Innovative Applications of O.R.

Evaluating different groups of mutual funds using a metafrontier approach: Ethical vs. non-ethical funds

Qianying Jin [a,*], Antonella Basso [b], Stefania Funari [c], Kristiaan Kerstens [d], Ignace Van de Woestyne [e]

[a] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
[b] Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, Venice, 30121, Italy
[c] Department of Management, Ca' Foscari University of Venice, Cannaregio 873, Venice, 30121, Italy
[d] Univ. Lille, CNRS, IESEG School of Management, UMR 9221 - LEM - Lille Économie Management, F-59000 Lille, France
[e] KU Leuven, Research Centre for Operations Research and Statistics (ORSTAT), Brussels Campus, Warmoesberg 26, Brussels B-1000, Belgium

A R T I C L E   I N F O

A B S T R A C T

Ethical mutual funds (MFs) have grown in popularity over the past few years. However, the investors generally have concerns about their profitability compared to the investment group of non-ethical MFs. Performance comparison could be a potential way to address this concern, but the differences in their essential investment objectives raise the issue of heterogeneity between the ethical and non-ethical investment groups. Motivated by addressing this heterogeneity, this article proposes a general nonconvex metafrontier framework for comparing different investment groups of MFs. Investment groups can exhibit heterogeneity from different perspectives, such as from regulations, resource constraints, to name a few. To provide a rather complete framework for estimating the frontiers, the diversified, convex and nonconvex evaluation approaches are adapted and presented in a multi-moment setting. The proposed metafrontier framework is then applied to an empirical example where the investment groups are heterogeneous from the ethical perspective. The empirical results suggest that the ethical constraint does not necessarily lead to a worse financial performance; quite the contrary, the results provide some evidence on the outperformance of ethical MFs over the non-ethical MFs.

1. Introduction

Ethical or socially responsible mutual funds (MFs) are playing an increasingly important role over the past decades, as there is a growing awareness of social engagement. The ethical MFs are defined by their compliance with the stated beliefs in environmental, religious, or political precepts. These ethical features encompassed by ethical MFs allow them to address the deep human needs sufficiently while perusing profitability. However, there is also a concern as to whether satisfying the ethical needs of MFs will necessarily harm their profitability. This concern arises because the fund manager of the ethical MFs are restricted to invest only in "ethical" assets, potentially hampering the market selection skills. To investigate this concern, lots of researchers have empirically compared the performance of ethical MFs with that of non-ethical MFs,

especially from a financial point of view.[1] In the literature, some of the studies conclude that the performance differences between the ethical and non-ethical MFs are not statistically significant (Basso & Funari, 2014a; 2014b; Statman, 2000), while some others find that the ethical MFs, on average, have a lower return than the non-ethical ones (Basso & Funari, 2008; Havemann & Webster, 1999). By now, no consensus has emerged about the above concern.

In this contribution, we are also interested in comparing the financial performance of ethical MFs with that of non-ethical MFs, but under a frontier-based framework. Since the seminal contribution of Murthi et al. (1997), the frontier-based framework has been amply applied in evaluating the performance of MFs (see Basso & Funari, 2016 for an early overview considering also different categories of mutual funds besides the conventional and ethical

---

* Corresponding author.
E-mail addresses: qianying.jin@nuaa.edu.cn (Q. Jin), basso@unive.it (A. Basso), funari@unive.it (S. Funari), k.kerstens@ieseg.fr (K. Kerstens), ignace.vandewoestyne@kuleuven.be (I. Van de Woestyne).

[1] In this contribution, the term "non-ethical MFs" is used as a counterpart to the term "ethical MFs". It describes all funds which do not comply with explicitly stated environmental, religious, political, or governmental precepts. One may also talk about traditional or conventional funds.

ones, e.g. Islamic funds). The main advantage to the use of frontier or extremum estimators is that it allows assessing the performance of each MF along a multitude of dimensions instead of using just some combination as in most financial performance ratios (Brandouy et al., 2015).

Under the frontier-based framework, previous contributions in the literature evaluate the ethical and non-ethical MFs with respect to the same frontier (e.g., Abdelsalam et al., 2014b; Basso & Funari, 2014a, among others). However, a frontier is supposed to be composed of similar entities. In evaluating the ethical MFs, a non-ethical MF does not count as a similar entity to constitute the frontier, since its potential investment in "non-ethical" companies. Therefore, the frontier for evaluating ethical MFs should consist of ethical MFs only. Likewise, the frontier of non-ethical MFs should be constructed from non-ethical MFs. Thus, separate frontiers should be constructed for evaluating ethical and non-ethical MFs, since their underlying assets could be different. Nevertheless, the efficiencies measured relative to the frontier of ethical MFs cannot be directly compared with the efficiencies measured relative to the frontier of non-ethical MFs. The only special case occurs when these two frontiers happen to be identical, which rarely happens. That is to say, the ethical constraint differentiates the MFs into independent investment groups, hence the problem of heterogeneity arises and the efficiencies are not directly comparable.

In practice, the ethical constraint is not the only restriction that categorizes the MFs into heterogeneous investment groups; the fund managers are commonly restricted from investing according to their stated fund objectives. The fund objectives can be distinguished by the primary type of investing securities (i.e., equity funds, bond funds, money market funds, etc.) and/or by the announced investment style, strategy and philosophy (i.e., growth fund, cash fund, income fund, etc.). It is believed and tested that these fund objectives would result in groupings that have homogeneous within-group and heterogeneous between-group risks (e.g., Klemkosky, 1976; Starks, 1987, among others). Thus, MFs under different fund objectives would require separate frontiers for evaluation, and similarly efficiencies derived under different frontiers are not directly comparable.

To realize the efficiency comparison across different investment groups, this contribution aims at handling the heterogeneity of investment groups among the MFs. In comparing the performance where there exists technology heterogeneity among the firms, one particular solution initiated by Hayami & Ruttan (1970) is known as the metafrontier method. Under the framework of metafrontier, the firms in one technology group are evaluated under its own group frontier, but are also evaluated with respect to a metafrontier. The metafrontier is defined as the envelop of producible input-output combinations across all feasible technology groups. The efficiencies evaluated under the metafrontier are comparable across different groups. This so-called metafrontier approach has been amply applied in production studies across sectors and disciplines. Examples include agriculture (e.g., Latruffe et al., 2012), banking (e.g., Casu et al., 2013), fisheries (e.g., Lee & Midani, 2015), hotels (e.g., Huang et al., 2013), schools (e.g., Thieme et al., 2013), and wastewater treatment plants (e.g., Sala-Garrido, Molinos-Senante, & Hernández-Sancho, 2011) to name but a few.

Yet, to date, there has been limited attention devoted to the application of metafrontier in finance, let alone its application in solving the heterogeneity in MF evaluation. The only exception that we are aware of is the application in comparing the efficiencies of the Islamic equity funds across six investment regions (Makni et al., 2015). However, the metafrontier approach they applied follows the work of Battese et al. (2004) where a convexification strategy is adopted. This convexification strategy has recently been

criticized for being self-contradictory (Kerstens et al., 2019). To some extent, the convexification strategy destroys the very idea of distinguishing among different groups. Empirically, the convexification strategy also leads to statistically significant differences and contradictions on estimating the efficiencies (Kerstens et al., 2019) and the productivity indices (Jin et al., 2020).

Therefore, the first purpose of this contribution is to propose a general nonconvex metafrontier framework for realizing the performance comparison across different investment groups. Specifically, a group frontier is the boundary of one restricted investment group, where restrictions are derived from resource limitations, regulatory or other environmental constraints, as discussed above. Then, a common metafrontier that envelops all possible group frontiers is defined for comparing the efficiencies across different investment groups. The difference between two efficiencies estimates the alternative investment gap. Importantly, the metafrontier should be nonconvex even if the group frontier can be convex. The proposed metafrontier framework is in principle suitable for any categorization of MFs that calls for investigating the relative performance between two or more groups of MFs.

Second, both the diversified and convex (C) and nonconvex (NC) evaluation approaches are adopted to build a fairly complete set of variants for estimating the frontiers. The former is transposed from the modern portfolio theory which explicitly considers the diversification effect on the risk. The seminal article on a diversified approach is Murthi et al. (1997), followed by Morey & Morey (1999), which triggered a series of new developments in MF evaluation, including McMullen & Strong (1998) or Premachandra et al. (1998) (see Glawischnig & Sommersguter-Reichmann (2010) for an early overview). The latter proposal was launched by Kerstens et al. (2011). The authors proposed to analyse MFs via hedonic price models by analogy to the characteristics' approach to heterogeneous consumer goods. They argue that MFs can be trivially interpreted as financial products for which the investor pays a variety of fees (entry and exit loads, among others) to have access to a managed fund whose price distribution is characterised by its moments. To the best of our knowledge, the diversified approach under the metafrontier framework is proposed for the first time; the C and NC evaluation approaches under a NC metafrontier framework have not been adapted for fund evaluation. Moreover, in deference to mixed risk aversion preferences, all approaches are presented in a multi-moment setting, rather than in the classical mean-variance (MV) setting.

Third, this contribution aims at offering the first empirical analysis on comparing the financial performance of ethical MFs with that of non-ethical MFs under a metafrontier framework. The empirical results turn out to support that the ethical MFs could outperform the non-ethical MFs from a financial point of view. Moreover, regarding those opposite conclusions on whether satisfying the ethical need of a MF may damage its profitability, our empirical results give an alternative explanation. Lastly, our empirical analysis shows that the proposed metafrontier framework enables identifying the relative advantages of investment groups.

