IÉSEG SCHOOL OF MANAGEMENT

May 2021

WORKING PAPER SERIES

2021-EQM-03

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

Kristiaan Kerstens
IESEG School of Management, CNRS-LEM (UMR 9221), Université de Lille,
3 rue de la Digue, F-59000 Lille, France, k.kerstens@ieseg.fr

Paolo Mazza
IESEG School of Management, CNRS-LEM (UMR 9221),
3 rue de la Digue, F-59000 Lille, France, p.mazza@ieseg.fr

Tiantian Ren
Corresponding author: School of Business Administration, Hunan University,
Changsha 410081, China, and IESEG School of Management, 3 rue de la Digue, F-59000
Lille, France, Tel: +33 320545892 (switch-board), Fax: +33 320574855, t.ren@ieseg.fr

Ignace Van de Woestyne
KU Leuven, Research Unit MEES, Warmoesberg 26, B-1000 Brussels, Belgium,
ignace.vandewoestyne@kuleuven.be

IÉSEG School of Management Lille Catholic University 3, rue de la Digue F-59000 Lille Tel: 33(0)3 20 54 58 92
www.ieseg.fr

Multi-Time and Multi-Moment Nonparametric Frontier-Based Fund Rating: Proposal and Buy-and-Hold Backtesting Strategy*

Kristiaan Kerstens[†], Paolo Mazza[‡], Tiantian Ren[§], Ignace Van de Woestyne[¶]

May 10, 2021

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.

JEL CODES: D24, G11

KEYWORDS: Shortage function; Frontier; Fund rating.

*We gratefully acknowledge the help of Mikael Petitjean in collecting the data.

†IESEG School of Management, CNRS-LEM (UMR 9221), Université de Lille, 3 rue de la Digue, F-59000 Lille, France, `k.kerstens@ieseg.fr`

‡IESEG School of Management, CNRS-LEM (UMR 9221), 3 rue de la Digue, F-59000 Lille, France, `p.mazza@ieseg.fr`

§Corresponding author: School of Business Administration, Hunan University, Changsha 410081, China, and IESEG School of Management, 3 rue de la Digue, F-59000 Lille, France, Tel: +33 320545892 (switchboard), Fax: +33 320574855, `t.ren@ieseg.fr`

¶KU Leuven, Research Unit MEES, Warmoesberg 26, B-1000 Brussels, Belgium, `ignace.vandewoestyne@kuleuven.be`

1 Introduction

The foundational work of Markowitz (1952) 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 (2008) and Feibel (2003) 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. 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 (2003)). At least since Scott and Horvath (1980), 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 (see Eeckhoudt and Schlesinger (2006)). 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 (2018) 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

[1] More rarely alternatives are proposed for the expected return: e.g., Benati (2015) focuses on the median as a location parameter of the distribution of returns.

approach is found in Lai (1991) 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, Liechty, Liechty, and Müller (2010)).

To our knowledge, Sengupta (1989) 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 (1992)). In consumer theory the distance function is dual to the expenditure function: it serves to characterize multiple commodity and single utility choice sets.[2] 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 (2008)). 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, Grosskopf, and Lovell (1994)).[3]

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, Kerstens, and Lesourd (2004) who establish duality between a distance function and MV utility functions, Briec, Kerstens, and Jokung (2007) 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.[4] Even more general, for the class of mixed risk-aversion utility functions, Briec and Kerstens (2010) 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. On the other hand, within a standard MV framework, Morey and Morey (1999) develop a multiple time horizon assessment: in particular, these authors use either a risk contraction or a return expansion efficiency measure

[2] This distance function has sometimes been employed to make welfare comparisons (e.g., Slesnick (1998)).

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

[4] Briec, Kerstens, and Van de Woestyne (2013) establish a relation between MVS portfolio optimisation using the shortage function and the far more popular Polynomial Goal Programming method proposed by Lai (1991).

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 (2009). [5] An empirical application is available in Ren, Zhou, and Xiao (2021).

Empirical applications of this diversified multi-moment approach are found in Adam and Branda (2020), Branda (2013), Branda and Kopa (2014), Branda (2015), Joro and Na (2006), Jurczenko, Maillet, and Merlin (2006), Khemchandani and Chandra (2014), Krüger (2020), Massol and Banal-Estañol (2014), among others. Furthermore, Bacmann and Benedetti (2009), Boudt, Cornilly, and Verdonck (2020), and Jurczenko and Yanou (2010), among others, are empirical diversified multi-moment contributions focusing on hedge funds (HF).

To the best of our knowledge, Murthi, Choi, and Desai (1997) 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 (1957) and Charnes, Cooper, and Rhodes (1978), 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 (2016).

Following Heffernan (1990) and Blake (1996), among others, Kerstens, Mounir, and Van de Woestyne (2011) 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 mod-

[5] Note 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.

els 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, Zhou, Liu, and Xiao (2015) 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, Mounir, and Van de Woestyne (2011) provides a conservative underestimation of the corresponding nonconvex diversified portfolio model within some common subspace of moments (see also Germain, Nalpas, and Vanhems (2011)). By contrast, the more widely used convex nonparametric production frontier specification may overestimate the corresponding nonconvex diversified 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, Sedzro, and Zhu (2005)).

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 (2008), Feibel (2003) and Caporin, Jannin, Lisi, and Maillet (2014)). 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. 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, Jannin, Lisi, and Maillet (2014) 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.[6]

The major objective of this contribution is to define new distance functions or efficiency measures that can simultaneously handle both multiple moments and multiple times. 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, Coen, and Hubner (2008), Menardi and Lisi (2012) and Grau-Carles, Doncel, and Sainz (2019), among others).

