal linkages among the MF assessments for each time period. Mathematically, one can solve the static mathematical program (2) for each time period separately and compute the objective function of model (4) based on the optimal solutions of these T sub-problems (see Briec et al. [14] and Briec and Kerstens [16]).

In the following Sections 4 and 5, we employ MF data to compare the proposed multi-time and multi-moment measures with traditional financial measures, as well as with single-time MV measures. These comparisons are aimed not only to illustrate the impact of multiple moments and multiple times on MF performance evaluation, but more importantly to further explore the potential benefits of the newly proposed performance measures for MF selection by means of backtesting. We now turn to explain the backtesting framework in Section 3.

3. Backtesting Framework

Our main objective in this contribution is to test that the multi-time and multi-moment performance measures can be expected to perform well for MF ratings and selection. To this end, a comparative approach based on a backtesting methodology is adopted. Backtesting refers to executing fictitious investment strategies using historical data to simulate how these strategies would have performed if they had actually been adopted by MF managers in the past.[15]

It is powerful for evaluating and comparing the performance of different investment strategies without using real capital. Some examples of a backtesting approach are found in DeMiguel et al. [26], Tu and Zhou [76], Brandouy et al. [13], Zhou et al. [79] and Lin and Li [57], among others.

For comparison, there are 15 fund rating methods in total being collected in our work. On the one hand, we test some popular traditional financial indicators: Sharpe ratio, Sortino ratio and Omega ratio. The exact definition for the Sharpe, Sortino and Omega ratios can be found in Feibel [35, p. 187 and p. 200] and Eling and Schuhmacher [30, p. 2635], respectively. Based on these definitions and notations introduced in Section 2, these three traditional financial ratios for MF j, $(j \in \{1,\ldots,n\})$ at time t are presented as follows:

$$\text{Sharpe}_j^t = \frac{E(R_j^t) - r_f}{\sigma(R_j^t)}, \tag{5}$$

$$\text{Sortino}_j^t = \frac{E(R_j^t) - r_f}{\sigma_-(R_j^t)}, \tag{6}$$

$$\text{Omega}_j^t = \frac{E(R_j^t) - L}{E[\max(L - R_j^t, 0)]} + 1, \tag{7}$$

where $E(R_j^t)$ and r_f represent the expected return and the risk-free rate, respectively; $\sigma(R_j^t)$ and $\sigma_-(R_j^t)$ denote the standard and lower semi-standard deviations, respectively; L is the loss threshold, in particular, above this threshold returns are considered gains, while below this threshold these are regarded as losses. Using the above three ratios, we obtain the financial indexes for the above n MFs, which can be use to measure their performance at time t, and the higher the value, the better the performance. The risk-free

[15] The use of a backtesting approach is implicitly linked to the hypothesis of efficient markets whereby participants in the financial market have no effect on prices: thus, a given investment strategy of one individual investor does not affect the observed results of the financial market in which he/she is operating.

Table 1
List of various rating models compared.

Classification	Methods
Traditional financial measures	Eff(Sharpe)
	Eff(Sortino)
	Eff(Omega)
Convex frontier rating methods	Single-time and MV framework
	Single-time and MVS framework
	Single-time and MVSK framework
	Multi-time and MV framework
	Multi-time and MVS framework
	Multi-time and MVSK framework
Nonconvex frontier rating methods	Single-time and MV framework
	Single-time and MVS framework
	Single-time and MVSK framework
	Multi-time and MV framework
	Multi-time and MVS framework
	Multi-time and MVSK framework

rate r_f and the loss threshold L are here specified as zero. Furthermore, in line with the properties of the shortage function used in the nonparametric frontier-based methods, we define the following traditional finance-based efficiency measures that bound the values between zero and unity and that make sure that the zero indicates full efficiency:

$$\text{Eff}(\text{Sharpe}_j^t) = \frac{\max\{\text{Sharpe}_j^t \mid j = 1,\ldots,n\} - \text{Sharpe}_j^t}{\max\{\text{Sharpe}_j^t \mid j = 1,\ldots,n\} - \min\{\text{Sharpe}_j^t \mid j = 1,\ldots,n\}}, \quad (8)$$

$$\text{Eff}(\text{Sortino}_j^t) = \frac{\max\{\text{Sortino}_j^t \mid j = 1,\ldots,n\} - \text{Sortino}_j^t}{\max\{\text{Sortino}_j^t \mid j = 1,\ldots,n\} - \min\{\text{Sortino}_j^t \mid j = 1,\ldots,n\}}, \quad (9)$$

$$\text{Eff}(\text{Omega}_j^t) = \frac{\max\{\text{Omega}_j^t \mid j = 1,\ldots,n\} - \text{Omega}_j^t}{\max\{\text{Omega}_j^t \mid j = 1,\ldots,n\} - \min\{\text{Omega}_j^t \mid j = 1,\ldots,n\}}. \quad (10)$$

On the other hand, we include convex and nonconvex nonparametric frontier-based ratings in different frameworks. All these 15 rating methods (3 traditional financial rating methods plus 12 frontier-based rating methods) are listed in Table 1.

To simplify names of the frontier-based methods, some notation indicates which frontier rating method is used for ranking MFs. This can be done in both single-time (ST) and multiple-time (MT) frameworks, using a convex (subscript c) or a nonconvex (subscript nc) frontier rating methods, and focusing on the first two (MV), three (MVS), or four moments (MVSK), respectively. For instance, MTMVSKc refers to the convex frontier model with the mean, variance, skewness and kurtosis over multiple times. Note that all the empirical results concerning these 15 rating methods are reported using these simplified notations.

We consider a simple buy-and-hold backtesting strategy consisting of buying in each time the 10, 20 and 30 best performing MFs ranked by rating method, respectively. Our work now is to empirically test the out-of-sample performance of these 15 buy-and-hold strategies. Since the Sharpe ratio and other relative performance measures are only suitable for the MV world, we opt for the shortage function as an absolute performance measure that is capable to assess the performance of these strategies in multiple dimensions simultaneously (i.e., mean, variance, skewness and kurtosis). Hence, the 15 buy-and-hold backtesting strategies are compared based on the MVSK performance of their holding values evaluated by combining shortage functions with the single-time and multi-moment frontiers (with convexity and nonconvexity).

Based on the fundamental logic of backtesting summarized so far, we design a backtesting analysis in detail for the buy-and-hold strategies constructed by the 15 rating methods. Our backtesting analysis is performed multiple times by rolling the time window. We first collect a sample of HFs with monthly return data starting from October 2006 till October 2020. The detailed description of this sample funds is presented in the following section (Section 4). Then, we split the period from the beginning of the sample period to the end of October 2015 in time windows of a given length, where the 5 years before the end of the sample period are kept apart to test the long-term holding performance of these strategies in the last backtesting period. Since the longest time period considered in our work is 5 years, it is appropriate to set the length of the rolling time window at 5 years. Therefore, the backtesting analysis is developed starting from November 2011, and is repeated 48 times (each time another month) with the rolling time window of 5 years till October 2015.