This contribution is structured as follows. In Section 2, a graphical example is used to clarify the intuitive idea of MF evaluation with a metafrontier. Then in Section 3, we explain the financial theory underpinning metafrontiers, and show how distances between the observed MF and the metafrontier can be decomposed into what we refer to as metatechnology differences and group-specific efficiency. In addition, both the efficiencies under the diversified approach and that under the C and NC evaluation frameworks are introduced. In Section 4, an empirical analysis is conducted to compare the efficiencies of ethical MFs with that of the non-ethical MFs. Section 5 summarizes the results and draws some conclusions.
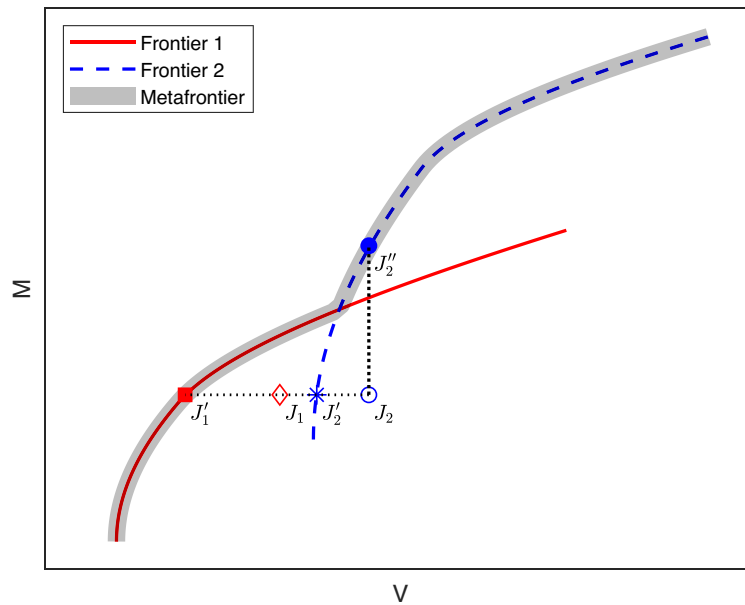
Fig. 1. Efficient frontiers and metafrontier under the MV case.

2. Metafrontier of mutual funds: a clarification

Before heading to the methodological part, it is essential to intuitively clarify the use of metafrontier in evaluating MFs. To fix our ideas, we start by an example where only two groups of MFs are available on the market, namely, investment group 1 and investment group 2. In Fig. 1, the red solid curve represents the efficient group frontier 1 and the blue dashed curve represents the efficient group frontier 2. The efficient group frontier indicates the best performance achievable when the fund managers are restricted to invest within their announced investment group. MF J_1, marked by a red diamond, is chosen from the investment group 1. MF J_2, marked by a blue circle, is chosen from the investment group 2. In the following, MFs J_1 and J_2 are used to explain the efficiency comparison across the investment groups.

The group-specific efficiency evaluates the performance under a specific group frontier. In Fig. 1, both MFs J_1 and J_2 are inefficient with respect to their own group frontier. The inefficient MF J_1 is projected onto the square J'_1 on the group frontier 1. Its group inefficiency is represented by the distance of $J'_1 J_1$. As for inefficient MF J_2, its group inefficiency is calculated by projecting it onto the asterisk J'_2 on the group frontier 2, and is represented by the distance of $J'_2 J_2$. Although $J'_1 J_1$ is larger than $J'_2 J_2$, it does not imply that MF J_1 performs worse than MF J_2. On the contrary, MF J_1 is observed to perform better than MF J_2, since it achieves the same level of mean return as MF J_2 but at a relatively lower risk. Thus, the group efficiencies derived from different group frontiers are not directly comparable. A direct comparison of group efficiencies under their own frontier may lead to confusing conclusions on their relative performance.

Let us now see how the metafrontier handles the efficiency comparison across different investment groups. The grey bold curve in Fig. 1 represents the metafrontier which provides a common frontier for encompassing all MFs from the two investment groups. It consists of part of the group frontier 1 and part of the group frontier 2. For both MFs J_1 and J_2, the projection on the metafrontier is J'_1. Therefore, the metafrontier inefficiencies of MFs J_1 and J_2 are represented by $J'_1 J_1$ and $J'_1 J_2$, respectively. Since $J'_1 J_1$ is smaller than $J'_1 J_2$, the comparison under the metafrontier concludes that MF J_1 is more efficient than MF J_2. This conclusion derived

from the metafrontier efficiencies coincides with the observation of their relative performance.

Except for reasonably comparing the MFs from different investment groups, the metafrontier framework also provides some insights on the choice of investment groups. The metafrontier inefficiency of MF J_2 consists of two parts. One part is the group inefficiency that evaluates the construction of MF within its own investment group. The other is the difference between the group (in)efficiency and metafrontier (in)efficiency. In Fig. 1, these two parts are represented by $J_2 J'_2$ and $J'_2 J'_1$, respectively. Within its original investment group 2, MF J_2 achieves the lowest risk at J'_2 while retaining its current mean value. A further risk reduction is only possible if MF J_2 changes to the investment group 1 where the lowest risk is achieved at J'_1. Therefore, if there exists an efficiency difference between the group (in)efficiency and metafrontier (in)efficiency, then the original investment group of the evaluated MF may not be the best choice to realize its targeted best performance.

While the existence of an efficiency difference always implies that the original investment group is not the best choice for realizing the targeted best performance, this implication works differently for investors of different types. We still take MF J_2 as an example. If MF J_2 corresponds to an investor, who has no preference for any investment group, then changing to the best-performing investment group is always possible and suits the need of realizing the targeted best performance, namely, J'_1. However, if MF J_2 corresponds to a fund manager or say an investor adhering to a specific investment group, then this change on the investment group may not be plausible. For a fund manager, most regulations require that the investment objectives must be adhered to and may only be changed with the approval of the shareholders by a majority vote (Najand & Prather, 1999). Similarly, for investors adhering to a specific investment group, i.e., the ethical group, changing the investment group is also not an option. In this case, the existence of an efficiency difference indicates that an alternative targeted best performance should be considered. For MF J_2, realizing its horizontal projection J'_1 requires a change of the investment group. However, if its vertical projection J''_2 is used as the targeted best performance, then there is no efficiency difference observed. J''_2, marked by a blue dot, locates on the group frontier 2, as well as on the

metafrontier. Hence, improving along J_2J_2'', the best performance is achieved without changing its original investment group.

3. Metafrontier methodology for mutual funds

In the following, we give the definitions of group frontier for the MFs from one specific investment group, as well as the definition of metafrontier to encompass all MFs from different investment groups. The frontiers are estimated with both diversified and C and NC evaluation models. Measures for characterizing the efficiencies and the efficiency difference are defined accordingly.

3.1. Group-specific frontier for certain group of mutual funds

In this subsection, we build the performance evaluation framework under a pre-announced investment group that is adapted in the next subsection to all available investment groups.

3.1.1. Group-specific diversified frontier

To evaluate the efficiency of MFs from the investment group τ, a diversified frontier is constructed. Specifically, given returns $R_{1,\tau}, \ldots, R_{n_\tau,\tau}$ of a collection of MFs from a specific investment group τ, a portfolio of funds is constructed from these n_τ MFs. The MFs are characterized by a set of moments, and normally by the expected return $E[R_{i,\tau}]$ for $i = 1, \ldots, n_\tau$ and the co-variance matrix $\Omega_{i,j,\tau} = Cov[R_{i,\tau}, R_{j,\tau}] = E[(R_{i,\tau} - E[R_{i,\tau}])(R_{j,\tau} - E[R_{j,\tau}])]$ for $i, j = 1, \ldots, n_\tau$. This contribution expands the focused moments into higher moments, such as the co-skewness matrix $CoS_{i,j,k,\tau} = E[(R_{i,\tau} - E[R_{i,\tau}])(R_{j,\tau} - E[R_{j,\tau}])(R_{k,\tau} - E[R_{k,\tau}])]$ for $i, j, k = 1, \ldots, n_\tau$ and the co-kurtosis matrix $CoK_{i,j,k,l,\tau} = E[(R_{i,\tau} - E[R_{i,\tau}])(R_{j,\tau} - E[R_{j,\tau}])(R_{k,\tau} - E[R_{k,\tau}])(R_{l,\tau} - E[R_{l,\tau}])]$ for $i, j, k, l = 1, \ldots, n_\tau$. For a more general class of moments, please refer to Briec & Kerstens (2010) which introduced the expressions in a portfolio selection setting.