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 (2003), Darolles and Gourieroux (2010), Eling and Faust (2010), among others, and especially El Kalak, Azevedo, and Hudson (2016) for a survey). 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. Given the complexities to assess the performance of HFs using traditional performance measures (e.g., see Smith (2017)), 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 non-

[6] Tammer and Zălinescu (2010) 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.

parametric production frontier studies: examples include, e.g., Gregoriou, Sedzro, and Zhu (2005), Kumar, Roy, Saranga, and Singal (2010), Germain, Nalpas, and Vanhems (2011), among others. However, to the best of our knowledge none of these studies appeal to the characteristics approach as proposed by Kerstens, Mounir, and Van de Woestyne (2011). Furthermore, all these existing nonparametric production frontier studies are single time: this contribution is the first to develop a multi-time evaluation framework.

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 ($j \in \{1, \ldots, n\}$) is characterized by m input-like values x_{ij}^t ($i \in \{1, \ldots, m\}$) and s output-like values y_{rj}^t ($r \in \{1, \ldots, s\}$). 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). Following Briec, Kerstens, and Vanden Eeckaut (2004), 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:

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

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 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 $S^t_\Lambda(z^t_o; g^t_o) > 0$ for the input-output combination $z^t_o = (x^t_o, y^t_o)$ of a specific MF at time t, then z^t_o 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 $S^t_\Lambda(z^t_o; g^t_o) = 0$, then z^t_o 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:

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

Note that Model (2) results in a linear programming (LP) problem under convexity and a mixed binary integer programming (MBIP) problem under nonconvexity.

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 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 $R_1^t, \ldots, R_n^t$ denote the random 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 $j \in \{1, \ldots, n\}$. 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 $K(R_j^t) = E[(R_j^t - E(R_j^t))^4]$. 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 $j \in \{1, \ldots, n\}$. 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. To develop the nonparametric frontier rating models in this multi-time framework, some definitions and properties are presented.

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 $\mathbf{P}_\Lambda^T$. If the time discounted multi-time shortage function value $S_\Lambda^T(Z; G) > 0$ 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:

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

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.

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 eplain 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. 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, Garlappi, and Uppal (2009), Tu and Zhou (2011), Brandouy, Kerstens, and Van de Woestyne (2015), Zhou, Xiao, Jin, and Liu (2018) and Lin and Li (2020), among others.

For comparison, there are 15 fund rating methods in total being collected in our work. On the one hand, we test some 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 (2003, p. 187 and p. 200) and Eling and Schuhmacher (2007, p. 2635), respectively. Based on these definitions, these three traditional financial ratios are presented as follows:

$$\text{Sharpe} = \frac{E(R) - r_f}{\sigma(R)}, \tag{5}$$

$$\text{Sortino} = \frac{E(R) - r_f}{\sigma_-(R)}, \tag{6}$$

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

where $E(R)$ and r_f represent the mean value of a random return R and the risk-free rate, respectively; $\sigma(R)$ and $\sigma_-(R)$ denote the standard and lower semi-standard deviations of a random return R, respectively; L is the loss threshold, in particular, above this threshold returns are considered gains, while below this treshold these are regarded as losses. Using the above three ratios, we obtain the financial indexes for the above n MFs (i.e., Sharpe_j, Sortino_j and Omega_j, where $j = 1, \ldots, n$) which can be use to measure their performance at the given time horizon T, and the higher the value, the better the performance. The risk-free rate R 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:

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

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

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

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.

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

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 non-convex (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.

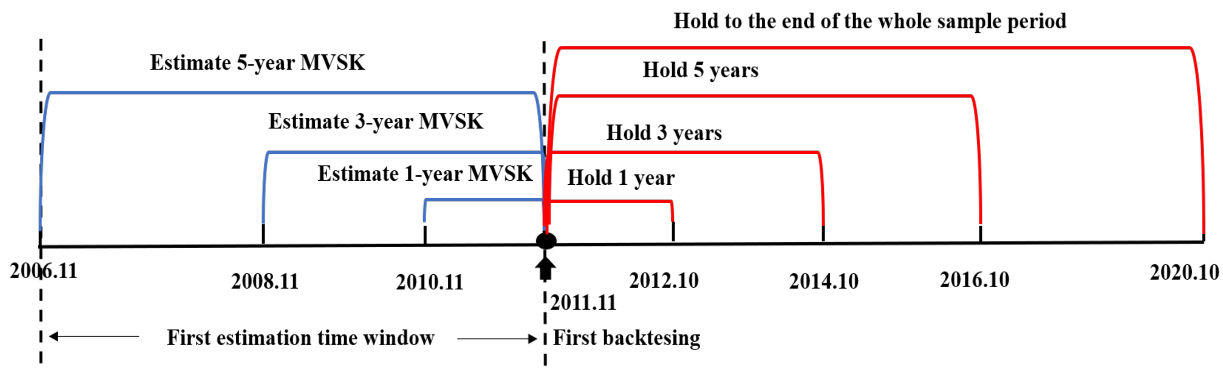

Figure 1: Process of the first backtesting window

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.

(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 convex and nonconvex VRS 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.

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. The data is all downloaded from Lipper for Investment Management made available by Hedge Funds database. It needs to be stated that we initially specify these nonparametric frontier rating methods following the idea of Kerstens, Mounir, and Van de Woestyne (2011) 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. 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 of the sample.

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

	Mean	Variance	Skewness	Kurtosis
Min.	−0.328	0.633	−621.506	3.866
Q1	0.306	8.764	−43.341	481.584
Median	0.447	14.971	−10.294	1293.516
Mean	0.480	26.810	210.182	34145.995
Q3	0.601	27.018	1.468	4267.635
Max.	1.733	521.156	22732.909	2655540.333

From the descriptive statistics of the monthly returns reported in Table 2, 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 A.1 in Appendix A. 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 (1996) and refined by Fan and Ullah (1999) and most recently by Li, Maasoumi, and Racine (2009). 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.[7] Table 3 provides Li-test statistics 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 by traditional financial performance measures and those computed by frontier-based rating methods are significantly different at the 1 % significance level.

Second, in both convex and nonconvex 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

[7] Matlab code developed by P.J. Kerstens based on Li, Maasoumi, and Racine (2009) is found at: `https://github.com/kepiej/DEAUtils`.

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(*).

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***

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, Mounir, and Van de Woestyne (2011).

