



Título artículo / Títol article: On the informativeness of persistence for evaluating mutual fund performance using partial frontiers

Autores / Autors Soler Domínguez, Amparo ; Matallín Sáez, Juan Carlos ; Tortosa-Ausina, Emili

Revista: Omega, 2014, Volume 42, Issue 1

Versión / Versió: Pre-print

Cita bibliográfica / Cita bibliogràfica (ISO 690): MATALLÍN-SÁEZ, Juan Carlos; SOLER-DOMÍNGUEZ, Amparo; TORTOSA-AUSINA, Emili. On the informativeness of persistence for evaluating mutual fund performance using partial frontiers. *Omega*, 2014, vol. 42, no 1, p. 47-64.

url Repositori UJI: <http://hdl.handle.net/10234/86469>

On the informativeness of persistence for evaluating mutual fund performance using partial frontiers^{*}

Juan Carlos Matallín[†]

Amparo Soler[†]

Emili Tortosa-Ausina[‡]

March 9, 2013

Abstract

The last few years have witnessed a rapid evolution in the literature evaluating mutual fund performance using frontier techniques. The instruments applied, mostly DEA (Data Envelopment Analysis) and, to a lesser extent, FDH (Free Disposal Hull), are able to encompass several dimensions of performance, but they also have some disadvantages that might be preventing a wider acceptance. The recently developed order- m and order- α partial frontiers overcome some of the disadvantages (they are robust with respect to extreme values and noise, and do not suffer from the well-known curse of dimensionality) while keeping the main virtues of DEA and FDH (they are fully-nonparametric). In this article we apply not only the non-convex counterpart of DEA (FDH) but also order- m and order- α partial frontiers to a sample of US mutual funds. The results obtained for both order- m and order- α are useful, since a full ranking of the mutual funds' performance can be obtained. We merge these methods with the literature on mutual fund performance persistence. By combining the two literatures we derive an algorithm which establishes how the choice of m and α parameters intrinsic to order- m and order- α (respectively) relate to the existence of performance persistence and the contrarian effect.

Keywords: efficiency, mutual funds, partial frontiers, persistence

JEL Classification: F15, F21, F36, Z13

Communications to: Emili Tortosa-Ausina, Departament d'Economia, Universitat Jaume I, Campus del Riu Sec, 12071 Castelló de la Plana, Spain. Tel.: +34 964387168, fax: +34 964728591, e-mail: tortosa@uji.es

^{*}The authors acknowledge the financial support provided by the Ivie (Instituto Valenciano de Investigaciones Económicas). This study is part of research project ECO2011-27227/ECON supported by the Spanish Ministry of Science and Technology, P1-1B2009-54 by Fundació Caixa Castelló-Bancaixa and Universitat Jaume I, and PROMETEO/2009/066 by the Generalitat Valenciana. We are also grateful for comments from participants at the International Risk Management Conference (Florence, June 2010), the EURO XXIV (Lisbon, July 2010) and the International Conference on Operations Research (Munich, September 2010).

[†]Universitat Jaume I.

[‡]Universitat Jaume I and Ivie.

1. Introduction

Investors are increasingly interested in sound performance evaluation of available investment funds and, in this regard, they rely on risk-adjusted measures to make their choices. The development of mutual fund industries has given rise to a large body of literature. In this specific field, one issue of particular interest to investors, managers, and academics, and which has been extensively analyzed is, precisely, the performance of funds. From a methodological point of view, the existing literature dates back to Treynor (1965), Jensen (1968) and Sharpe (1966). Since these pioneering contributions, the literature has evolved to propose newer approaches to performance measurement. Some of them have been surveyed by Ippolito (1993), Grinblatt and Titman (1995), Cesari and Panetta (2002) or, in the particular field of hedge funds, Eling and Schuhmacher (2007).

In contrast to what we might call *traditional* approaches to mutual fund evaluation, since the late 1990s interest has been growing in applying the so-called frontier techniques, both parametric and nonparametric (see Murillo-Zamorano, 2004, for a survey) to evaluate the performance of mutual funds. The number of proposals, both from theoretical and empirical points of view, is now substantial, including Murthi et al. (1997), McMullen and Strong (1998), Morey and Morey (1999), Wilkens and Zhu (2001), Basso and Funari (2001), or Choi and Murthi (2001), among others. Indeed, due to the now remarkable number of proposals, some initiatives have been taken to review early contributions, such as Eling (2006), Glawischnig and Sommersguter-Reichmann (2010), or the monograph by Gregoriou and Zhu (2005) in the specific fields of hedge fund and commodity trading advisor (CTAs) performance evaluation.

These studies examine the advantages and disadvantages of applying nonparametric frontier techniques—mostly Data Envelopment Analysis (DEA)—to evaluate the performance of mutual funds.¹ The main advantage of these approaches is one of the features which has led to the popularity of DEA, i.e. the ability to deal simultaneously with several inputs and outputs, and to combine them in a single performance indicator—namely, the so-called *efficiency score*. This ability fits conveniently into the context of mutual fund performance evaluation, where one may be interested in extending the approach to include other dimensions apart from mean and variance, thus allowing the inclusion not only of skewness, but also of other relevant dimensions. DEA also has the ability to weight easily, by selecting the optimal weight for each dimension.

However, one disadvantage may have prevented some academics and practitioners from using DEA and related techniques such as Free Disposable Hull (FDH), namely, the so-called “curse of dimensionality”, which is related to problems associated with a low number of DMUs (Decision Making Units) relative to the number of input-output variables. This phenomenon not only affects both FDH and DEA estimators, but is also shared by other nonparametric approaches in statistics and econometrics. Although the issue was reported a while ago, few empirical applications have actually acknowledged its severity. However, some authors have taken the problem very seriously, claiming that “a number of applied papers using

¹Some recent contributions to the DEA literature include, Dulá and López (2013), Samoilenko and Osei-Bryson (2013), Lamb and Tee (2012), or Olesen and Petersen (2013), among others. In the particular case of applications of DEA to the mutual funds literature see, for instance, Pérez Gladish et al. (2007).

relatively small numbers of observations with many dimensions have appeared in the literature, but we hope that no more will appear” (Simar and Wilson, 2008, p.441). The curse of dimensionality severely affects those cases in which the number of inputs and outputs might vary, as well as the number of units under analysis.

From a theoretical point of view, the literature has evolved to provide solutions to the curse of dimensionality. The order- m (Cazals et al., 2002) and order- α (Daouia and Simar, 2007) estimators are robust indicators not only to the curse of dimensionality itself, but also to the presence of outliers and noise in the data, to which both DEA and FDH are particularly sensitive. Neither order- m nor order- α require convexity assumptions and, in addition, they both have several desirable properties that are useful for drawing inferences about efficiency. As Wheelock and Wilson (2009) indicate, while keeping the fundamental advantages of DEA and FDH (i.e., being fully-nonparametric), they overcome some of their shortcomings, since they are \sqrt{n} consistent, do not suffer from the curse of dimensionality, and are robust to outliers and noise.

However, empirical applications are still scarce. In the particular context of mutual fund performance evaluation, only Daraio and Simar (2005, 2006, 2007b) have considered not only FDH but also these robust methods. Although their theoretical contributions are highly valuable, they confine their analysis almost entirely to order- m estimators. In our paper, we update the contributions by Daraio and Simar (2005, 2006, 2007b) in several directions. First, we stretch the data to more recent dates, i.e., we focus on the period 2001–2011 whereas Daraio and Simar (2005, 2006, 2007b) consider sample periods for the early 2000s only; in addition, we have a much tighter focus on the application than on the specific details of the techniques. Second, our analysis is not confined to order- m techniques only. Taking into account the recent developments in the theoretical literature on efficiency and productivity analysis, we perform a comparison of *classical* approaches (FDH) with the new contributions, considering both Cazals et al.’s order- m estimators and Daouia and Simar’s (2007) order- α estimators. This robustness analysis has relevant implications, since the analyst (especially from a practitioner’s point of view) might be puzzled if different methodologies yield different results.

Related to this, one of the main contributions of the paper is to analyze the robustness, in terms of persistence, of results when applying partial frontiers. Should the applied methodologies be robust, it would be possible to forecast mutual fund efficiencies where they persist over time. Given that the partial frontiers methodologies provide us with funds’ rankings, the practitioner, or an individual investor, could use this information to buy the best (winner) and sell the worst (loser) funds. It would therefore be possible to evaluate which method is best able to discriminate between best and worst funds. This approach stands along with the large body of literature devoted to measuring whether certain fund managers consistently achieve higher (or lower) returns than their competitors. As one key component of the fund selection process, most individual investors and their advisors spend a significant amount of time studying historical performance of mutual funds, since it contains useful information about future performance, i.e. studying their *persistence*. As indicated by Droms (2006), “winners in one year tend to remain winners in the following year and losers have an even stronger tendency to remain losers” (Droms, 2006, p.60). This

particular topic has gained importance in the mutual fund performance evaluation literature, and several significant studies have been published since the early 1990s acknowledging this reality (see, for instance Grinblatt and Titman, 1992; Brown and Goetzmann, 1995; Carhart, 1997; Hendricks et al., 1993; Elton et al., 1996; Hendricks et al., 1993, among others). More recently, Pätäri (2009) has provided an extensive literature review of mutual fund performance persistence, and Cremers and Petajisto (2009) and Loon (2011) have proposed new methods to report evidence of persistence, and also on how investors respond to previous performance rankings.

As Pätäri (2009) points out, analysis of persistence is often sensitive to methodological choices, especially in the case of equity funds. These choices are either parametric or nonparametric methods that focus on the analysis of persistence as a static association between the performance of different time periods. To avoid this sensitivity to the method and provide enhanced robustness, our paper focuses on the economic relevance of mutual fund persistence rather than adopting a static approach. Related to this, Carhart (1997) proposed a framework in which the most relevant result is the economic value added of persistence. Following this approach, we will construct equally weighted portfolios that follow a buy-and-hold strategy based on the past efficiencies of mutual funds obtained using partial frontiers. This strategy will help guide investors' choices, based on the assumption that a good methodology to measure mutual fund efficiency is one that provides investment recommendations which, when followed, yield good results; in other words, a methodology that captures the *persistence* of managers' skills over time.

The paper is structured as follows. In section 2 we discuss the advantages and disadvantages of the most popular nonparametric techniques for efficiency measurement, namely, DEA and FDH, along with the new partial frontiers. Section 3 presents the underpinnings of the persistence analysis. Section 4 and 5 report the data and results, respectively. Section 6 concludes.

2. Mutual fund evaluation using frontier techniques

As noted above, the literature on the evaluation of mutual fund performance using frontier techniques has grown considerably. Apart from the nonparametric approaches referred to in the previous section, contributions have also come from the parametric field, where the most popular method is Stochastic Frontier Analysis, SFA (Lovell and Kumbhakar, 2000). These approaches have to specify a functional form for the frontier, and choose a distribution for the inefficiency. None of these requirements have to be met in the case of nonparametric frontier methods. Studies applying parametric frontier analysis methods to mutual funds include, for instance, Annaert et al. (2003), who considered stochastic Bayesian techniques. Although these approaches have several advantages, their drawbacks (not only having to specify a functional form for the frontier and distributions for the inefficiency, but also the assumption of independence for the inefficiency term) have led many authors to lean towards nonparametric methods.

Within the nonparametric field, we can distinguish between a theoretical view (Sengupta, 1991; Sengupta and Park, 1993; Briec and Kerstens, 2010) or a more applied perspective (apart from the references provided in the introduction, see also Sengupta, 2000). From a theoretical point of view, Sengupta (1991)

and Sengupta and Park (1993) provide links between the Capital Asset Pricing Model (CAPM) and nonparametric estimation of frontiers, whereas Briec and Kerstens (2010) analyze the relation between the hypothesis of the basic Markowitz (1952) model and efficiency analysis theory, by developing a dual framework for assessing the degree to which investors' preferences are satisfied. From a more applied perspective, the first specific application of DEA for evaluating the performance of mutual funds was Murthi et al. (1997), whose main motivation was to overcome the shortcomings of the *classical* two dimensional (mean-variance) performance measures.