Within the investment group τ, a portfolio of funds $\omega_\tau = (\omega_{1,\tau}, \ldots, \omega_{n_\tau,\tau})$ is represented by a vector of proportions invested in each of these n_τ MFs with $\sum_{i=1}^{n_\tau} \omega_{i,\tau} = 1$. If no shorting is allowed, then all $\omega_{i,\tau}$ are non-negative. But, this assumption can be easily relaxed. In general, the set of admissible funds portfolios is written as follows:

$$\Im_\tau = \left\{ \omega_\tau \in \mathbb{R}^{n_\tau} : \sum_{i=1}^{n_\tau} \omega_{i,\tau} = 1, \omega_{i,\tau} \geq 0, i = 1, \ldots, n_\tau \right\} \quad (1)$$

The return of portfolio ω_τ is then defined as $R_\tau(\omega_\tau) = \sum_{i=1}^{n_\tau} \omega_{i,\tau} R_{i,\tau}$. In order to characterize a portfolio of funds, different measures are incorporated including the moments of $R_\tau(\omega_\tau)$, the cost, etc. The cost can be different according to the underpinning MFs. The commonly used moments are calculated in a standard way as follows:

$$E[R_\tau(\omega_\tau)] = \sum_{i=1}^{n_\tau} \omega_{i,\tau} E[R_{i,\tau}], \quad (2)$$

$$V[R_\tau(\omega_\tau)] = E\left[\left(R_\tau(\omega_\tau) - E[R_\tau(\omega_\tau)] \right)^2 \right] = \sum_{i,j=1}^{n_\tau} \omega_{i,\tau} \omega_{j,\tau} \Omega_{i,j,\tau}, \quad (3)$$

$$Sk[R_\tau(\omega_\tau)] = E\left[\left(R_\tau(\omega_\tau) - E[R_\tau(\omega_\tau)] \right)^3 \right]$$
$$= \sum_{i,j,k=1}^{n_\tau} \omega_{i,\tau} \omega_{j,\tau} \omega_{k,\tau} CoS_{i,j,k,\tau}^\tau, \quad (4)$$

$$K[R_\tau(\omega_\tau)] = E\left[\left(R_\tau(\omega_\tau) - E[R_\tau(\omega_\tau)] \right)^4 \right]$$
$$= \sum_{i,j,k,l=1}^{n_\tau} \omega_{i,\tau} \omega_{j,\tau} \omega_{k,\tau} \omega_{l,\tau} CoK_{i,j,k,l,\tau}^\tau. \quad (5)$$

Eqs. (2)–(5) correspond to the expected return of a funds portfolio, its variance, skewness and its kurtosis, respectively.

To condense the notation for a funds portfolio, we introduce the function $\Phi_\tau : \Im_\tau \to \mathbb{R}^4$ defined by

$$\Phi_\tau(\omega_\tau) = \left(E[R_\tau(\omega_\tau)], V[R_\tau(\omega_\tau)], Sk[R_\tau(\omega_\tau)], K[R_\tau(\omega_\tau)] \right) \quad (6)$$

to represent its expected return, variance, skewness and kurtosis. In the remainder, the image by Φ_τ of a portfolio of funds is called a Mean-Variance-Skewness-Kurtosis (MVSK) point. Moreover, an arbitrary MVSK point v is denoted by its coordinates (v_M, v_V, v_S, v_K), where v_M, v_V, v_S and v_K are its mean, variance, skewness and kurtosis components, respectively. Correspondingly, the MVSK representation of the set $\Im_\tau$ is the image of Φ_τ on $(\Im_\tau)$:

$$\Phi_\tau(\Im_\tau) = \left\{ \Phi_\tau(\omega_\tau) : \omega_\tau \in \Im_\tau \right\}. \quad (7)$$

The above set can be extended by defining an MVSK *disposal representation set* through

$$\mathscr{DR}_\tau = \Phi_\tau(\Im_\tau) + (\mathbb{R}_- \times \mathbb{R}_+ \times \mathbb{R}_- \times \mathbb{R}_+)$$
$$= \left\{ (v_M, v_V, v_S, v_K) \in \mathbb{R}^4 : \exists \, \omega_\tau \in \Im_\tau, \right.$$
$$(v_M, -v_V, v_S, -v_K) \leq (E[R_\tau(\omega_\tau)], -V[R_\tau(\omega_\tau)],$$
$$\left. Sk[R_\tau(\omega_\tau)], -K[R_\tau(\omega_\tau)]) \right\}. \quad (8)$$

The boundary of this MVSK *disposal representation group set* is called a *group-specific diversified frontier*.

Before generalizing the diversified evaluation model, we introduce the shortage function under the investment group τ. Briec & Kerstens (2010) introduce a general procedure allowing for general higher moments in portfolio choice respecting a mixed risk aversion preference structure. These authors transpose the generalized shortage function to the multi-moment portfolio problem to account for a preference for odd moments (that need to be increased) and an aversion to even moments (that need to be reduced). This ability of the shortage function to seek for improvements in multiple directions simultaneously makes it an excellent tool for gauging financial product performances concurring with general investor preferences. Kerstens & Van de Woestyne (2011) show that a slight variation on the shortage function offers a more general method to handle negative data values which are common with financial data. Following Kerstens & Van de Woestyne (2011), the shortage function is defined as follows:

Definition 3.1. Let $g = (g_E, -g_V, g_S, -g_K) \in \mathbb{R}_+ \times (-\mathbb{R}_+) \times \mathbb{R}_+ \times (-\mathbb{R}_+)$. The shortage function $d_{g,\mathscr{DR}_\tau}$ for MF v in the direction of vector g under the investment set $\mathscr{DR}_\tau$ is defined as $d_{g,\mathscr{DR}_\tau}(v) = \sup\{\delta : v + \delta g \in \mathscr{DR}_\tau\}$.

The representation set $\mathscr{DR}_\tau$, defined by (8), can be directly used to compute the shortage function. The shortage function for a MF v_0 from the investment group τ, is denoted by $d_{g,\mathscr{DR}_\tau}(v_0)$. Consider this MF $v_0 = (v_{M,0}, v_{V,0}, v_{S,0}, v_{K,0})$ under evaluation, its shortage function can then be computed by solving the following program:

$$\max \; \delta_\tau$$
$$s.t. \; v_{M,0} + \delta_\tau g_E \leq E[R(\omega)]$$
$$v_{V,0} - \delta_\tau g_V \geq V[R(\omega)]$$
$$v_{S,0} + \delta_\tau g_S \leq S[R(\omega)]$$
$$v_{K,0} - \delta_\tau g_K \geq K[R(\omega)]$$

$$\sum_{i=1}^{n_\tau} \omega_i = 1, \omega_i \geq 0, i = 1, \ldots, n_\tau \qquad (9)$$

Making use of (2) to (5), the above program is rewritten as follows:

$$\max \quad \delta_\tau$$

$$s.t. \quad v_{M,0} + \delta_\tau g_E \leq \sum_{i=1}^{n_\tau} \omega_i E[R_i]$$

$$v_{V,0} - \delta_\tau g_V \geq \sum_{i,j=1}^{n_\tau} \omega_i \omega_j \Omega_{i,j}$$

$$v_{S,0} + \delta_\tau g_S \leq \sum_{i,j,k=1}^{n_\tau} \omega_i \omega_j \omega_k CoS_{i,j,k}$$

$$v_{K,0} - \delta_\tau g_K \geq \sum_{i,j,k,l=1}^{n_\tau} \omega_i \omega_j \omega_k \omega_l CoK_{i,j,k,l}$$

$$\sum_{i=1}^{n_\tau} \omega_i = 1, \omega_i \geq 0, i = 1, \ldots, n_\tau \qquad (10)$$

This diversified evaluation model provides a concrete evaluation of MF v_0. However, it faces a fundamental difficulty to evaluate a MF under a multi-moment setting: even for a small sample of MFs, the computational burden may be extremely high when adding higher moments. This computational problem inhibits the practical use of this diversified evaluation approach.

3.1.2. Convex and nonconvex group-specific evaluation frontiers

In this subsection, the nonparametric evaluation models developed under the framework of hedonic price theory are introduced to estimate the group frontier. We follow Kerstens et al. (2011) who list a variety of specification issues that have often been ignored in the early MF efficient frontier literature. Summarising their analysis, these authors argue convincingly that the most relevant returns to scale assumption when assessing MF with frontier models is to impose variable returns to scale (VRS). While most nonparametric frontier articles measuring MF performance impose convexity, these authors put forward some reasons to also consider nonconvexity. Based on their discussion, we examine both the C and NC VRS models.

To condense the notation for a MF v, let $x \in \mathbb{R}^m$ denote vectors of input-like measures (i.e., variance, kurtosis, cost, etc.) and let $y \in \mathbb{R}^s$ denote vectors of output-like measures (i.e., expected return, skewness, etc.). That is, $v = (x, y)$ where a MF is denoted by its input-like and output-like measures, respectively. Assume that N MFs $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ are observed, and that investment group τ is determined by $N_\tau \leq N$ of these observed MFs. To identify these particular MFs, consider the one-to-one index function ϕ_τ mapping the set $\{1, \ldots, N_\tau\}$ into the set $\{1, \ldots, N\}$. Then, $(x_{\phi_\tau(j)}, y_{\phi_\tau(j)})$ denotes the jth observation of τ. To illustrate these notations, consider the case where investment group τ is determined by the three observations (x_1, y_1), (x_3, y_3) and (x_8, y_8). Then, $N_\tau = 3$ and $\phi_\tau : \{1, 2, 3\} \to \{1, \ldots, n\}$ with $\phi_\tau(1) = 1$, $\phi_\tau(2) = 3$ and $\phi_\tau(3) = 8$.

For a specific investment group, a *group-specific investment set* is introduced to represent all admissible investments. If each group-specific investment set is C and exhibits VRS, then it is defined:

$$T_{\tau,C} = \left\{ (x, y) \in \mathbb{R}^m \times \mathbb{R}^s : \sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} x_{\phi_\tau(j)} \leq x, \right.$$

$$\left. \sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} y_{\phi_\tau(j)} \geq y, \sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} = 1, \lambda_{\phi_\tau(j)} \in \mathbb{R}_+ \right\} \qquad (11)$$

The boundary of the group-specific investment set is called a *group-specific efficient frontier*. When the disposability assumption applies, then $T_{\tau,C}$ is represented by the group-specific shortage function $d_{g,T_{\tau,C}}(v)$ as follows:

$$d_{g,T_{\tau,C}}(v) = \sup \left\{ \beta : (x + \beta g_X, y + \beta g_Y) \in T_{\tau,C} \right\} \qquad (12)$$

where $g = (g_X, g_Y) \in \mathbb{R}_- \times \mathbb{R}_+$ and $g \neq 0$.

