Non-parametric frontier estimates of mutual fund performance using C- and L-moments: Some specification tests

Kristiaan Kerstens [a,*], Amine Mounir [b], Ignace Van de Woestyne [c]

[a] CNRS-LEM (UMR 8179), IESEG School of Management, 3 rue de la Digue, F-59000 Lille, France
[b] Faculte de Management, MUNDIAPOLIS, Aeropole de formation, Nouaceur, Maroc
[c] Hogeschool-Universiteit Brussel, Stormstraat 2, B-1000 Brussel, Belgium

A B S T R A C T

There is a burgeoning literature using non-parametric frontier methods to measure mutual fund performance. These articles measure the relationship between the various characteristics (mainly return information and some costs of ownership) of these specialized financial products to establish a ranking using some efficiency measure. We argue in favor of the use of the shortage function, which is compatible with general investor preferences, and question some of the often maintained hypotheses in this line of research. The empirical part employs a large database of US and European mutual funds to offer extensive tests of the underlying modeling assumptions using various frontier estimators.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

In the USA, it is well-known that mutual funds have been the most rapidly growing non-depository financial institutions. Mutual funds have even become second in size only to commercial banks (see Haslem, 2003). The European mutual fund industry has meanwhile become the second largest in the world (accounting for about 30% of world fund assets). In the investment process, investors pay attention to mutual fund performances established by several rating agencies in the different industries around the world (such as Lipper, Morningstar, Standard & Poor's, Fitch Ratings, etc. in the USA, EuroPerformance in Europe and ASSIRT in Australia).

The typical investor is likely to look at highly rated mutual funds (e.g., five stars according to the Morningstar taxonomy) and make a choice among these. Although such industry ratings enjoy quite some popularity, the rating agencies' methodologies are subject to controversy. For instance, Blake and Morey (2000) examine the Morningstar rating system as a predictor of mutual fund performance for USA domestic equity funds and find little statistical evidence that Morningstar's highest-rated funds outperform the next-to-highest and median-rated funds. Sharpe (1998) investigates the properties of Morningstar's measure and shows that the risk-adjusted rating (RAR) produced by Morningstar leads to results similar to the traditional excess return Sharpe ratio.

In general, a whole variety of empirical regularities, including some behavioral biases, have been documented for both investors and fund managers in the mutual fund industry (see, e.g., Haslem, 2003; Zheng, 2008). It is known that investors typically have little knowledge about the funds they are investing in (Capon et al., 1996). For instance, Sirri and Tufano (1998) find that equity mutual fund investors select funds with the highest recent returns. Furthermore, Del Guercio and Tkac (2008) show that mutual funds receiving a five star Morningstar rating experience a high inflow of funds.

The limits of existing fund rating methodologies are to some extent related to the foundations of traditional performance measurement (Markowitz, 1952). Modern portfolio theory heavily relies on combining the information of expected return and variance (or some other variations of a risk measure) to assess performance. From the beginning, the theoretical foundations of this approach have been subject to several criticisms. From a theoretical point of view, this work maintains strong assumptions on probability distributions and Von Neumann–Morgenstern utility functions. From a practical point of view, since the computational

cost for computing the covariances was historically substantial for samples of even modest size, Sharpe (1963) proposed a simpler "diagonal" model. The same concern also motivated Sharpe (1964) and Lintner (1965) to develop the capital asset pricing model (CAPM), an equilibrium model assuming that all agents have similar expectations about the market. Meanwhile, an enormous literature has emerged on portfolio performance measurement.

Over time, portfolio performance appraisal has moved from total-risk foundations (e.g., variance or standard deviation) to performance indexes where returns in excess of the risk-free rate are compared to some risk measure (e.g., Sharpe and Treynor ratios are among the classics). Both in the academic and professional literature, literally dozens of performances gauges have been proposed over the years. A recent survey of this vast literature is found in the book of Bacon (2008) (see Farinelli et al., 2008 or Zakamouline and Koekebakker, 2009 for even more recent proposals). In the specific literature on mutual fund performance measurement, Haslem (2003) and especially Gregoriou (2007) and Lehmann and Timmermann (2008) summarize what is known about both the issues of the definition of performance benchmarks and performance attribution, i.e., the market timing and asset selection skills of active managers (see, e.g., Alexander and Baptista, 2010 or Brandouy et al., 2010 for even more recent proposals).

Relatively recently, probably in view of widespread criticisms of traditional performance measures, several authors have been introducing non-parametric frontier methods to open up new perspectives in mutual fund performance assessment. Following the seminal articles of Farrell (1957) and Charnes et al. (1978), efficiency measures and their extremum estimators from production theory are being transposed into the financial literature in an effort to provide alternative mutual funds ratings. In general, based on a sample of observed units, one estimates non-parametric frontiers of any multi-dimensional choice set and positions each of these observations with respect to the boundary of such choice set using an efficiency measure (see Ray, 2004). The use of these frontier or extremum estimators has the advantage of allowing to rate the performance of mutual funds along a multitude of dimensions (rather than mean and variance solely). For instance, the seminal article of Murthi et al. (1997) employs return as a desirable output to be increased and risk and a series of transaction costs as an input to be reduced, and measures the performance of each mutual fund with respect to a piecewise linear frontier established on the universe of funds under consideration (rather than a traditional non-linear portfolio frontier). Immediately thereafter, similar models started appearing in the literature (e.g., McMullen and Strong, 1998; Premachandra et al., 1998).

Meanwhile, one can speak of a burgeoning literature with about thirty or so related articles that have appeared in a variety of outlets. For instance, extensions have been proposed to the evaluation of pension funds (Barros and Garcia, 2006), ethical mutual funds (Basso and Funari, 2003), and hedge funds (Gregoriou, 2003), while lower and/or upper partial moments are utilized instead or in combination with ordinary moments in, e.g., the work of Eling (2006). Quite often, these efficiency measures are compared to more traditional portfolio performance measures. For instance, employing a similar framework Choi and Murthi (2001) compare the resulting efficiency measures to the traditional Sharpe ratio. Glawischnig and Sommersguter-Reichmann (2010) offer an up-to-date and fairly comprehensive review of the literature that has developed along these lines. The book of Gregoriou (2007) offers an international perspective and a mix of traditional and frontier-based performance studies.

One of the main advantages of these non-parametric frontier or extremum estimators (often known under the moniker Data Envelopment Analysis (DEA)) is that these can handle multiple dimensions simultaneously and that these yield a single real number

performance index with respect to a frontier composed of similar entities. Especially the need for multiple moments is apparent for certain type of funds (e.g., hedge funds) which are known to follow non-normal distributions (e.g., Darolles and Gourieroux, 2010; Eling and Faust, 2010; Gregoriou, 2003). However, while a growing body of contributions use these methods to assess mutual fund performance, no consensus has emerged about a variety of model specification issues.

The aim of this contribution is threefold. First, we summarize theoretical arguments for the use of the shortage function as an efficiency measure compatible with general investor preferences. The existing literature on non-parametric frontiers to gauge mutual fund performance has so far employed less general efficiency measures. Second, we focus on the main specification issues surrounding the application of non-parametric frontier methods when assessing the financial performance of mutual funds. We therefore employ a test-statistic developed by Li (1996) and refined by Fan and Ullah (1999) to assess these key methodological choices. Finally, since obtaining more robust estimates for portfolio management received a lot of attention in the recent finance literature (e.g., Fabozzi et al., 2007), this paper is the first frontier contribution gauging mutual funds using the robust L-moments (Hosking, 1990) rather than classical moments.

This contribution is structured as follows. The next section develops the basic theoretical framework for the paper. After introducing the shortage function and some arguments for its application in mutual fund performance assessment, we list the in our view most important model specification issues. An empirical application on a large sample of USA and European funds provides a testing ground for these methodological choices at hand. A concluding section summarizes the key results.

2. Non-parametric frontier models: methodology

2.1. Shortage function: definition

Even though the transposition of these non-parametric frontier estimators from a production context into finance seems straight-forward, a variety of specification issues arise that have largely been ignored in the literature so far. Basically, there are two main issues: (i) the choice of an efficiency measure and (ii) the specification of the model linking the different dimensions involved in the frontier benchmarking model. The main contribution of this paper lays with respect to the second issue. We first settle the choice of efficiency measure using a series of theoretical arguments.

Following recent insights from Eeckhoudt and Schlesinger (2006) into the validity of the assumption that the derivatives of the expected utility (EU) function alternate in sign to characterize behavior towards risk, Briec and Kerstens (2010) introduce a general procedure allowing for general higher moments in portfolio choice respecting this mixed risk aversion preference structure.[1] These authors transpose the generalized distance or shortage function to the multi-moment portfolio problem to account for a preference for odd moments (that need to be increased) and an aversion to even moments (that need to be reduced).[2] While Briec et al. (2004, 2007) demonstrated that the shortage function can project any (in) efficient portfolio exactly onto Mean–Variance (MV) and

---

[1] This class of mixed risk aversion utility functions was initially proposed in Brockett and Golden, 1987. It is well-known that these moment orderings meet necessary, not sufficient conditions for EU maximization under strong additional assumptions on probability distributions and investor's preferences.