4.3 Backtesting Results

We analyze the backtesting scenarios with a selection of the 10, 20 or 30 best performing HFs, respectively. 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 convex and nonconvex VRS frontiers in single-time and multi-moment frameworks (with convexity and nonconvexity). 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 convex VRS frontier in single-time and multi-moment framework (VRSc), while the last two columns report these results in the nonconvex case (VRSnc).

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

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 B.1 in Appendix B 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 B.

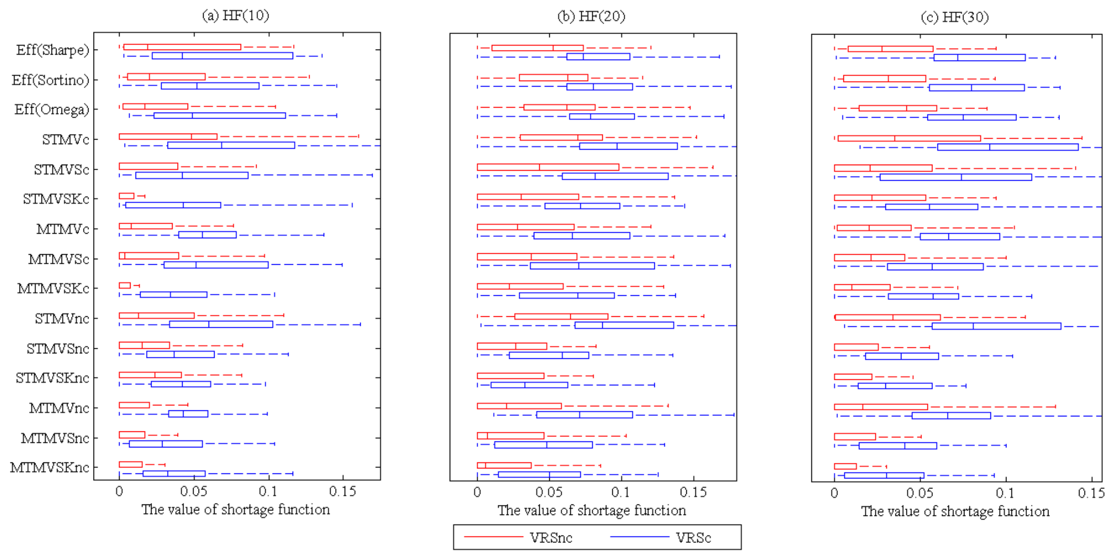

Figure 2: Distributions of inefficiency scores for 15 buy-and-hold backtesting strategies

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 convex VRS frontier are in blue, and those based on the nonconvex VRS 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 B.1, B.2 and B.3 in Appendix B, respectively. From Figures B.1, B.2 and B.3, 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 B for details).

5 Conclusion

Inspired by recent nonparametric frontier rating methods contributing to assessing MF performance (e.g., Kerstens, Mounir, and Van de Woestyne (2011)), 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, Mounir, and Van de Woestyne (2011).

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 fund of funds. This is one of the main avenues for future research. Another desirable extension is to transfer the current methodological framework and perform a backtesting analysis in diversified models. It is worthwhile to compare the performance in MF selection between the backtesting strategies driven by diversified frontier rating methods and those driven by nondiversified frontier rating methods (i.e., convex and nonconvex frontier rating methods).

References

Adam, L., and M. Branda (2020): "Risk-aversion in Data Envelopment Analysis Models with Diversification," *Omega*, 102, 102338.

Bacmann, J.-F., and S. Benedetti (2009): "Optimal Bayesian Portfolios of Hedge Funds," *International Journal of Risk Assessment and Management*, 11(1-2), 39–58.

Bacon, C. (2008): *Practical Portfolio Performance Measurement and Attribution*. Wiley, New York, 2nd edn.

Basso, A., and S. Funari (2016): "DEA Performance Assessment of Mutual Funds," in *Data Envelopment Analysis: A Handbook of Empirical Studies and Applications*, ed. by J. Zhu, pp. 229–287. Springer, Berlin.

Benati, S. (2015): "Using Medians in Portfolio Optimization," *Journal of the Operational Research Society*, 66(5), 720–731.

Blake, D. (1996): "Financial Intermediation and Financial Innovation in a Characteristics Framework," *Scottish Journal of Political Economy*, 43(1), 16–31.

Bodson, L., A. Coen, and G. Hubner (2008): "How Stable are the Major Performance Measures?," *Journal of Performance Measurement*, 13(1), 21–30.

Boudt, K., D. Cornilly, and T. Verdonck (2020): "A Coskewness Shrinkage Approach for Estimating the Skewness of Linear Combinations of Random Variables," *Journal of Financial Econometrics*, 18(1), 1–23.

Branda, M. (2013): "Diversification-consistent data envelopment analysis with general deviation measures," *European Journal of Operational Research*, 226(3), 626–635.

Branda, M. (2015): "Diversification-consistent Data Envelopment Analysis based on Directional-Distance Measures," *Omega*, 52, 65–76.

Branda, M., and M. Kopa (2014): "On Relations between DEA-Risk Models and Stochastic Dominance Efficiency Tests," *Central European Journal of Operations Research*, 22(1), 13–35.

Brandouy, O., K. Kerstens, and I. Van de Woestyne (2015): "Frontier-based vs. Traditional Mutual Fund Ratings: A First Backtesting Analysis," *European Journal of Operational Research*, 242(1), 332–342.

Briec, W., and K. Kerstens (2009): "Multi-Horizon Markowitz Portfolio Performance Appraisals: A General Approach," *Omega: The International Journal of Management Science*, 37(1), 50–62.

——— (2010): "Portfolio Selection in Multidimensional General and Partial Moment Space," *Journal of Economic Dynamics and Control*, 34(4), 636–656.

Briec, W., K. Kerstens, and O. Jokung (2007): "Mean-Variance-Skewness Portfolio Performance Gauging: A General Shortage Function and Dual Approach," *Management Science*, 53(1), 135–149.