The group-specific investment set $T_{\tau,C}$, defined by (11), can be directly used to compute (12). Consider a specific MF v_0 from the investment group τ under evaluation. The shortage function for this MF $v_0 = (x_0, y_0)$ is computed by solving the following program:

$$\max \quad \beta_{\tau,C}$$

$$s.t. \quad \sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} x_{\phi_\tau(j)} \leq x_0 + \beta_{\tau,C} g_X$$

$$\sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} y_{\phi_\tau(j)} \geq y_0 + \beta_{\tau,C} g_Y$$

$$\sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} = 1$$

$$\lambda_{\phi_\tau(j)} \geq 0 \qquad (13)$$

If each group-specific investment set is NC and exhibits VRS, then it is defined with the same N_τ observations as follows:

$$T_{\tau,NC} = \left\{ (x, y) \in \mathbb{R}^m \times \mathbb{R}^s : \sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} x_{\phi_\tau(j)} \leq x, \right.$$

$$\left. \sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} y_{\phi_\tau(j)} \geq y, \sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} = 1, \lambda_{\phi_\tau(j)} \in \{0, 1\} \right\} \qquad (14)$$

Similarly, when the disposability assumption applies, $T_{\tau,NC}$ is represented by the following group-specific shortage function:

$$d_{g,T_{\tau,NC}}(v) = \sup \left\{ \beta : (x + \beta g_X, y + \beta g_Y) \in T_{\tau,NC} \right\} \qquad (15)$$

where $g = (g_X, g_Y) \in \mathbb{R}_- \times \mathbb{R}_+$ and $g \neq 0$.

The group-specific investment set $T_{\tau,NC}$, defined by (14), can be directly used to compute (15). Consider a specific MF v_0 from the investment group τ under evaluation. The shortage function for this MF $v_0 = (x_0, y_0)$ is computed by solving the following program:

$$\max \quad \beta_{\tau,NC}$$

$$s.t. \quad \sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} x_{\phi_\tau(j)} \leq x_0 + \beta_{\tau,NC} g_X$$

$$\sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} y_{\phi_\tau(j)} \geq y_0 + \beta_{\tau,NC} g_Y$$

$$\sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} = 1$$

$$\lambda_{\phi_\tau(j)} \in \{0, 1\} \qquad (16)$$

Both the C evaluation model (13) and the NC evaluation model (16) are introduced in this subsection. Ideally, the evaluation models should provide a conservative estimation of the corresponding diversified models. In other words, we assume that ideally the evaluation model is a proper subset of the diversified model. Thus, the evaluation models should underestimate the diversified models if we are talking about inefficiencies (the distance to the frontier).

Assuming the diversified model is C (e.g., the two-dimensional MV case), then both the C and NC evaluation models provide a

reasonable approximation to it. Obviously, the C evaluation model is closer to the C diversified frontier than the NC evaluation model.

Assuming the diversified model is NC (e.g., the three-dimensional MVS case), then the NC evaluation model provides a reasonable approximation to it. However, in this case the C evaluation model may provide inefficiencies overestimating the diversified model (see Appendix A for an illustrative example), making it unsuitable for our approximation purpose.

To conclude, the NC evaluation model always provides a conservative approximation of the diversified models, while the C evaluation model only provides such a conservative approximation for the C diversified models.

3.2. Metafrontier for different groups of mutual funds

In practice, MFs can be distinguished and categorized into different groups. The set of available investment groups are recorded as Γ. In this subsection, the metatechnology and metafrontier are introduced to obtain comparable efficiencies for MFs from different investment groups.

3.2.1. Diversified metafrontier

Having the MVSK *disposal representation group set* $\mathscr{DR}_\tau$ defined for the investment group τ, the MVSK *disposal representation set* that represents the admissible funds portfolios under the whole set of available investment groups Γ is defined through

$$
\begin{aligned}
\mathscr{DR}_\Gamma &= \cup_{\tau\in\Gamma}\Phi(\mathfrak{I}_\tau) + (\mathbb{R}_- \times \mathbb{R}_+ \times \mathbb{R}_- \times \mathbb{R}_+) \\
&= \Big\{ (v_M, v_V, v_S, v_K) \in \mathbb{R}^4 : \exists\, \tau \in \Gamma \text{ and } \exists\, \omega_\tau \in \mathfrak{I}_\tau, \\
&\quad (v_M, -v_V, v_S, -v_K) \leq (E[R(\omega_\tau)], -V[R(\omega_\tau)], Sk[R(\omega_\tau)], \\
&\quad -K[R(\omega_\tau)]) \Big\}
\end{aligned} \tag{17}
$$

Obviously, we have that $\mathscr{DR}_\Gamma = \cup_{\tau\in\Gamma}\mathscr{DR}_\tau$.

Definition 3.2. Let $g = (g_E, -g_V, g_S, -g_K) \in \mathbb{R}_+ \times (-\mathbb{R}_+) \times \mathbb{R}_+ \times (-\mathbb{R}_+)$. The shortage function for MF v in the direction of vector g under the whole investment pool is defined as $D_{g,\mathscr{DR}_\Gamma}(v) = \sup\{\delta : v + \delta g \in \mathscr{DR}_\Gamma\}$.

Correspondingly, $D_{g,\mathscr{DR}_\Gamma}(v) = \max_{\tau\in\Gamma}\{D_{g,\mathscr{DR}_\tau}(v)\}$.

In this case, if the representation set $\mathscr{DR}_\Gamma$, defined by expression (17), is directly used to compute the shortage function, then a mixed integer non-linear program is required to be solved. Consider a specific MF v_0 from the investment group τ under evaluation. The shortage function for this MF, i.e. $D_{g,\mathscr{DR}_\Gamma}(v_0)$, is computed by solving the following program:

$$
\max \sum_\tau \delta_\tau
$$

$$
s.t. \quad \zeta_\tau v_{M,0} + \delta_\tau g_E \leq \zeta_\tau \sum_{i=1}^{n_\tau} \omega_i E[R_i] \qquad \forall \tau
$$

$$
\zeta_\tau v_{V,0} - \delta_\tau g_V \geq \zeta_\tau \sum_{i,j=1}^{n_\tau} \omega_i \omega_j \Omega_{i,j} \qquad \forall \tau
$$

$$
\zeta_\tau v_{S,0} + \delta_\tau g_S \leq \zeta_\tau \sum_{i,j,k=1}^{n_\tau} \omega_i \omega_j \omega_k CoS_{i,j,k} \qquad \forall \tau
$$

$$
\zeta_\tau v_{K,0} - \delta_\tau g_K \geq \zeta_\tau \sum_{i,j,k,l=1}^{n_\tau} \omega_i \omega_j \omega_k \omega_l CoK_{i,j,k,l} \qquad \forall \tau
$$

$$
\zeta_\tau \sum_{i=1}^{n_\tau} \omega_i = \zeta_\tau, \omega_i \geq 0, i = 1,\ldots,n_\tau \qquad \forall \tau
$$

$$
\sum_\tau \zeta_\tau = 1
$$

$$
\zeta_\tau \in \{0,1\} \qquad \forall \tau \tag{18}
$$

Obviously, the computational problem would inhibit the practical use of the diversified evaluation model (18).

3.2.2. Convex and nonconvex evaluation metafrontiers

The *metatechnology* Γ is the set of all investment groups that exist and are available for investments. The set of all input-like and output-like vectors that are feasible using a given metatechnology Γ (i.e., using some investment group that is contained in Γ) is labelled a *metatechnology investment possibilities set* (MIPS).

If every group-specific investment set is C, denoted as $T_{\tau,C}$, then the C MIPS is:

$$
T_{\Gamma,C} = \cup_{\tau\in\Gamma}T_{\tau,C}. \tag{19}
$$

When the disposability assumption applies, $T_{\tau,C}$ is represented by the following metatechnology shortage function:

$$
D_{g,T_{\Gamma,C}}(v) = \sup\left\{ \beta : (x + \beta g_X, y + \beta g_Y) \in T_{\Gamma,C} \right\} \tag{20}
$$

where $g = (g_X, g_Y) \in \mathbb{R}_- \times \mathbb{R}_+$ and $g \neq 0$.