A careful review of the literature assessing performance of traditional and alternative investment funds using DEA is provided by Glawischnig and Sommersguter-Reichmann (2010). Their survey implicitly suggests that the amount of studies applying nonparametric frontier methods such as DEA greatly outnumbers others using parametric methods. They conclude that DEA applications in the investment fund industry can be classified into two categories, namely, *traditional* and *alternative* fund performance evaluation studies. Their survey also implicitly recognizes that the studies applying FDH to mutual fund evaluation are virtually non-existent. The paper by Daraio and Simar (2006) is also mentioned in their survey, but only to briefly indicate that their approach was “computationally demanding” (Glawischnig and Sommersguter-Reichmann, 2010, p.297).²

2.1. Data Envelopment Analysis and Free Disposable Hull

Several measures of efficiency can be used comprehensively to carry out rigorous comparative efficiency analysis. The DEA efficiency score is a performance indicator obtained by comparing each mutual fund with the best performers of its objective group. The same underpinnings of DEA are shared by the FDH estimator. In his early proposals, Tulkens (1993) stressed the relevance of the main difference between DEA and FDH, namely, DEA rests on the hypothesis of convexity of the attainable set, whereas FDH does not. If the convexity hypothesis is questionable, DEA might be a wrong measure (i.e. statistically inconsistent). Both DEA and FDH allow definition of mutual fund performance indexes that can take into account different risk measures and the costs of investment (e.g., fees). Following Banker et al. (1984) few assumptions are required for DEA, among which we may highlight the convexity of the efficient frontier (convexity implies that any convex combination of inputs and outputs is feasible in the production function), and strong input and output disposability. The efficiency of a fund can be determined by the relative distance between the observed output and the efficient frontier. Thus, a fund is classified as inefficient if its output (e.g. return) and input (e.g., risk) are below the best practice frontier.

Banker and Maindiratta (1986) compared the advantages of DEA over parametric methods. In the context of mutual fund performance evaluation, DEA has the advantage of being a nonparametric analysis and, as such, does not require any theoretical model as a benchmark, such as the Capital Asset Pricing Model (CAPM) or the Arbitrage Pricing Theory (APT). Instead, DEA measures how well a fund performs relative to the best funds. Furthermore, it can address the problem of endogeneity of transaction costs in

²Thanks to the FEAR package for R by Paul W. Wilson this claim is no longer valid (Wilson, 2008). See also URL:<http://http://www.clemson.edu/economics/faculty/wilson/>, accessed October 2012.

the analysis by simultaneously considering expense ratios, turnover, and loads, as well as returns. Basso and Funari (2001) measured the efficiency of a sample of mutual funds between 1997 and 1999. Their contribution was to develop a generalized DEA-based performance measure that can integrate both classic performance measures (such as Sharpe, Treynor, and Jensen) and the approach of Murthi et al. (1997).

Another positive characteristic is that DEA evaluates the performance of a fund in reference to the best set of funds within the declared objective category. In Banker and Morey (1986) and Kamakura (1988) controllable categorical variables in the form of outputs are only treated as hierarchically ordered, e.g. outputs are classified in categories or similar orderings, according to attributes. Basso and Funari (2003) proposed a DEA categorical variable model in order to find an appropriate model to obtain an indicator of ethical fund performance. Fund performance is a combination of multiple fund attributes such as mean returns (outputs), risk (total or systematic) and expenses, and sometimes even fund size, turnover speed and minimum initial investment (inputs). Employing essentially basic DEA models like CCR (Charnes et al., 1978) or BCC (Banker et al., 1984), they sought to compare the efficiency of funds within a category or among several different categories of funds.

This nonparametric approach allows one to estimate an efficient frontier combining mean-variance and cost efficiency and to further estimate returns to scale for each mutual fund, implying that with DEA the effect of returns to scale on performance is controlled for (Choi and Murthi, 2001). The next advantage is that DEA measures efficiency with respect to the efficient frontier, which evaluates the best performance that can be achieved in practical terms. Another important point is the consideration that DEA provides an efficient index (the so-called *efficiency scores*) for each mutual fund, which enables calculation of the optimal weights for each attribute. And finally, DEA not only measures inefficiency, but also the magnitude of the inefficiency in the different dimensions. This is considered the greatest advantage of the DEA method over other approaches for measuring fund performance: namely that DEA reveals the reason why a fund is inefficient and shows how to restore the fund to its optimum level of efficiency. Choi and Murthi (2001) and Kuosmanen et al. (2006) argue that economic insights are provided by the slack variables in the optimization, as they indicate the extent to which each input can be reduced to achieve an efficiency score of one. Therefore DEA not only measures efficiency, but can also provide guidance as to how to improve the efficiency of inefficient funds.

In the first stage of the estimation process of this paper we evaluate the performance of the mutual funds in our sample considering the common non-convex FDH frontier. Although the preceding paragraphs have focused more closely on DEA, we have chosen its non-convex counterpart, i.e. the FDH frontier because of its higher flexibility and its asymptotic properties (Park et al., 2000). The set of attainable combinations of inputs and outputs, which delimits the frontier of the set of possibilities, must first be defined. To define the efficiency of a given fund we will then measure the distance between the observed value of the fund variables and the frontier. The Ψ set of possibilities is the set of attainable points (\mathbf{x}, \mathbf{y}) , defined as:

$$\Psi = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}_+^{p+q} | (\mathbf{x}, \mathbf{y}) \text{ are attainable}\} \quad (1)$$

where $\mathbf{x} \in \mathbb{R}_+^p$ is the vector of inputs and $\mathbf{y} \in \mathbb{R}_+^q$ is the vector of outputs. In the particular case of output efficiency scores on which we focus, we will characterize the production set Ψ considering output feasibility sets that we define for all $\mathbf{x} \in \mathbb{R}_+^p$. Hence, for all possible input values the set of possible values of \mathbf{y} is

$$Y(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}_+^q \mid (\mathbf{x}, \mathbf{y}) \in \Psi\} \quad (2)$$

In this particular setting, the Farrell (1957) measure of output-oriented efficiency in this output-oriented setting for a given mutual fund (\mathbf{x}, \mathbf{y}) is defined as

$$\tilde{\theta}(\mathbf{x}, \mathbf{y}) = \sup\{\theta : (\mathbf{x}, \theta\mathbf{y}) \in \Psi\} = \max\{\theta : \theta\mathbf{y} \in Y(\mathbf{x})\}, \quad (3)$$

where $\theta(\mathbf{x}, \mathbf{y}) \geq 1$ is the proportionate expansion of outputs required for a mutual fund with the input-output mix (\mathbf{x}, \mathbf{y}) to become efficient, i.e. to achieve the value of 1, since the efficient frontier corresponds to those funds whose $\theta(\mathbf{x}, \mathbf{y}) = 1$. The Farrell (1957) input-oriented efficiency score would be defined analogously.

If we were considering either DEA or FDH, the efficiency scores would be obtained by comparing to the full frontier of all funds, defining the maximum output that is technically feasible for a given level of inputs—recall we are adopting an output orientation. Therefore, in expression (3), $\theta(\mathbf{x}, \mathbf{y}) \geq 1$ is the proportionate expansion of outputs required for a mutual fund with the input-output mix (\mathbf{x}, \mathbf{y}) to become efficient, i.e. to achieve the value of 1, since the efficient frontier corresponds to those funds whose $\theta(\mathbf{x}, \mathbf{y}) = 1$. The Farrell (1957) input-oriented efficiency score would be defined analogously. The FDH estimator of Ψ , based on a sample of n observations $(\mathbf{x}_i, \mathbf{y}_i)$ is the free disposal closure of the reference set $\{(\mathbf{x}_i, \mathbf{y}_i) \mid i = 1, \dots, n\}$, and it can be defined as:

$$\hat{\Psi}_{\text{FDH}} = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq \mathbf{x}_i, \mathbf{y} \leq \mathbf{y}_i, i = 1, \dots, n\}. \quad (4)$$

2.2. Order- m and order- α estimators

The sensitivity of the deterministic DEA and FDH to measurement errors, outliers, sampling errors, and missing variables is an ongoing concern. In this regard, it is worth noting that the return data from financial markets are typically much more reliable and accurate than empirical production data usually studied with DEA. Therefore the problem of measurement error could seem *a priori* a less serious concern in the present context. However, the problem of outliers can actually occur in this setting if the return possibilities set includes assets that, for whatever reason, are infeasible investment alternatives for the fund manager. By careful modeling of the investment alternatives as well as the investment criteria and constraints facing the fund managers, the problem of outliers can be alleviated. For instance, Kuosmanen (2007) constructs benchmark portfolios directly from stocks and other assets, and his results indicate that heterogeneity of the evaluated funds was not obscuring the efficiency measures—although it could affect their ranking. Neither does sampling error seem to be a major problem: return data for stocks, bonds, and

other investment alternatives are available, and modeling the fund manager's entire investment universe is technically feasible. Moreover, the sampling theory of the DEA and FDH estimators is nowadays well understood and those insights can also be directly applied in the present context.

The sensitivity of both DEA and FDH to the presence of outliers is caused by the fact that both methods envelop all observation points quite closely. All the nonparametric envelopment estimators of frontiers are particularly sensitive to extreme observations, or outliers, which may disproportionately influence the evaluation of mutual fund performance. In addition, both DEA and FDH estimators, like other nonparametric measures, are affected by the curse of dimensionality due to their slow convergence rate (Simar and Wilson, 2008, p.441). As indicated in the introduction, the problem is especially severe when the number of inputs and outputs is high with respect to the sample size, or when the number of inputs and outputs is unclear. On this particular point, Tulkens (1993) also argues that, although sensitivity to outliers increased with the strength of the postulates made in constructing the production reference set, it is lower with FDH.

Taken together, the aforementioned problems may be serious enough to jeopardize the FDH estimates. To solve these problems, some additional procedures are required in order to make FDH estimates more robust. Several approaches have already been proposed in the literature. For instance, Wilson (1993, 1995) introduced descriptive methods to detect influential observations in nonparametric efficiency calculations. More recently, Cazals et al. (2002), Daraio and Simar (2005), Aragon et al. (2005) and Daouia and Simar (2007) have developed robust alternatives to the DEA and FDH estimators. Specifically, the nonparametric estimation order- m method developed by Cazals et al. (2002) is much more robust to both outliers and the curse of dimensionality. These authors introduced the concept of expected maximum output (or minimum input) frontier. It reflects a more realistic benchmark because it is constructed by comparing the performance of each fund (in terms of its use of inputs) not with the best performing funds of the group, but considering the *expected* value of the minimum level of inputs of m funds drawn from the distribution of funds with a level of output equal to or higher than that of the analyzed fund. The order- m also allows for statistical inference while keeping its nonparametric nature. We briefly describe this approach below.

Let us consider the conditional distribution function $F_{\mathbf{y}|\mathbf{x}}(\mathbf{y}_0|\mathbf{x}_0) = Pr(\mathbf{y} \leq \mathbf{y}_0|\mathbf{x} \geq \mathbf{x}_0)$. For a given level of inputs \mathbf{x}_0 in the interior of the support of \mathbf{x} , consider the m i.i.d. random variables $\{\nu_j\}_{j=1}^m$, drawn from the conditional distribution $F_{\mathbf{y}|\mathbf{x}}(\cdot|\mathbf{x}_0)$.³

Formally, the proposed algorithm (algorithm I) to compute the order- m estimator has the following steps:

1. For a given level of \mathbf{x}_0 , draw a random sample of size m with replacement among those \mathbf{y}_i , such that $\mathbf{x}_i \leq \mathbf{x}_0$.
2. Obtain the efficiency measures, $\tilde{\theta}_i$.

³Full technical details can be found in, for instance, Daraio and Simar (2007a).

3. Repeat steps 1 and 2 B times and obtain B efficiency coefficients $\tilde{\theta}_i^b (b = 1, 2, \dots, B)$. The quality of the approximation can be tuned by increasing B , but in most applications $B = 200$ seems to be a reasonable choice.
4. Compute the empirical mean of B samples as:

$$\bar{\theta}_i^m = \frac{1}{B} \sum_{b=1}^B \tilde{\theta}_i^b \quad (5)$$

As m increases, the number of observations considered in the estimation approaches the observed units that meet the condition $\mathbf{x}_i \geq \mathbf{x}_0$ and the expected order- m estimator in each one of the b iterations ($\tilde{\theta}_i^b$) tends toward the FDH estimator. Thus, m is an arbitrary positive integer value, but it is always convenient to observe the fluctuations of the $\tilde{\theta}_i^b$ coefficients depending on the level of m . For acceptable values of m , θ_i^m will normally present values higher than unity. When $\theta_i^m < 1$, the i unit can be labeled as superefficient (Andersen and Petersen, 1993). In addition, from an economic perspective, the order- m efficiency score has its own interest, since it does not provide the output-efficient frontier, but rather another reasonable benchmark value of the output for a fund with an \mathbf{x}_0 level of input: it is the *expected* value of the maximum level of output among a fixed number of m funds drawn from the population of funds with at most the same \mathbf{x}_0 level of input (Simar, 2003). Please note that these interpretations would correspond to an output orientation.

As indicated above, three aspects of the FDH methodology deserve special attention, namely, *efficiency by default*,⁴ the presence of *outliers* and the *curse of dimensionality*. In the absence of a sufficient number of *similar* mutual funds for a comparison, a particular fund is labeled as efficient *by default*. This efficiency ranking does not result from any effective superiority, but rather is due to the lack of information that would allow pertinent comparisons. In addition to this, both the DEA and FDH concepts of efficiency, by construction, apply both to the fund that presents the lowest level of inputs and to those with the highest values for at least one output indicator. This extreme form of the sparsity bias that characterizes the FDH technique ultimately leads to a lack of discrimination (i.e., an inability to *rank*) among production units, and constitutes a shortcoming of the FDH approach.