Briec, W., K. Kerstens, and J. Lesourd (2004): "Single Period Markowitz Portfolio Selection, Performance Gauging and Duality: A Variation on the Luenberger Shortage Function," *Journal of Optimization Theory and Applications*, 120(1), 1–27.

Briec, W., K. Kerstens, and I. Van de Woestyne (2013): "Portfolio Selection with Skewness: A Comparison of Methods and a Generalized One Fund Result," *European Journal of Operational Research*, 230(2), 412–421.

Briec, W., K. Kerstens, and P. Vanden Eeckaut (2004): "Non-Convex Technologies and Cost Functions: Definitions, Duality and Nonparametric Tests of Convexity," *Journal of Economics*, 81(2), 155–192.

Caporin, M., G. Jannin, F. Lisi, and B. Maillet (2014): "A Survey on the Four Families of Performance Measures," *Journal of Economic Surveys*, 28(5), 917–942.

Charnes, A., W. Cooper, and E. Rhodes (1978): "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research*, 2(6), 429–444.

Cornes, R. (1992): *Duality and Modern Economics*. Cambridge University Press, Cambridge.

Darolles, S., and C. Gourieroux (2010): "Conditionally Fitted Sharpe Performance with an Application to Hedge Fund Rating," *Journal of Banking & Finance*, 34(3), 578–593.

DeMiguel, V., L. Garlappi, and R. Uppal (2009): "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?," *Review of Financial Studies*, 22(5), 1915–1953.

Eeckhoudt, L., and H. Schlesinger (2006): "Putting Risk in its Proper Place," *American Economic Review*, 96(1), 280–289.

El Kalak, I., A. Azevedo, and R. Hudson (2016): "Reviewing the Hedge Funds Literature II: Hedge Funds' Returns and Risk Management Characteristics," *International Review of Financial Analysis*, 48, 55–66.

Eling, M., and R. Faust (2010): "The Performance of Hedge Funds and Mutual Funds in Emerging Markets," *Journal of Banking & Finance*, 34(8), 1993–2009.

Eling, M., and F. Schuhmacher (2007): "Does the Choice of Performance Measure Influence the Evaluation of Hedge Funds?," *Journal of Banking & Finance*, 31(9), 2632–2647.

Fan, Y., and A. Ullah (1999): "On Goodness-Of-Fit Tests for Weakly Dependent Processes Using Kernel Method," *Journal of Nonparametric Statistics*, 11(1-3), 337–360.

Färe, R., S. Grosskopf, and C. Lovell (1994): *Production Frontiers*. Cambridge University Press, Cambridge.

Farrell, M. (1957): "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society Series A: General*, 120(3), 253–281.

Feibel, B. (2003): *Investment Performance Measurement.* Wiley, New York.

Germain, L., N. Nalpas, and A. Vanhems (2011): "Nonparametric Hedge Funds and Replication Indices Performance Analysis: A Robust Directional Application," in *Hedge Funds Replication*, ed. by G. Gregoriou, and M. Kooli, pp. 90–105. Palgrave Macmillan, Basingstoke.

Grau-Carles, P., L. Doncel, and J. Sainz (2019): "Stability in Mutual Fund Performance Rankings: A New Proposal," *International Review of Economics & Finance*, 61, 337–346.

Gregoriou, G. (2003): "Performance Appraisal of Funds of Hedge Funds Using Data Envelopment Analysis," *Journal of Wealth Management*, 5(4), 88–95.

Gregoriou, G., K. Sedzro, and J. Zhu (2005): "Hedge Fund Performance Appraisal Using Data Envelopment Analysis," *European Journal of Operational Research*, 164(2), 555–571.

Hackman, S. (2008): *Production Economics: Integrating the Microeconomic and Engineering Perspectives.* Springer, Berlin.

Harvey, C., J. Liechty, M. Liechty, and P. Müller (2010): "Portfolio Selection with Higher Moments," *Quantitative Finance*, 10(5), 469–485.

Heffernan, S. (1990): "A Characteristics Definition of Financial Markets," *Journal of Banking & Finance*, 14(2–3), 583–609.

Jondeau, E., and M. Rockinger (2003): "Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements," *Journal of Economic Dynamics and Control*, 27(10), 1699–1737.

Joro, T., and P. Na (2006): "Portfolio Performance Evaluation in Mean-Variance-Skewness Framework," *European Journal of Operational Research*, 175(1), 446–461.

Jurczenko, E., B. Maillet, and P. Merlin (2006): "Hedge Funds Portfolio Selection with Higher-order Moments: A Nonparametric Mean-Variance-Skewness-Kurtosis Efficient Frontier," in *Multi-moment Asset Allocation and Pricing Models*, ed. by E. Jurczenko, and B. Maillet, pp. 51–66. Wiley, New York.

Jurczenko, E., and G. Yanou (2010): "Fund of Hedge Funds Portfolio Selection: A Robust Non-parametric Multi-moment Approach," in *The Recent Trend of Hedge Fund Strategies*, ed. by Y. Watanabe, pp. 21–56. Nova Science, New York.

Kerstens, K., A. Mounir, and I. Van de Woestyne (2011): "Non-Parametric Frontier Estimates of Mutual Fund Performance Using C- and L-Moments: Some Specification Tests," *Journal of Banking & Finance*, 35(5), 1190–1201.

Khemchandani, R., and S. Chandra (2014): "Efficient Trading Frontier: A Shortage Function Approach," *Optimization*, 63(10), 1533–1548.

Krüger, J. (2020): "Nonparametric Portfolio Efficiency Measurement with Higher Moments," *Empirical Economics*, p. forthcoming.

Kumar, U. D., A. B. Roy, H. Saranga, and K. Singal (2010): "Analysis of Hedge Fund Strategies Using Slack-Based DEA Models," *Journal of the Operational Research Society*, 61(12), 1746–1760.

Lai, T. Y. (1991): "Portfolio Selection with Skewness: A Multiple Objective Approach," *Review of Quantitative Finance and Accounting*, 1(3), 293–305.