[2] This shortage function is introduced in production theory by Luenberger (1995): this general distance function simultaneously looks for reductions in inputs and expansions in outputs and is dual to the profit function. While this same function has been labeled the directional distance function in Chambers et al. (1996), we prefer sticking to the original name.

Mean–Variance–Skewness (MVS) portfolio frontiers and that this function is connected via duality to an indirect MV and MVS utility function, Briec and Kerstens (2010) generalize this same result for general moments. Thus, the ability of the shortage function to seek for improvements in two directions simultaneously makes it an excellent tool for gauging financial product performances concurring with general investor preferences. By contrast, most current articles employ a more restrictive distance function as an efficiency measure that is not compatible with such general investor preferences. For instance, Glawischnig and Sommersguter-Reichmann (2010) only look for reductions in some of the dimensions in the frontier model.

One important difference between production theory and portfolio theory is that production data are normally restricted to the positive data domain while financial information (e.g., odd moments) may contain negative numbers. The issue of handling negative data has attracted some research attention in the frontier literature. For instance, proposals have been made to translate the data, though in many models this data transformation may have an impact on the efficiency measure (see, e.g., Lovell and Pastor, 1995 on the lack of translation-invariance of some of the non-parametric frontier models). An example of a fund rating contribution opting for this solution of translating the data is Glawischnig and Sommersguter-Reichmann, 2010. A variety of other solutions have been proposed in this frontier literature (see Pastor and Ruiz (2007) for a summary). Kerstens and Van de Woestyne (forthcoming) show that a slight variation on the shortage function offers a more general method to handle negative data values while maintaining a proportional interpretation (which has some convenience for practitioners).

Thus, summarizing the above discussion, the shortage function is a perfectly general efficiency measure that is compatible with general investor preferences and it can easily be extended to handle negative data occurring in a financial context. Therefore, we can now focus on the mathematical formulation of this shortage function.

If $U(x)$ is a utility function defined over a choice set X of combinations of m input-like variables (to be reduced) combined in the input vector I and s output-like variables (to be expanded) combined in the output vector O (i.e., $x = (I, O) \in X$, whereby $I \in \mathbb{R}^m$ and $O \in \mathbb{R}^s$), and g is a direction vector or reference bundle used for utility comparisons ($g \in -\mathbb{R}^m_+ \times \mathbb{R}^s_+$, with $g \neq 0$), then the shortage function value of $x \in X$ with reference g and reference utility value u is defined by[3]:

$$B(x, g, u) = \sup \{\beta : x + \beta g \in X \text{ and } U(x + \beta g) \geqslant u\}. \qquad (1)$$

In (1), the variables to be reduced (inputs) and variables to be expanded (outputs) can, e.g., be a vector of even and odd moment characteristics. The shortage function simultaneously permits the enhancement of multiple output-like variables and the reduction of multiple input-like variables. The shortage function defined in (1) uses a general direction vector g. However, sometimes one considers for positive data the special case $g = (-I, O)$ which yields a proportional interpretation that is convenient for practitioners. In view of the occurrence of negative data, we follow Kerstens and Van de Woestyne (forthcoming) and choose the direction vector $g = (-|I|, |O|)$, whereby $|I|$ and $|O|$ denote the vectors obtained by taking absolute values of the individual components of I and O, respectively.

In the next subsection, we focus on the specification issues related to the technology linking the different variables involved in a frontier benchmarking model for mutual fund rating.

2.2. Non-parametric frontier models: key specification issues

In the literature using non-parametric methods to gauge the financial performance of mutual funds, the evaluation is done using models which are directly derived from the production context. For instance, concentrating on just one issue and ignoring any other differences between both contributions, Murthi et al. (1997) evaluate mutual fund performance assuming a constant returns to scale (CRS) model, while McMullen and Strong (1998) use a variable returns to scale (VRS) model. Returns to scale are a basic notion in production theory, but it is a priori unclear what it means when modeling mutual fund performance. As it turns out, there are also other basic modeling choices that have hardly been systematically discussed in the current literature and to which we return below.

In our view, to meaningfully discuss specification issues it is important to first establish what kind of theoretical framework one can adopt to adequately model mutual fund performance. We basically see two options: (i) portfolio theory and (ii) hedonic price theory. Morey and Morey, 1999, is the only non-parametric-frontier based article explicitly adopting a traditional portfolio theory framework to evaluate an efficiency measure for mutual funds in a traditional MV model combining return and risk information over three different time horizons.[4] This approach differs from all other DEA-based models gauging mutual funds, since it explicitly considers the diversification effect on portfolio risk. Most other authors do not seem to have an explicit portfolio theory framework in mind when applying non-parametric frontier models to evaluate mutual funds. However, one cannot preclude that this is the implicit source of inspiration.

The "characteristics" approach to consumer theory developed by Lancaster (1966) writes utility not as a function of a vector of goods, but as a function of their characteristics.[5] Building upon this characteristics approach, Rosen (1974) developed a theoretical economic framework to study market equilibria for differentiated products differing along multiple characteristics. One basically seeks to derive an implicit price for the vector of observed characteristics to aggregate these into a single measure of market value. The recent literature in this area shows that (i) implicit price functions for characteristics are in general non-linear and (ii) the market need not provide a continuum of products in equilibrium, but rather product clusters exhibiting similar combinations of characteristics whereby products with certain other combinations of characteristics may well be hardly available (e.g., Ekeland et al., 2004). Hedonic price functions are mostly estimated using traditional econometric methods focusing on average relations between observed prices and characteristics. But, more recently a series of frontier specifications have aimed to estimate a price-quality correspondence and to explicitly measure any eventual consumer inefficiencies (see the seminal article of Kamakura et al. (1988)).[6]

While the evaluation of product efficiency has become quite common (e.g., Lee et al., 2005), we are unaware of any article applying non-parametric frontiers to mutual funds that explicitly refers to this theoretical framework. In fact, the "characteristics" approach has not been intensively used in the finance literature. For example, Heffernan (1990) has argued to interpret all financial products in terms of three main characteristics: (i) expected rate of return, (ii) security, and (iii) liquidity. Her analysis remains limited

---

[3] In consumer theory, Luenberger (1992) introduced a similar benefit function: this other name for a distance function positions consumption bundles relative to a reference utility level and it is dual to the expenditure function.

[4] This article is corrected and expanded in Briec and Kerstens (2009).

[5] Characteristics are objective features of products, rather than subjective attributes as analyzed in psychology and marketing.

[6] While explanations in terms of inefficiencies are often suspect in economics, there exist models of consumer choice explicitly accounting for the consumer's imperfect ability to choose due to limited information-processing capacities: e.g., De Palma et al., 1994.

to individual financial products, even though it is stated that ideally one would also like to analyze portfolios in terms of their characteristics. A number of other authors has developed this "characteristics" approach to finance (see, e.g., Blake (2000) for a collection of articles in finance). Therefore, similar to other services, mutual funds can in our view be trivially interpreted as a fee-based (loads) financial product that is characterized by some distributional characteristics, as summarized by a combination of moments (see also below). This disregards any repercussions at the portfolio level.

Based on a careful analysis of the available non-parametric frontier-based articles assessing mutual funds, we have identified three crucial specification issues arising when specifying a non-parametric frontier model to gauge mutual funds: (i) nature of returns to scale; (ii) inclusion of higher moments and cost components; and (iii) convexity or not.

The discussion of each of these issues in turn below allows to return to and elaborate upon the reasons to opt for a hedonic price theory framework rather than a portfolio theory perspective. Exploiting the relation between efficiency measures and goodness-of-fit measures used for hypothesis testing (see Färe and Grosskopf, 1995), the comparison of efficiency measures based upon the shortage function (1) computed relative to two non-parametric frontier models amounts to a test of the null hypothesis distinguishing both models involved. We employ a non-parametric test-statistic developed by Li (1996) that compares two entire distributions and that is valid for both dependent and independent variables alike.[7] Since efficiency measured using non-parametric frontier models is relative in nature and thus depends on the sample size, this test is in principle suitable for our purposes.[8] By testing for the difference between two densities of shortage function values computed relative to two model specifications, we test the null hypothesis that distinguishes these models.