Even though each group-specific investment set can be a C set, Kerstens et al. (2019) emphasize that the MIPS defined as the union of such sets is generally not convex. In this sense, assessing the metatechnology shortage function is equivalent to finding the maximum of the group-specific shortage function across all investment groups. This suggests implementing an enumeration algorithm where each step requires solving the linear program (13). Alternatively, Huang et al. (2013) proposed a mixed integer linear program (MILP) for calculating the input-oriented metafrontier efficiency which is effectively a compact statement of the enumeration algorithm. Recently, Afsharian & Podinovski (2018) show that the metafrontier efficiency can also be calculated by solving the following single linear program.

$$
\max \sum_\tau \beta_{\tau,C}
$$

$$
s.t. \quad \sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} x_{\phi_\tau(j)} \leq \zeta_\tau x_0 + \beta_{\tau,C} g_X \qquad \forall \tau
$$

$$
\sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} y_{\phi_\tau(j)} \geq \zeta_\tau y_0 + \beta_{\tau,C} g_Y \qquad \forall \tau
$$

$$
\sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} = \zeta_\tau \qquad \forall \tau
$$

$$
\sum_\tau \zeta_\tau = 1
$$

$$
\lambda_{\phi_\tau(j)} \geq 0 \qquad \forall j, \tau
$$

$$
\zeta_\tau \geq 0 \qquad \forall \tau \tag{21}
$$

If every group-specific investment set is NC, denoted as $T_{\tau,NC}$, then the NC MIPS is:

$$
T_{\Gamma,NC} = \cup_{\tau\in\Gamma}T_{\tau,NC}. \tag{22}
$$

Equivalently,

$$
\begin{aligned}
T_{\Gamma,NC} = \Big\{ &(x, y) \in \mathbb{R}^m \times \mathbb{R}^s : \sum_{\tau\in\Gamma}\sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} x_{\phi_\tau(j)} \leq x, \\
&\sum_{\tau\in\Gamma}\sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} y_{\phi_\tau(j)} \geq y, \\
&\sum_{\tau\in\Gamma}\sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} = 1, \lambda_{\phi_\tau f(j)} \in \{0,1\} \Big\}. \tag{23}
\end{aligned}
$$

When the disposability assumption applies, $T_{\Gamma,NC}$ is represented by the following metatechnology shortage function:

$$D_{g,T_{\Gamma,NC}}(v) = \sup \left\{ \beta : (x + \beta g_X, y + \beta g_Y) \in T_{\Gamma,NC} \right\} \qquad (24)$$

where $g = (g_X, g_Y) \in \mathbb{R}_- \times \mathbb{R}_+$ and $g \neq 0$.

The NC MIPS $T_{\Gamma,NC}$, defined by expression (23), can be directly used to compute the above shortage function. Consider that a specific MF v_0 from the investment group τ is under evaluation. The shortage function for this MF $v_0 = (x_0, y_0)$ is computed by solving the following program:

$$\max \quad \beta_{\Gamma,NC}$$

$$s.t. \quad \sum_{\tau \in \Gamma} \sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} x_{\phi_\tau(j)} \leq x_0 + \beta_{\Gamma,NC} g_X$$

$$\sum_{\tau \in \Gamma} \sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} y_{\phi_\tau(j)} \geq y_0 + \beta_{\Gamma,NC} g_Y$$

$$\sum_{\tau \in \Gamma} \sum_{j=1}^{N_\tau} \lambda_{\phi_\tau(j)} = 1$$

$$\lambda_{\phi_\tau(j)} \in \{0, 1\} \qquad (25)$$

3.3. Technical efficiencies and metatechnology difference

In this contribution, the metatechnology technical efficiency (MTE) of a MF in an investment group $\tau \in \Gamma$ is calculated from the shortage function as follows:

$$\text{MTE}_{g,\Lambda_\Gamma}(v) = 1 - D_{g,\Lambda_\Gamma}(v) \qquad (26)$$

where Λ_Γ corresponds to different meta sets under different frameworks. To be specific, $\Lambda_\Gamma \equiv \mathscr{DR}_\Gamma$ if a diversified evaluation framework is applied, thereby model (18) is adopted to calculate the $\text{MTE}_{g,\mathscr{DR}_\Gamma}(v)$. If a C evaluation framework is applied, then $\Lambda_\Gamma \equiv T_{\Gamma,C}$ and the corresponding $\text{MTE}_{g,T_{\Gamma,C}}(v)$ is calculated by solving model (21). If a NC evaluation framework is applied, then $\Lambda_\Gamma \equiv T_{\Gamma,NC}$ and the corresponding $\text{MTE}_{g,T_{\Gamma,NC}}(v)$ is calculated by solving model (25).

This MTE measure lies in the closed unit interval and indicates the maximum performance value that is achievable under some investment group $\tau' \in \Gamma$. A MF is perceived as being metafrontier-efficient when MTE = 1.

The associated measure of group-specific technical efficiency (GTE) is

$$\text{GTE}_{g,\Lambda_\tau}(v) = 1 - d_{g,\Lambda_\tau}(v) \qquad (27)$$

where Λ_τ corresponds to different group sets under different frameworks. To be specific, $\Lambda_\tau \equiv \mathscr{DR}_\tau$ if a diversified evaluation framework is applied, thereby model (10) is adopted to calculate the $\text{GTE}_{g,\mathscr{DR}_\tau}(v)$. If a C evaluation framework is applied, then $\Lambda_\tau \equiv T_{\tau,C}$ and the corresponding $\text{GTE}_{g,T_{\tau,C}}(v)$ is calculated by solving model (13). If a NC evaluation framework is applied, then $\Lambda_\tau \equiv T_{\tau,NC}$ and the corresponding $\text{GTE}_{g,T_{\tau,NC}}(v)$ is calculated by solving model (16).

A MF is perceived as being group-efficient if and only if its GTE measure is 1. If it is smaller than 1, then it implies that the performance can be improved by a better construction within the investment group τ.

If Γ contains more than one investment group, then there could be a difference between MTE and GTE. Relative to the whole investment group set Γ, the metatechnology difference (MD) of a MF characterized by v under the investment group τ is defined as:

$$\text{MD}_{g,\Lambda_\Gamma,\Lambda_\tau}(v) = D_{g,\Lambda_\Gamma}(v) - d_{g,\Lambda_\tau}(v) \qquad (28)$$

where $\Lambda_\Gamma \equiv \mathscr{DR}_\Gamma$ and $\Lambda_\tau \equiv \mathscr{DR}_\tau$ if a diversified evaluation framework is applied, $\Lambda_\Gamma \equiv T_{\Gamma,C}$ and $\Lambda_\tau \equiv T_{\tau,C}$ if a C evaluation framework is applied, and $\Lambda_\Gamma \equiv T_{\Gamma,NC}$ and $\Lambda_\tau \equiv T_{\tau,NC}$ if a NC evaluation framework is applied.

This measure also lies in the closed unit interval. It can be interpreted as a measure of whether the best investment group that is available has been selected. A MF is perceived as being efficient with respect to MD when MD = 0.

Finally, Eqs. (26), (27) and (28) imply that

$$\text{MTE}_{g,\Lambda_\Gamma}(v) = \text{GTE}_{g,\Lambda_\tau}(v) - \text{MD}_{g,\Lambda_\Gamma,\Lambda_\tau}(v). \qquad (29)$$

Hence, MTE can be decomposed into GTE and MD: the first measures how efficient an investor is operating as to the group-specific frontier, while the second measures how close a group-specific frontier is to the metafrontier. If MTE equals 1, then GTE = 1 and MD = 0.

In the following, the MV case where only two investment groups are available is employed to further illustrate the intuition underlying the metafrontier approach in evaluating the MF. Within each investment period, one can only choose one of these two groups to invest in.

Fig. 2 shows the investment sets and the estimated frontiers under the NC evaluation model. MFs $A_1, B_1, C_1, D_1, E_1, F_1, G_1$ and H_1, represented by the red squares, are from the investment group 1. The other 8 MFs ($A_2, B_2, C_2, D_2, E_2, F_2, G_2, H_2$), represented by the blue dots, are from the investment group 2. The estimated frontiers of two investment groups are represented by the red solid polyline and the blue dashed polyline, respectively. They are the group-specific frontiers estimated under the NC evaluation model. The metatechnology is the union of two group-specific investment sets. It generates a NC metafrontier which is represented by the grey bold polyline. This metafrontier consists of part of the NC frontier 1 and part of the NC frontier 2. Moreover, the metafrontier is slightly more determined by group 1, since 4 out of 7 metafrontier-efficient MFs belong to the investment group 1.

Let us first see the graphical representations of the proposed measures. Take MF H_1 in Fig. 2 as an example. Its proportional projection to the NC frontier 1 is $\bar{H}_1$, marked by the red asterisk. The distance between H_1 and $\bar{H}_1$ describes the measure $1 - $ GTE of MF H_1. The larger this distance is, the smaller the GTE gets and thus the MF is less group efficient. Meanwhile, the proportional projection of MF H_1 to the NC metafrontier is $\hat{H}_1$. The distance $H_1\hat{H}_1$ describes the measure $1 - $ MTE of MF H_1. Similarly, the larger this distance is, the smaller the MTE gets and thus the MF is less metafrontier efficient. Apparently for MF H_1, its MTE and GTE are not the same. The difference between MTE and GTE implies a positive MD for MF H_1 and it is represented by $\bar{H}_1\hat{H}_1$. As for MF H_2, its proportional projection onto the NC frontier 2 is $\hat{H}_2$, which is also the proportional projection onto the metafrontier. Hence, MTE = GTE holds for MF H_2 and its MD is 0.