Li, Q. (1996): "Nonparametric Testing of Closeness Between Two Unknown Distribution Functions," *Econometric Reviews*, 15(3), 261–274.

Li, Q., E. Maasoumi, and J. Racine (2009): "A Nonparametric Test for Equality of Distributions with Mixed Categorical and Continuous Data," *Journal of Econometrics*, 148(2), 186–200.

Lin, R., and Z. Li (2020): "Directional Distance Based Diversification Super-Efficiency DEA Models for Mutual Funds," *Omega*, 97, 102096.

Liu, W., Z. Zhou, D. Liu, and H. Xiao (2015): "Estimation of Portfolio Efficiency via DEA," *Omega*, 52(1), 107–118.

Markowitz, H. (1952): "Portfolio Selection," *Journal of Finance*, 7(1), 77–91.

Massol, O., and A. Banal-Estañol (2014): "Export Diversification through Resource-Based Industrialization: The Case of Natural Gas," *European Journal of Operational Research*, 237(3), 1067–1082.

Menardi, G., and F. Lisi (2012): "Are Performance Measures Equally Stable?," *Annals of Finance*, 8(4), 553–570.

Morey, M., and R. Morey (1999): "Mutual Fund Performance Appraisals: A Multi-Horizon Perspective With Endogenous Benchmarking," *Omega*, 27(2), 241–258.

Murthi, B., Y. Choi, and P. Desai (1997): "Efficiency of Mutual Funds and Portfolio Performance Measurement: A Non-Parametric Approach," *European Journal of Operational Research*, 98(2), 408–418.

Ren, T., Z. Zhou, and H. Xiao (2021): "Estimation of Portfolio Efficiency Considering Social Responsibility: Evidence from the Multi-Horizon Diversification DEA," *RAIRO-Operations Research*, 55(2), 611–637.

Scott, R., and P. Horvath (1980): "On the Direction of Preference for Moments of Higher Order than the Variance," *Journal of Finance*, 35(4), 915–919.

Sengupta, J. (1989): "Nonparametric Tests of Efficiency of Portfolio Investment," *Journal of Economics*, 50(3), 1–15.

Slesnick, D. (1998): "Empirical Approaches to the Measurement of Welfare," *Journal of Economic Literature*, 36(4), 2108–2165.

Smith, D. (2017): "Evaluating Hedge Fund Performance," in *Hedge Funds: Structure, Strategies, and Performance*, ed. by H. Baker, and G. Filbeck, pp. 415–438. Oxford University Press, Oxford.

Tammer, C., and C. Zălinescu (2010): "Lipschitz Properties of the Scalarization Function and Applications," *Optimization*, 59(2), 305–319.

Trautmann, S., and G. van de Kuilen (2018): "Higher Order Risk Attitudes: A Review of Experimental Evidence," *European Economic Review*, 103, 108–124.

Tu, J., and G. Zhou (2011): "Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies," *Journal of Financial Economics*, 99, 204–215.

Zhou, Z., H. Xiao, Q. Jin, and W. Liu (2018): "DEA Frontier Improvement and Portfolio Rebalancing: An Application of China Mutual Funds on Considering Sustainability Information Disclosure," *European Journal of Operational Research*, 269(1), 111–131.

Appendices: Supplementary Material

A Sample Description: Further Details

The descriptive statistics on the first four moments of the 187 HF sample over 1-year time period (sample period: Nov 2019 to Oct 2020), 3-year time period (sample period: Nov 2017 to Oct 2020) and 5-year time period (sample period: Nov 2015 to Oct 2020) are provided in Table A.1.

Table A.1: Descriptive statistics for all 187 HFs over 1-, 3- and 5-year time periods

1 year monthly return: From Nov 2019 to Oct 2020

	Mean	Variance	Skewness	Kurtosis
Min.	−3.913	0.241	−4765.126	0.305
Q1	−0.467	8.580	−135.113	259.251
Median	0.356	18.270	−12.409	1277.641
Mean	0.423	45.237	414.608	71518.803
Q3	0.930	45.155	14.606	6685.385
Max.	10.857	937.351	60038.381	5381641.258

3 year monthly return: From Nov 2017 to Oct 2020

	Mean	Variance	Skewness	Kurtosis
Min.	−2.836	0.258	−1820.695	0.217
Q1	−0.134	7.925	−65.187	344.122
Median	0.175	14.259	−17.723	945.365
Mean	0.248	27.364	240.996	36348.550
Q3	0.463	28.983	2.409	3974.036
Max.	4.623	533.743	30000.036	2865464.753

5 year monthly return: From Nov 2015 to Oct 2020

	Mean	Variance	Skewness	Kurtosis
Min.	−1.618	0.228	−1142.152	0.196
Q1	−0.005	7.290	−45.376	275.106
Median	0.254	12.959	−12.149	715.167
Mean	0.345	26.036	243.423	39749.791
Q3	0.609	25.073	0.299	3002.459
Max.	3.943	705.232	27466.851	3289535.317

As observed from column 4 in Table A.1, we find that for each time, the series composed by the skewness of 187 HFs shows positive mean and negative median, as well as a large dispersion. From column 5, it can be seen that all 187 HFs have positive kurtosis in each time, and also exhibit a high dispersion. These results are in line with the ones reported in the main body of the text. This partly indicates that the stability and persistence of these return characteristics for the HF sample is maintained over different times. In addition, there are certain differences among the 1-year, 3-year and 5-year MVSK of this HF sample. To some extent, the addition of multiple moments and multiple times may provide a more accurate picture to describe HF's return characteristics compared to only considering the mean and variance at a single time.

B Backtesting Results: Sensitivity Analysis

To develop a sensitivity analysis with respect of the holding period, this Appendix focuses on testing the short-, medium- and long-term holding performance of the buy-and-hold backtesting strategies based on the proposed multi-moment and multi-time rating methods. The performance of strategies held for only 1 year is regarded as a short-term holding performance, for 3 years as a medium-term holding performance, and for 5 years as a long-term holding performance. For each of the three holding scenarios, the 15 strategies are compared in terms of the MVSK performances of their holding values that are always evaluated by the shortage function based on the convex and nonconvex VRS frontiers in single-time and multi-moment framework. Table B.1 reports the summarized results with respect to the performance per buy-and-hold backtesting strategy held for 1, 3 and 5 years.