In view of the first and third specification issue indicated above, we need to specify three types of non-parametric frontier models capable of testing on the one hand variable returns to scale (VRS) versus constant returns to scale (CRS), and on the other hand convexity (C) versus non-convexity (NC). This requires three specifications that can be introduced formally as follows (see, e.g., Ray, 2004). Assume there is a set of mutual funds to be evaluated (indexed by $j = 1, \ldots, n$), where each mutual fund is characterized by m input-like values $x_{ij}, i = 1, \ldots, m$ and s output-like values $y_{rj}, r = 1, \ldots, s$. To compute the shortage function, one needs to solve the following mathematical program for unit $o \in \{1, \ldots, n\}$ in need of assessment:

$$\max \beta$$

$$\text{s.t.} \sum_{j=1}^{n} \lambda_j y_{rj} \geqslant y_{ro} + \beta|y_{ro}| \qquad r = 1, \ldots, s$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} \leqslant x_{io} - \beta|x_{io}| \qquad i = 1, \ldots, m$$

$$\beta \geqslant 0, \Gamma \text{ with :}$$

(i) $\Gamma^{C,CRS} = \{\lambda_j \geqslant 0; j = 1, \ldots, n\}$

(ii) $\Gamma^{C,VRS} = \{\lambda_j \geqslant 0; j = 1, \ldots, n\} \cup \left\{ \sum_{j=1}^{n} \lambda_j = 1 \right\}$

(iii) $\Gamma^{NC,VRS} = \{\lambda_j \in \{0,1\}; j = 1, \ldots, n\} \cup \left\{ \sum_{j=1}^{n} \lambda_j = 1 \right\}$,

where the specification of the direction vector follows Kerstens and Van de Woestyne (forthcoming) to allow for negative data values in general. The application of these three different models allows to project any observed mutual fund (o) being evaluated in the direction $g = (-|x_{1o}|, \ldots, -|x_{mo}|, |y_{1o}|, \ldots, |y_{so}|)$ such that the output-like values $y_{ro}, (r = 1, \ldots, s)$, are increased and the input-like values $x_{io}, (i = 1, \ldots, m)$, are decreased in proportion to the initial position of o. The value of β measures the resulting proportional amount of inefficiency representing the shortage function. Depending on the exact specification of the activity or weight vector $\lambda = (\lambda_1, \ldots, \lambda_n)$ we impose (i) C and CRS, (ii) C and VRS, or (iii) NC and VRS, whereby the latter simply adds binary constraints to (ii). Now, we turn to the details relating to the three specification issues mentioned above.

2.2.1. Returns to scale issue

A first specification issue that has seldom been explicitly discussed in the non-parametric frontier articles gauging mutual fund performance is the nature of returns to scale of the underlying "technology". There seem to be about as much articles specifying variable returns to scale (VRS) as there are contributions adopting constant returns to scale (CRS). Limiting ourselves to the seminal articles (listed above), while Murthi et al. (1997) and Premachandra et al. (1998) opt for CRS, McMullen and Strong (1998) imposes VRS in their rating model. This controversy continues among more recent contributions: Anderson et al. (2004) and Basso and Funari (2001), among others, impose CRS, while Haslem and Scheraga (2006), Glawischnig and Sommersguter-Reichmann (2010) and Wilkens and Zhu (2001) do specify VRS.

We see three potential sources of justification for the use of CRS in portfolio theory. However, none of these stands the test of scrutiny. First, Choi and Murthi (2001, pp. 862–867) indicate that the Sharpe index is similar to a CRS convex frontier model applied to a single input (risk) and a single output (return). But, this does not justify imposing CRS when including additional variables (e.g., higher order moments, loads, etc.) in a mutual fund rating model, as is done in the majority of articles in this literature.

Second, some articles even combine both assumptions and infer information about local returns to scale for the individual funds (see, e.g., Choi and Murthi, 2001 and Galagedera and Silvapulle, 2002) and then sometimes continue the analysis by investigating the reasons for performance differences between funds. For instance, Galagedera and Silvapulle (2002) attempt to explain mutual fund performance differences by fund-specific characteristics like age, size and the 12-month net asset flow as proxies for experience, scale of operation, and level of investor confidence, respectively. However, in our view, this type of approach confounds the analysis of mutual fund performance from the investor perspective, which is the main aim in this literature, with an analysis of the reasons of good and bad performance at the level of investment management companies (i.e., the firms creating, organizing, staffing, and managing these mutual funds for fund shareholders).[9] Furthermore, assuming looking for returns-to-scale properties at any point on the frontier is a relevant item, it is worthwhile realizing that the CRS model is always just an artificial auxiliary frontier permitting to examine different points on the VRS frontier with regard to their returns-to-scale properties (see Ray (2004) for details).[10]
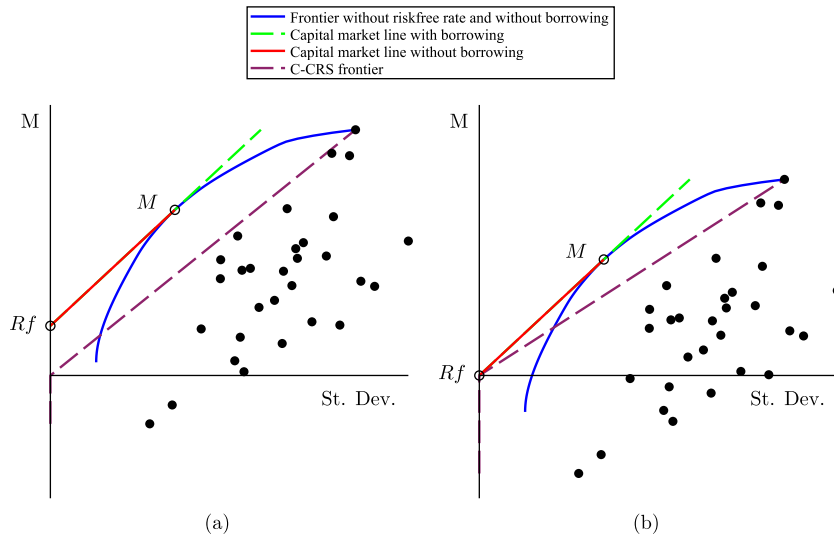
Third, another potential source of inspiration justifying the application of CRS in this literature is based on the apparent

---

[7] This test has been refined in Fan and Ullah, 1999.

[8] However, none of the statistical tests known to us seem to be designed to test for the effect of convexity in particular.

[9] Smith (2009) summarizes the economics of these investment management companies and distinguishes between returns to scale, economies of scope, and other empirical regularities.

[10] Note that VRS models the most general technology representation allowing for increasing, constant, or diminishing returns to scale at different points on the frontier, and that CRS is unlikely to hold globally in many realistic settings (see Chapter 3 in Ray (2004)).

Fig. 1. Comparison of the non-parametric convex CRS frontier, traditional portfolio frontier, and capital market line (CML). (a) Frontiers obtained from the original mutual fund positions. (b) Frontiers after translating the mutual fund in a vertical direction for positioning the risk-free rate (*Rf*) in the origin. Even then, the C-CRS frontier and the CML do not necessarily coincide unless the market portfolio is included.

similarity between the capital market line (CML) and the shape of a convex CRS efficient frontier. When representing the CML and the CRS frontier in the same two-dimensional plane of mean and standard deviation (see Fig. 1), then the CRS assumption yields a straight line going through the origin. First, notice that the connection between CML and the CRS frontier is only valid when we can translate the data such that the positive risk-free rate is located in the origin.[11] Furthermore, assuming that the CRS efficient frontier and CML coincide (which from a practical point of view would require adding the market portfolio (*M*) among the observations to be evaluated), then the efficiency of an undervalued fund must be the same with respect to both frontiers. However, almost all non-parametric frontier models assume non-negative weights: in portfolio theory this can be interpreted as borrowing restrictions. Under such restrictions, one is only able to invest in asset combinations lying on the CML between the risk-free rate and the market portfolio indicated by point *M* (by investing one part in the risk-free rate and the remainder in the market portfolio). Thus, when borrowing is not allowed, the CML is the segment starting from the risk-free rate and ending at market portfolio *M*. In Fig. 1, the CRS frontier continues beyond this CML segment. Only when one can borrow money, it is possible to move beyond point *M* on the CML, but this requires negative weights which are not allowed in typical non-parametric frontier models.

In terms of hedonic price theory, in view of the generally nonlinear nature of the implicit price functions for characteristics (see above), one needs to impose VRS rather than CRS (which amounts to linearity). Hence, from both theoretical perspectives, one can safely conclude that the most relevant assumption with regard to the nature of returns to scale when assessing mutual funds using non-parametric frontier models is VRS.[12]

2.2.2. Higher-order moments and cost components issues

While the application of non-parametric frontier methods to assess the performance of financial products has become rather

widespread because it allows combining multiple variables into a single aggregate efficiency score, this methodology has a major drawback as well: when various candidates of input-like and output-like variables coexist, then there is no evident rule for their selection. This leads various mutual fund frontier articles to focus on different aspects. For example, Galagedera and Silvapulle (2002) put the spotlight on various transaction costs, while Eling (2006) combines lower and upper partial moments through different models and tries combining some of these. Lozano and Guttiérez (2008) summarize a wide variety of specifications in a selection of 14 articles using non-parametric frontier models to assess mutual fund performance (see their Table 1). Thus, while this burgeoning literature seems to introduce an ever growing variety of new specifications, the question as to the correct specification remains hitherto unanswered. To the best of our knowledge, none of these existing articles has systematically implemented statistical tests to discover a proper specification.[13]

To systematically discuss specification issues, it is useful to distinguish between the return characteristics of the mutual fund's share price and the shareholder transaction costs related to the buying and selling of mutual fund shares above the net asset value per share as well as the expenses for fund administration and portfolio management. Both from the viewpoint of portfolio and hedonic pricing theories, the former return characteristics define the nature of the financial product, while the latter determine the cost of mutual fund ownership. All articles in the current non-parametric frontier literature include both return characteristics and some cost components in their models.