Using the first 5 year time window of data (from November 2006 to October 2011) to obtain the rankings for different rating methods, we determine the first buy-and-hold backtesting strategies in November 2011. These strategies are held for four holding scenarios: the end of October 2012 (for 1 year); the end of October 2014 (for 3 years); the end of October 2016 (for 5 years); and until the end of October 2020 (the end of the whole sample period). The process of the first backtesting is represented in Figure 1.[16]

Then, the time window is shifted with a step of a single month to develop the next backtesting analysis. For each time window or each backtesting event, the steps can be detailed as follows:

(1) Adopt the 5-year time window of data to compute the single-time frontier rankings, as well as the traditional financial rankings. In combination with the other two time periods (i.e., 1-year and 3-year) of data from this time window, the multi-time frontier ratings are computed.
(2) Depending on the ranking computed by this time window of data for each rating method, the 10, 20 or 30 best performing HFs are selected for the backtesting exercise, and then one holds these selected HFs for 1 year, for 3 years, for 5 years, and till the end of the whole sample period, respectively.[17]
(3) In each of the above four holding period scenarios, we compute and store the complete historical track record of the holding value per buy-and-hold backtesting strategy, and then we calculate the mean, variance, skewness and kurtosis of these holding value series.

The above steps for backtesting are repeated over 48 time windows in total. For each of the four holding period scenarios, the performance of these MVSK observations (15 times 48 observations) that are generated by the 15 strategies over 48 backtesting exercises are all evaluated by the shortage functions in the single-time and multi-moment frameworks (with convexity and nonconvexity). In particular, we first establish the VRSc and VRSnc nonparametric frontiers in the single-time and multi-moment framework for these MVSK observations, and then measure their efficiency scores using the shortage functions. Clearly, each buy-and-hold strategy yields the efficiency scores of 48 MVSK observations. The average efficiency score and the number of efficient units, as well as the distribution of inefficiency scores across these 48 observations, are adopted to evaluate the 15 strategies. For the four holding scenarios, the same pattern is used to compare the 15 strategies based on the different rating methods.

[16] In contrast to a rebalancing portfolio strategy, the buy-and-hold strategy is a long-term passive investment strategy whereby investors maintain a relatively stable portfolio over time. For this reason, our backtesting analysis sets up several fairly long holding windows to test the performance of these buy-and-hold backtesting strategies.

[17] Appendix discusses the empirical results pertaining to the performance of 15 buy-and-hold backtesting strategies held for 1 year, for 3 years, for 5 years, respectively.

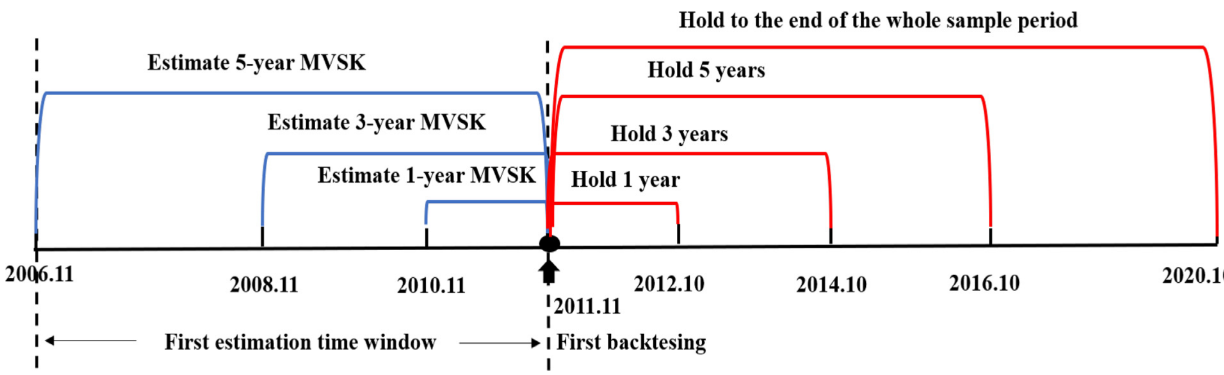

Fig. 1. Process of the first backtesting window.

Table 2
Descriptive statistics for all 187 HFs over the whole sample period.

	Mean	Variance	Skewness	Kurtosis	Scaled skewness	Scaled kurtosis
Min.	-0.328	0.633	-621.506	3.866	-5.989	2.469
Q1	0.306	8.764	-43.341	481.584	-0.785	3.874
Median	0.447	14.971	-10.294	1293.516	-0.320	5.385
Mean	0.480	26.810	210.182	34145.995	-0.510	7.283
Q3	0.601	27.018	1.468	4267.635	0.077	7.417
Max.	1.733	521.156	22732.909	2655540.333	1.913	59.538

4. Empirical Backtesting Results

As previously mentioned, the purpose of the empirical analysis is twofold. First, we examine whether the consideration of multiple moments and multiple times has an impact on both the efficiencies and the rankings of HFs. Second, we aim to further illustrate the eventual superiority of the proposed multi-time and multi-moment frontier rating methods by the backtesting analysis.

4.1. Sample Description

Considering the use of backtesting in the newly proposed multi-time and multi-moment ratings, the sample data collected requires the availability of continuous data for at least 14 years. Hence, we choose 187 HFs with monthly returns from October 2006 to October 2020 to test the 15 rating methods.[18] The data is all downloaded from Lipper for Investment Management made available by Hedge Funds database. According to the Lipper classification regarding HFs strategies, our sample database includes twelve different strategies.[19] Indeed, HFs with different strategies tend to have different trade-offs between risk and return when building their optimal portfolios (see Racicot and Théoret [65,66]). In our buy-and-hold backtesting analysis, we do not account for this variation since most individual investors are normally free to select a number of HFs across strategies from the universe of HFs to invest in. It needs to be stated that we initially specify these non-parametric frontier rating methods following the idea of Kerstens et al. [48] that higher order moments and cost components are included. But, since HF cost data is unavailable in this database, our empirical analysis is limited to focus on the characteristics of the return distributions for these HFs without considering cost factors.[20] In the following, we make a basic analysis of the monthly return characteristics of the 187 HF sample over the whole sample period. Table 2 reports descriptive statistics on the first four moments in columns 2 to 5 as well as on the scaled versions of skewness and kurtosis in columns 6 and 7 for the sample.[21]

Several studies mention that both scaled skewness and scaled kurtosis are not independent of one another for asymmetric distributions. Wilkins [78] proves a lower statistical bound for scaled kurtosis which links it to the squared value of scaled skewness (i.e., $kurtosis \geq 1 + skewness^2$). More recent theoretical and empirical research discussing the relation between these two measures is found in Schopflocher and Sullivan [68] and Racicot and Théoret [65,66]. Clearly, the descriptive statistics for the scaled versions of skewness and kurtosis in Table 2 differ from those of the unscaled versions of skewness and kurtosis (i.e., the third and fourth central moments of the asset return distribution).