Table B.1 is organized as follows: the three series consisting of four columns list the performance results for holding the selected HFs over 1, 3 and 5 years, respectively. Within each holding period scenario, the first two columns report the average inefficiency scores and the number of efficient units for each method when evaluated using the convex VRS frontier in single-time and multi-moment framework (VRSc), while the last two columns report these results in the nonconvex case (VRSnc). Horizontally, each block of rows contains the results of the selection of the 10, 20 or 30 best performing HFs, respectively.

We now analyze the results on the three holding scenarios presented in Table B.1, following the same basic logic of analysing the 15 strategies in the main text. Thus, the performance of strategies generated by two family of ratings (frontier vs. finance) is compared first, and then the comparison between the frontier families of ratings is developed separately (i.e., multi-moments vs. MV; multi-times vs. single time; convexity vs. nonconvexity).

We first discuss the short-term holding performance of the 15 buy-and-hold strategies, as shown in columns 2-6 of Table B.1. First, it can be observed that minor difference on the short-term holding performance is observed between the strategies depending on the multi-time frontier ratings and those depending on traditional financial ratings, and both their performances are superior over other frontier-based strategies. Second, in most cases, the strategies based on the multi-moment ratings do not show superiority compared to those based on the MV ratings when these strategies are held for only 1 year. This result is somewhat at odds with the one reported in the main text. Third, combining average inefficiency scores and the number of efficient units, the strategies constructed in the multi-time rating framework perform better over those in the single-time rating framework under the 1-year holding scenario. Finally, in terms of short-term holding performance, the strategies determined by the nonconvex frontier-based ratings outperform those determined by the convex frontier-based ratings in the majority of cases. The latter finding is in line with the one shown in the main text. It needs to be mentioned that some of the findings may be somewhat unstable with respect to the 1-year holding period due to the limited data for testing the short-term holding performance of the 15 buy-and-hold strategies.

Looking at columns 7-10 of Table B.1 for the medium-term holding performance of the 15 strategies, one can draw the following observations. The frontier-based strategies with consideration of multi-moments and multi-times (separately or jointly) largely outperform the finance-based strategies. It is easy to observe that the strategies driven by the multi-moment and multi-time frontier ratings generally yield lower average inefficiency scores and more efficient units compared to Sharpe-, Sortino- and Omega-driven strategies. Turning to the comparisons between various frontier-based rating methods, the buy-and-hold strategies based on the multi-moment ratings (MVS & MVSK) perform better than those based on the basic MV ratings. This is confirmed in both single-time and multi-time rating frameworks. Moreover, consistent with the finding on considering multiple times in the 1-year holding scenario (see the third finding), the multi-time frontier-based strategies outperform the single-time frontier-based strategies in most cases in the 3-year holding scenario. Again, when comparing convex and nonconvex frontier-based strategies in the medium-term holding scenario, the same coherent finding emerges as in the short-term holding scenario (see the final finding analyzed in the 1-year holding context).

Following up the results regarding the 5-year holding scenario as reported in columns 11-14 of Table B.1, the above four findings emerging in the 3-year holding period are also evidenced in this holding scenario. These results in the medium- and long-term holding

Table B.1: Performance results for 15 buy-and-hold backtesting strategies held for 1, 3 and 5 years: Descriptive statistics of the values of shortage function