Let us now read the implications of the proposed measures. The GTE measure evaluates the performance of MFs within the corresponding investment group. The values of GTE derived from different group frontiers are not comparable for assuring a better performance across the investment groups. In Fig. 2, although the GTE of MF H_1 is higher than that of MF H_2, it does not assure that MF H_1 outperforms MF H_2. It is the MTE that makes the performance comparison across the investment groups possible. Since the MTE of MF H_1 is larger than that of MF H_2, MF H_1 is perceived to perform relatively better than MF H_2. It is worth noting that although MF H_1 from the investment group 1 performs better with respect to MTE, it does not signify that there is a relative advantage of the investment group 1 over the investment group 2. On the contrary, MF H_1 represents a case where the relative advantage of the investment group 2 could be observed. This is because the best performance of MF H_1 under its own group frontier 1 could be further improved along $\bar{H}_1\hat{H}_1$ by investing in the investment group 2. That is, a positive MD reveals the relative advantage of one investment group over the other. If MD is zero, like in the case of MF H_2, then the corresponding investment group is already the best choice.
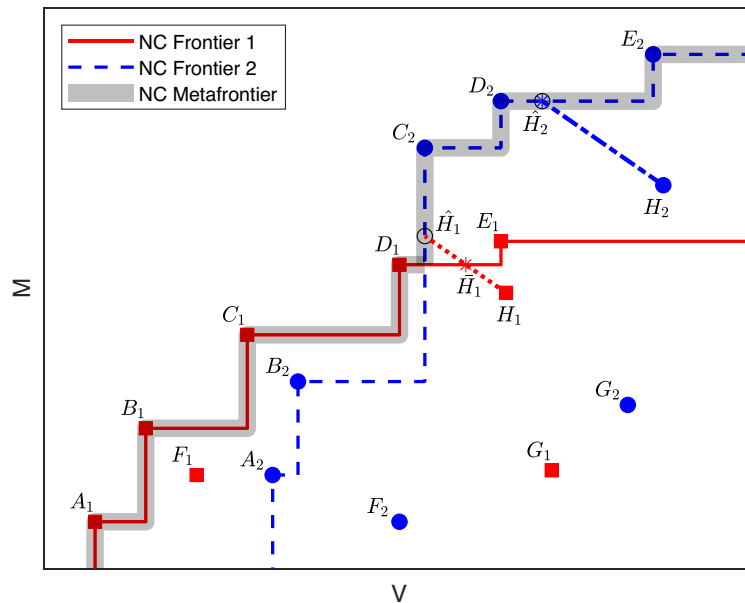
Fig. 2. NC group technologies and metatechnology under the MV case.

4. Empirical analysis

4.1. Data description

To illustrate how metafrontier can be used, we focus on a set of MFs belonging to 2 different groups: ethical MFs (MF group 1), which are restricted to invest in companies that meet some ethical requirements generally set out in the MF regulations, and non-ethical MFs (MF group 2), which are not subject to such a constraint.

The analysis is carried out on 445 open equity MFs with domicile in Western Europe, 227 ethical and 218 non-ethical MFs, chosen by matching their financial features and investment style. The source of data is the Bloomberg platform and we have considered the closing prices adjusted for dividends in the 5 year period 29/11/2013 to 30/11/2018.

As for the characterization of the MF returns, in this contribution we use a set of lower and higher order central moments, namely the mean and the variance but also the skewness and the kurtosis. Notice that most empirical studies in the literature just focus on the first two central moments. Specifically, for a MF j whose random returns are denoted by $R_{j,\tau}$, the calculations of its mean, variance, skewness and kurtosis are expressed as follows: $E(R_{j,\tau})$, $V(R_{j,\tau}) = E[(R_{j,\tau} - E(R_{j,\tau}))^2]$, $S(R_{j,\tau}) = E[(R_{j,\tau} - E(R_{j,\tau}))^3]$ and $K(R_{j,\tau}) = E[(R_{j,\tau} - E(R_{j,\tau}))^4]$. In addition, in the performance analysis we include also the MF costs, namely the entry and exit fees (front-end and back-end loads) and the annual management fees. These cost data can be directly downloaded from the Bloomberg platform. The descriptive statistics for these selected variables along with the comments are reported in Appendix B.

Considering both the higher moments and the investment costs, our empirical analysis is conducted under the framework of both C and NC evaluation models. The estimates of MTE, MD and GTE are calculated from Eqs. (26) to (28) under the C and NC settings, respectively. Specifically, for the measures characterizing a MF, the variance, kurtosis and costs are considered as input variables, and the mean, skewness are the output variables. In calculating the estimates, a general direction vector $g = (-|x|, |y|)$ is used for realizing a proportional interpretation that is convenient for practitioners. Note that $|x|$ and $|y|$ denote the vectors obtained by taking the absolute values of the individual components of input measures x and output measures y, respectively. To make a comparison of the contributions of two measures, a Li-test is applied. This test has been first proposed by Li (1996) and has been refined by Fan & Ullah (1999) and by others: one of the most recent developments is found in Li et al. (2009). This nonparametric test analyzes the differences between two distributions by comparing the differences between two kernel-based estimates of density functions. The null hypothesis suggests that the two distributions are equal.[2]

4.2. Empirical results

Before analyzing the details, a general impression on the distributions of the estimates for ethical MFs and those for non-ethical MFs is reported in Table 1. In Table 1, the three columns correspond to the Li-test results under the MV&Cost, MVS&Cost, MVSK&Cost cases, respectively. Horizontally, the first block of the rows contains the results under the C setting while the second block is under the NC setting. Within each of these two horizontal blocks, we report the results of the Li-test for MTE, GTE and MD, respectively. For example, at the cross of the C-MTE row and the MV&Cost column, the result tells if there is a significant difference between the MTE estimates of the ethical MFs and that of the non-ethical MFs under the MV&Cost case. The value "-0.0355" is the Tn value and the value "0.4735" in parentheses is the p-value. If the p-value is smaller than 1%, then the null hypothesis is rejected at the 1% significance level and the Tn value is marked with "***". If the p-value is smaller than 5%, then the null hypothesis is rejected at the 5% significance level and the Tn value is marked with "**".

From the Li-test results under the MV&Cost case in Table 1, 3 observations could be made. Note that these observations hold for both the C and NC settings. First, there is no statistical difference between the MTE measure of the ethical MFs and that of the non-ethical ones. To some extent, this implies that bearing an ethical interest does not harm the profitability of ethical MFs. Second, if a direct comparison of group efficiencies, namely GTE, is ap-

---

Table 1
The results of Li-test for the comparison between the ethical and non-ethical MFs.

		MV&Cost	MVS&Cost	MVSK&Cost
C	MTE	−0.0355	−0.0336	−0.0095
		(0.4735)	(0.4570)	(0.3620)
	GTE	0.5715***	0.0721	0.3343**
		(0.0055)	(0.1600)	(0.0320)
	MD	21.2776***	8.1462***	10.1051***
		(0.0000)	(0.0000)	(0.0000)
NC	MTE	−0.0454	0.0195	0.0497
		(0.5025)	(0.2760)	(0.2110)
	GTE	0.6572***	0.4590**	0.7301***
		(0.0000)	(0.0145)	(0.0005)
	MD	14.2993***	4.5628***	6.5538***
		(0.0000)	(0.0000)	(0.0000)

Table 2
Descriptive statistics for the estimates of MTE, GTE and MD: MV&Cost.

		MTE		GTE		MD	
		C	NC	C	NC	C	NC
All Obs.	# Eff. Obs.	15	72	20	97	240	267
	Mean	0.5158	0.6753	0.5411	0.7218	0.0253	0.0465
	Std. Dev.	0.2620	0.2658	0.2626	0.2568	0.0481	0.0987
	Min	0.0352	0.0500	0.0352	0.0563	0.0000	0.0000
	Max	1.0000	1.0000	1.0000	1.0000	0.5487	0.7377
	Li-test	8.1663***		10.6335***		3.9655***	
	p-value	(0.0000)		(0.0000)		(0.0000)	
Ethical	# Eff. Obs.	9	47	9	54	220	204
		(3.96%)	(20.70%)	(3.96%)	(23.79%)	(96.92%)	(89.87%)
	Mean	0.5069	0.6795	0.5080	0.6878	0.0011	0.0083
	Std. Dev.	0.2657	0.2782	0.2660	0.2814	0.0069	0.0389
	Min	0.0352	0.0563	0.0352	0.0563	0.0000	0.0000
	Max	1.0000	1.0000	1.0000	1.0000	0.0624	0.2968
	Li-test	3.6604***		3.8978***		0.1402**	
	p-value	(0.0000)		(0.0000)		(0.0340)	
Non-Ethical	# Eff. Obs.	6	25	11	43	20	63
		(2.75%)	(11.47%)	(5.05%)	(19.72%)	(9.17%)	(28.90%)
	Mean	0.5250	0.6708	0.5756	0.7572	0.0505	0.0864
	Std. Dev.	0.2585	0.2528	0.2552	0.2236	0.0585	0.1233
	Min	0.0424	0.0500	0.0555	0.0718	0.0000	0.0000
	Max	1.0000	1.0000	1.0000	1.0000	0.5487	0.7377
	Li-test	2.7236***		4.6002***		5.2850***	
	p-value	(0.0000)		(0.0000)		(0.0000)	

plied, the result suggests that the ethical MFs are performing significantly different from the non-ethical MFs. Obviously, this confirms that a direct comparison of GTE could lead to an opposite conclusion as to the comparison of MTE. Third, further improvement with respect to the choice of investment group exists for both the ethical and non-ethical MFs. Moreover, the comparison of the MD between the ethical and non-ethical MFs implies a significant difference. The same three observations can also be made under the MVSK&Cost case. For the MVS&Cost case, similar observations can be made, too. A slightly different observation is under the C setting, where the GTE of the ethical MFs and that of the non-ethical MFs do not show a significant difference.