While several streams in the financial literature do use scaled skewness and kurtosis, we adapt the third and fourth central moments as the input-like and output-like variables in the proposed rating methods. The main concerns are twofold. First, using the third and fourth central moments as output and input allows performance gauging of HFs consistent with general mixed risk-aversion investor preferences, i.e., a preference for odd moments

[18] As introduced in Section 3, each backtesting exercise needs the return data of the 5 previous years to calculate the statistics of HFs for computing the finance-based and frontier-based ratings, and at least the return data of 5 years ahead to evaluate the out-of-sample performance of these fund ratings. This process is repeated for 48 successive months. Therefore, this requires that the selected sample of HFs contains at least 14 ($= 5 + 5 + 4$) years of monthly return data. Thus, our sample runs from October 2006 to October 2020 for a total of 187 HFs in Lipper that are available over the sample period.

[19] In detail: Long/Short Equity (71 obs.), Managed Futures/CTAs (41 obs.), Multi Strategies (20 obs.), Event Driven (18 obs.), Emerging Markets (16 obs.), Global Macro (7 obs.), Credit Focus (5 obs.), Long Bias (4 obs.), Convertible Arbitrage (1 obs.), Equity Market Neutral (1 obs.), Fixed Income Arbitrage (1 obs.), and Other Hedge (2 obs.).

[20] Kerstens et al. [48] argue that MF can be trivially interpreted as a cost-based (loads) financial product that is identified by the characteristics of the return distribution, as summarized by some common subspace of moments. For example, when two MFs are identical in terms of the return distribution as summarized by the four moments, then the rational investor chooses for the MF with the lowest cost components. In the absence of data on the cost characteristics of MF, there is an implicit assumption that these costs are identical across the sample: therefore, their effect on the evaluation of the rating methods in our work can be ignored.

[21] In finance, the scaled versions of skewness and kurtosis are often used to characterize the asymmetry of return distributions. Consistent with the notations in Section 2, the scaled skewness and scaled kurtosis can be computed as: $E[(R_j^t - E(R_j^t))^3]/(E[(R_j^t - E(R_j^t))^2])^{\frac{3}{2}}$, and $E[(R_j^t - E(R_j^t))^4]/(E[(R_j^t - E(R_j^t))^2])^2$, respectively.

and an aversion for even moments. Second, the association between the third and fourth central moments of the return distribution with the Taylor approximation of the expected utility function for a mixed risk-averse investor have been argued in the literature (see, e.g., Briec et al. [18], and Krüger [51]). The use of ratios in the scaled skewness and kurtosis imposes a proportionality between moments that is not present in the above investor preferences. For instance, when using the scaled kurtosis, while we know that investors want a reduction in kurtosis and variance separately, we do not know their preference for the ratio of both. Therefore, all discussions and computations hereafter make use of central moments.

From the descriptive statistics of the monthly returns reported in Table 2, we see that some HFs are characterised by negative return and/or negative skewness. Such HFs are handled by taking absolute values of the output-like variables. Furthermore, we find that the series consisting of 187 HFs'skewness present positive mean and negative median, while the dispersion is quite large. Furthermore, all 187 HFs display positive kurtosis and also have a high dispersion. It is evident that some HFs do not perform well in terms of skewness and kurtosis. Therefore, for investors seeking non-negative skewness with small positive kurtosis, the multi-moment rating methods can be of great importance to select well-performing HFs from a large and heterogeneous HF universe. To assess the stability and persistence of these return characteristics over time, we further report the first four moments of the sample over three time periods: a 1-year, a 3-year and a 5-year time periods, respectively, is presented in Table in Appendix. Fundamentally, the same results regarding the return characteristics are available for these three time periods.

4.2. Evaluation Results

For the first aim of the empirical analysis, we compare both the efficiency distributions and the rankings of the 187 HFs calculated by the 15 rating methods. In the single-time rating framework, we extract the monthly returns of these samples for the past 5 years to date to calculate the efficiency and ranking. While in the multi-time rating framework, the monthly returns for the past 1 year, 3 years and 5 years to date are integrated and applied to evaluate the performance of these funds.

First, the efficiency distributions computed for the 15 rating methods are compared by means of nonparametric tests comparing two entire distributions initially developed by Li [54] and refined by Fan and Ullah [31] and most recently by Li et al. [55]. It tests for the eventual statistical significance of differences between two kernel-based estimates of density functions f and g of a random variable x. The null hypothesis maintains the equality of both density functions almost everywhere: $H_0 : f(x) = g(x)$ for all x; while the alternative hypothesis negates this equality of both density functions: $H_1 : f(x) \neq g(x)$ for some x.[22] Table 3 provides Li-test statistics using 2000 bootstrap replications for all rating methods considered in this contribution: in total, we report 105 relevant rating methods comparisons.

Several observations can be made regarding the results in Table 3. First, it is clear that the efficiency distributions computed

Table 3
Li-test statistics comparing the efficiency distributions computed by different rating methods.

	Eff(Sortino)	Eff(Omega)	STMVc	STMVSc	STMVSKc	MTMVc	MTMVSc	MTMVSKc	STMVnc	STMVSnc	STMVSKnc	MTMVnc	MTMVSnc	MTMVSKnc
Eff(Sharpe)	13.105***	52.572***	32.974***	26.262***	28.112***	28.856***	19.95***	10.787***	37.267***	29.058***	33.588***	22.357***	12.428***	14.98***
Eff(Sortino)		36.715***	34.775***	27.804***	26.431***	15.464***	9.735***	4.772***	34.704***	28.537***	32.088***	11.644***	8.669***	11.052***
Eff(Omega)			30.826***	23.574***	25.33***	31.387***	28.876***	24.233***	31.146***	32.792***	36.324***	20.163***	25.316***	27.69***
STMVc				-5.818	-2.186	43.486***	40.519***	39.723***	-0.011	4.092***	6.917***	43.546***	37.76***	38.265***
STMVSc					-0.969	41.352***	36.06***	32.862***	0.162	2.763***	5.764***	38.787***	31.598***	32.38***
STMVSKc						6.845***	4.645***	28.251***	0.434	0.172	1.62*	36.439***	23.893***	23.503***
MTMVc							0.629	4.607***	47.637***	49.282***	52.157***	1.699**	12.015***	13.962***
MTMVSc								1.311*	39.891***	38.643***	41.47***	0.68	8.052***	10.026***
MTMVSKc									36.62***	28.562***	30.288***	-1.789	1.09	2.774***
STMVnc										0.206	2.272**	44.249***	33.723***	33.033***
STMVSnc											-1.476	39.818***	21.488***	19.946***
STMVSKnc												41.843***	21.709***	19.273***
MTMVnc													4.327***	6.209***
MTMVSnc														-2.158

Li test: critical values at 1% level= 2.33(***); 5% level= 1.64(**); 10%level= 1.28(*).