In our analysis, following Briec and Kerstens (2010), we focus on the main statistical characteristics of mutual funds return distributions (i.e., classical moments such as return, variance, skewness, kurtosis, or even higher order moments). While traditional investment performance measures only take into account the return and risk attributes for investment decision making, it is worthwhile checking the need for higher order moments in view of the return patterns of mutual funds exhibiting significant levels of skewness and kurtosis (see Glawischnig and Sommersguter-Reichmann, 2010). Therefore, using traditional measures based on some combination

---

[11] This requires opting for translation-invariant non-parametric frontier models to avoid any effect on the resulting efficiency measure, which is not trivial: see supra.

[12] The criticism of CRS need logically not imply VRS in a production context, since alternative, but more special returns to scale assumptions are available (i.e., increasing and decreasing returns to scale). However, in both a hedonic pricing and a portfolio framework the latter alternatives do not seem to make much sense.

[13] For example, Galagedera and Silvapulle (2002) report on 11 different specifications, but no statistical test is employed to discriminate between these.

Table 1
Descriptive statistics.

Number of mutual funds	1068
Average minimum return	−0.11103
Average expected return	−0.00017
Average maximum return	0.11134
Average variance	0.00022
Average skewness	−0.54312
Average kurtosis	16.72734
Average annual net expense ratio	1.43523
Average maximum front loads	2.67464
Average deferred loads	0.15801
Minimum Jarque–Bera test statistic	160.040

of the first two moments may well be irrelevant to evaluate mutual funds. Thus, using higher order moments as input–output specification allows performance gauging of mutual funds consistent with general investor preferences (see Briec and Kerstens, 2010). Since it is well-known that the quality of moment approximations representing EU functions is an empirical issue (see Hlawitschka, 1994), one can also expect that the approximation quality of a partial series of a Taylor expansion of the shortage function needs to be empirically assessed. Using the non-parametric test-statistic developed by Li (1996), we are the first article to test for the exact sequence of moments needed to empirically assess mutual fund performance. In the empirical part, this is done by systematically adding moments up to the sixth and testing for their impact. In the existing literature, all published articles we are aware of include at most up to the third moment (see, e.g., Glawischnig and Sommersguter-Reichmann (2010) for ordinary moments and Eling (2006) for partial moments).

This focus on an adequate description of the mutual funds return distributions does not deny that other variables may be of relevance in assessing mutual fund performance using non-parametric frontier models. For instance, Galagedera and Silvapulle (2002) include the minimum initial investment as an additional variable. Other contributions add variables related to the composition of the mutual fund: for example, Haslem and Scheraga (2006) include the percentage of stocks, Premachandra et al. (1998) add a variable indicating the total amount that is invested risk-free, etc. However, in the non-parametric frontier literature, it is not always clear whether a certain variable should be included in the model computing the efficiency measure, or rather should be used in some second stage analysis attempting to explain the observed variations in the efficiency measures computed at the initial stage. Therefore, we refrain from extending our specification search beyond the moment characteristics.

In the finance literature, recently a lot of attention has been drawn on the need for robust portfolio management. This mainly academic literature appearing in engineering, statistical and operations research publications has meanwhile provided a rich source of inspiration for finance professionals (see Fabozzi et al., 2007). The reluctance of investment managers to apply traditional quantitative portfolio optimization models is, among others, related to their unreliability in practice. The main goal of the development of robust portfolio optimization tools is to explicitly incorporate the statistical precision of individual estimates in the portfolio allocation process.

Therefore, we decided to duplicate the analysis of the classical moments (C-moments) with the more robust univariate L-moments introduced by Hosking (1990).[14] In view of handling heavy-tailed distributions, these L-moments are defined in terms of

strategically selected linear functions of the expectations of order statistics. Therefore, L-moments provide more robust estimators of moments than their corresponding sample C-moments. But, L-moments can also characterize a wider range of distributions since they exist (i.e., they are finite for all orders) whenever there is a finite first moment. While L-moments have found their way in fields like meteorology and hydrology, their empirical application in finance is of recent date (e.g., Tolikas and Gettinby, 2009).

While the price of a mutual fund is its net asset value per share, as determined by the financial markets, many mutual funds impose transaction costs in addition to the expenses paid by shareholders for fund administration and portfolio management and the costs of trading fund portfolios (see, e.g., Haslem, 2003 or Khorana et al., 2008, the latter providing an international overview). In fact, mutual fund investors may pay a variety of monetary costs: front-end and/or deferred loads, operating expenses (management fees, transfer agent fees, legal fees, auditor fees, etc.), account fees (redemption fees, account maintenance fees, etc.), and trading costs (brokerage fees, bid-ask spreads, etc.). The complexity and the large variety of practices in the mutual fund market tends to blur the total cost of ownership.[15] While attempts have been made to define a total shareholder cost (e.g., Khorana et al., 2008), this notion depends upon time horizon and other parameters. While sometimes just a single cost variable is included (for instance, Haslem and Scheraga (2006) focus on the expense ratio), most non-parametric frontier applications specify multiple cost components (e.g., Galagedera and Silvapulle, 2002). In view of the "characteristics" approach, we include the multiple cost components of the mutual fund in the empirical specification and explicitly test for their relevance. Hence, following the majority of the mutual fund frontier literature, we are forced to handle these cost aspects in some ad hoc way by including the main components available in our sample as separate variables.

2.2.3. Convexity issue

By far most non-parametric frontier articles measuring the financial performance of mutual funds impose the traditional convexity assumption. Examples include the articles of Anderson et al. (2004), Basso and Funari (2001), Gregoriou (2006), Haslem and Scheraga (2006), Lozano and Guttiérez (2008) and McMullen and Strong (1998). We are only aware of one contribution that does not impose convexity to gauge efficiency. One rather popular non-convex non-parametric frontier model is the Free Disposal Hull (FDH). Daraio and Simar (2006) use order-m based FDH models to derive robust non-parametric results to evaluate and explain mutual fund performance. Despite the apparent consensus towards the use of convex non-parametric frontier models in mutual fund rating, the almost unanimous use of convexity does in itself not guarantee the validity of the assumption. Therefore, we develop the arguments showing that from both theoretical perspectives convexity is an assumption to be scrutinized.

While in portfolio theory convexity seems like a natural assumption, even the basic MV portfolio set is not convex under all circumstances (e.g., think about cardinality constraints on the number of assets, transaction round lot restrictions, etc.). But, there is a related, more specific criticism to the use of convex non-parametric frontier models to gauge mutual funds.

Convexity is not always useful to guide investors in terms of choosing among funds even in terms of return and risk only, since the projection point resulting from applying the efficiency measure may yield a MV combination that need not be feasible. Indeed, the optimal projection is normally some hypothetical combination of several mutual funds that cannot be achieved in practice. The basic

---

[14] Recently, also L-comoments have been defined. Computations have been done using the `Lmomco` and `Lmoments` packages in the software `R`.

[15] Obviously, the problem of determining the cost of ownership for other consumer durables (e.g., cars or houses) is somewhat similar.
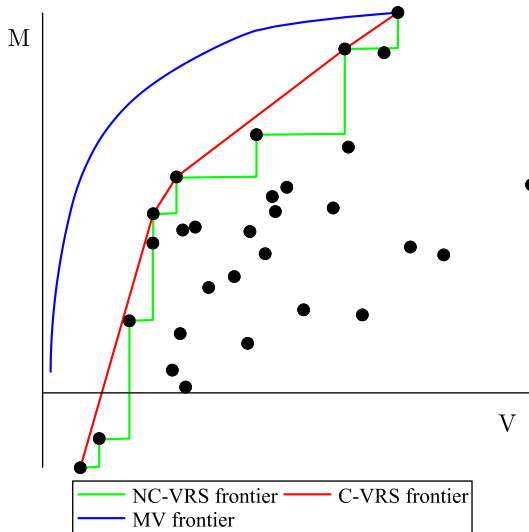
Fig. 2. MV frontier contrasted with non-parametric C-VRS and NC-VRS frontiers.

reason is that the combination of mutual funds would require explicitly accounting for the diversification effect.