In the following, detailed comparisons on the measures under the MV&Cost, MVS&Cost, MVSK&Cost cases are displayed in Tables 2 to 4 to give some further implications. These tables are structured in a similar way. Vertically, the two columns under MTE list the descriptive statistic results of the metafrontier efficiencies. The following two columns under GTE report the descriptive statistic results of group-specific efficiencies, and the last two columns under MD report the efficiency differences. A further distinction is related to whether the convexity is applied in the group-specific investment set or not: C indicates that the GTE is calculated under a C frontier, while NC indicates that the GTE is calculated under a NC frontier. Horizontally, the first block of rows contains the descriptive statistics for all 445 MFs. The second and third blocks correspond to the descriptive statistics for the 227 ethical MFs and the 218 non-ethical MFs, respectively. Within each of these three horizontal blocks, we report the results on the number of efficient observations, mean, standard deviation, minimum and maximum values of the corresponding estimates. Note that a MFs is MD-efficient if its MD equals 0. Lastly, a nonparametric Li-test is applied to test the null hypothesis that the distributions of the estimates under C and NC are equal.

Table 2 contains the descriptive statistics of the three estimates under the MV&Cost case. In the following, we will take the NC setting as an example to derive the main remarks regarding the comparison between the ethical and non-ethical MFs.

First, although more ethical MFs are group-efficient, we observe that the GTE of ethical MFs is on average lower than that of non-ethical MFs. In Table 2, 23.79% (54/227) of the ethical MFs are group-efficient, while only 19.72% (43/218) of the non-ethical MFs are group-efficient. However, the average GTE of ethical MFs is lower than that of the non-ethical MFs. The inefficient ethical MFs could be located more distant from the efficient ethical MFs, therefore, decreasing the average of GTE. This disparity in the group performance of the ethical MFs could be revealed by the standard deviation of GTE. The standard deviation of GTE for the ethical MFs is indeed higher than that for the non-ethical MFs.

Second, the results of MTE show that in the NC case the metafrontier is more determined by the ethical MFs. Among

Table 3
Descriptive statistics for the estimates of MTE, GTE and MD: MVS&Cost.

		MTE		GTE		MD	
		C	NC	C	NC	C	NC
All Obs.	# Eff. Obs.	16	135	25	161	253	310
	Mean	0.5307	0.7607	0.5598	0.7950	0.0291	0.0343
	Std. Dev.	0.2608	0.2576	0.2655	0.2460	0.0756	0.0807
	Min	0.0352	0.0563	0.0352	0.0563	0.0000	0.0000
	Max	1.0000	1.0000	1.0000	1.0000	0.8964	0.5089
	Li-test	16.4117***		17.3550***		1.9862***	
	p-value	(0.0000)		(0.0000)		(0.0000)	
Ethical	# Eff. Obs.	9	72	14	82	192	193
		(3.96%)	(31.72%)	(6.17%)	(36.12%)	(84.58%)	(85.02%)
	Mean	0.5212	0.7480	0.5443	0.7661	0.0231	0.0181
	Std. Dev.	0.2639	0.2760	0.2748	0.2778	0.0942	0.0563
	Min	0.0352	0.0563	0.0352	0.0563	0.0000	0.0000
	Max	1.0000	1.0000	1.0000	1.0000	0.8964	0.4118
	Li-test	6.1671***		5.7208***		−0.1297	
	p-value	(0.0000)		(0.0000)		(0.8910)	
Non-Ethical	# Eff. Obs.	7	63	11	79	61	117
		(3.21%)	(28.90%)	(5.05%)	(36.24%)	(27.98%)	(53.67%)
	Mean	0.5406	0.7739	0.5760	0.8250	0.0354	0.0511
	Std. Dev.	0.2579	0.2369	0.2551	0.2041	0.0489	0.0973
	Min	0.0424	0.0574	0.0570	0.0718	0.0000	0.0000
	Max	1.0000	1.0000	1.0000	1.0000	0.5138	0.5089
	Li-test	6.7403***		8.2528***		1.9862***	
	p-value	(0.0000)		(0.0000)		(0.0000)	

Table 4
Descriptive statistics for the estimates of MTE, GTE and MD: MVSK&Cost.

		MTE		GTE		MD	
		C	NC	C	NC	C	NC
All Obs.	# Eff. Obs.	45	143	55	170	259	308
	Mean	0.5790	0.7670	0.6044	0.8022	0.0253	0.0352
	Std. Dev.	0.2755	0.2584	0.2743	0.2456	0.0498	0.0836
	Min	0.0357	0.0563	0.0357	0.0563	0.0000	0.0000
	Max	1.0000	1.0000	1.0000	1.0000	0.5111	0.5089
	Li-test	11.1468***		12.6765***		1.5094***	
	p-value	(0.0000)		(0.0000)		(0.0000)	
Ethical	# Eff. Obs.	22	77	25	85	204	197
		(9.69%)	(33.92%)	(11.01%)	(37.44%)	(89.87%)	(86.78%)
	Mean	0.5662	0.7521	0.5765	0.7670	0.0103	0.0148
	Std. Dev.	0.2801	0.2773	0.2854	0.2783	0.0410	0.0527
	Min	0.0357	0.0563	0.0357	0.0563	0.0000	0.0000
	Max	1.0000	1.0000	1.0000	1.0000	0.3417	0.4118
	Li-test	4.11826***		4.2173***		−0.0066	
	p-value	(0.0000)		(0.0000)		(0.3695)	
Non-Ethical	# Eff. Obs.	23	66	30	85	55	111
		(10.55%)	(30.28%)	(13.76%)	(38.99%)	(25.23%)	(50.92%)
	Mean	0.5924	0.7826	0.6333	0.8389	0.0409	0.0564
	Std. Dev.	0.2707	0.2369	0.2596	0.2002	0.0534	0.1027
	Min	0.0424	0.0574	0.0570	0.0718	0.0000	0.0000
	Max	1.0000	1.0000	1.0000	1.0000	0.5111	0.5089
	Li-test	4.6086***		6.0496***		1.4135***	
	p-value	(0.0000)		(0.0000)		(0.0000)	

the ethical MFs, 20.70% (47/227) are metafrontier-efficient, while for the non-ethical MFs, only 11.47% (25/218) of them are metafrontier-efficient. Focusing on the MFs located on the metafrontier, 65.28% (47/72) of them are ethical while only 34.72% (25/72) are non-ethical. This implies that the metafrontier is more determined by the ethical MFs.

Third, the MD results of the ethical MFs imply that having an additional ethical constraint does not necessarily lead to a worse financial performance. Under the metafrontier framework, a positive MD suggests the existence of a group-related improvement. If the ethical constraint do harm the performance, then there should be frequent group-related improvements, i.e., positive MD. However, for the group-efficient ethical MFs, 87.04% (47/54) of them are already metafrontier-efficient. Their MD equals 0 and no group-

related improvement is possible. Only the remaining 12.96% of the group-efficient ethical MFs could further improve their performance if they choose to move to the non-ethical group. Considering both the group-efficient and group-inefficient ethical MFs, 89.87% (204/227) of them are already MD-efficient, implying that no group-related improvement is possible. Only 10.13% of the ethical MFs are actually dominated by non-ethical MFs. Thus, for most of the ethical MFs, their ethical constraint does not necessarily lead to a worse financial performance.

Fourth, as for the MD results of the non-ethical MFs, we observe that for the group-efficient non-ethical MFs, only 58.14% (25/43) of them are also metafrontier-efficient. Nearly half of the group-efficient non-ethical MFs are dominated by ethical MFs. Being dominated by ethical MFs implies that their best achievable

performance could be further improved by a better management of the fund, possibly drawing inspiration from the management of some ethical MFs. This group-related improvement is more obvious when considering all non-ethical MFs, since only 28.90% (63/218) of the non-ethical MFs have a MD of 0 while the remaining 71.10% have a positive MD.

In Table 2, all the estimates under the C setting are statistically different from those under the NC setting. Thus, it is always important to take into account the model features when choosing between the C and NC settings as discussed in Section 3.1.2. Although the estimates under the C setting are statistically different from those under the NC setting, the above four observations derived under the NC setting remain true for the C setting.

The analysis presented for the comparison between the ethical and non-ethical MFs is conducted also for the MVS&Cost and the MVSK&Cost cases. The descriptive statistics of the corresponding estimates are displayed in Tables 3 and 4, respectively. Similar observations as those discussed for the MV&Cost case can be made also for both the MVS&Cost and the MVSK&Cost cases. To sum up, the non-ethical MFs are generally performing better under their own frontier than the ethical MFs. However, in determining the metafrontier, the ethical MFs contribute more, since more metafrontier-efficient MFs comes from the ethical group. Moreover, the results of the MD measure show that the ethical constraint does not necessarily lead to a worse financial performance; quite the contrary, non-ethical investors could sometimes improve their performance by choosing MFs belonging to the ethical group.

5. Conclusion

This contribution introduces a general method for comparing MFs across various investment groups using the metafrontier framework. Comparing to the alternative financial ratios, the metafrontier framework can account for multiple input and output dimensions. The metafrontier efficiency of a MF is evaluated by looking for improvements in a space where all available investment groups are considered, regardless of the investment group it belongs to. The metafrontier efficiency could be decomposed into two components. One is the group-specific efficiency which is measured relative to its group frontier. The other is the metatechnology difference measuring the difference between the group-specific efficiency and the metafrontier efficiency. Exploring this allows to differentiate between the inefficiency within the investment group and the inefficiency from the choice of the investment group.