[22] Matlab code developed by P.J. Kerstens based on Li et al. [55] is found at: https://github.com/kepiej/DEAUtils. In fact, we use the so-called Simar-Zelenyuk adaptation of this test statistic for nonparametric frontier estimators to circumvent the problem of spurious mass at the boundary by considering two algorithms: Algorithm I ignores the boundary estimates and Algorithm II smooths these estimates by adding a uniform noise of order of magnitude less than the order of magnitude of the noise added by the nonparametric frontier estimator. The Monte Carlo evidence indicates that Algorithm II is more robust when the dimensions of the specification are increased. Therefore, we employ the Li-test version of Li et al. [55] amended with Algorithm II.

by traditional financial performance measures and those computed by frontier-based rating methods are significantly different at the 1 % significance level.

Second, in both c and nc frontier ratings, the single-time and multi-time rating methods yield significantly different efficiency distributions. This implies that the consideration of multiple times has a significant impact on the efficiency distributions.

Third, the effect of adding multiple moments on the efficiency distributions are somewhat different in single-time and multi-time ratings. For instance, in the case of convexity, adding skewness and kurtosis jointly has a significant effect on the efficiency distributions at the 1 % significance level in multi-time ratings. In single-time ratings, adding higher moments does not contribute in a significant way. Furthermore, the nonconvex frontier rating methods are more discriminatory in the impact of adding multiple moments. Compared to the above results in the case of convexity, in the case of nonconvexity, both adding skewness in itself and adding skewness and kurtosis jointly have significant effects on the efficiency distributions at 1 % significance level in multi-time ratings, and adding these jointly has a significant impact at 5 % significance level in single-time ratings.

Fourth, for multi-time ratings, imposing convexity always has a significant impact on the efficiency distributions. The efficiency distributions obtained by convex and nonconvex frontier ratings in MV, MVS and MVSK cases all yield differences at 1 % significance level, respectively. For the single-time ratings, the efficiency distributions of the convex and the nonconvex models are different at the 1 % and 10 % significance level in MVS and MVSK cases, respectively.

We further determine the Kendall rank correlations to test the degree of concordance in rankings determined by these performance measures. Table 4 shows the rank correlation between different HF ratings. In this table, *** indicates that the correlation coefficient between the rankings is significantly different from zero at 1 % significance level. The following key findings are revealed from Table 4. First, it is clear that the traditional financial ratings present a consistently low correlation (around 0.39-0.43) with the multi-time and multi-moment (MVS & MVSK) frontier ratings, but a high correlation (more than 0.8) with the single-time MV ratings. Second, turning to the comparisons between frontier ratings in single-time and multi-time frameworks, the single-time frontier rating and multi-time frontier rating show a low correlation overall. Third, the MV frontier rating exhibits a lower correlation with the multi-moment (MVS & MVSK) frontier ratings in multi-time framework compared in single-time framework. Moreover, the MV frontier rating has a lower correlation with the MVSK frontier rating compared with the MVS frontier rating. Finally, regarding comparisons between the rating models with convexity and nonconvexity, both the second and third findings tend to be more pronounced in the nonconvex case compared to the convex case.

From these analyses, we can conclude that the multiple moments and multiple times both separately and jointly have an impact on the HF efficiency and ranking for our data, and this impact is more significant when the two factors are considered jointly. Furthermore, nonconvexity may prove to be a more modest hypothesis in the proposed multi-time and multi-moment ratings since it exhibits a stronger discriminatory power with respect to the effect of adding multiple moments. This confirms earlier comparative results between the convex and nonconvex models with higher order moments in the contribution of Kerstens et al. [48].

Table 4
Kendall rank correlations comparing the rankings computed by different rating methods.

	Eff(Sortino)	Eff(Omega)	STMVc	STMVSc	STMVSKc	MTMVc	MTMVSc	MTMVSKc	STMVnc	STMVSnc	STMVSKnc	MTMVnc	MTMVSnc	MTMVSKnc
Eff(Sharpe)	0.956***	0.961***	0.833***	0.778***	0.627***	0.718***	0.639***	0.398***	0.824***	0.605***	0.592***	0.728***	0.406***	0.401***
Eff(Sortino)		0.933***	0.813***	0.776***	0.641***	0.737***	0.663***	0.427***	0.812***	0.620***	0.608***	0.745***	0.437***	0.431***
Eff(Omega)			0.823***	0.770***	0.620***	0.711***	0.632***	0.392***	0.822***	0.604***	0.589***	0.722***	0.400***	0.394***
STMVc				0.908***	0.749***	0.782***	0.668***	0.423***	0.948***	0.703***	0.690***	0.775***	0.435***	0.427***
STMVSc					0.836***	0.793***	0.741***	0.506***	0.875***	0.785***	0.772***	0.789***	0.520***	0.513***
STMVSKc						0.700***	0.706***	0.624***	0.725***	0.886***	0.887***	0.692***	0.630***	0.627***
MTMVc							0.857***	0.579***	0.770***	0.676***	0.665***	0.938***	0.580***	0.573***
MTMVSc								0.713***	0.661***	0.690***	0.680***	0.832***	0.713***	0.705***
MTMVSKc									0.418***	0.602***	0.606***	0.562***	0.905***	0.901***
STMVnc										0.732***	0.712***	0.775***	0.438***	0.429***
STMVSnc											0.967***	0.677***	0.651***	0.643***
STMVSKnc												0.668***	0.651***	0.657***
MTMVnc													0.594***	0.589***
MTMVSnc														0.977***

Table 5
Performance results for 15 buy-and-hold backtesting strategies: Descriptive statistics of the values of shortage function.

Methods	HF(10)				HF(20)				HF(30)			
	VRSc		VRSnc		VRSc		VRSnc		VRSc		VRSnc	
	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.
Eff(Sharpe)	0.064	0	0.040	9	0.081	2	0.047	10	0.078	0	0.034	9
Eff(Sortino)	0.063	1	0.034	10	0.084	2	0.055	7	0.077	1	0.037	9
Eff(Omega)	0.064	0	0.031	10	0.084	1	0.059	4	0.077	0	0.040	7
STMVc	0.077	0	0.045	17	0.101	1	0.064	5	0.096	0	0.047	11
STMVSc	0.059	7	0.027	28	0.090	2	0.055	14	0.076	4	0.033	16
STMVSKc	0.044	6	0.014	31	0.070	4	0.039	17	0.059	1	0.031	15
MTMVc	0.061	1	0.020	22	0.075	1	0.038	14	0.078	2	0.032	11
MTMVSc	0.063	4	0.025	22	0.078	2	0.044	14	0.065	2	0.028	16
MTMVSKc	0.041	9	0.008	30	0.065	1	0.033	17	0.053	1	0.020	17
STMVnc	0.068	2	0.031	20	0.100	0	0.062	8	0.090	0	0.038	11
STMVSnc	0.042	5	0.023	16	0.054	4	0.029	19	0.039	5	0.014	25
STMVSKnc	0.042	4	0.026	13	0.040	6	0.022	27	0.035	7	0.012	26
MTMVnc	0.047	3	0.013	26	0.075	0	0.035	18	0.074	0	0.030	15
MTMVSnc	0.034	9	0.010	27	0.049	9	0.024	19	0.039	6	0.013	28
MTMVSKnc	0.039	5	0.012	31	0.047	7	0.021	21	0.032	7	0.009	28

4.3. Backtesting Results

We analyze the backtesting scenarios with a selection of the 10, 20 or 30 best performing HFs, respectively.[23] As stated previously, the 15 buy-and-hold strategies are compared in terms of the MVSK performances of their holding value series that are evaluated by the shortage functions based on the VRSc and VRSnc frontiers in single-time and multi-moment frameworks. Table 5 presents an overall analysis with respect to the performances of the MVSK observations generated per strategy held until the end of the whole sample period. This table is structured as follows: the first series of four columns list the results with regard to the 10 best HFs selected for the backtesting exercise, and the second and third series of four columns present the results for selecting 20 and 30 best HFs, respectively. Within each selecting (buying) scenario, the first two columns report the average inefficiency scores and the number of efficient units for each strategy when evaluated using the VRSc frontier in single-time and multi-moment framework, while the last two columns report these results in the VRSnc case.