To illustrate this argument in a simple way, we generate three efficient frontiers based upon 30 mutual funds in MV space. We represent the traditional MV portfolio frontier as well as the convex (C-VRS) and non-convex (NC-VRS) VRS non-parametric efficient frontiers in Fig. 2. Clearly, the MV frontier dominates both convex and non-convex VRS non-parametric frontiers. This is due to the importance of covariances among mutual funds, which is ignored by the non-parametric frontier models. While the convex frontier takes advantage of its ability to combine mutual funds, thereby dominating the non-convex frontier, we cannot discriminate between the latter two methods since both are still far below the traditional MV frontier because they ignore the covariance structure in the return data.

From a hedonic price theory perspective, the theoretical concern about non-convexity goes back to the seminal article of Lancaster (1966) who warned about the impact of indivisibilities ("combinable" and "non-combinable" goods in his parlance).[16] In a hedonic non-parametric frontier context, the need for assuming non-convexity has been argued for, but not tested, by Lee et al. (2005), among others. In the same context, Chumpitaz et al. (2010) are the first to test for the impact of convexity using the above convex and non-convex non-parametric estimators imposing VRS and their empirical results clearly reject convexity. In finance, Heffernan (1990) stresses the imperfect divisibility of many financial products and interprets the selection of several of those products in a portfolio as an attempt to overcome this imperfection. Therefore, in hedonic price models one cannot simply adopt a traditional convex approach, but one must ideally test it against a non-convex frontier method.

2.2.4. Conclusions

Summarizing this extensive discussion so far, we observe that the CRS assumption is suspect from both the portfolio and hedonic pricing perspective. The inclusion of higher moments and cost components (mainly loads) is also compatible with these two the-

---

[16] In economics, Shephard (1978) axiomatically analyzes Lancaster's household production theory: he stresses that the transformation of goods into characteristics space need not result in some convex frontier, let alone a twice differentiable function (p. 454). In marketing, Hjorth-Andersen (1983) is first to use a simple vector dominance criterion distinguishing efficient and inefficient product varieties to test Lancaster's principle of efficient choice. While no explicit efficiency measure is involved, this amounts to adopting a non-convex frontier model.

Table 2
Average C- and L-moments up to order four.

Order	1	2	3	4
C-moments[*]	−0.16524	0.22303	−0.00352	0.00242
L-moments[*]	−0.16524	7.00280	−0.41391	2.06081

[*] $\times 10^{-3}$.

oretical perspectives. Finally, even a standard convexity assumption is questionable from both these theories: non-convexity may prove to be a more modest hypothesis to be maintained.

3. Empirical testing

3.1. Data description and outlier analysis

We illustrate the three main specification issues discussed above by using a sample of European and United States mutual funds drawn from the reputed Morningstar database and completed with historical prices from the Datastream database (Thomson Reuters). The dataset contains daily prices of 1070 mutual funds from January 2, 2004 to February 28, 2009 spread over six categories: on the one hand funds domiciliated in Europe (i) OE (=open end) Large-Cap Blend Equity, (ii) OE US Large-Cap Growth Equity, (iii) OE US Large-Cap Value Equity, and on the other hand funds domiciliated in the United States, (iv) OE Large Growth, (v) OE Large Value, and (vi) OE Europe Stock.

Table 1 reports some descriptive statistics of the overall dataset. Prices are converted into a common currency (i.e., Euro) from which the daily returns are computed. The return distributions have been tested for normality using Jarque–Bera tests. Normality is rejected for all mutual funds in the dataset, since the minimal value of the Jarque–Bera test statistic of 160.04 is much larger than the value of 9.21 corresponding with a 1% significance level. Notice from the moment characteristics that skewness and kurtosis seem non-negligible. In particular, the average value for kurtosis is higher than the benchmark value of 3 for a normal distribution. This observation is in line with the Jarque–Bera test results. Furthermore, the Morningstar database contains the following information on the cost of ownership: (i) maximum front loads, (ii) deferred loads, and (iii) annual net expense ratio.

Table 2 reports the average C- and L-moments up to order four. Notice that the first order C- and L-moments coincide since these are merely the average expected return. The C-moments of order two to four are centralized with respect to the average expected return. In this respect, the moments of order three should not be confused with skewness and kurtosis, the latter being normalized by taking proper ratios involving the variance. Consequently, the average values for skewness and kurtosis reported in Table 1 differ from those of the C-moments of order three and four provided in Table 2. Also for the L-moments, non-normalized definitions have been used. All computations executed hereafter make use of non-normalized moments, especially since moments up to order six are considered.

In the empirical analysis based on non-parametric frontier models, it is not uncommon to start with an outlier analysis to check for the existence of atypical observations. Indeed, it is widely acknowledged that the occurrence of extreme observations may influence the shape of non-parametric frontier models, thereby having a potentially substantial influence on the relative performance based efficiency estimates. Among the available outlier detection methods in the non-parametric frontier literature, we follow Simar (2003) who proposed a more robust frontier estimator (known as the order-m frontier) that is less sensitive to the

Table 3
Number of extreme points (potential outliers) in basic MV-Loads model.

Fund categories	C-moments	L-moments
C1	3	6
C2	2	2
C3	3	4
C4	2	4
C5	2	2
C6	2	2
All funds	5	9

influence of outliers.[17] Intuitively speaking, the order-m frontier draws samples of size m from the initial data using a bootstrapping strategy (2000 replications) and tries to fit a local non-convex frontier on these selected data.[18]

To detect outliers, we apply the Simar (2003) methodology to what we consider a basic non-parametric frontier model to gauge mutual funds. Borrowing from the current literature on frontier models for mutual funds, we include the above mentioned three cost variables (maximum front loads, deferred loads, and annual net expense ratio). Starting from a traditional portfolio framework, we account for just the first two moments. Thus, the basic MV and three-dimensional cost model (with the moniker MV-Loads) imposing convexity and VRS is the basis for any further testing. Throughout the empirical section, all tests are always performed on the two variations: C- and L-moments. Thus, also these first two moments are estimated using both C- and L-moments.

Applying this outlier methodology to the six Morningstar categories separately and across all categories, extreme observations are defined by two criteria. First, the efficiencies based on order-m must converge to the FDH estimates: in particular, the difference between these local and global non-convex estimators should be less than 0.1.[19] Second, the number of observations sampled with replication that dominate the observation under evaluation must be less than 5% of the sample size of its fund category or of the whole sample. Extreme observations are suspect and are in need for some close attention.

Table 3 reports the number of extreme points in the above specified basic model. The following conclusions emerge. First, while the number of extreme points detected with L-moments is higher than with C-moments, the amount of extreme observations is less than 1% of the total sample. Second, the larger set of extreme points detected with L-moments contains all extreme points detected with the C-moments. Hence, the outlier analysis results are robust with regard to the choice of moment estimator. Third, it is only after scrutinizing these potentially extreme observations in detail that one can determine whether these are real outliers. We pay especially close attention to the underlying price distributions and find two types of anomalies. One mutual fund in the first category shows an abnormal price for just one trading day.[20] Two funds seem to be dead: one in the first category for which the price distribution exhibits constant net asset value for the considered time

series, except for one observation; another one in the second category exhibits a constant net asset value in its price distribution from February 2005 onwards.

It is important to mention that we checked the price distributions of the three concerned funds several times with other internet sources. However, the errors seem to be specific for this financial data provider (Datastream). Being sure now about the outlier status of these three funds, we delete the two (almost) dead funds and we correct the price of the third one for the single day concerned.[21] Thus, after the outlier analysis our sample consists now of 1068 instead of 1070 mutual funds. These results also show that the Simar (2003) methodology does constitute a useful tool for detecting extreme observations that are potentially outliers.[22]

After correcting and deleting the real outliers, we now have the following agenda for specification testing. First, we shed light on the nature of returns to scale by testing whether CRS can ever prevail in a mutual fund context. Then, we test for the inclusion of higher order moments and check whether all cost components contribute equally well to the fit of the models. Finally, we tackle the issue of imposing convexity or not. Each time, we basically follow the same test methodology: we formally test for differences between the kernel density estimates of the efficiency distributions based on the shortage function of the two models by employing the non-parametric test-statistic developed by Li (1996) (refined by Fan and Ullah, 1999) that is valid for both dependent and independent variables. The null hypothesis is simply that the efficiency distributions of both non-parametric frontier models are identical, while the alternative hypothesis is that these distributions are different. Testing is performed at a 5% significance level.

3.2. Specification issues

3.2.1. Variable versus constant returns to scale

In this subsection, we shed light on the question whether it is useful to assess mutual fund performance using a convex non-parametric frontier model imposing the particular CRS assumption rather than the far more general VRS hypothesis. Again starting from the basic non-parametric frontier model including the first two moments and three cost variables (maximum front loads, deferred loads, and annual net expense ratio), we apply the Li-test statistic for both models to the traditional C- as well as the more robust L-moments.