	Methods	Hold 1 year				Hold 3 years				Hold 5 years			
		VRSc		VRSnc		VRSc		VRSnc		VRSc		VRSnc	
		Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.	Average	#Ef. Obs.
HF(10)	Eff(Sharpe)	0.056	1	0.041	12	0.066	2	0.043	6	0.060	0	0.034	14
	Eff(Sortino)	0.056	2	0.041	13	0.064	3	0.036	13	0.057	2	0.030	15
	Eff(Omega)	0.052	3	0.039	9	0.062	3	0.035	9	0.056	2	0.030	19
	STMVc	0.079	2	0.063	6	0.085	6	0.062	14	0.077	3	0.045	9
	STMVSc	0.098	5	0.055	16	0.074	1	0.038	15	0.081	2	0.039	17
	STMVSKc	0.077	4	0.045	16	0.066	4	0.040	18	0.060	4	0.028	24
	MTMVc	0.059	3	0.048	7	0.069	2	0.044	16	0.057	2	0.033	14
	MTMVSc	0.076	3	0.063	7	0.086	4	0.057	13	0.076	4	0.045	14
	MTMVSKc	0.070	3	0.055	7	0.064	0	0.037	22	0.053	7	0.025	18
	STMVnc	0.073	2	0.058	4	0.072	6	0.045	14	0.061	3	0.031	18
	STMVSnc	0.050	3	0.033	11	0.044	3	0.027	18	0.042	9	0.026	17
	STMVSKnc	0.043	3	0.031	8	0.038	2	0.017	21	0.041	4	0.024	10
	MTMVnc	0.051	4	0.042	7	0.066	4	0.037	22	0.051	3	0.023	18
	MTMVSnc	0.067	4	0.050	12	0.066	2	0.040	26	0.049	5	0.024	19
	MTMVSKnc	0.059	3	0.045	14	0.049	3	0.027	31	0.046	7	0.020	21
HF(20)	Eff(Sharpe)	0.044	2	0.035	6	0.095	2	0.048	10	0.058	1	0.034	13
	Eff(Sortino)	0.044	0	0.034	8	0.091	1	0.049	14	0.059	0	0.033	8
	Eff(Omega)	0.042	2	0.034	9	0.089	0	0.042	7	0.059	0	0.035	9
	STMVc	0.060	1	0.049	7	0.118	0	0.071	9	0.072	2	0.045	10
	STMVSc	0.080	3	0.053	12	0.102	1	0.071	5	0.088	0	0.067	8
	STMVSKc	0.060	4	0.037	13	0.086	4	0.058	10	0.067	3	0.044	12
	MTMVc	0.038	4	0.027	9	0.102	1	0.076	6	0.070	3	0.046	17
	MTMVSc	0.046	3	0.032	12	0.109	1	0.073	8	0.081	3	0.056	8
	MTMVSKc	0.046	1	0.036	9	0.078	0	0.045	18	0.058	2	0.039	6
	STMVnc	0.057	3	0.049	8	0.116	0	0.070	8	0.069	1	0.042	13
	STMVSnc	0.060	2	0.037	14	0.070	4	0.051	17	0.052	0	0.034	7
	STMVSKnc	0.055	2	0.037	11	0.066	5	0.046	22	0.043	5	0.024	20
	MTMVnc	0.034	3	0.024	12	0.092	2	0.065	10	0.067	2	0.043	16
	MTMVSnc	0.051	1	0.042	4	0.074	3	0.046	13	0.053	2	0.032	7
	MTMVSKnc	0.043	4	0.032	11	0.066	4	0.039	18	0.047	5	0.027	11
HF(30)	Eff(Sharpe)	0.038	1	0.029	6	0.072	2	0.035	9	0.068	0	0.046	8
	Eff(Sortino)	0.038	3	0.028	8	0.070	2	0.031	12	0.068	1	0.042	8
	Eff(Omega)	0.038	0	0.028	7	0.069	2	0.038	7	0.071	1	0.044	9
	STMVc	0.055	1	0.046	6	0.101	0	0.064	4	0.079	1	0.053	7
	STMVSc	0.068	2	0.043	8	0.095	1	0.065	4	0.090	2	0.070	8
	STMVSKc	0.053	2	0.034	15	0.082	1	0.053	9	0.083	2	0.065	6
	MTMVc	0.039	0	0.028	10	0.085	1	0.059	3	0.065	5	0.043	19
	MTMVSc	0.042	2	0.025	13	0.085	5	0.065	7	0.085	5	0.058	12
	MTMVSKc	0.039	3	0.027	10	0.064	2	0.044	9	0.071	1	0.040	9
	STMVnc	0.053	1	0.041	8	0.096	0	0.058	3	0.072	2	0.050	6
	STMVSnc	0.051	1	0.029	14	0.072	1	0.050	12	0.066	1	0.040	6
	STMVSKnc	0.051	1	0.029	13	0.065	5	0.044	12	0.060	2	0.041	8
	MTMVnc	0.033	3	0.021	12	0.076	1	0.051	6	0.065	2	0.038	16
	MTMVSnc	0.042	3	0.031	11	0.060	3	0.037	16	0.064	3	0.039	10
	MTMVSKnc	0.038	5	0.027	9	0.052	4	0.036	17	0.055	4	0.035	13

scenarios are rather in line with the ones reported in the main body. We basically conclude that the buy-and-hold backtesting strategies based on the proposed multi-time and multi-moment models show a superior performance in different holding period scenarios.

Apart from comparing the performance of the 15 buy-and-hold strategies vertically for each of three holding scenarios, we have also run a horizontal analysis on the consistency and stability of the performance per strategy over different holding periods. Looking at the evolution of the average inefficiency scores and the number of efficient units per strategy held for 1, 3 and 5 years allows to infer two new and interesting observations. First, the strategies consisting of the best HFs selected by financial indicators and basic MV frontier rating methods tend to exhibit worse performance in medium- and long-term holding periods compared to their performances in a short-term holding period. By contrast, the strategies with the consideration of multiple moments and multiple times usually exhibit favorable and consistent short-, medium- and long-term holding performance. Second, focusing on the MVS and MVSK settings in the multi-moment rating framework, it can be noticed that compared to the strategies based on the multi-time ratings adding skewness only, the ones based on the ratings adding both skewness and kurtosis show better and more consistent short-, medium-, and long-term holding performance. This finding reveals the necessity for the addition of kurtosis in HF rating and selection. Indeed, including the kurtosis reduces the disturbance of certain extreme values to the fund ratings, and therefore the funds selected tend to present both better and more stable returns. These results somewhat suggest that the performance persistence of the best-performing HFs rated by the multi-moment and multi-time performance measure is well maintained over time.

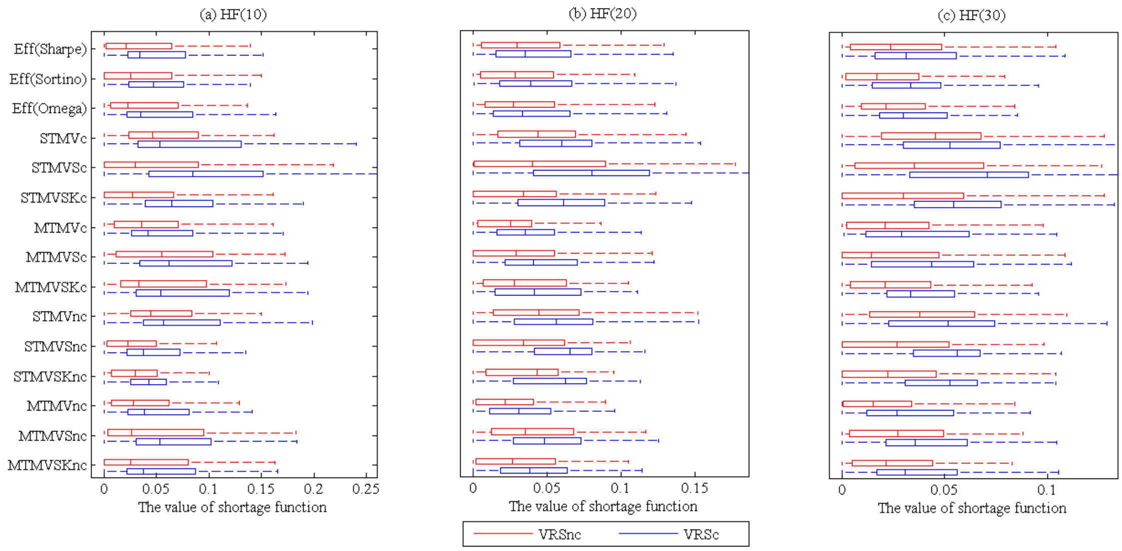

Figure B.1: Distributions of inefficiency scores for 15 buy-and-hold backtesting strategies held for 1 year