We first analyze the main findings in the context of buying and holding until the end of the whole sample period, as presented in Table 5. From these results, there are four main conclusions.

The first key finding is that all the frontier-based strategies outperform the strategies based on traditional financial indicators, except the strategies constructed by the single-time MV frontier rating methods. From the average inefficient scores reported in Table 5, it is easy to see that the average inefficiency scores of all strategies based on the multi-moment and/or the multi-time frontier ratings are lower than those of Sharpe-, Sortino- and Omega-driven strategies. This result is valid when buying the 10, 20 and 30 best HFs. Combining the numbers of efficient units given in Table 5, the frontier-based strategies clearly yield more efficient units compared to those based on traditional indicators.

The second key result is that the buy-and-hold strategies according to the multi-moment ratings present superior results compared to those based on the MV ratings. Again, this result is confirmed when buying the 10, 20 and 30 best HFs. Both in the single-time and multi-time rating frameworks, we find that the strategies driven by the multi-moment ratings yield lower average inefficiency scores and a higher number of efficient units over strategies driven by the MV ratings.

Third, combining the two evaluation indicators of average inefficiency scores and the number of efficient units, it is found that in the majority of cases the buy-and-hold strategies consisting of the HFs selected by the multi-time rating methods perform better than strategies consisting of the HFs selected by the single-time rating methods. This result remains valid when buying the 10, 20 and 30 best HFs.

A last key finding is that strategies determined by the nonconvex frontier-based ratings always outperform those determined by the convex frontier-based ratings. Moreover, by comparing the average inefficiency scores and the number of efficient units between the two in MVS and MVSK frameworks, it can be seen that when multiple moments are considered, the strategies based on the nonconvex frontier-based ratings usually display a more significant advantage. The reason for this finding is that skewness and kurtosis imply nonconvexities in diversified portfolio optimisation. As stated above, nonconvex production frontier models used for fund rating underestimate the nonconvex diversified portfolio models, while the convex production frontier models may tend to overestimate these same nonconvex diversified portfolio models.

Thus, this backtesting analysis shows that the buy-and-hold strategies constructed by our proposed multi-moment and multi-time rating methods exhibit superior performance in most scenarios. We therefore believe that the joint consideration of multi-moments and multi-times provides additional useful information for HF selection in practice.

As a sensitivity analysis, we test the performance of the 15 buy-and-hold backtesting strategies held for 1 year, 3 years and 5 years, which can be regarded as their short-, medium- and long-term holding performance. Table in Appendix summarizes the performance results of the 15 strategies held for these three alternative holding periods. The above four findings are also evidenced in most cases for these three holding period scenarios. Moreover, the buy-and-hold backtesting strategies consisting of the best HFs rated by the multi-moment and multi-time performance measure tend to show a consistent performance over the different holding periods. We basically conclude that the buy-and-hold strategies driven by the multi-moment and multi-time ratings exhibit favorable and consistent short-, medium- and long-term holding performance, somewhat implying that the performance of the best-performing HFs rated by the proposed multi-moment and multi-time performance measure would be sustained over time. A more detailed discussion on the sensitivity analysis is provide in Appendix.

Furthermore, we also report the performance of the 15 buy-and-hold backtesting strategies held during the COVID-19 period

[23] All HFs with an efficiency of 0 are ranked as 1 in our calculations. As a consequence, when we have to take a certain amount of funds among these ties then we take these randomly among the tied units (as in Brandouy et al. [13]).

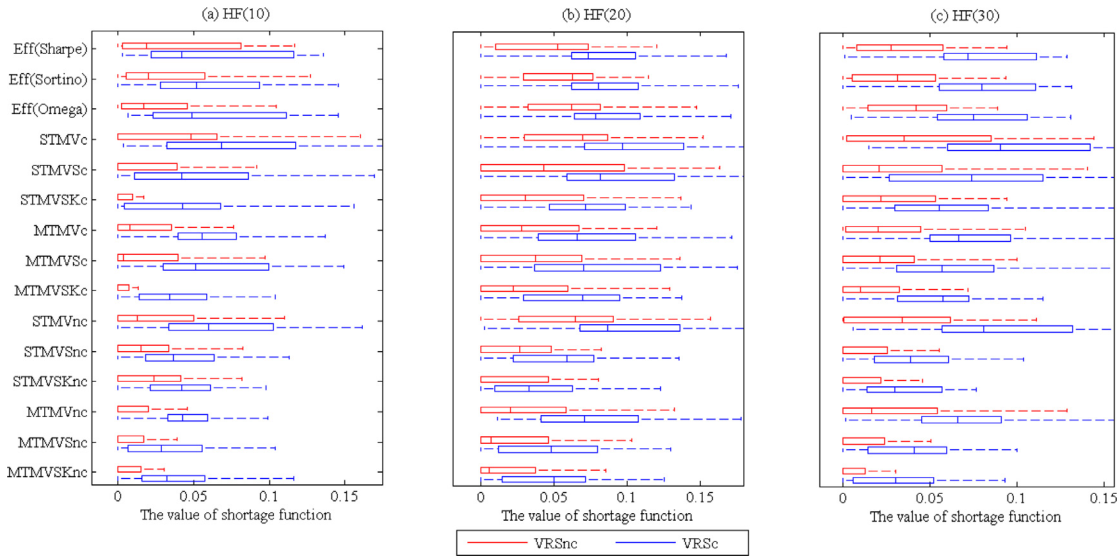

Fig. 2. Distributions of inefficiency scores for 15 buy-and-hold backtesting strategies.eps.

which for our purpose ranges from February 2020 to February 2022 to explore whether the proposed multi-time and multi-moment frontier ratings perform well in such a holding phase with high market volatility. To add this very harsh holding period, we complement our initial sample with data from November 2020 till February 2022.[24] This question pertains to the more general issue that the behavior of HF strategies are linked to the phases of the economic cycle, i.e., economic expansion or crisis (see, e.g., Racicot and Théoret [65,66]). The performance results of the 15 strategies held during the COVID-19 crisis are presented in Table of Appendix. Overall, one finds that the buy-and-hold strategies determined by the multi-time and multi-moment frontier ratings maintain a mild advantage over those determined by the other existing ratings, though this advantage is not so pronounced as in other holding scenarios with lower market volatility. To save space, the details are presented and discussed in Appendix.