Regarding outliers, nonparametric frontiers are defined by the extreme values of the dimensional space of inputs and outputs. Therefore, the existence of outliers (atypical observations that differ significantly from the rest of the data) may considerably influence the estimation of efficiency, so it is important to verify that the divergences do not result from evaluation errors. Due to the trimming nature of the order- m frontier, this estimator does not envelop all the observed data points (even for large values of m) and, therefore, it is more robust to outliers and/or extreme values. Finally, order- m estimators are much less affected by the curse of dimensionality than either DEA or FDH because of some of their statistical properties—they are \sqrt{n} -consistent and asymptotically normal.

Apart from the order- m estimators, another family of partial frontiers has been proposed to overcome

⁴The reader interested in the details of this concept can consult, for instance, Tulkens (1993).

both the influence of outliers and the curse of dimensionality, namely, the order- α quantile frontiers (Daouia and Simar, 2007). The idea behind the order- α quantile-type frontier is to go the other way round, i.e., to determine the frontier by first fixing the probability $(1 - \alpha)$ of observing points below this order- α frontier. Therefore, the order- α quantile frontiers reverse the causation of order- m and choose the proportion of the data lying directly below the frontier.

Order- α estimators also have better properties than the usual nonparametric frontier estimators (either DEA or FDH). They are consistent estimators of the full frontier, since the “order” (in this case the α order) of the frontier is allowed to grow with sample size. They have also the advantage, shared with order- m , that the asymptotic properties are the same as those of FDH. But perhaps the main advantage, also shared with order- m , is that in finite samples, the new estimators do not envelop all the data, and they are therefore more robust to outliers than FDH or DEA. As indicated by Simar and Wilson (2008), they have the side benefit of detecting outliers (Simar and Wilson, 2008, p.480). The order- m ideas can easily be adapted to order- α quantile type-frontiers. The underpinnings of order- α were initially developed for the univariate case by Aragon et al. (2005) and extended to the multivariate setting by Daouia and Simar (2007), and are similar to those of quantile regression (Koenker, 2001).

Recall that in the context of order- m partial frontiers, a mutual fund operating at $(\mathbf{x}_0, \mathbf{y}_0)$ is benchmarked against the expected maximum output among m peers drawn randomly from the population of funds with input levels of at least \mathbf{x}_0 . In contrast, order- α quantile frontiers benchmark the mutual fund considered at $(\mathbf{x}_0, \mathbf{y}_0)$ against the output level not exceeded by $(1 - \alpha) \times 100\%$ of funds among the population of funds using less input than \mathbf{x}_0 .

Following Simar and Wilson (2008), for $\alpha \in (0, 1]$, the α -quantile output efficiency score for the mutual fund operating at $(\mathbf{x}_0, \mathbf{y}_0) \in \Psi$ can be defined as

$$\theta_\alpha(\mathbf{x}_0, \mathbf{y}_0) = \sup\{\theta|F_{\mathbf{y}|\mathbf{x}}(\theta\mathbf{y}_0|\mathbf{x}_0) > 1 - \alpha\} \quad (6)$$

Clearly, $\theta_\alpha(\mathbf{x}_0, \mathbf{y}_0)$ converges to the usual Farrell-Debreu output efficiency score $\theta(\mathbf{x}, \mathbf{y})$ (i.e. to the FDH estimator) when $\alpha \rightarrow 1$. As Daraio and Simar (2007a) point out, the order- α efficiency score has an interesting interpretation. In cases where $\theta_\alpha(\mathbf{x}_0, \mathbf{y}_0) = 1$, the fund is “efficient” at the level $\alpha \times 100\%$, since it is dominated by mutual funds using less input than \mathbf{x}_0 with probability $1 - \alpha$. In those cases where $\theta_\alpha(\mathbf{x}, \mathbf{y}) > 1$, the unit $(\mathbf{x}_0, \mathbf{y}_0)$ has to increase its output to achieve the output efficient frontier of level $\alpha \times 100\%$. Analogously to the case of the partial order- m frontiers, the case where $\theta_\alpha(\mathbf{x}, \mathbf{y}) < 1$ is feasible, indicating that a particular fund (\mathbf{x}, \mathbf{y}) can decrease its output by a factor $\theta_\alpha(\mathbf{x}, \mathbf{y})$ to reach the same frontier—a case in which this fund would be labeled as super-efficient with respect to the order- α frontier level. Finally, we can apply the plug-in principle to obtain an intuitive nonparametric estimator of $\theta_\alpha(\mathbf{x}, \mathbf{y}) = 1$ by replacing $F_{\mathbf{y}|\mathbf{x}}(\cdot|\cdot)$ with its empirical counterpart to obtain:

$$\hat{\theta}_{\alpha,n}(\mathbf{x}, \mathbf{y}) = \sup\{\theta|\hat{F}_{\mathbf{y}|\mathbf{x},n}(\theta\mathbf{y}|\mathbf{x}) > 1 - \alpha\} \quad (7)$$

Again, it is clear that when $\alpha \rightarrow 1$, then $\hat{\theta}_{\alpha,n}(\mathbf{x}, \mathbf{y})$ converges to the FDH input efficiency score.

As Daouia and Simar (2007) indicate, in practice the choice of the “tuning” parameters, both m and α , may be governed by their economic interpretation. Whereas in the case of order- m the benchmark could be the best of m virtual competitors, in the case of order- α it would be against a level of output with a probability $(1 - \alpha) \times 100\%$ of being dominated.

Regarding the choice of partial frontier estimator, Daouia and Simar (2007) conclude that both approaches (order- m and order- α) provide nonparametric estimators of the efficient frontier which are more robust than the usual envelopment estimators (like FDH/DEA estimators). It could be argued that the α -quantile approach is easier to interpret, since the parameter α is just the selected level of the quantile. The choice of the m parameter is more intricate, although in our particular setting it can be interpreted as the number of potential funds against which the benchmark is set to determine the performance of a particular fund. As Daouia and Simar (2007) indicate, although the choice of m can also be indirectly piloted by the percentage of observed funds staying above the frontier for a given m , the α -quantile approach seems to be more direct.

3. Partial frontiers and persistence analysis of mutual fund performance

3.1. Measuring the performance persistence of partial frontiers

In the previous sections we have proposed and described the order- m and order- α estimators to evaluate mutual fund efficiency. We now propose a method for testing the performance of these methods in guiding the selection of funds, i.e. how appropriate these methods are to choose among several investment alternatives. For this, we must evaluate the performance of each method itself (order- m and order- α) along with the choice of the corresponding *tuning* parameters, i.e., we will also evaluate how choosing different levels of m and α might influence the results. This would provide additional usefulness to these parameters, i.e. not only determining the number and importance of outliers (which is always difficult and, up to certain point, subjective), but also what its implications are when selecting among different funds.

In this task, we must bear in mind that an important use of mutual fund performance measures is to assess the possible value added by fund managers. Since the seminal contributions of Sharpe (1966) and Jensen (1968), many publications have contributed to the literature analyzing the efficiency of mutual funds. This literature has evolved hand in hand with the evolution of asset pricing models. Accordingly, nowadays it is common practice to the use multifactor models including those factors proposed by Fama and French (1993) and Carhart (1997). According to this literature, the return of the fund is compared with what one could expect depending on the asset pricing model considered. The gap between the two returns is usually attributed to the managers’ value added. Sharpe (1991, 1992) defines a general framework for evaluating fund efficiency, according to which the efficiency will be the gap between the fund’s performance and that yielded by a set of benchmarks or risk factors which reproduces the investment style of the same fund. In this particular framework, the fund’s results are thought to be those one might expect from

active management—i.e. by managers—whereas those obtained theoretically are only attributable to *passive* management. If a given manager provides value added to the mutual fund, its results should be better than those yielded by the simple and passive investment in some particular benchmarks.

However, even more important than the existence of value added today is the fact that we can expect this value to hold in the future. Indeed, both individual and institutional investors will be interested in a methodology for efficiency evaluation which helps them to choose those funds with the best result prospects. Accordingly, a broad literature focused on the analysis of mutual fund performance persistence has emerged over the last few years (see, for instance Carhart, 1997; Bollen and Busse, 2005). In other words, the attempt is not only to measure managers' (likely) value added, but also to analyze if it persists over time. Only if this is the case can a method to measure mutual fund efficiency offer useful guidance for investors.

Therefore, we will analyze the performance of FDH, order- m and order- α methods taking into account their ability to select efficient funds in the future. In fact, an appropriate methodology for measuring mutual funds' efficiency is one offering investment recommendations that, when followed, provide better performance; in other words, a methodology that captures the persistence of managers' skills over time. Following some relevant contributions in the performance persistence literature (see, for instance Carhart, 1997; Bollen and Busse, 2005) we will analyze whether order- m and order- α methods have the ability to rank funds according to their performance. Specifically, we construct equally-weighted portfolios (Grinblatt and Titman, 1992) that follow investment strategies based on the past efficiency of mutual funds using results yielded by FDH, order- m and order- α methods. Consecutively, at the end of each year portfolios are rebalanced, investing (selling) the best (worst) mutual funds according to the efficiency ranking obtained using either FDH, order- m or order- α methods.

Previous contributions have also considered isolating the performance persistence of best and worst past mutual funds, with the aim of analyzing possible asymmetries in managers' abilities to handle performance persistence. For instance, Carhart (1997), Lynch and Musto (2003) and Bollen and Busse (2005) show different levels of persistence for the best and worst mutual funds. The rebalancing of the portfolio and the approach of different quantile levels is a useful procedure in finance for making the fund more robust. For instance, Fama and French (1993), or Carhart's (1997) momentum factor are constructed following similar procedures.

In this line, in order to distinguish among the best and worst funds we consider different quantile levels for the top and bottom of the mutual fund efficiency ranking. Specifically, we consider strategies for the 5%, 10%, 20% and 40% of the mutual funds at the top and bottom of the ranking—according to the previous year's efficiency. We therefore build 96 portfolios for order- m and order- α methods (2 methods \times 8 quantiles \times 6 parameters) and 8 portfolios for FDH (one for each quantile at the top and bottom). The following step is to compute the daily return for each one of these 104 portfolios, and to measure their efficiency according to the method proposed in the following subsection.

3.2. Measuring strategy performance

Once the portfolios have been built according to the strategies described in the previous paragraphs, the next step is to analyze the results. If the manager effect *persists* over time, a good methodology for measuring mutual fund efficiency will be able to isolate that effect, thus leading to good results in the future. We should bear in mind that each portfolio is linked to an investment strategy on selected funds according to their past efficiency, which we estimate by selecting different parameters for the partial frontiers— m , in the case of order- m , and α , in the case of order- α .

The objective is now to evaluate the efficiency of these portfolios, which we could do via partial frontiers. However, this might entail an endogeneity problem, because this methodology involves evaluating the results of portfolios constructed considering the information obtained with the same methodology. In order to avoid this, we evaluate the portfolios with an “independent” method which has been intensely used by the financial literature on portfolio management.⁵ The multifactor linear model considered would be represented as:

$$r_{p,t} = \alpha_p + \beta_{p,m}r_{m,t} + \beta_{p,smb}r_{smb,t} + \beta_{p,hml}r_{hml,t} + \beta_{p,wml}r_{wml,t} + \varepsilon_{p,t} \quad (8)$$

where $r_{p,t}$ is the excess return of the portfolio. The coefficients on the right-hand side correspond to the Fama and French (1993) factors: excess market return ($r_{m,t}$), the return of small stocks minus the return of large stocks ($r_{smb,t}$), and the difference of returns between higher and lower book-to-market ratio stocks ($r_{hml,t}$). The remaining factor is the momentum factor, i.e. the return of past winners minus past losers ($r_{wml,t}$), proposed by Carhart (1997); $\varepsilon_{p,t}$ is the error term.

According to this model, the efficiency or performance of the portfolio is measured by the α_p parameter. If the model is used for evaluating mutual funds, when α_p is positive (negative) it implies that the fund yields higher (lower) return than that expected by the model, which can be interpreted as positive (negative) performance, i.e. the managers are (are not) able to *add value*. In our case, model (8) is used to evaluate the efficiency of those portfolios using investment strategies based on past efficiency, estimated using FDH, order- m and order- α methods. If these methods are able to reflect managers’ value added, and this persists over time, the corresponding portfolios will have efficiencies (as measured by α_p) significantly different from zero. In other words, the method to measure efficiency (either FDH, order- m or order- α) is a useful aid for the investor in selecting the mutual fund in which to invest.