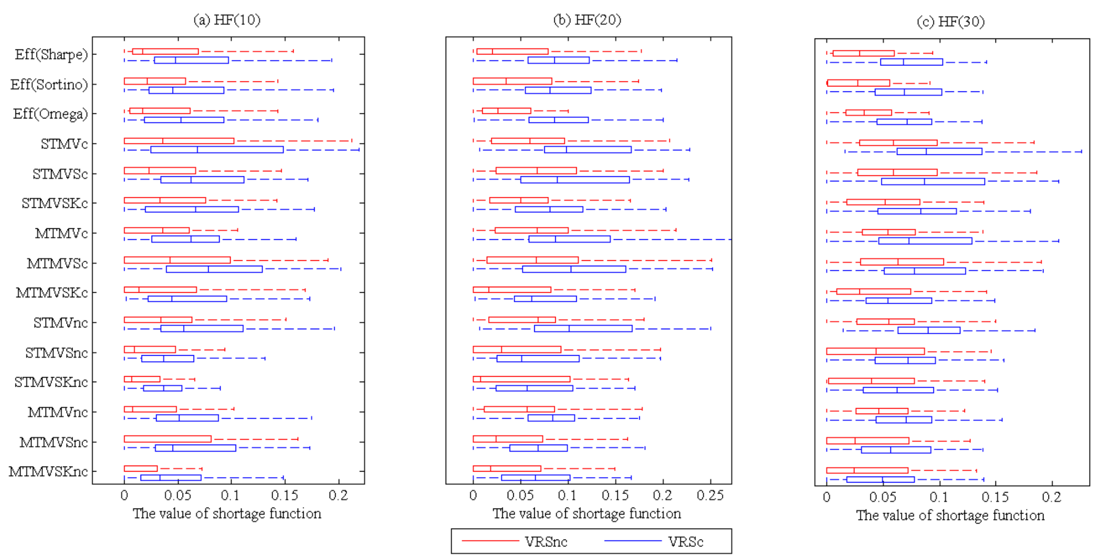

Figure B.2: Distributions of inefficiency scores for 15 buy-and-hold backtesting strategies held for 3 years

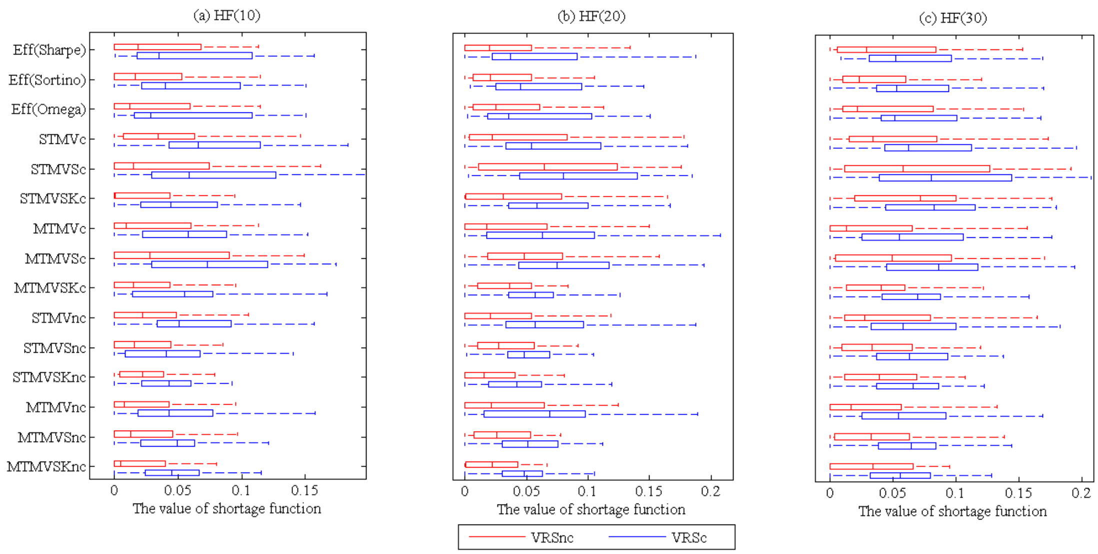

Figure B.3: Distributions of inefficiency scores for 15 buy-and-hold backtesting strategies held for 5 years

To compare the 15 buy-and-hold strategies intuitively, Figures B.1, B.2 and B.3 offer box-plots to describe the entire distributions of the inefficiency scores per strategy held for 1, 3 and 5 years, respectively. In each figure, the sub-figures (a) to (c) correspond to the performance results of the buying scenarios with 10, 20 and 30 best HFs selected, whereby the performance of strategies based on the convex VRS frontier are depicted in blue, and those based on the nonconvex VRS frontier are displayed in red. As introduced in the main text, the box of these box-plots indicates the interquartile range where the small vertical line reports the location of the median. Straightforwardly, the location of the median closer to the left indicates that the entire distribution of inefficiency scores for one strategy is somewhat skewed to the left, which signals that the strategy performs better in the backtesting analysis because the probability mass of the inefficiency is closer to zero. Two major observations can be made with regard to these results in Figures B.1, B.2 and B.3. First, although the buy-and-hold strategies constructed by the multi-moment and multi-time frontier rating methods do not exhibit a significant superiority in the 1-year holding scenario, they establish a clear dominance over the other strategies in both the 3- and 5-year holding scenarios (see sub-figures (a) and (b) of Figures B.1, B.2 and B.3). Second, concentrating on Figures B.1, B.2 and B.3 individually, the good performance of the buy-and-hold strategies depending on the multi-moment and multi-time ratings tends to be consistent and stable over time.