Besides evaluating strategies based on the two summarized indicators reported in Table 5, we further provide the entire distribution of the inefficiency scores per strategy to compare these intuitively. Figure 2 presents a graphical overview of the performance of all strategies by integrating the box-plot per strategy held to end in the buying scenarios with 10, 20 and 30 HFs selected. In this figure, the sub-figures (a) to (c) correspond to the performance results of these three buying scenarios. The box-plots for the performance of strategies based on the VRSc frontier are in blue, and those based on the VRSnc frontier are in red. In these box-plots, the box indicates the interquartile range where the small vertical lines reporting the location of the median. Their locations closer to the left suggests that the entire distribution of inefficiency scores for the strategy is at a lower level, which implies that the strategy has a better performance in backtesting analysis. As we can observe from Figure 2, comparing the performance of these strategies in each buying (backtesting) scenario, the buy-and-hold strategies constructed by the multi-moment and multi-time frontier rating methods are superior to strategies constructed by the existing rating methods in most cases.

Equally so, the entire distributions of the inefficiency scores for the 15 strategies held for 1, 3 and 5 years are presented in Figures, and in Appendix, respectively. From Figures, and, one can observe that the dominance of the strategies driven by the multi-moment and multi-time ratings over other strategies remains valid and that this relation is strengthened as the holding period increases. It is therefore clear that the good performance of the strategies driven by the proposed frontier-based performance measures including multiple moments and multiple times exhibits good stability (see Appendix for details). In addition, we also report the entire distribution of the inefficiency scores per strategy held during the COVID-19 period in Figure provided in Appendix. From Figure, it can be observed that the multi-time and multi-moment ratings somewhat do better than the traditional financial ratings and several existing frontier ratings in general during the holding period when market volatility is relatively high.

Finally, we provide some minimal sensitivity analysis with respect to one popular alternative risk measure: CVaR. VaR is defined as the maximum loss that investors may suffer over a given time horizon at a specified confidence level. CVaR (one popular variation on VaR) corresponds to the expected loss conditional on the loss exceeding VaR. CVaR risk measures are coherent (see Branda [11]): this leads to a convex mean-CVaR portfolio optimization problem. Thus, the nonparametric production frontier specifications have the potential to provide a conservative approximation for this diversified meanCVaR frontier. Therefore, we first opt to add another 4 buy-and-hold strategies consisting of the HFs selected by on the one hand the convex and nonconvex and on the other hand the single time and multiple time nonparametric mean-CVaR frontier ratings. In addition, Branda [11, p. 75, equation (18)] proposes an equivalent linearized version of the diversified mean-CVaR model, which refines the computation of the diversified mean-CVaR model (see also Branda [10] and see Mansini et al. [59] for a survey of linear programming approaches to diversified portfolio models). Hence, we also opt to add another 2 buy-and-hold strategies determined by the diversified mean-CVaR models in both single-time and multi-time frameworks.

The four nondiversified mean-CVaR ratings are obtained by solving model (2) and model (4) with CVaR as the input-like variable and return as the output-like variable (denoted as STMCVaRc, STMCVaRnc, MTMCVaRc and MTMCVaRnc, respectively). The two diversified mean-CVaR ratings are based on solving the equivalent linearized model of the diversified mean-CVaR portfolio proposed by Branda [11], and the multi-time one is an extension of this diversified mean-CVaR model in our multi-time framework (denoted

[24] Note that if certain HFs are liquidated during this extended period, then the corresponding return data are all set to zero.

as STMCVaRdiv and MTMCVaRdiv, respectively).[25] The performance results of these 21 buy-and-hold backtesting strategies held until the end of the whole sample period are presented in Table and Figure of Appendix . Overall, it can be concluded that the proposed multi-time and multi-moment ratings in this contribution have a clear dominance over both nondiversified and diversified mean-CVaR ratings.[26]

5. Conclusion

Inspired by recent nonparametric frontier rating methods contributing to assessing MF performance (e.g., Kerstens et al. [48]), this contribution has aimed to define a new shortage function or performance measure for rating MFs that can simultaneously handle both multiple moments and multiple times. Furthermore, we have explored the potential benefits of this new performance measure for selecting the best performing MF. We are now in a position to summarize the main contributions.

First, we establish a series of nonparametric convex and nonconvex frontier rating methods with multi-moments and multi-times. The proposed rating methods are capable of not only assessing to which extent a MF performs well in the several moments following mixed risk-aversion preferences, but it simultaneously measures to which extent a MF performs well in all these moments in different times as well. These new multi-time and multi-moment performance measures are suitable for handling mixed risk-aversion preferences of investors which aim at time persistence.

Second, the proposed rating methods are empirically applied to HFs, given that HFs tend to exhibit strong asymmetric and long-tail return characteristics compared to other MFs. Using Li-test and Kendall rank correlation, the multi-time and multi-moment ratings are compared with traditional financial indicators and basic single-time MV rating methods to examine the impact of multiple moments and multiple times. From the comparison among 15 various rating methods, we find that in both convex and nonconvex cases, the multiple moments and multiple times both separately and jointly have an impact on the HF efficiency and ranking, and this impact is more significant when the two factors are considered jointly. Furthermore, the nonconvex rating models have stronger discriminatory power with respect to the effect of adding multiple moments over the convex rating models. This confirms earlier comparative results between convex and nonconvex models with higher order moments in Kerstens et al. [48].

Third, having the impact of the multi-moments and multi-times in mind, we develop a simple buy-and-hold backtesting strategy to test whether the new ratings perform any better than more traditional financial ratings and single-time MV ratings in HF selection. In most backtesting exercises, the buy-and-hold strategies based on the multi-time and multi-moment ratings exhibit a superiority over those based on traditional financial ratings and single-time MV ratings. This superiority is clearly confirmed by comparing the MVSK performance of holding values with respect to various buy-and-hold backtesting strategies. The multi-time and multi-moment strategies tend to exhibit more stable and favorable short-, medium- and long-term holding performance than the other strategies. Equally so, we focus on the comparison of these multi-time and multi-moment strategies in the convex and nonconvex cases. The strategies based on the nonconvex frontier ratings usually display a more significant advantage over the convex frontier ratings probably for reasons of a closer fit with the nonconvex skewness and kurtosis in diversified portfolio optimisation.