In addition to measuring the efficiency of portfolios using α_p , we should also measure whether this is significant or not. The attempt is to distinguish whether the efficiency obtained is actually due to the investment strategy based on *past* rankings of mutual funds or, on the contrary, it is simply due to a spurious effect of a dynamic investment strategy. Therefore, for each quantile referred to in section 3.1 we generate 2,000 portfolios which follow a rebalanced and random strategy of investment in the mutual funds of the sample. For any portfolio chosen randomly we run model (8) and estimate a value for α_p .

⁵See, for instance, Kosowski et al. (2007), Kacperczyk and Seru (2007), Huij and Verbeek (2007), Fama and French (2010), Busse et al. (2010), or Barras et al. (2010), among others.

Then, the set of 2,000 α_p 's for each of the 8 quantiles is the expected distribution of efficiencies for any dynamic strategy that invests in the mutual funds of the sample. If there is persistence in managers' value added and it is captured by the FDH, order- m or order- α methods, then the efficiency (α_p) obtained by portfolios based on past efficiency will be outside the central 95% of this distribution. In other cases, the results for the portfolios will not be very different from those one might obtain randomly.

4. Data

4.1. Mutual fund sample

The empirical analysis used a sample of 1,450 US mutual funds from Morningstar. The sample period runs from March 1 2001, to May 31 2011. Mutual funds are grouped according to their investment styles as defined by the Morningstar Style Box. Previous studies such as Teo and Woo (2004) have also demonstrated the suitability of using the Style Box.

4.2. Inputs and outputs selection

As indicated in previous sections, one of the main benefits of using frontier techniques to evaluate the performance of mutual funds is their ability to handle multiple inputs and outputs in the model. As indicated by Basso and Funari (2001), the "DEA approach allows defining mutual fund performance indexes that can take into account several inputs and thus consider different risk measures (standard deviation, standard-semi deviation and beta) and redemption cost."

DEA, FDH or the partial frontiers order- m and order- α approaches may include other outputs apart from the traditional mean return measure. In computing their portfolio efficiency index Murthi et al. (1997) considered the standard deviation of returns, expense ratio, loads and turnover as inputs, and mean gross return as output. Choi and Murthi (2001) applied the same inputs and outputs as Murthi et al. (1997) although they adopted a different DEA formulation. Wilkens and Zhu (2001) performed their study with standard deviation and percentage of periods with negative returns as inputs, and mean return, minimum return and skewness as outputs. Chang (2004) proposed a new non-standard DEA formulation based on minimum convex input requirement set: the standard deviation, β , total assets and loads, while the output was the traditional mean return.

The *right* selection of inputs and outputs is crucial when using frontier techniques. Some investors might be more concerned with central tendencies (mean, standard deviation), while others may care more about extreme values (skewness, kurtosis). Briec et al. (2004) developed a quadratic-constrained (mean-variance) DEA model applying a mean-variance approach with variance as input and mean return as output. And Lozano and Gutiérrez (2007) proposed a quadratic-constrained DEA model consistent with Third-degree Stochastic Dominance (TSD) in order to obtain an optimal portfolio benchmark for any rational risk-averse investor. Briec and Kerstens (2009) present a quadratic program that extends the multi-horizon analysis by Morey and Morey (1999) in several ways. Joro and Na (2006) suggested a cubic-constrained mean-variance-skewness framework similarly to Briec et al. (2007), who consider

both skewness and mean return as outputs. In the particular case of higher moments, the literature is not entirely conclusive on the role of kurtosis, although some authors such as Guo et al. (2012) argue convincingly for its inclusion, based on the more effective performance evaluation for investors when including not only skewness but also kurtosis (as an input).

To apply our methodological approach we must therefore define some input and output variables. We consider the daily mean return (gross return, y_1) over the sample period as the main output. The other output, i.e. skewness (y_2) has also been computed from the daily returns distribution. Mutual fund daily net returns are computed by comparing the NAV (the net asset value of the fund) for daily dates and considering any distributed gain. Gross returns are subsequently estimated by adding daily fund expenses. As inputs, the risk of the fund (x_1) is measured by the standard deviation of the daily returns, as well as kurtosis (x_2), also computed from the daily returns. In some of the proposed models the management costs of the fund are also considered as input. In order to include these costs, we consider the fees the fund pays to managers as well as the loads, including fees and other costs incurred for operational management, e.g. for turnover. This variable is the expense ratio (x_3), and it is measured as percentage (average of the sample period) of costs over the managed portfolio size. Finally, each fund's beta (x_4) with respect to the market return is also included as input. Descriptive statistics for inputs and outputs are reported in Table 1.

4.3. Other financial data

To assess the performance of the partial frontier methods we apply model (8) to the portfolios that follow investment strategies based on past efficiency and to the random portfolios that are used to measure their significance. To apply this model a set of data on risk factors is needed. Specifically, we use the daily data of the returns of the three Fama and French (1993) and the Carhart (1997) momentum factor. This data—together with the data for the risk free asset, the one-month Treasury bill rate, to compute excess returns for mutual funds—were obtained from French's website (2011).⁶

5. Results

5.1. FDH, order- m and order- α efficiency measures

Table 2 reports summary statistics for FDH efficiency. Results are reported for all mutual funds evaluated jointly and also for different style categories. The joint evaluation, for all 1,450 mutual funds (for each one of the 11 years, totalling 15,950 observations), is reported in the last row of Table 2.

As one might expect, given the characteristics of FDH, the number of efficient funds is relatively high. This is partly suggested by the overall mean efficiency value, which indicates that the average fund has an efficiency value of 116.21%. This result indicates that the total amount of outputs could be expanded, on average, by 16.21% to catch up with the best practice funds. This result holds for all style categories, whose mean efficiencies range from 107.65%, for the index funds (which would be the most efficient),

⁶http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, accessed July 2011.

to 130.61%, for those funds with no style (which would be the most inefficient). When dividing for the different categories, the discrepancies are lower.

The relatively low discrepancies (which imply higher efficiencies) for the different style categories could have been caused by multi-modal distributions, with many efficient funds (in the vicinity of 1) and many inefficient funds, with a thinning in the middle. In order to control for this, Table 2 reports additional summary statistics of the distributions of efficiency so as to provide a more comprehensive view of the results. Although it could *a priori* seem remarkable that, for the first quartile backwards, several style categories are efficient (index, mid blend, mid growth and small blend), this trend is actually not surprising if one takes into account that FDH drops the convexity assumption of DEA which, in practical terms, implies that when a given fund cannot be compared with others because of their input/output combinations, it is classified as efficient *by default*. Therefore, the useful property of FDH of being more flexible than DEA comes at the cost of a lower ability to discriminate among efficient funds. All this implies that, despite the attractive asymptotic properties of FDH referred to in previous sections, and despite being more flexible than DEA, FDH has difficulties in both *discriminating* and, more importantly (especially in the context of mutual fund performance evaluation) in *ranking* units. As indicated earlier, it is very well suited to those contexts in which the analyst wants to ascertain the most obvious cases of inefficiency.

These patterns are more apparent in through the violin plots depicted in Figure 1.⁷ The first three sub-figures correspond to the large (Figure 1a), mid (Figure 1b) and small (Figure 1c) funds, whereas the last one corresponds to all funds, as well as the index and no style funds (Figure 1d). In this last case, it is patent that the funds in the index category are the most efficient, since the probability mass concentrates very close to unity. In contrast, for the funds in the “no style” category, probability mass stretches along the vertical axis, indicating not only lower efficiencies but also a great deal of heterogeneity. The sub-figures in Figure 1 also reveal more distinctly some features partly hidden in Table 2, such as the relatively fine performance of mid value funds, whose average efficiency is not among the best; however, the tight concentration of probability mass in the vicinity of one suggests homogeneous behavior within this sub-category of funds.

The order- m estimators overcome the difficulties of both FDH and DEA for ranking *efficient* funds—i.e. those with a value of 1. We have computed the order- m estimates for different values of m , from $m = 25$ to $m = 150$. These frontiers are nested and, therefore, for $m' > m$, the order- m frontier is below the order- m' frontier. Although the choice of the m parameter might seem somewhat arbitrary, it can be shown that the impact of the decision might not be so relevant when plotting the order- m efficiencies for different values of m , which usually show they are highly correlated. Indeed, the choice should not be complex if one follows Cazals et al. (2002), who suggest that “a few values of m could be used to guide the manager of the production unit to evaluate its own performance”. In addition, as Simar (2003) indicate, it is also important to notice the difference between m and n . Whereas m is a “trimming” parameter fixed

⁷Violin plots are a combination of a box plot and a kernel density plot, where the box plot is depicted in the center of the violin, and then a rotated kernel density plot is added to each side of the box plot—i.e. unfolded. The central point inside the box corresponds to the median.

at any desired level defining the level of the benchmark, n is the sample size and accordingly, there are no *a priori* links between m and n . This idea of trimming is not new in statistics, although its use in boundary estimation is.

Tables 3 and 4 report summary statistics for efficiencies estimated using order- m , for $m = 25$ and $m = 150$, respectively. Although computations were performed for more values of m , only results for $m = 25$ and $m = 150$ are reported, for reasons of space. Regardless of this m value, i.e. the trimming parameter which allows one to tune the percentage of points that lie above the order- m (partial) frontier (i.e. points whose efficiencies will be lower than 1), the mean is always higher than for the FDH case—here 99.51% and 110.18% for $m = 25$ and $m = 150$, respectively. This could suggest a superior ability of order- m to rank observations. For instance, for $m = 25$ (Table 3), the efficiencies corresponding to the first quartile are always different from 1—as opposed to the FDH case (Table 2).

One of the results reported in Tables 3 and 4 which may surprise the reader unfamiliar with partial frontiers is the presence of efficiency scores lower than 1. Although these cases of super-efficiency (Andersen and Petersen, 1993) are present regardless of the m parameter, most of the summary statistics displaying such values correspond to $m = 25$ (Table 3). This occurs because as m increases, the order- m estimator converges to the FDH estimator and, therefore, order- m efficiencies become more similar to FDH efficiencies and those cases below unity tend to disappear, as indicated by the results corresponding to $m = 150$ (Table 4). As explained in Simar (2003), for large values of m the two frontiers—FDH and order- m —coincide.

The discrepancies between results yielded by FDH and order- m are quite apparent when comparing the violin plots for the former method (Figure 1) with those for the latter (Figures 2 and 3). The ability to yield a full ranking using order- m methods, regardless of the trimming parameter considered, is shown through the different “violins” (density traces), whose probability mass is not as concentrated in the vicinity of 1 as in the FDH case (Figure 1). In addition, some interesting findings are corroborated, such as the homogeneous performance of the funds in the mid value style category (see Figures 2b and 3b), most of which are quite efficient—as shown by the high concentration of probability mass in the vicinity of 1.

In light of these results, although some readers might be tempted to reject these methods because an m parameter has to be selected somewhat arbitrarily, the order- m technique has the remarkable virtue of ranking the mutual funds with the best performance—i.e. to discriminate among *efficient* funds. Although results are not entirely coincidental for all values of m considered, the correlation is very high. In these circumstances, a suitable criterion for selecting funds could consist of selecting only those that are classified in a given percentile according to different values of m , which in practical terms implies making these robust methods even more robust. We will consider a different way to do this based on an analysis of persistence.

Results for the order- α partial frontier are reported in Tables 5 and 6 for $\alpha = .90$ and $\alpha = .99$, respectively, as well as Figures 4 and 5, for the violin plots.⁸ These tables and figures report analogous

⁸As in the order- m case, although results were also obtained for more values of the trimming parameter (α), we only

information to that reported in Tables 3 and 4, and Figures 2 and 3 in the case of order- m frontiers. In this case of order- α frontiers, the impact of the α parameter could seem *a priori* stronger, but this only occurs because the range of variation is not exactly equivalent to that chosen for m (Daouia and Gijbels, 2011). Therefore, as indicated in both Tables 5 and 6, the average efficiency for $\alpha = .90$ is quite high (actually, much lower than unity, 87.61%), which implies the presence of super-efficient funds whose efficiencies are remarkably low (recall that the output orientation implies that the most inefficient funds are those with the highest values). In addition, some of the results obtained for order- m , such as the homogeneous performance of mid value funds, is corroborated under order- α as shown by Figures 4b and 5b.

One of the main differences one may perceive between results for order- m and order- α is the great impact of α , which makes the order- α converge to FDH efficiencies much faster. But the same bottom line should apply for order- α results: those funds which perform better regardless of the α parameter considered should be the best candidates for selection by the investor. The analysis in the next section provides further insights on this point.

5.2. On the links between partial frontiers and persistence analysis for mutual fund efficiency

The preceding sections have estimated both FDH and partial frontiers (order- m and order- α) to evaluate mutual fund efficiency. In accordance with the description in Section 3, we will now analyze the informativeness of combining these methods with an analysis of persistence for fund efficiency. Specifically, we build portfolios that follow an investment strategy based on past efficiency, estimated by FDH, order- m and order- α methods. Subsequently, model (8) is used to assess the efficiency of these portfolios. If these methods are able to capture the value added by the mutual fund managers, and it persists over time, the corresponding portfolios will achieve a value for efficiency, as measured by model (8), significantly different from zero. Thus, we evaluate the performance of each method, focusing on how the choice of the tuning parameters— m in the case of order- m and α in the case of order- α —might influence the results on persistence.