For this MV-Loads model and for all fund categories taken together, we find that we can safely reject the null hypothesis that both distributions are equal. This conclusion holds for C- as well as L-moments. Hence, the VRS assumption fits our mutual fund data better than the much stronger CRS hypothesis. Consequently, it seems hard to justify the use of CRS for gauging mutual funds performance as is done in at least part of the non-parametric frontier models, some of which have been cited above. In the remainder of our analysis, we maintain the VRS assumption.

3.2.2. Higher-order moments and cost components

While keeping VRS as a maintained hypothesis throughout, a series of other methodological choices remain available. In this

---

[17] See Ray (2004) for a review of the developments in stochastic versions of non-parametric frontier estimators of which the Simar (2003) outlier detection method is an extension.

[18] For technical details, see Simar (2003). Computations are done using the FEAR package for the software R: see http://www.clemson.edu/economics/faculty/wilson/Software/FEAR/fear.html.

[19] In fact, a sequence of order-m estimators is computed with m equal to 25, 50, 100, 200 and 500, respectively.

[20] On average, the price of this fund is about €20, except for that single day where it shows up at €600. This is clearly anomalous.

[21] In particular, we use the information from the internet site http://www.boursorama.com.

[22] In practice, it should be added that the list of extreme observations detected may somewhat depend on the choice of the parameter m and the minimal % of observations we select to be dominating the observations being evaluated (see Simar, 2003 for details). Another potential application of this methodology in a financial performance assessment context may be the detection of superior efficient funds. Indeed, the extreme points that are no outliers in any sense after scrutiny may simply signal atypical price distributions reflecting special fund management strategies.

Table 4
Li-test statistics comparing C-VRS and NC-VRS efficiencies of MV-Loads models with different loads (for both C- and L-moments).

C-VRS									NC-VRS								
	–	E	F	D	EF	ED	FD	EFD		–	E	F	D	EF	ED	FD	EFD
C-moments																	
–	0.00	157.81	185.59	131.60	197.71	157.65	199.49	199.86	–	0.00	183.10	203.34	134.04	202.93	183.10	203.22	202.93
E		0.00	26.64	3.49	54.74	0.00	58.09	60.09	E		0.00	26.39	28.39	25.60	0.00	26.26	25.60
F			0.00	43.24	5.83	26.71	7.50	8.42	F			0.00	88.71	0.71	26.39	0.02	0.71
D				0.00	75.21	3.79	78.31	80.68	D				0.00	88.57	28.39	88.62	88.57
EF					0.00	54.66	0.23	0.39	EF					0.00	25.60	0.56	0.00
ED						0.00	58.20	60.17	ED						0.00	26.26	25.60
FD							0.00	0.04	FD							0.00	0.56
EFD								0.00	EFD								0.00
L-moments																	
–	0.00	200.66	92.44	0.07	230.08	200.71	92.76	230.08	–	0.00	228.15	214.04	1.85	274.60	228.16	213.73	274.60
E		0.00	63.96	200.16	36.61	0.00	64.47	36.61	E		0.00	37.25	208.57	97.52	0.00	38.08	97.52
F			0.00	91.00	46.01	64.03	0.00	46.01	F			0.00	190.23	67.58	37.25	0.02	67.58
D				0.00	228.55	200.18	91.31	228.55	D				0.00	255.07	208.57	189.69	255.07
EF					0.00	36.66	46.16	0.00	EF					0.00	97.55	68.01	0.00
ED						0.00	64.54	36.66	ED						0.00	38.08	97.55
FD							0.00	46.16	FD							0.00	68.01
EFD								0.00	EFD								0.00

subsection, we explore the impact of adding higher moments up to order 6 to the basic MV-Loads model on the one hand, and the effect of systematically checking for the relevant cost components on the other hand. These two issues are systematically explored against the background of two types of maintained hypotheses: C-moments versus L-moments on the one hand, and the traditional convexity assumption versus the far less utilized non-convexity hypothesis on the other hand. We always first summarize results for the convex model, and then look at the non-convex results.

Note first that there is a natural order in the moments. This can be used for constructing higher order moment models. Obviously, since the quality of moment approximations representing EU functions is an empirical issue (Hlawitschka, 1994), the eventual irrelevance of a single moment does not imply that higher moments beyond the irrelevant one can be ignored. This leads to a sequential approach whereby each time some moment is added. This moment is retained if it leads to a significantly different efficiency distribution, otherwise it is eliminated and one tries adding the next moment.

Second, and by contrast, there is no natural order in the cost components. This order is determined using a basic MV-model solely. For this model, all possible combinations of cost components have been added. Then, the efficiency distributions computed for all these models are compared by means of Li-tests. The Li-test statistics for all models considered (both C-VRS and NC-VRS in combination with C- and L-moments) are given in Table 4 which contains in total 112 relevant model comparisons. Since the Li-test statistic measures the deviation between two efficiency distributions, it can be used to measure the impact of including a specific cost component. The higher the value of the Li-test statistic, the bigger the differences between both efficiency distributions and, hence, the larger the impact of the component included.

Using this methodology, it is possible to rank the three cost components. We explain the procedure in the case of C-VRS in combination with C-moments. Carefully checking Table 4 for this combination leads to the following ranking: the front loads (F) are the most important component, followed by the deferred loads (D) and the net expense ratio (E). Indeed, by looking at the first row with Li-test statistics in the upper left part of the table (combination of C-VRS and C-moments), the highest value (185.59) for adding one single cost component is realized by including the front loads. Therefore, we consider this to be the most important variable. Then, starting from the MV-F model (row F) and comparing it with the MV-EF and MV-FD models (columns EF and FD) being

the only models that have one additional cost component with respect to the curent MV-F model, the highest Li-test statistic is found in column FD (7.50), thereby making the deferred loads the second most important component. The remaining net expense ratio then turns out to be the least important component. Consequently, the natural decreasing order established here includes the front loads, deferred loads and net expense ratio. This ranking is denoted by FDE. A similar exercise can be performed for the other maintained hypotheses. This leads to the following rankings: FED for NC-VRS in combination with C-moments and EFD for both C-VRS and NC-VRS using L-moments.

The effect of adding significant higher order moments up to order 6 is reported in Table 5. Starting from a minimal model (i.e., a model with at least the expected return (M)) and accounting for the results regarding the three cost components established above, the higher order moments have been added one by one according to the natural order.[23,24] A moment is only retained if it results in a statistically significant different efficiency distribution with respect to the model not including it. This is again measured with a Li-test at a 5% significance level. All of these computations are again done for different maintained hypotheses (C-VRS and NC-VRS for both C- and L-moments).

Analyzing Table 5 yields the following key conclusions. First, moments of order higher than two contribute significantly for different hypotheses, except for the case of NC-VRS with L-moments and the choice of having net expense ratio and front loads in the model. For instance, in the case of NC-VRS and C-moments, the M-F model can be significantly expanded to include the second up to the sixth moment, with an exception for the fourth moment. A similar pattern emerges for other combinations of cost components.

Second, in the L-moments case, the effects of higher order moments are somewhat more modest (probably due to the increased robustness). For instance, from the absence of the sixth moment in the sequences of models with L-moments in Table 5, it can be concluded that this moment does not contribute in a significant way. Furthermore, the skewness only seems to have a significant impact

---

[23] In Table 5, we refer to the order of the moments (e.g., "123" stands for a model including mean, variance and skewness).

[24] Note that it has not been tested whether the effect of adding more than one additional moment has an impact, simply because there are too many possibilities. For instance, if adding skewness in itself has no effect, we do not test whether adding skewness and kurtosis jointly has any impact.

Table 5
Adding significant higher order moments to models with different loads.

C-VRS									NC-VRS										
C-moments									*C-moments*										
F	1	→	12	→	125	→	1256		F	1	→	12	→	123	→	1235	→	12356
FD	1	→	12	→	125	→	1256		FE	1	→	12	→	123	→	1235	→	12356
FDE	1	→	12	→	125	→	1256		FED	1	→	12	→	123	→	1235	→	12356
L-moments																		
E	1	→	12	→	125				E	1	→	12	→	123	→	1235		
EF	1	→	12	→	125				EF	1	→	12						
EFD	1	→	12	→	125				EFD	1	→	12	→	125				

Table 6
Effect of reduction of loads on different multi-moment models.

C-VRS			NC-VRS		
C-moments					
Order	FDE → FD	FD → F	Order	FED → FE	FE → F
12	E	NE	12	E	E
123	E	NE	123	E	NE
1234	E	NE	1234	E	NE
12345	E	NE	12345	E	NE
123456	E	NE	123456	E	NE
1256	E	NE	12356	E	NE
L-moments					
Order	EFD → EF	EF → E	Order	EFD → EF	EF → E
12	E	NE	12	E	NE
123	E	NE	123	E	NE
1234	E	NE	1234	E	NE
12345	E	NE	12345	E	NE
123456	E	NE	123456	E	NE
125	E	NE	125	E	NE
			1235	E	NE

Table 7
Distributional comparison of the best C-VRS and NC-VRS models.