Overall, the proposed multi-time and multi-moment performance measures provide a novel idea into the important topic of rating and selecting MF. From the basic backtesting setup in our empirical analysis, further extensive backtesting studies can be developed to exploit the potential of the new performance measures in constructing a fund of funds. Another desirable avenue for future research is to transfer the current methodological framework and to perform a backtesting analysis using diversified portfolio models. It is worthwhile to compare the performance in MF selection between the backtesting strategies driven by diversified and convex and nonconvex nondiversified frontier rating methods. But, this in principle calls for overcoming the computational difficulties of extending the diversified models to the multi-moment and multi-time framework. Furthermore, one can account for other popular risk measures that assess tail risk as an alternative to variance (i.e., the second moment of the return distribution): e.g., VaR or CVaR. While mean-CVaR models are common, their extension to include in addition higher moments is a bit unusual. The fundamental methodologies developed in this work can be extended to the mean-CVaR (or VaR)-skewness-kurtosis framework. Also the eventual use of alternative risk measures in conjunction with higher moments is left for future work. Due to certain regulatory and strategic limitations, fund managers may only be able to construct portfolios among several specific categories (e.g., strategies) of HFs. It could be intriguing to develop HFs rating across categories by incorporating the current efforts with the so-called nonparametric metatechnology (e.g., Kerstens et al. [49]). Clearly, it is also useful to apply the current methodology and the further extensions for tackling the evaluation and comparison across categories for other traditional MFs (e.g., equity MFs, bond MFs, and mixed asset MFs, etc.).

CRediT authorship contribution statement

Kristiaan Kerstens: Conceptualization, Methodology, Writing – original draft, Supervision. **Paolo Mazza:** Data curation, Writing – review & editing. **Tiantian Ren:** Methodology, Software, Writing – original draft. **Ignace Van de Woestyne:** Visualization, Writing – review & editing.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.omega.2022.102718.

References

[1] Adam L, Branda M. Risk-aversion in Data Envelopment Analysis models with diversification. Omega 2021;102:102338.
[2] Back K. A characterization of the coskewness–cokurtosis pricing model. Economics Letters 2014;125(2):219–22.
[3] Bacmann J-F, Benedetti S. Optimal Bayesian portfolios of hedge funds. International Journal of Risk Assessment and Management 2009;11(1-2):39–58.
[4] Bacon C. Practical Portfolio Performance Measurement and Attribution. 2nd. New York: Wiley; 2008.
[5] Basso A, Funari S. DEA performance assessment of mutual funds. In: Zhu J, editor. Data Envelopment Analysis: A Handbook of Empirical Studies and Applications. Berlin: Springer; 2016. p. 229–87.
[6] Benati S. Using medians in portfolio optimization. Journal of the Operational Research Society 2015;66(5):720–31.
[7] Blake D. Financial intermediation and financial innovation in a characteristics framework. Scottish Journal of Political Economy 1996;43(1):16–31.
[8] Bodson L, Coen A, Hubner G. How stable are the major performance measures? Journal of Performance Measurement 2008;13(1):21–30.
[9] Boudt K, Cornilly D, Verdonck T. A coskewness shrinkage approach for estimating the skewness of linear combinations of random variables. Journal of Financial Econometrics 2020;18(1):1–23.

[25] We simply set the confidence level as 95% as is commonly used to calculate the CVaR measure and the mean-CVaR diversified model, while Branda [11] sets a whole range of confidence levels in his computations.

[26] In Appendix the 21 buy-and-hold strategies driven by both VRSc and VRSnc nonparametric as well as diversified mean-CVaR frontier ratings in single-time and multi-time frameworks are included in our backtesting exercise.