Following our approach, we first split the sample period into 11 sub-periods. The first one runs from March 1 2001 to December 31 2001, the next nine sub-periods correspond to the following calendar years, and the last sub-period runs from January 1 2011 to May 31 2011. For each of the 11 periods the efficiencies were measured following the three methods considered, namely, FDH, order- m and order- α ; different tuning parameters (m and α) were also selected for the two partial frontier methods. Specifically, for both order- m and order- α we considered six possible parameters. Therefore, for each mutual fund and in each sub-period the efficiency is estimated 13 times—i.e. 2×6 parameters and methods, plus 1 for FDH, resulting into a total number of estimates for each fund of 13×11 . For each efficiency score estimated in each sub-period, and for each method-parameter combination, we form a ranking of the funds. Then, for each of these 13 method-parameter combination we build equally-weighted portfolios that invest in the

report two of the choices for space reasons.

mutual funds in the sample following an investment strategy based on past efficiency.

Specifically, there are 8 portfolios for each method-parameter combination, which invest in different quantiles from the ranking of the mutual funds—i.e. four in the selected top quantiles, and four in the selected bottom quantiles. The selected top and bottom quantiles correspond to the 5%, 10%, 20% and 40% of the upper and lower tails of the distribution of efficiencies, respectively. Therefore, we estimate 104 portfolios (13 method-parameter combinations \times 8 quantiles). These portfolios are *rebalanced* at the end of each sub-period according to the mutual fund efficiency ranking obtained from the past sub-period. Then we obtain 104 series of daily returns for the last ten sub-periods, i.e. from January 1 2001 to May 31 2011. Next we apply the financial model in (8) to these series in order to estimate the efficiency, α_p , for each portfolio, i.e. for each strategy and method-parameter combination.

In order to measure the significance of the efficiency of previous portfolios it is necessary to distinguish between the efficiency effectively obtained by an investment strategy based on past efficiency and that obtained simply due to a dynamic investment in the sample of mutual funds. To do this, we first form 2,000 equally-weighted portfolios for each of the eight quantiles listed above. These portfolios also invest in the mutual funds of the sample. The portfolios are built following the same procedure as that for the 104 portfolios, but now considering a random ranking for all the sub-periods. Second, we estimate the efficiency of these random portfolios by means of model (8). Then, for each of the analyzed quantiles, the 2,000 values obtained define a distribution of efficiency useful to test the significance of the efficiency of the portfolios that follow an investment strategy based on past efficiency. It will be significant if the efficiency of the portfolio lies outside the central 95% of this distribution.

Figure 6 shows the results for the portfolios built following an investment strategy based on past efficiency when it was estimated using FDH. The vertical axis measures the annualized efficiency estimated by model (8), whereas the horizontal axis represents the quantile analyzed. The two dashed-lines connect the efficiency values for those portfolios investing respectively in the top (triangles) or bottom (inverted triangles) funds of the 5%, 10%, 20% and 40% quantiles based on past FDH efficiency. The three solid lines connect the 95% percentile (crosses), the mean (circles), and the 5% percentile (inverted triangles) of the distribution efficiency of the random portfolios. It can be seen that, for most quantiles, following a strategy based on past efficiency does not yield very different results from those obtained from a random investment. However, for the 40% quantile, results are significant. In this case in which the random portfolios invest in 40% of the sample mutual funds, the efficiency distribution is more centered around the mean and, therefore, the results of the portfolios based on past efficiency lie in the regions of significance (outside the central 95% of the distribution). In other cases, for instance, investing randomly in 5% of the mutual funds, the results achieved are more disperse around the mean. Consequently, the results of the portfolios based on past FDH-efficiency lie inside the limits of the central 95% of the distribution of these random values, i.e. they could most likely have been achieved using a random investment strategy.

It is also interesting to note that the strategy of investing in the past top mutual funds yields worse results than investing in the bottom past mutual funds. This result is contrary to what one might expected if there were persistence. In other words, if past top (bottom) mutual funds lie at the top (bottom) in the

next periods, it would be expected that a portfolio that follows an investment strategy according to past efficiency will achieve best (worst) efficiency. Then, the results show how an opposite strategy may lead to improved results. There is a large body of literature on contrarian investment strategies (see, for instance, De Bondt and Thaler, 2012; Chan, 1988; De Haan and Kakes, 2011; Yao, 2012, among others), showing the performance of contrarian investors, defined as those that buy past losers and sell past winners. Our results support the evidence of a contrarian effect in the mutual funds for the sample period analyzed. Therefore, when the FDH method is used to measure the efficiency of mutual funds, it is a good predictor of this contrarian behavior, but it is only significant when the investment is in 40% of the worst or best mutual funds.

In this vein, Figures 7 and 8 show the results when the portfolios are built according to past efficiency measured by order- m and order- α , respectively. The interpretation of these figures is similar to Figure 6, but now the horizontal axis represents the selected trimming parameters, either for order- m or order- α , i.e. m and α . Each sub-figure in Figures 7 and 8 show the results for the portfolios built according to past efficiency that are investing in 5%, 10%, 20% and 40% quantiles of the best and worst mutual funds of the sample, respectively, for both methods.

The panels in Figure 7 show the results for efficiency as measured by model (8) of portfolios built according to past efficiency estimated using order- m . We find evidence of the contrarian effect. However, it is only significant for panels 7a and 7b, corresponding to the 5% and 10% quantiles, and for low values of order- m . In general, when the value of the m parameter increases, the significance is lower, except for panel 7d, corresponding to the 40% quantile where for $m = 150$ the efficiency achieved by the portfolios based on past efficiency is significant, following a similar pattern to that in Figure 6—where FDH is used to estimate past efficiency. This result could be partly expected, since efficiencies yielded by order- m converge to FDH as $m \rightarrow \infty$.

The panels in Figure 8 show the results for the order- α case. Interestingly, in all panels, as in Figures 6 and 7, the contrarian effect on the efficiency persistence of mutual funds emerges again. Furthermore, a comparison of the different panels shows how investing in smaller quantities provides more significant results. Indeed, panels 8a and 8b show how there are larger values outside the solid lines marking the border of significance, while in Figure 8d, corresponding to 40%, values lie inside the band and are closer to the mean line. Consequently, unlike the FDH case (Figure 6), the order- α method is more suited to capture the extreme behavior of investment funds. This indicates that persistence in mutual fund efficiency generally focuses in the extreme mutual funds, and not in the middle of the distribution of past efficiency. This evidence is similar to findings in the financial literature such as Carhart (1997), Lynch and Musto (2003), or Loon (2011), among others. Therefore, an interesting point to highlight is that the results yielded by partial frontiers, which maintain the advantages of both DEA and FDH (i.e. higher flexibility to handle several inputs and outputs) are consistent with the evidence found in the financial literature, even when we focus on studies that used other fund samples, periods and, especially, applied very different methodologies to measure mutual fund efficiency. In contrast, the panels in Figure 8 also show that the results based on past efficiency according to order- α are sensitive to the α parameter selected. In fact,

the efficiency on the vertical axis is improved for the higher values of α .

In brief, Figures 6, 7 and 8 show the efficiency, measured by the annualized intercept of the model (8), achieved by portfolios following an investment strategy based on past efficiency when it was estimated respectively by FDH, order- α and order- m methods. For all three figures, the values of the efficiencies yielded by these portfolios range approximately between -0.4% and $+0.7\%$ of excess annual return. This range does not imply a relevant economic impact. Furthermore, in some cases these results are not very different from those achieved by portfolios that follow a random investment strategy. In short, we do not find any clear evidence of persistence in the efficiencies of mutual funds that we might use to obtain abnormal significant returns. But, for the case of some of the selected parameters in the efficiency methods, we are able to capture a contrarian effect on the persistence of mutual fund efficiency.

6. Conclusions

The mutual fund industry is immersed in a process of continuous expansion and changes and, therefore, its analysis is gaining importance over time. A financial investor has to consider not only the actual number of competitors (i.e. the number of funds in the same objective category) but also the fact that it could vary up and down. Like other financial industries it is subject to expansion, acceleration and contraction cycles that greatly affect the performance of firms managing the assets. Interest comes from academics and investment industry participants alike. Since the traditional methodologies were initially proposed, many studies have been developed around the evaluation of mutual funds. In more recent years the importance of using nonparametric approaches for mutual fund performance evaluation has been stressed because of the key benefits they offer. These tools provide a single value of efficiency and have the great advantage of allowing one to include a high number of inputs and outputs in the model specified. Although these techniques, basically DEA and its non-convex sibling, FDH, are not free from disadvantages, the literature applying them to evaluate the performance of mutual funds is growing rapidly.

Some of the disadvantages of both DEA and FDH are the sensitivity to both outliers and the curse of dimensionality. This has been recognized by the literature (Dyson et al., 2001). In recent times, however, some methods have been proposed to overcome these pitfalls. The order- m (Cazals et al., 2002) and order- α (Daouia and Simar, 2007) partial frontiers are more robust both to the presence of outliers and the curse of dimensionality. Although they require selecting some parameters, which may be difficult, their advantages outweigh their disadvantages. In the specific case of mutual fund performance evaluation, they have notable ability to rank all mutual funds.

Yet applications of these techniques are still scarce. Only Daraio and Simar (2005, 2006, 2007b) have considered order- m techniques, although their contributions are theoretical. We extend their applications by considering not only order- m but also order- α partial frontiers. We also have a much tighter focus on the application. Specifically, we measure the performance of a sample of US mutual funds for the 2001–2011 period. Applying both order- m and order- α methods to our sample enables us to provide a full ranking. We can corroborate this through the violin plots. In contrast, FDH faces more difficulties

for discriminating among efficient funds. Although some readers might be puzzled by the fact that results vary for different m and α values, most funds rank very well regardless of the value of m or α considered. The results for the different types of funds indicate there are remarkable performance differences among them depending on the type of fund considered, which would beg for a more involved analysis exploring whether the differences among types of funds are actually significant or not.

We combine these methods and results with the literature on mutual fund performance persistence. Specifically, we propose a method for testing the performance of these methods in guiding the selection of funds, i.e. how appropriate order- m and order- α methods are to choose among several investment alternatives. This provides additional usefulness to these parameters (m and α), i.e. not only determining the number and importance of outliers, but also to ascertain which the implications might be for selecting different funds in terms of their results' prospects—i.e. in terms of the fund's *persistence*.

In other words, we analyze how FDH, order- m and order- α methods perform in terms of their ability for selecting efficient funds in the future. This is, precisely, what determines if a given methodology is appropriate or not, i.e. its power for offering investment recommendations that, when followed, provide better performance. Although similar attempts had been pursued by the literature on mutual fund performance persistence, there had been no initiatives to derive a guiding algorithm which combines this literature with that on partial frontier methods. In addition, this algorithm is strongly rooted in the literature on mutual fund performance persistence—specifically, its construction is based on the multifactor linear model.

Although the algorithm is useful, our results are relatively modest. Results range approximately in the $(-0.4\%, 0.7\%)$ of annualized excess return which, in addition, do not differ a great deal from those that could be obtained following a random investment strategy. This would imply that there is no conclusive evidence which could be used for obtaining abnormal *significant* returns. This modest result could be partly the result of conducting the analysis for all types of funds simultaneously. However, for some of the selected parameters, we do find evidence of a contrarian effect on mutual fund performance persistence.

References

- Andersen, P. and Petersen, N. C. (1993). A procedure for ranking efficient units in Data Envelopment Analysis. *Management Science*, 39(10):1261–1264.
- Annaert, J., Van den Broeck, J., and Vander Vennet, R. (2003). Determinants of mutual fund underperformance: a Bayesian stochastic frontier approach. *European Journal of Operational Research*, 151(3):617–632.
- Aragon, Y., Daouia, A., and Thomas-Agnan, C. (2005). Nonparametric frontier estimation: A conditional quantile-based approach. *Econometric Theory*, 21(2):358–389.
- Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science*, 30:1078–1092.
- Banker, R. D. and Maindiratta, A. (1986). Piecewise loglinear estimation of efficient production surfaces. *Management Science*, 32(1):126–135.
- Banker, R. D. and Morey, R. C. (1986). The use of categorical variables in Data Envelopment Analysis. *Management Science*, 32:1613–1627.
- Barras, L., Scaillet, O., and Wermers, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *Journal of Finance*, 65(1):179–216.
- Basso, A. and Funari, S. (2001). A Data Envelopment Analysis approach to measure the mutual fund performance. *European Journal of Operational Research*, 135(3):477–492.
- Basso, A. and Funari, S. (2003). Measuring the performance of ethical mutual funds: a DEA approach. *Journal of the Operational Research Society*, 54(5):521–531.
- Bollen, N. P. B. and Busse, J. A. (2005). Short-term persistence in mutual fund performance. *Review of Financial Studies*, 18(2):569–597.
- Briec, W. and Kerstens, K. (2009). Multi-horizon Markowitz portfolio performance appraisals: a general approach. *OMEGA*, 37(1):50–62.
- Briec, W. and Kerstens, K. (2010). Portfolio selection in multidimensional general and partial moment space. *Journal of Economic Dynamics and Control*, 34(4):636–656.
- Briec, W., Kerstens, K., and Jokung, O. (2007). Mean-variance-skewness portfolio performance gauging: A general shortage function and dual approach. *Management Science*, 53(1):135–149.
- Briec, W., Kerstens, K., and Lesourd, J. B. (2004). Nonparametric tests of portfolio investment efficiency: A shortage function generalization. *Journal of Optimization Theory and Applications*, 120(1):1–27.
- Brown, S. J. and Goetzmann, W. N. (1995). Performance persistence. *Journal of Finance*, pages 679–698.
- Busse, J. A., Goyal, A., and Wahal, S. (2010). Performance and persistence in institutional investment management. *Journal of Finance*, 65(2):765–790.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82.