Both the diversified model and the C and NC (depending on the assumption made on convexity) evaluation models are adopted to approximate the "true" group-specific frontier within our metafrontier framework. Moreover, in deference to mixed risk aversion preferences, these approaches are also presented in a multi-moment setting, rather than in the classical MV setting. When considering higher moments like skewness and kurtosis, the diversified approach experiences some difficulties in solving the optimization program, although we can assume that its efficient frontier better approximates the "true" one. Neither the C nor the NC evaluation models present any computational difficulties, even including the higher moments. When a further choice is to be made between the C and NC evaluation models, then it is important to take into account the model features of the diversified models. The C evaluation model is preferred for approximating a C diversified frontier, even though the NC evaluation model also does the job. The NC evaluation model is preferred for approximating a NC diversified frontier. However, with respect to this NC diversified frontier, the C evaluation model may overestimate the corresponding NC diversified frontier.

Our empirical analysis on the comparison of ethical and non-ethical MFs illustrates the use of the proposed metafrontier framework. The empirical analysis is conducted under different settings, namely the MV&Cost, MVS&Cost and MVSK&Cost cases. The results obtained under the different settings are quite similar. In general, the empirical results show that the non-ethical MFs have a relatively better performance within their own investment group. However, the metafrontier turns out to be more determined by ethical MFs than by the non-ethical MFs. Moreover, investing in ethical MFs does not necessarily lead to a penalty on the financial performance, compared to the non-ethical MFs; on the contrary, the ethical MFs can even outperform the non-ethical MFs.

There remain some open challenges for future investigation. Analyzing the performance of MF, its persistence and stability over time is of increasing interest to academics and practitioners (Abdelsalam et al., 2014a; Kerstens et al., 2022). Thus, from a methodological perspective it is worthwhile to further extend the proposed metafrontier framework for better exploiting the time dimension. Empirically, a first interesting extension of the metafrontier analysis may consider encompassing an additional output variable measuring the ethical level (Basso & Funari, 2014a; 2014b) such that one can reward the ethical objectives pursued by socially responsible investors besides the financial performance. Another desirable extension is to change some of the empirical settings to check the robustness of our empirical conclusions. It can be interesting to test how the empirical conclusions will develop with different data sets and with different years of the sample. Moreover, instead of the purely financial matching used in our empirical setting, further advanced matching methods can be exploited for a (quasi-)experiment, e.g., nearest neighbour matching with a logistic regression-based propensity score as suggested by a referee (see, e.g., Ho et al., 2007; Stuart, 2010).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2023.07.019.

References

Abdelsalam, O., Duygun, M., Matallín-Sáez, J. C., & Tortosa-Ausina, E. (2014a). Do ethics imply persistence? The case of Islamic and socially responsible funds. *Journal of Banking and Finance, 40*, 182–194.

Abdelsalam, O., Fethi, M. D., Matallín, J. C., & Tortosa-Ausina, E. (2014b). On the comparative performance of socially responsible and Islamic mutual funds. *Journal of Economic Behavior and Organization, 103*, S108–S128.

Afsharian, M., & Podinovski, V. V. (2018). A linear programming approach to efficiency evaluation in nonconvex metatechnologies. *European Journal of Operational Research, 268*(1), 268–280.

Basso, A., & Funari, S. (2008). DEA models for ethical and non ethical mutual funds with negative data. *Mathematical Methods in Economics and Finance, 2*, 21–40.

Basso, A., & Funari, S. (2014a). Constant and variable returns to scale DEA models for socially responsible investment funds. *European Journal of Operational Research, 235*(3), 775–783.

Basso, A., & Funari, S. (2014b). DEA models with a constant input for SRI mutual funds with an application to European and Swedish funds. *International Transactions in Operational Research, 21*(6), 979–1000.

Basso, A., & Funari, S. (2016). DEA performance assessment of mutual funds. In Zhu J. (Ed.), *Data Envelopment Analysis: A handbook of empirical studies and applications* (pp. 229–287). Berlin: Springer.

Battese, G. E., Rao, D. S. P., & O'Donnell, C. J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis, 21*(1), 91–103.

Brandouy, O., Kerstens, K., & Van de Woestyne, I. (2015). Frontier-based vs. traditional mutual fund ratings: A first backtesting analysis. *European Journal of Operational Research, 242*(1), 332–342.

Briec, W., & Kerstens, K. (2010). Portfolio selection in multidimensional general and partial moment space. *Journal of Economic Dynamics and Control, 34*(4), 636–656.

Casu, B., Ferrari, A., & Zhao, T. (2013). Regulatory reform and productivity change in Indian banking. *Review of Economics and Statistics, 95*(3), 1066–1077.

Fan, Y., & Ullah, A. (1999). On goodness-of-fit tests for weakly dependent processes using kernel method. *Journal of Nonparametric Statistics, 11*(1), 337–360.

Glawischnig, M., & Sommersguter-Reichmann, M. (2010). Assessing the performance of alternative investments using non-parametric efficiency measurement approaches: Is it convincing? *Journal of Banking and Finance, 34*(2), 295–303.

Havemann, R., & Webster, P. (1999). Does ethical investment pay. EIRIS Report.

Hayami, Y., & Ruttan, V. W. (1970). Factor prices and technical change in agricultural development: The United States and Japan, 1880–1960. *Journal of Political Economy, 78*(5), 1115–1141.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*(3), 199–236.

Huang, C. W., Ting, C. T., Lin, C. H., & Lin, C. T. (2013). Measuring non-convex metafrontier efficiency in international tourist hotels. *Journal of the Operational Research Society, 64*(2), 250–259.

Jin, Q., Kerstens, K., & Van de Woestyne, I. (2020). Metafrontier productivity indices: Questioning the common convexification strategy. *European Journal of Operational Research, 283*(2), 737–747.

Kerstens, K., Mazza, P., Ren, T., & Van de Woestyne, I. (2022). Multi-time and multi-moment nonparametric frontier-based fund rating: Proposal and buy-and-hold backtesting strategy. *Omega, 113*, 102718.

Kerstens, K., Mounir, A., & Van de Woestyne, I. (2011). Non-parametric frontier estimates of mutual fund performance using C- and L-moments: Some specification tests. *Journal of Banking and Finance, 35*(5), 1190–1201.

Kerstens, K., O'Donnell, C., & Van de Woestyne, I. (2019). Metatechnology frontier and convexity: A restatement. *European Journal of Operational Research, 275*(2), 780–792.

Kerstens, K., & Van de Woestyne, I. (2011). Negative data in DEA: A simple proportional distance function approach. *Journal of the Operational Research Society, 62*(7), 1413–1419.

Klemkosky, R. C. (1976). Additional evidence on the risk level discriminatory powers of the Wiesenberger classifications. *The Journal of Business, 49*(1), 48–50.

Latruffe, L., Fogarasi, J., & Desjeux, Y. (2012). Efficiency, productivity and technology comparison for farms in Central and Western Europe: The case of field crop and dairy farming in Hungary and France. *Economic Systems, 36*(2), 264–278.

Lee, S.-G., & Midani, A. R. (2015). Comparison of efficiency levels using meta-frontier analysis of global fisheries for the period 1960–2010. *Fisheries Science, 81*(2), 247–254.

Li, Q. (1996). Nonparametric testing of closeness between two unknown distribution functions. *Econometric Reviews, 15*(1), 261–274.

Li, Q., Maasoumi, E., & Racine, J. S. (2009). A nonparametric test for equality of distributions with mixed categorical and continuous data. *Journal of Econometrics, 148*(2), 186–200.

Makni, R., Benouda, O., & Delhoumi, E. (2015). Large scale analysis of Islamic equity funds using a meta-frontier approach with data envelopment analysis. *Research in International Business and Finance, 34*, 324–337.

McMullen, P. R., & Strong, R. A. (1998). Selection of mutual fund using data envelopment analysis. *Journal of Business and Economic Studies, 4*(1), 1–14.

Morey, M. R., & Morey, R. C. (1999). Mutual fund performance appraisals: A multi--horizon perspective with endogenous benchmarking. *Omega, 27*(2), 241–258.

Murthi, B. P. S., Choi, Y. K., & Desai, P. (1997). Efficiency of mutual funds and portfolio performance measurement: A non-parametric approach. *European Journal of Operational Research, 98*(2), 408–418.

Najand, M., & Prather, L. J. (1999). The risk level discriminatory power of mutual fund investment objectives: Additional evidence. *Journal of Financial Markets, 2*(3), 307–328.

Premachandra, I. M., Powell, J. G., & Shi, J. (1998). Measuring the relative efficiency of fund management strategies in New Zealand using a spreadsheet-based stochastic data envelopment analysis model. *Omega, 26*(2), 319–331.

Sala-Garrido, R., Molinos-Senante, M., & Hernández-Sancho, F. (2011). Comparing the efficiency of wastewater treatment technologies through a DEA metafrontier model. *Chemical Engineering Journal, 173*(3), 766–772.

Starks, L. T. (1987). Performance incentive fees: An agency theoretic approach. *Journal of Financial and Quantitative Analysis, 22*(1), 17–32.

Statman, M. (2000). Socially responsible mutual funds. *Financial Analysts Journal, 56*(3), 30–39.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics, 25*(1), 1.

Thieme, C., Prior, D., & Tortosa-Ausina, E. (2013). A multilevel decomposition of school performance using robust nonparametric frontier techniques. *Economics of Education Review, 32*, 104–121.