[10] Branda M. Diversification-consistent data envelopment analysis with general deviation measures. European Journal of Operational Research 2013;226(3):626–35.
[11] Branda M. Diversification-consistent Data Envelopment Analysis based on directional-distance measures. Omega 2015;52:65–76.
[12] Branda M, Kopa M. On relations between DEA-risk models and stochastic dominance efficiency tests. Central European Journal of Operations Research 2014;22(1):13–35.
[13] Brandouy O, Kerstens K, Van de Woestyne I. Frontier-based vs. traditional mutual fund ratings: A first backtesting analysis. European Journal of Operational Research 2015;242(1):332–42.
[14] Briec W, Comes C, Kerstens K. Temporal technical and profit efficiency measurement: Definitions, duality and aggregation results. International Journal of Production Economics 2006;103(1):48–63.
[15] Briec W, Dumas A, Mekki A. Directional distance functions and social welfare: Some axiomatic and dual properties. Mathematical Social Sciences 2021;113:181–90.
[16] Briec W, Kerstens K. Multi-horizon Markowitz portfolio performance appraisals: A general approach. Omega: The International Journal of Management Science 2009;37(1):50–62.
[17] Briec W, Kerstens K. Portfolio selection in multidimensional general and partial moment space. Journal of Economic Dynamics and Control 2010;34(4):636–56.
[18] Briec W, Kerstens K, Jokung O. Mean-variance-skewness portfolio performance gauging: A general shortage function and dual approach. Management Science 2007;53(1):135–49.
[19] Briec W, Kerstens K, Lesourd J. Single period Markowitz portfolio selection, performance gauging and duality: A variation on the Luenberger shortage function. Journal of Optimization Theory and Applications 2004;120(1):1–27.
[20] Briec W, Kerstens K, Van de Woestyne I. Portfolio selection with skewness: A comparison of methods and a generalized one fund result. European Journal of Operational Research 2013;230(2):412–21.
[21] Briec W, Kerstens K, Vanden Eeckaut P. Non-convex technologies and cost functions: Definitions, duality and nonparametric tests of convexity. Journal of Economics 2004;81(2):155–92.
[22] Caporin M, Jannin G, Lisi F, Maillet B. A survey on the four families of performance measures. Journal of Economic Surveys 2014;28(5):917–42.
[23] Charnes A, Cooper W, Rhodes E. Measuring the efficiency of decision making units. European Journal of Operational Research 1978;2(6):429–44.
[24] Cornes R. Duality and Modern Economics. Cambridge: Cambridge University Press; 1992.
[25] Darolles S, Gourieroux C. Conditionally fitted sharpe performance with an application to hedge fund rating. Journal of Banking & Finance 2010;34(3):578–93.
[26] DeMiguel V, Garlappi L, Uppal R. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? Review of Financial Studies 2009;22(5):1915–53.
[27] Eeckhoudt L, Schlesinger H. Putting risk in its proper place. American Economic Review 2006;96(1):280–9.
[28] El Kalak I, Azevedo A, Hudson R. Reviewing the hedge funds literature ii: Hedge funds' returns and risk management characteristics. International Review of Financial Analysis 2016;48:55–66.
[29] Eling M, Faust R. The performance of hedge funds and mutual funds in emerging markets. Journal of Banking & Finance 2010;34(8):1993–2009.
[30] Eling M, Schuhmacher F. Does the choice of performance measure influence the evaluation of hedge funds? Journal of Banking & Finance 2007;31(9):2632–47.
[31] Fan Y, Ullah A. On goodness-of-fit tests for weakly dependent processes using kernel method. Journal of Nonparametric Statistics 1999;11(1-3):337–60.
[32] Fang H, Lai T-Y. Co-kurtosis and capital asset pricing. Financial Review 1997;32(2):293–307.
[33] Färe R, Grosskopf S, Lovell C. Production Frontiers. Cambridge: Cambridge University Press; 1994.
[34] Farrell M. The measurement of productive efficiency. Journal of the Royal Statistical Society Series A: General 1957;120(3):253–81.
[35] Feibel B. Investment Performance Measurement. New York: Wiley; 2003.
[36] Friend I, Westerfield R. Co-skewness and capital asset pricing. Journal of Finance 1980;35(4):897–913.
[37] Germain L, Nalpas N, Vanhems A. Nonparametric hedge funds and replication indices performance analysis: A robust directional application. In: Gregoriou G, Kooli M, editors. Hedge Funds Replication. Basingstoke: Palgrave Macmillan; 2011. p. 90–105.
[38] Grau-Carles P, Doncel L, Sainz J. Stability in mutual fund performance rankings: A new proposal. International Review of Economics & Finance 2019;61:337–46.
[39] Gregoriou G. Performance appraisal of funds of hedge funds using Data Envelopment Analysis. Journal of Wealth Management 2003;5(4):88–95.
[40] Gregoriou G, Sedzro K, Zhu J. Hedge fund performance appraisal using Data Envelopment Analysis. European Journal of Operational Research 2005;164(2):555–71.
[41] Hackman S. Production Economics: Integrating the Microeconomic and Engineering Perspectives. Berlin: Springer; 2008.
[42] Harvey C, Liechty J, Liechty M, Müller P. Portfolio selection with higher moments. Quantitative Finance 2010;10(5):469–85.
[43] Heffernan S. A characteristics definition of financial markets. Journal of Banking & Finance 1990;14(2–3):583–609.
[44] Jondeau E, Rockinger M. Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. Journal of Economic Dynamics and Control 2003;27(10):1699–737.
[45] Joro T, Na P. Portfolio performance evaluation in mean-variance-skewness framework. European Journal of Operational Research 2006;175(1):446–61.
[46] Jurczenko E, Maillet B, Merlin P. Hedge funds portfolio selection with higher-order moments: A nonparametric mean-variance-skewness-kurtosis efficient frontier. In: Jurczenko E, Maillet B, editors. Multi-moment Asset Allocation and Pricing Models. New York: Wiley; 2006. p. 51–66.
[47] Jurczenko E, Yanou G. Fund of hedge funds portfolio selection: A robust non-parametric multi-moment approach. In: Watanabe Y, editor. The Recent Trend of Hedge Fund Strategies. New York: Nova Science; 2010. p. 21–56.
[48] Kerstens K, Mounir A, Van de Woestyne I. Non-parametric frontier estimates of mutual fund performance using C- and L-moments: Some specification tests. Journal of Banking & Finance 2011;35(5):1190–201.
[49] Kerstens K, O'Donnell C, Van de Woestyne I. Metatechnology frontier and convexity: A restatement. European Journal of Operational Research 2019;275(2):780–92.
[50] Khemchandani R, Chandra S. Efficient trading frontier: A shortage function approach. Optimization 2014;63(10):1533–48.
[51] Krüger J. Nonparametric portfolio efficiency measurement with higher moments. Empirical Economics 2021;61(3):1435–59.
[52] Kumar UD, Roy AB, Saranga H, Singal K. Analysis of hedge fund strategies using slack-based DEA models. Journal of the Operational Research Society 2010;61(12):1746–60.
[53] Lai TY. Portfolio selection with skewness: A multiple objective approach. Review of Quantitative Finance and Accounting 1991;1(3):293–305.
[54] Li Q. Nonparametric testing of closeness between two unknown distribution functions. Econometric Reviews 1996;15(3):261–74.
[55] Li Q, Maasoumi E, Racine J. A nonparametric test for equality of distributions with mixed categorical and continuous data. Journal of Econometrics 2009;148(2):186–200.
[56] Lin R, Chen Z, Hu Q, Li Z. Dynamic network DEA approach with diversification to multi-period performance evaluation of funds. OR Spectrum 2017;39(3):821–60.
[57] Lin R, Li Z. Directional distance based diversification super-efficiency DEA models for mutual funds. Omega 2020;97:102096. doi:10.1016/j.omega.2019.08.003.
[58] Liu W, Zhou Z, Liu D, Xiao H. Estimation of portfolio efficiency via DEA. Omega 2015;52(1):107–18.
[59] Mansini R, Ogryczak W, Speranza MG. Twenty years of linear programming based portfolio optimization. European Journal of Operational Research 2014;234(2):518–35.
[60] Markowitz H. Portfolio selection. Journal of Finance 1952;7(1):77–91..
[61] Massol O, Banal-Estañol A. Export diversification through resource-based industrialization: The case of natural gas. European Journal of Operational Research 2014;237(3):1067–82.
[62] Menardi G, Lisi F. Are performance measures equally stable? Annals of Finance 2012;8(4):553–70.
[63] Morey M, Morey R. Mutual fund performance appraisals: A multi-horizon perspective with endogenous benchmarking. Omega 1999;27(2):241–58.
[64] Murthi B, Choi Y, Desai P. Efficiency of mutual funds and portfolio performance measurement: A non-parametric approach. European Journal of Operational Research 1997;98(2):408–18.
[65] Racicot F-É, Théoret R. Multi-moment risk, hedging strategies, & the business cycle. International Review of Economics & Finance 2018;58:637–75.
[66] Racicot F-É, Théoret R. Hedge fund return higher moments over the business cycle. Economic Modelling 2019;78:73–97.
[67] Ren T, Zhou Z, Xiao H. Estimation of portfolio efficiency considering social responsibility: Evidence from the multi-horizon diversification DEA. RAIRO-Operations Research 2021;55(2):611–37.
[68] Schopflocher T, Sullivan P. The relationship between skewness and kurtosis of a diffusing scalar. Boundary-layer Meteorology 2005;115(3):341–58.
[69] Scott R, Horvath P. On the direction of preference for moments of higher order than the variance. Journal of Finance 1980;35(4):915–19.
[70] Sears RS. Asset pricing, higher moments, and the market risk premium: A note. Journal of Finance 1985;40(4):1251–3.
[71] Sengupta J. Nonparametric tests of efficiency of portfolio investment. Journal of Economics 1989;50(3):1–15.
[72] Slesnick D. Empirical approaches to the measurement of welfare. Journal of Economic Literature 1998;36(4):2108–65.
[73] Smith D. Evaluating hedge fund performance. In: Baker H, Filbeck G, editors. Hedge Funds: Structure, Strategies, and Performance. Oxford: Oxford University Press; 2017. p. 415–38.
[74] Tammer C, Zălinescu C. Lipschitz properties of the scalarization function and applications. Optimization 2010;59(2):305–19.
[75] Trautmann S, van de Kuilen G. Higher order risk attitudes: A review of experimental evidence. European Economic Review 2018;103:108–24.
[76] Tu J, Zhou G. Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. Journal of Financial Economics 2011;99:204–15.
[77] Von Neumann J, Morgenstern O. Theory of Games and Economic Behavior. Princeton: Princeton University Press; 1953.
[78] Wilkins JE. A note on skewness and kurtosis. Annals of Mathematical Statistics 1944;15(3):333–5.
[79] Zhou Z, Xiao H, Jin Q, Liu W. Dea frontier improvement and portfolio rebalancing: An application of china mutual funds on considering sustainability information disclosure. European Journal of Operational Research 2018;269(1):111–31.