- Cazals, C., Florens, J.-P., and Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106:1–25.
- Cesari, R. and Panetta, F. (2002). The performance of Italian equity funds. *Journal of Banking & Finance*, 26(1):99–126.
- Chan, K. C. (1988). On the contrarian investment strategy. *Journal of Business*, 61(2):147–163.
- Chang, K. P. (2004). Evaluating mutual fund performance: an application of minimum convex input requirement set approach. *Computers and Operations Research*, 31(6):929–940.
- Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429–444.
- Choi, Y. K. and Murthi, B. P. S. (2001). Relative performance evaluation of mutual funds: A non-parametric approach. *Journal of Business Finance & Accounting*, 28(7 & 8):853–876.
- Cremers, K. J. and Petajisto, A. (2009). How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies*, 22(9):3329–3365.
- Daouia, A. and Gijbels, I. (2011). Robustness and inference in nonparametric partial frontier modeling. *Journal of Econometrics*, 161(2):147–165.
- Daouia, A. and Simar, L. (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics*, 140:375–400.
- Daraio, C. and Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of Productivity Analysis*, 24:93–121.
- Daraio, C. and Simar, L. (2006). A robust nonparametric approach to evaluate and explain the performance of mutual funds. *European Journal of Operational Research*, 175(1):516–542.
- Daraio, C. and Simar, L. (2007a). *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and Applications*. Studies in Productivity and Efficiency. Springer, New York.
- Daraio, C. and Simar, L. (2007b). Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. *Journal of Productivity Analysis*, 28(1):13–32.
- De Bondt, W. F. M. and Thaler, R. H. (2012). Does the stock market overreact? *Journal of Finance*, 40(3):793–805.
- De Haan, L. and Kakes, J. (2011). Momentum or contrarian investment strategies: evidence from Dutch institutional investors. *Journal of Banking & Finance*, 35(9):2245–2251.
- Droms, W. G. (2006). Hot Hands, Cold Hands: Does Past Performance Predict Future Returns? *Journal of Financial Planning*, 19(5):60–69.
- Dulá, J. H. and López, F. J. (2013). DEA with streaming data. *OMEGA*, 41(1):41–47.
- Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., and Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132(2):260–273.

- Eling, M. (2006). Performance measurement of hedge funds using data envelopment analysis. *Financial Markets and Portfolio Management*, 20(4):442–471.
- Eling, M. and Schuhmacher, F. (2007). Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking & Finance*, 31(9):2632–2647.
- Elton, E. J., Gruber, M. J., and Blake, C. R. (1996). The persistence of risk-adjusted mutual fund performance. *Journal of Business*, 69(2):133–157.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.
- Fama, E. F. and French, K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance*, 65(5):1915–1947.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society, Ser.A*, 120:253–281.
- Glawischmig, M. and Sommersguter-Reichmann, M. (2010). Assessing the performance of alternative investments using non-parametric efficiency measurement approaches: Is it convincing? *Journal of Banking & Finance*, 34(2):295–303.
- Gregoriou, G. N. and Zhu, J. (2005). *Evaluating Hedge Fund and CTA performance: Data Envelopment Analysis Approach*. Wiley, Hoboken, NJ.
- Grinblatt, M. and Titman, S. (1992). The persistence of mutual fund performance. *Journal of Finance*, 47(5):1977–1984.
- Grinblatt, M. and Titman, S. (1995). Performance evaluation. In Jarrow, R., Maksimovic, V., and Ziemba, W. T., editors, *Handbooks in OR & MS: Finance*, volume 9, pages 581–609. Elsevier, Amsterdam.
- Guo, J., Ma, C., and Zhou, Z. (2012). Performance evaluation of investment funds with dea and higher moments characteristics: Financial engineering perspective. *Systems Engineering Procedia*, 3:209–216.
- Hendricks, D., Patel, J., and Zeckhauser, R. (1993). Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988. *Journal of Finance*, 48(1):93–130.
- Huij, J. and Verbeek, M. (2007). Cross-sectional learning and short-run persistence in mutual fund performance. *Journal of Banking & Finance*, 31(3):973–997.
- Ippolito, R. A. (1993). On studies of mutual fund performance, 1962–1991. *Financial Analysts Journal*, 49(1):42–50.
- Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *Journal of Finance*, 23:389–416.
- Joro, T. and Na, P. (2006). Portfolio performance evaluation in a mean–variance–skewness framework. *European Journal of Operational Research*, 175(1):446–461.
- Kacperczyk, M. and Seru, A. (2007). Fund manager use of public information: New evidence on managerial skills. *Journal of Finance*, 62(2):485–528.

- Kamakura, W. A. (1988). A Note on “The Use of Categorical Variables in Data Envelopment Analysis”. *Management Science*, 34(10):1273–1276.
- Koenker, R. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4):143–156.
- Kosowski, R., Timmermann, A., Wermers, R., and White, H. (2007). Can mutual fund “stars” really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance*, 61(6):2551–2595.
- Kuosmanen, T. (2007). Performance measurement and best-practice benchmarking of mutual funds: combining stochastic dominance criteria with data envelopment analysis. *Journal of Productivity Analysis*, 28(1):71–86.
- Kuosmanen, T., Cherchye, L., and Sipiläinen, T. (2006). The law of one price in data envelopment analysis: Restricting weight flexibility across firms. *European Journal of Operational Research*, 170(3):735–757.
- Lamb, J. D. and Tee, K.-H. (2012). Data Envelopment Analysis models of investment funds. *European Journal of Operational Research*, 216(3):687–696.
- Loon, Y. C. (2011). Model uncertainty, performance persistence and flows. *Review of Quantitative Finance and Accounting*, 36(2):153–205.
- Lovell, C. A. K. and Kumbhakar, S. C. (2000). *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge.
- Lozano, S. and Gutiérrez, E. (2007). Tsd-consistent performance assessment of mutual funds. *Journal of the Operational Research Society*, 59(10):1352–1362.
- Lynch, A. W. and Musto, D. K. (2003). How investors interpret past fund returns. *Journal of Finance*, 58(5):2033–2058.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.
- McMullen, P. R. and Strong, R. A. (1998). Selection of mutual funds using Data Envelopment Analysis. *Journal of Business and Economic Studies*, 4(1):1–14.
- Morey, M. R. and Morey, R. C. (1999). Mutual fund performance appraisals: a multi-horizon perspective with endogenous benchmarking. *OMEGA*, 27(2):241–258.
- Murillo-Zamorano, L. R. (2004). Economic efficiency and frontier techniques. *Journal of Economic Surveys*, 18(1):33–77.
- Murthi, B. P. S., Choi, Y. K., and Desai, P. (1997). Efficiency of mutual funds and portfolio performance measurement: a nonparametric measurement. *European Journal of Operational Research*, 98:408–418.
- Olesen, O. B. and Petersen, N. C. (2013). Imposing the regular ultra passum law in DEA models. *OMEGA*, 41(1):16–27.
- Park, B. U., Simar, L., and Weiner, C. (2000). The FDH estimator for productivity efficiency scores. *Econometric Theory*, 16(6):855–877.
- Pérez Gladish, B., Jones, D. F., Tamiz, M., and Bilbao Terol, A. (2007). An interactive three-stage model for mutual funds portfolio selection. *OMEGA*, 35(1):75–88.

- Pătări, E. J. (2009). Do hot hands warm the mutual fund investor? The myth of performance persistence phenomenon. *International Research Journal of Finance and Economics*, 34:117–139.
- Samoilenko, S. and Osei-Bryson, K.-M. (2013). Using Data Envelopment Analysis (DEA) for monitoring efficiency-based performance of productivity-driven organizations: Design and implementation of a decision support system. *OMEGA*, 41(1):131–142.
- Sengupta, J. K. (1991). Maximum probability dominance and portfolio theory. *Journal of Optimization Theory and Applications*, 71(2):341–357.
- Sengupta, J. K. (2000). *Dynamic and Stochastic Efficiency Analysis: Economics of Data Envelopment Analysis*. World Scientific Publishers, Singapore.
- Sengupta, J. K. and Park, H. S. (1993). Portfolio efficiency tests based on stochastic dominance and co-integration. *International Journal of Systems Science*, 24(11):2135–2158.
- Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, 39(1):119–138.
- Sharpe, W. F. (1991). The arithmetic of active management. *Financial Analysts Journal*, 47(1):7–9.
- Sharpe, W. F. (1992). Asset allocation: management style and performance measurement. *Journal of Portfolio Management*, 18(2):7–19.
- Simar, L. (2003). Detecting outliers in frontier models: A simple approach. *Journal of Productivity Analysis*, 20(3):391–424.
- Simar, L. and Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: Recent developments and perspectives. In Fried, H., Lovell, C. A. K., and Schmidt, S. S., editors, *The Measurement of Productive Efficiency*, chapter 4, pages 421–521. Oxford University Press, Oxford, 2nd edition.
- Teo, M. and Woo, S. J. (2004). Style effects in the cross-section of stock returns. *Journal of Financial Economics*, 74(2):367–398.
- Treynor, J. L. (1965). How to rate management of investment funds. *Harvard Business Review*, 43(1):63–75.
- Tulkens, H. (1993). On FDH efficiency analysis: Some methodological issues and applications to retail banking, courts, and urban transit. *Journal of Productivity Analysis*, 4:183–210.
- Wheelock, D. C. and Wilson, P. W. (2009). Robust nonparametric quantile estimation of efficiency and productivity change in U.S. commercial banking, 1985–2004. *Journal of Business and Economic Statistics*, 27(3):354–368.
- Wilkins, K. and Zhu, J. (2001). Portfolio evaluation and benchmark selection. *Journal of Alternative Investments*, 4(1):9–19.
- Wilson, P. W. (1993). Detecting outliers in deterministic nonparametric frontier models with multiple outputs. *Journal of Business and Economic Statistics*, 11(3):319–23.
- Wilson, P. W. (1995). Detecting influential observations in Data Envelopment Analysis. *Journal of Productivity Analysis*, 6(1):27–45.

- Wilson, P. W. (2008). FEAR: A software package for frontier efficiency analysis with R. *Socio-Economic Planning Sciences*, 42(4):247–254.
- Yao, Y. (2012). Momentum, contrarian, and the January seasonality. *Journal of Banking & Finance*, 36(10):2757–2769.