Best C-VRS	Best NC-VRS	Result	*t*-Value
C-moments			
1256-FD	12356-FE	NE	79.12
L-moments			
125-EF	12-EF	NE	29.36

in the case of NC-VRS and L-moments and in combination with having the net expense ratio in the model.

Having initially determined the relevance of the three cost components for a given basic MV model, we now turn to test whether some of these cost components are redundant for different higher order moment models. The effect of reducing the cost components in some basic models is reported in Table 6. Again, we treat the four hypotheses: C-VRS and NC-VRS, with both C- and L-moments. The efficiency distributions of different models are also compared in this case with a Li-test at a 5% significance level. Note that we do not list the exact test statistic and its significance level in Table 6 for reasons of readability, but rather simply report whether the efficiency distributions are equal (denoted by "E") or different from one another (i.e., not equal, as denoted by "NE").

Starting with some maintained hypothesis and a model including all three cost components in the preferred order of importance, it is possible to eliminate the least important component without noticing a significant change in the efficiency distribution. Further removal of an additional cost component, however, does lead to noticeable changes in the majority of cases and models. For example, when starting in C-VRS and C-moments with the MVS-FDE model (row "123"), the net expense ratio can be removed safely. The deferred loads, however, do have an impact on the efficiency distribution and should be retained in the model. In the case of NC-VRS and C-moments, the MV-FED model can even be reduced to the MV-F model (see row "12" in the appropriate part of Table 6). Note that Tables 5 and 6 involve in total 110 Li-test statistics comparing the different modeling variations.

3.2.3. Convexity or not

Having explored the impact of higher moments up to order six and some of the cost components using moment information

based on either C- or L-moments and for given VRS models that impose either convexity or non-convexity, in this subsection we focus specifically on testing for the traditional convexity assumption. Given the different results in the previous subsection conditional on maintaining the assumption of either convexity or non-convexity, we develop the following strategy to test for this hypothesis. On the one hand, based on the results in the previous subsection, we define an encompassing model that includes all variables that turn out to be relevant in either the convex or non-convex model. On the other hand, we follow the double strategy of picking the best fitting convex (non-convex) model and then comparing it to the corresponding non-convex (convex) model results.

Picking the best C-VRS model in the case of C-moments is relatively easy. From Table 5, we notice that the first, second, fifth and sixth moment should be included. In addition, looking at Table 6 one notices in the line representing the "1256" model that one always needs the front loads and the deferred loads (FD).[25] Hence, the model "1256-FD" seems to be performing best.

The best C-VRS model in combination with L-moments and the best NC-VRS model with C-moments is determined using a similar reasoning. One obtains the models "125-EF" and "12356-FE" respectively. Determining the best NC-VRS model with L-moments is somewhat more difficult. Indeed, from Table 5, we end up with three different candidates: "1235-E", "12-EF", and "125-EFD". All three models are now checked against Table 6. First, since model "1235-E" cannot be achieved by reducing loads from the "1235-EFD" model, it is rejected as best model. Second, the model "125-EFD" can be reduced to the model "125-EF". However, since this model is not retained in Table 5, this candidate is also rejected. Third, the model "12-EF" can actually be achieved by reduction of cost components from the "12-EFD" model, as can be observed in Table 6. Consequently, this model is the only one compatible with both Tables 5 and 6. Therefore, it is considered being the best choice.

Table 7 summarizes these results and reports the Li-test statistics representing the distributional differences between the best convex and non-convex models for the C- and L-moments respectively. In both cases, convexity is clearly rejected at any conventional level of statistical significance (e.g., a 5% significance level corresponds with a *t*-value of 1.64).

---

[25] To avoid introducing additional abbreviations for the moments beyond MV, we refer to the order of the moments as these appear in Tables 5 and 6.

4. Conclusions

Starting from the current non-parametric frontier contributions assessing mutual fund performance, this contribution focuses on a series of specification issues that hitherto have been somewhat neglected in this literature. We first present and discuss the shortage function that can account for a preference for odd moments (to be increased) and an aversion to even moments (to be reduced) in multi-moment portfolio problems. This makes the shortage function a general tool for gauging mutual fund performances in line with general investor preferences (exhibiting a mixed risk aversion preference structure).

From both a portfolio theory and hedonic price function framework, we identify and discuss three crucial specification issues arising in the context of mutual fund performance gauging with non-parametric frontier models: (i) the nature of returns to scale (VRS vs. CRS), (ii) the inclusion of higher moments and cost components, and (iii) imposing convexity or not. The influence of higher order moments is tested for using both classical moments and the more robust L-moments (which are still rarely used in finance). Exploiting the relation between efficiency measures and goodness-of-fit measures, the comparison of efficiency measured using the shortage function computed relative to two non-parametric frontier models using the non-parametric test-statistic of Li (1996) amounts to a test of the hypothesis distinguishing both models involved.

Adopting the Simar (2003) methodology based upon robust frontier estimators of order-m, we manage to detect two (almost) dead funds and to correct a typo in a mutual fund price. Having deleted and corrected these real outliers, we next test VRS versus CRS in a mutual fund context: CRS is firmly rejected in both C- and L-moments. Next, we establish the order of importance of the different cost components and observe that it differs depending on the chosen hypothesis regarding convexity, except when using L-moments. Thereafter, we test for the inclusion of higher moments and check whether all cost components contribute equally well to the fit of the models.

Looking at the need for higher order moments, we find that both C-VRS and NC-VRS models need higher order moments when using C-moments but less so using L-moments, with only one exception. When trying to reduce the cost components in a structured way, all models can do without the least important cost component. In one particular case, even the second least important component can be removed as well.

Finally, we have tackled the issue of imposing convexity or not. Selecting the best fitting models for the convex and non-convex approaches, it turns out that convexity cannot be maintained. This questions the large majority of current articles in the stream of literature measuring mutual fund performance using frontier methods (except Daraio and Simar (2006)). Note that overall 224 specification tests are performed on these different model variations.

We add some final remarks as to the use of this non-convex model in a mutual fund context. While this preferred non-convex frontier specification imposes minimal assumptions, this model may yield relatively lower inefficiency scores compared to more demanding specifications. Furthermore, the probability of being relatively efficient is for this model known to increase with both decreasing sample size and/or an increasing number of variables. However, in a mutual fund context the population is potentially huge and sample sizes tend to be big. But, even if one would be interested to analyse mutual fund performance for a rather small category of mutual funds, there is always the possibility to use a local non-convex model to increase the discriminatory power (the latter model is known as the order-m estimator: see Daraio and Simar, 2006).

Obviously, the proposed methodology and the resulting empirical results have some further limitations one should be aware of. First, the mutual fund rating resulting from these non-parametric frontier models should ideally be put extensively to a comparison test with alternative rating schemes provided by companies like Lipper Analytical Services, Morningstar etc. This testing of their relative predictive power should ideally be done for a wide variety of investment horizons and other critical parameter variations. In this respect, it could be useful crossing the current framework with the proposals in Briec and Kerstens (2009) to employ a time-discounted temporal shortage function with multiple time horizons to closer mimic the different time horizons included in, e.g., the Morningstar ratings. Second, it is worthwhile looking for further developments in the robust estimation of financial time series to come up with estimates for the moment distribution that are even more robust than the L-moments adopted here (e.g., trimmed L-moments and the like).

References

Alexander, G., Baptista, A., 2010. Active portfolio management with benchmarking: a frontier based on alpha. Journal of Banking & Finance 34 (9), 2185–2197.

Anderson, R., Brockman, C., Christos, G., McLeod, R., 2004. A non-parametric examination of real estate mutual fund efficiency. International Journal of Business and Economics 3 (3), 225–238.

Bacon, C., 2008. Practical Portfolio Performance Measurement and Attribution, second ed. Wiley, New York.

Barros, C., Garcia, M., 2006. Performance evaluation of pension funds management companies with Data Envelopment Analysis. Risk Management and Insurance Review 9 (2), 165–188.

Basso, A., 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., Funari, S., 2003. Measuring the performance of ethical mutual funds: a DEA approach. Journal of the Operational Research Society 54 (5), 521–531.

Blake, C., Morey, M., 2000. Morningstar ratings and mutual fund performance. Journal of Financial and Quantitative Analysis 35 (3), 451–483.

Blake, D. (Ed.), 2000. Finance: A Characteristics Approach. Routledge, London.

Brandouy, O., Briec, W., Kerstens, K., Van de Woestyne, I., 2010. Portfolio performance gauging in discrete time using a Luenberger productivity indicator. Journal of Banking & Finance 34 (8), 1899–1910.