Table 1: Descriptive statistics for inputs and outputs, mutual funds (2001–2011)

		Mean	Std.dev.
Index funds	Std.dev. daily return (x_1)	0.012800	0.005800
	Kurtosis daily returns (x_2)	1.150800	1.903690
	Expense ratio (x_3)	0.005016	0.002853
	Beta (x_4)	1.044612	0.138546
	Gross return (y_1)	0.000296	0.000787
	Skewness daily returns (y_2)	-0.046711	0.263136
Large blend	Std.dev. daily return (x_1)	0.011762	0.005735
	Kurtosis daily returns (x_2)	1.522866	3.340281
	Expense ratio (x_3)	0.011362	0.004842
	Beta (x_4)	0.944727	0.179112
	Gross return (y_1)	0.000277	0.000796
	Skewness daily returns (y_2)	-0.059292	0.390656
Large growth	Std.dev. daily return (x_1)	0.012621	0.005798
	Kurtosis daily returns (x_2)	1.319098	2.906607
	Expense ratio (x_3)	0.012407	0.004875
	Beta (x_4)	1.021110	0.207333
	Gross return (y_1)	0.000253	0.000903
	Skewness daily returns (y_2)	-0.004013	0.404106
Large value	Std.dev. daily return (x_1)	0.011626	0.005918
	Kurtosis daily returns (x_2)	1.504187	3.319983
	Expense ratio (x_3)	0.011292	0.003933
	Beta (x_4)	0.921112	0.153864
	Gross return (y_1)	0.000311	0.000745
	Skewness daily returns (y_2)	-0.102307	0.370758
Mid blend	Std.dev. daily return (x_1)	0.013072	0.006377
	Kurtosis daily returns (x_2)	1.613086	5.528361
	Expense ratio (x_3)	0.012994	0.005265
	Beta (x_4)	1.023785	0.270429
	Gross return (y_1)	0.000422	0.000945
	Skewness daily returns (y_2)	-0.095342	0.573141
Mid growth	Std.dev. daily return (x_1)	0.013402	0.005666
	Kurtosis daily returns (x_2)	1.228198	3.704283
	Expense ratio (x_3)	0.013501	0.004080
	Beta (x_4)	1.065699	0.230991
	Gross return (y_1)	0.000385	0.000981
	Skewness daily returns (y_2)	-0.058791	0.391875
Mid value	Std.dev. daily return (x_1)	0.012492	0.006058
	Kurtosis daily returns (x_2)	1.171120	1.267680
	Expense ratio (x_3)	0.013122	0.003198
	Beta (x_4)	0.984405	0.162196
	Gross return (y_1)	0.000476	0.000905
	Skewness daily returns (y_2)	-0.155976	0.298815
No style	Std.dev. daily return (x_1)	0.017118	0.012227
	Kurtosis daily returns (x_2)	0.999274	1.291682
	Expense ratio (x_3)	0.015279	0.006317
	Beta (x_4)	1.368721	0.736967
	Gross return (y_1)	0.000352	0.001188
	Skewness daily returns (y_2)	-0.114423	0.270223
Small blend	Std.dev. daily return (x_1)	0.013374	0.005651
	Kurtosis daily returns (x_2)	1.098105	4.193503
	Expense ratio (x_3)	0.013491	0.004881
	Beta (x_4)	1.058855	0.266314
	Gross return (y_1)	0.000496	0.000806
	Skewness daily returns (y_2)	-0.101677	0.402850
Small growth	Std.dev. daily return (x_1)	0.014150	0.005041
	Kurtosis daily returns (x_2)	0.686445	1.956662
	Expense ratio (x_3)	0.014357	0.006789
	Beta (x_4)	1.128632	0.217808
	Gross return (y_1)	0.000430	0.000993
	Skewness daily returns (y_2)	-0.087289	0.241455
Small value	Std.dev. daily return (x_1)	0.013548	0.006004
	Kurtosis daily returns (x_2)	0.875483	1.202491
	Expense ratio (x_3)	0.013454	0.003412
	Beta (x_4)	1.065292	0.247812
	Gross return (y_1)	0.000503	0.000844
	Skewness daily returns (y_2)	-0.138225	0.238704

Table 2: Descriptive statistics for FDH efficiencies, mutual funds (2001–2011)

Fund style	# of funds ^a	Mean	1 st quartile	Median	3 rd quartile	Std.dev.
Index	1,111	1.0765	1.0000	1.0268	1.1202	0.1076
Large blend	2,739	1.1619	1.0235	1.1225	1.2462	0.1690
Large growth	3,993	1.1636	1.0084	1.1062	1.2514	0.1872
Large value	1,881	1.1590	1.0441	1.1290	1.2500	0.1387
Mid blend	704	1.1616	1.0000	1.1109	1.2534	0.1866
Mid growth	1,782	1.1680	1.0000	1.1027	1.2552	0.1998
Mid value	352	1.1718	1.0204	1.1470	1.2592	0.1657
No style	77	1.3061	1.0425	1.2302	1.3966	0.3550
Small blend	935	1.1604	1.0000	1.1222	1.2541	0.1693
Small growth	1,804	1.1956	1.0209	1.1463	1.2949	0.2058
Small value	572	1.1838	1.0266	1.1521	1.2938	0.1716
All funds	15,950	1.1621	1.0073	1.1142	1.2513	0.1795

^a The number of funds corresponds to the entire period 2001–2011. For each single year the numbers in this column have to be divided by 11.

Table 3: Descriptive statistics for order- m ($m = 25$) efficiencies, mutual funds (2001–2011)

Fund style	# of funds ^a	Mean	1 st quartile	Median	3 rd quartile	Std.dev.
Index	1,111	0.8672	0.7851	0.9078	0.9876	0.1649
Large blend	2,739	0.9891	0.9250	0.9998	1.0793	0.1667
Large growth	3,993	1.0083	0.9453	1.0061	1.0817	0.1550
Large value	1,881	0.9963	0.9359	1.0039	1.0746	0.1294
Mid blend	704	0.9975	0.9308	0.9993	1.0894	0.1865
Mid growth	1,782	1.0088	0.9280	0.9989	1.0951	0.1742
Mid value	352	1.0285	0.9554	1.0226	1.0934	0.1198
No style	77	1.0506	0.9656	1.0196	1.1329	0.2343
Small blend	935	1.0049	0.9269	0.9981	1.0993	0.1650
Small growth	1,804	1.0122	0.9214	1.0027	1.1173	0.1925
Small value	572	1.0334	0.9476	1.0371	1.1302	0.1550
All funds	15,950	0.9951	0.9251	0.9992	1.0819	0.1677

^a The number of funds corresponds to the entire period 2001–2011. For each single year the numbers in this column have to be divided by 11.

Table 4: Descriptive statistics for order- m ($m = 150$) efficiencies, mutual funds (2001–2011)

Fund style	# of funds ^a	Mean	1 st quartile	Median	3 rd quartile	Std.dev.
Index	1,111	1.0161	0.9772	1.0018	1.0578	0.1001
Large blend	2,739	1.0969	1.0000	1.0738	1.1799	0.1517
Large growth	3,993	1.1050	1.0000	1.0746	1.1809	0.1570
Large value	1,881	1.0974	1.0083	1.0862	1.1776	0.1206
Mid blend	704	1.0998	1.0000	1.0712	1.1860	0.1829
Mid growth	1,782	1.1095	0.9997	1.0685	1.1918	0.1781
Mid value	352	1.1186	1.0038	1.1018	1.1934	0.1335
No style	77	1.1980	1.0196	1.1392	1.2727	0.2819
Small blend	935	1.1096	1.0000	1.0819	1.1997	0.1549
Small growth	1,804	1.1294	1.0000	1.0900	1.2233	0.1854
Small value	572	1.1383	1.0009	1.1195	1.2375	0.1515
All funds	15,950	1.1018	1.0000	1.0726	1.1826	0.1586

^a The number of funds corresponds to the entire period 2001–2011. For each single year the numbers in this column have to be divided by 11.

Table 5: Descriptive statistics for order- α ($\alpha = .90$) efficiencies, mutual funds (2001–2011)

Fund style	# of funds ^a	Mean	1 st quartile	Median	3 rd quartile	Std.dev.
Index	1,111	0.6753	0.5038	0.7044	0.9033	0.2541
Large blend	2,739	0.8733	0.8132	0.9224	0.9969	0.2083
Large growth	3,993	0.9049	0.8383	0.9328	1.0000	0.1676
Large value	1,881	0.8894	0.8354	0.9208	0.9916	0.1588
Mid blend	704	0.8926	0.8110	0.9254	1.0000	0.1840
Mid growth	1,782	0.8966	0.8095	0.9134	1.0000	0.1827
Mid value	352	0.9271	0.8527	0.9427	1.0000	0.1318
No style	77	0.9017	0.7990	0.9476	1.0000	0.2540
Small blend	935	0.8820	0.7961	0.9081	1.0000	0.1874
Small growth	1,804	0.8772	0.7679	0.9079	1.0000	0.2179
Small value	572	0.9056	0.8177	0.9467	1.0226	0.1881
All funds	15,950	0.8761	0.8027	0.9180	1.0000	0.1994

^a The number of funds corresponds to the entire period 2001–2011. For each single year the numbers in this column have to be divided by 11.

Table 6: Descriptive statistics for order- α ($\alpha = .99$) efficiencies, mutual funds (2001–2011)

Fund style	# of funds ^a	Mean	1 st quartile	Median	3 rd quartile	Std.dev.
Index	1,111	0.9719	0.9634	1.0000	1.0206	0.1243
Large blend	2,739	1.0543	1.0000	1.0404	1.1315	0.1563
Large growth	3,993	1.0720	1.0000	1.0487	1.1367	0.1445
Large value	1,881	1.0565	1.0000	1.0511	1.1268	0.1236
Mid blend	704	1.0699	1.0000	1.0440	1.1441	0.1759
Mid growth	1,782	1.0771	1.0000	1.0432	1.1574	0.1775
Mid value	352	1.0842	1.0000	1.0714	1.1503	0.1239
No style	77	1.1293	1.0000	1.0829	1.1868	0.2600
Small blend	935	1.0776	1.0000	1.0455	1.1667	0.1583
Small growth	1,804	1.0904	1.0000	1.0551	1.1894	0.1884
Small value	572	1.1078	1.0000	1.0946	1.1946	0.1538
All funds	15,950	1.0649	1.0000	1.0402	1.1392	0.1581

^a The number of funds corresponds to the entire period 2001–2011. For each single year the numbers in this column have to be divided by 11.

Figure 1: Violin plots, FDH efficiencies of mutual funds (2001–2011)

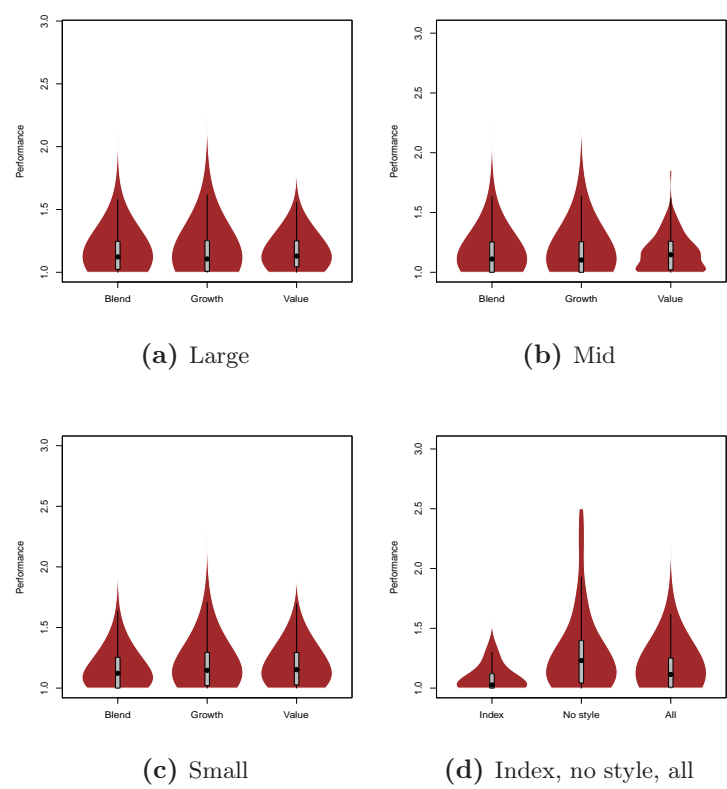


Figure 2: Violin plots, order- m ($m = 25$) efficiencies of mutual funds (2001–2011)

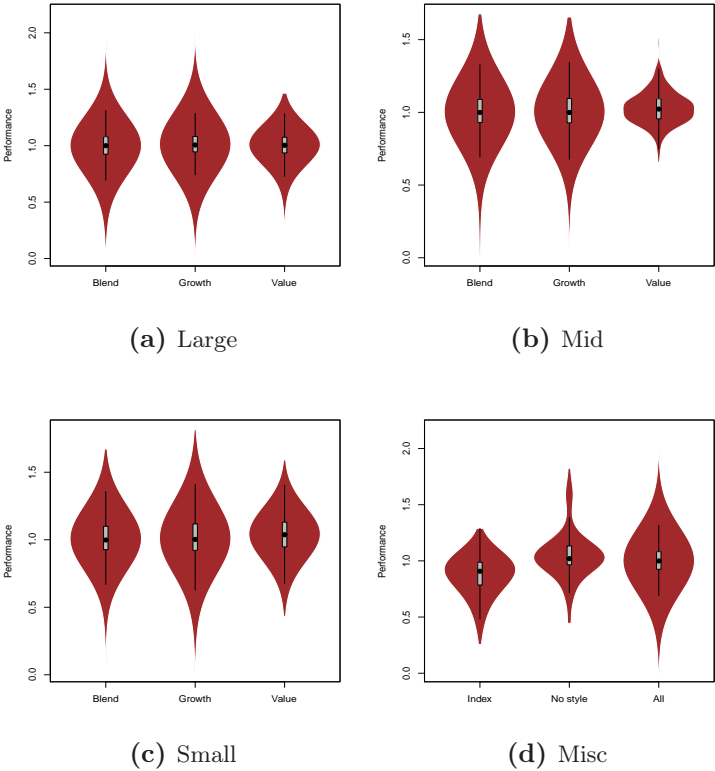


Figure 3: Violin plots, order- m ($m = 150$) efficiencies of mutual funds (2001–2011)

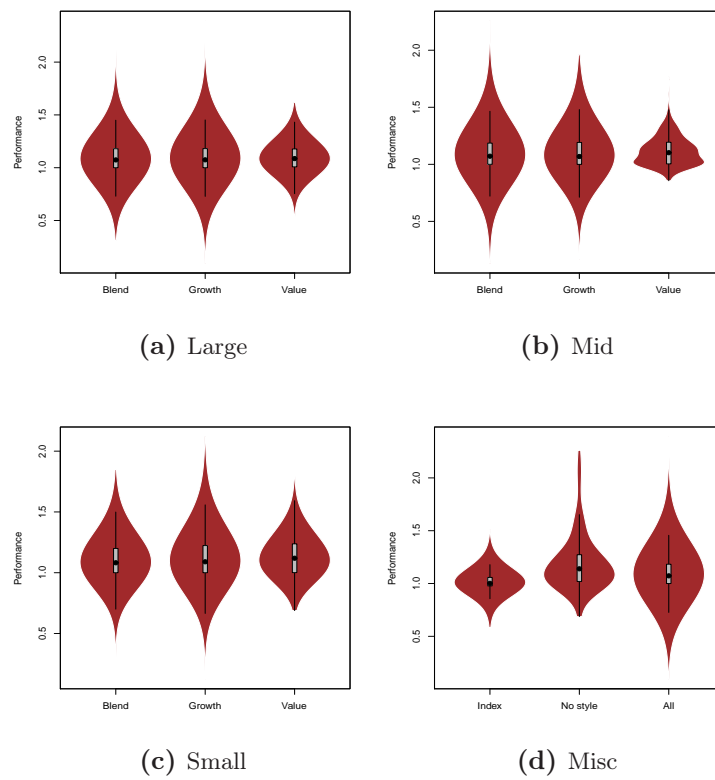


Figure 4: Violin plots, order- α ($\alpha = .90$) efficiencies of mutual funds (2001–2011)

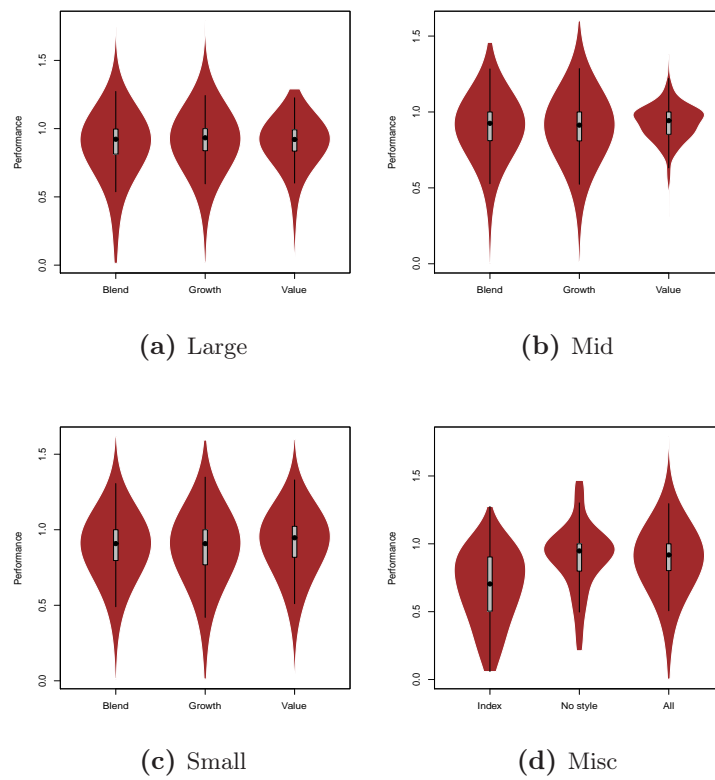
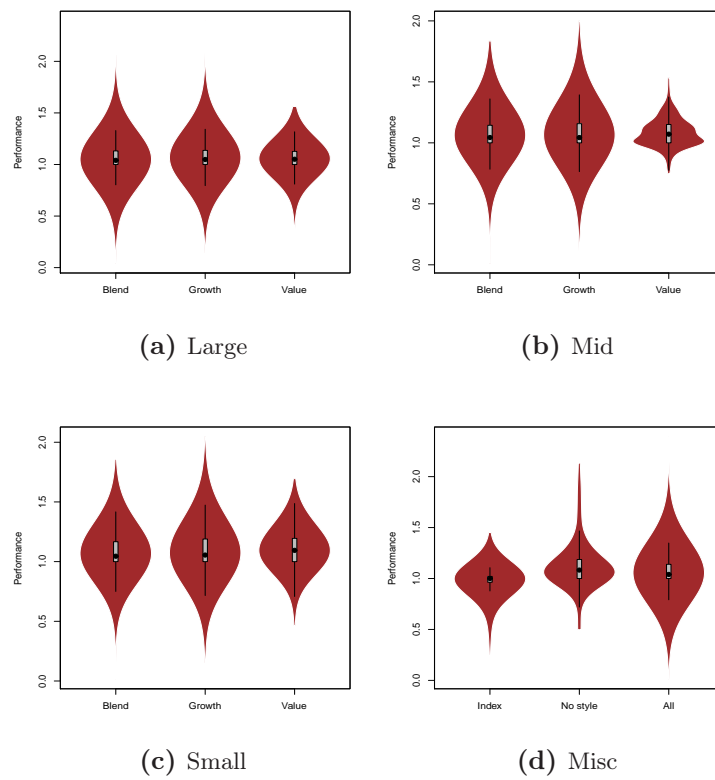


Figure 5: Violin plots, order- α ($\alpha = .99$) efficiencies of mutual funds (2001–2011)



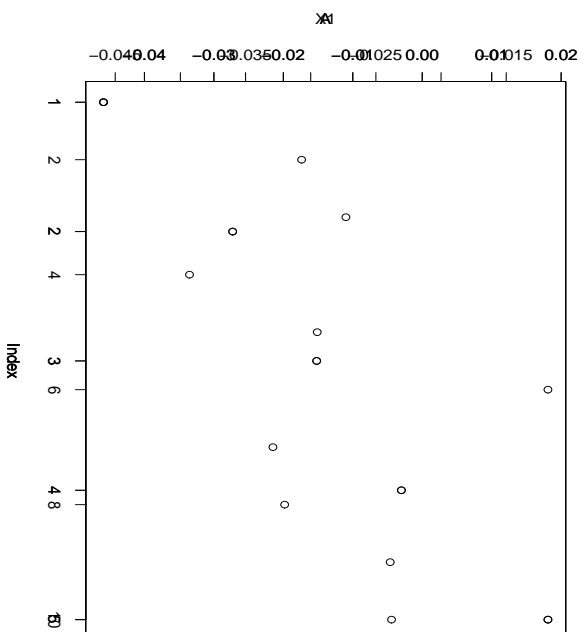
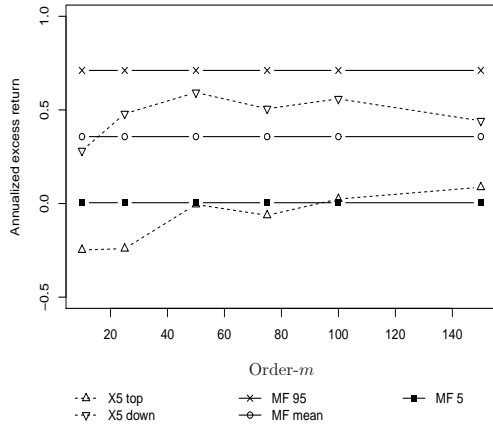
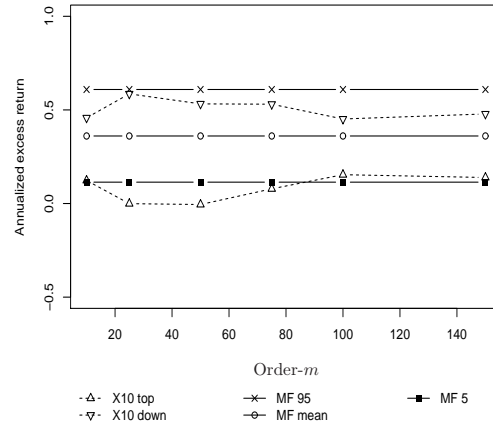


Figure 6: Results of performance persistence, FDH

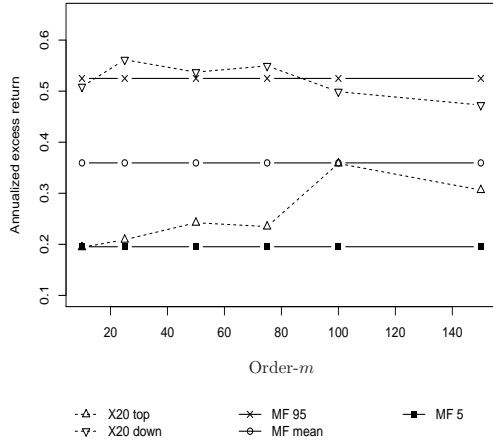
Figure 7: Results of performance persistence, order- m



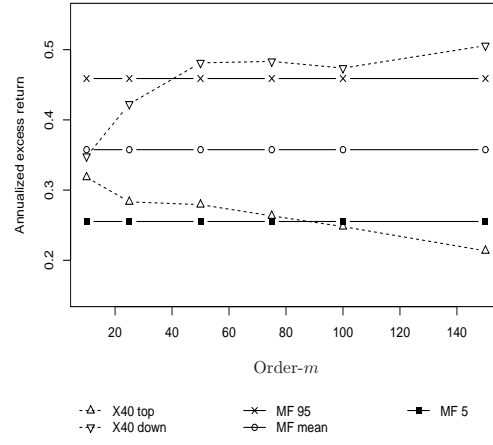
(a) 5%



(b) 10%

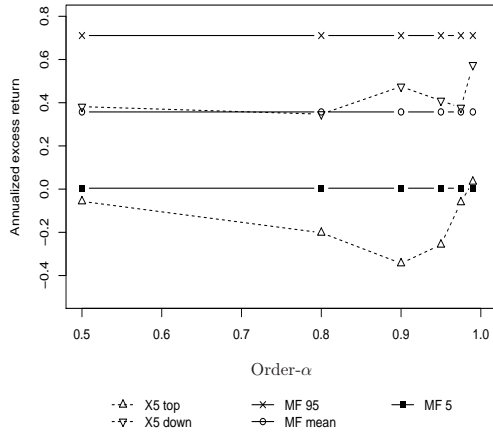


(c) 20%

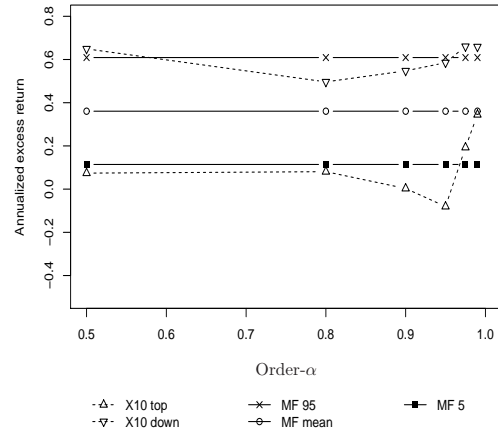


(d) 40%

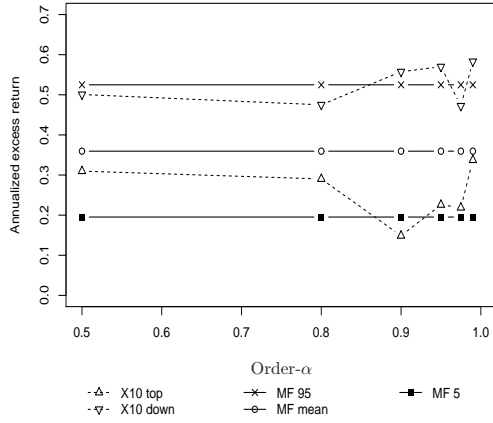
Figure 8: Results of performance persistence, order- α



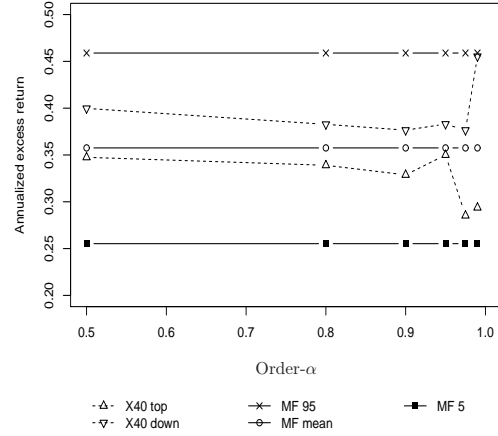
(a) 5%



(b) 10%



(c) 20%



(d) 40%