Briec, W., Kerstens, K., 2009. Multi-horizon Markowitz portfolio performance appraisals: a general approach. Omega 37 (1), 50–62.

Briec, W., 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., Jokung, K., 2007. Mean–variance–skewness portfolio performance gauging: a general shortage function and dual approach. Management Science 53 (1), 135–149.

Briec, W., Kerstens, K., Lesourd, J., 2004. Single-period Markowitz portfolio selection, performance gauging, and duality: a variation on the Luenberger shortage function. Journal of Optimization Theory and Applications 120 (1), 1–27.

Brockett, P., Golden, L., 1987. A class of utility functions containing all the common utility functions. Management Science 33 (8), 955–964.

Capon, N., Fitzsimons, G., Prince, R., 1996. An individual level analysis of the mutual fund investment decision. Journal of Financial Services Research 10 (1), 59–82.

Chambers, R., Chung, Y., Färe, R., 1996. Benefit and distance functions. Journal of Economic Theory 70 (2), 407–419.

Charnes, A., Cooper, W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of operational Research 3 (4), 338–339.

Choi, Y., Murthi, B., 2001. Relative performance evaluation of mutual funds: a nonparametric approach. Journal of Business Finance and Accounting 28 (7–8), 853–876.

Chumpitaz, R., Kerstens, K., Paparoidamis, N., Staat, M., 2010. Hedonic price function estimation in economics and marketing: revisiting Lancaster's issue of "noncombinable" goods. Annals of Operations Research 173 (1), 145–161.

Daraio, C., 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.

Darolles, S., Gourieroux, C., 2010. Conditionally fitted Sharpe performance with an application to hedge fund rating. Journal of Banking & Finance 34 (3), 578–593.

De Palma, A., Myers, G., Papageorgiou, Y., 1994. Rational choice under an imperfect ability to choose. American Economic Review 84 (3), 419–440.

Del Guercio, D., Tkac, P., 2008. Star power: the effect of Morningstar ratings on mutual fund flows. Journal of Financial and Quantitative Analysis 43 (4), 907–936.

Eeckhoudt, L., Schlesinger, H., 2006. Putting risk in its proper place. American Economic Review 96 (1), 280–289.

Ekeland, I., Heckman, J., Nesheim, L., 2004. Identification and estimation of hedonic models. Journal of Political Economy 112 (1), 60–109.

Eling, M., 2006. Performance measurement of hedge funds using Data Envelopment Analysis. Financial Markets and Portfolio Management 20 (1), 442–471.

Eling, M., Faust, R., 2010. The performance of hedge funds and mutual funds in emerging markets. Journal of Banking & Finance 34 (8), 1993–2009.

Fabozzi, F., Kolm, P., Pachamanova, D., Focardi, S., 2007. Robust Portfolio Optimization and Management. Wiley, New York.

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

Färe, R., Grosskopf, S., 1995. Nonparametric tests of regularity, Farrell efficiency and goodness of fit. Journal of Econometrics 69 (1), 415–425.

Farinelli, S., Ferreira, M., Rossello, D., Thoeny, M., Tibiletti, L., 2008. Beyond Sharpe ratio: optimal asset allocation using different performance ratios. Journal of Banking & Finance 32 (10), 2057–2063.

Farrell, M., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society 120A (3), 253–281.

Galagedera, D., Silvapulle, P., 2002. Australian mutual fund performance appraisal using Data Envelopment Analysis. Managerial Finance 28 (9), 60–73.

Glawischnig, M., 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., 2003. Performance appraisal of funds of hedge funds using Data Envelopment Analysis. Journal of Wealth Management 5 (4), 88–95.

Gregoriou, G., 2006. Optimisation of the largest US mutual funds using Data Envelopment Analysis. Journal of Asset Management 6 (6), 445–455.

Gregoriou, G. (Ed.), 2007. Performance of Mutual Funds: An International Perspective. Palgrave, New York.

Haslem, J., 2003. Mutual Funds: Risk and Performance Analysis for Decision Making. Blackwell, Oxford.

Haslem, J., Scheraga, C., 2006. Data Envelopment Analysis of Morningstar's small-cap mutual funds. Journal of Investing 15 (1), 87–92.

Heffernan, S., 1990. A characteristics definition of financial markets. Journal of Banking & Finance 14 (2–3), 583–609.

Hjorth-Andersen, C., 1983. Lancasters principle of efficient choice: an empirical note. International Journal of Industrial Organization 1 (2), 287–295.

Hlawitschka, W., 1994. The empirical nature of Taylor-series approximations to expected utility. American Economic Review 84 (3), 713–719.

Hosking, J., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society 52B (1), 105–124.

Kamakura, W., Ratchford, T., Agrawal, J., 1988. Measuring market efficiency and welfare loss. Journal of Consumer Research 15 (3), 289–302.

Kerstens, K., Van de Woestyne, I., 2010. Negative data in DEA: a simple proportional distance function approach. Journal of the Operational Research Society. doi:10.1057/jors.2010.108.

Khorana, A., Servaes, H., Tufano, P., 2008. Mutual funds fees around the world. Review of Financial Studies 22 (3), 1279–1310.

Lancaster, K., 1966. A new approach to consumer theory. Journal of Political Economy 74 (1), 132–157.

Lee, J., Hwang, S., Kim, T., 2005. The measurement of consumption efficiency considering the discrete choice of consumers. Journal of Productivity Analysis 23 (1), 65–83.

Lehmann, B., Timmermann, A., 2008. Performance measurement and evaluation. In: Thakor, A., Boot, A. (Eds.), Handbook of Financial Intermediation and Banking. Elsevier, Amsterdam, pp. 191–258.

Li, Q., 1996. Nonparametric testing of closeness between two unknown distribution functions. Econometric Reviews 15 (3), 261–274.

Lintner, J., 1965. Security prices, risk, and maximal gains from diversification. Journal of Finance 20 (4), 587–615.

Lovell, C., Pastor, J., 1995. Units invariant and translation invariant DEA models. Operations Research Letters 18 (3), 147–151.

Lozano, S., Guttiérez, E., 2008. TSD-consistent performance assessment of mutual funds. Journal of the Operational Research Society 59 (10), 1352–1362.

Luenberger, D., 1992. Benefit functions and duality. Journal of Mathematical Economics 21 (5), 461–481.

Luenberger, D., 1995. Microeconomic Theory. McGraw-Hill, Boston.

Markowitz, H., 1952. Portfolio selection. Journal of Finance 7 (1), 77–91.

McMullen, P., Strong, R., 1998. Selection of mutual fund using Data Envelopment Analysis. Journal of Business and Economic Studies 4 (1), 1–14.

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

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

Pastor, J., Ruiz, J., 2007. Variables with negative values in DEA. In: Zhu, J., Cook, W. (Eds.), Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis. Springer, Berlin, pp. 63–84.

Premachandra, I., Powell, J., Shi, J., 1998. Measuring the relative efficiency of fund management strategies in New Zealand using a spreadsheet-based stochastic Data Envelopment Analysis model. Omega 26 (2), 319–331.

Ray, S., 2004. Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research. Cambridge University Press, Cambridge.

Rosen, S., 1974. Hedonic prices and implicit markets: production differentiation in pure competition. Journal of Political Economy 82 (1), 34–55.

Sharpe, W., 1963. A simplified model for portfolio analysis. Management Science 9 (2), 277–293.

Sharpe, W., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 19 (3), 425–442.

Sharpe, W., 1998. Morningstar's risk-adjusted ratings. Financial Analysts Journal 54, 21–33.

Shephard, R., 1978. On household production theory. In: Eichhorn, W. et al. (Eds.), Theory and Applications of Economic Indices. Physica Verlag, Wurzburg, pp. 449–485.

Simar, L., 2003. Detecting outliers in frontier models: a simple approach. Journal of Productivity Analysis 20 (3), 391–424.

Sirri, E., Tufano, P., 1998. Costly search and mutual fund flows. Journal of Finance 53 (5), 1589–1622.

Smith, D., 2009. The economics of mutual funds. In: Haslem, J. (Ed.), Mutual Funds: Portfolio Structures, Analysis, Management, and Stewardship. Wiley, New York, pp. 33–49.

Tolikas, K., Gettinby, G., 2009. Modelling the distribution of the extreme share returns in Singapore. Journal of Empirical Finance 16 (2), 254–263.

Wilkens, K., Zhu, J., 2001. Portfolio evaluation and benchmark selection: a mathematical programming approach. Journal of Alternative Investments 4 (1), 9–19.

Zakamouline, V., Koekebakker, S., 2009. Portfolio performance evaluation with generalized Sharpe ratios: beyond the mean and variance. Journal of Banking & Finance 33 (7), 1242–1254.

Zheng, L., 2008. The behavior of mutual fund investors. In: Thakor, A., Boot, A. (Eds.), Handbook of Financial Intermediation and Banking. Elsevier, Amsterdam, pp. 259–283.