Efficiency and Productivity Growth

STATISTICS IN PRACTICE

Statistics in Practice is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceutics; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges.

It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series appears at the end of the volume.

Efficiency and Productivity Growth
Modelling in the Financial Services Industry

Edited by

Fotios Pasiouras

University of Surrey, UK and
Technical University of Crete, Greece

WILEY
A John Wiley & Sons, Ltd., Publication

This edition first published 2013

Registered Office
John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

Pasiouras, Fotios.
Efficiency and productivity growth : modelling in the financial services industry / edited by Fotios Pasiouras.
pages cm
Includes bibliographical references and index.
ISBN 978-1-119-96752-1 (cloth)
1. Banks and banking–Econometric models. 2. Financial services industry–Econometric models. I. Title.
HG1601.P29 2013
332.101′5195–dc23
2012045881

A catalogue record for this book is available from the British Library.

ISBN: 978-1-119-96752-1

Set in 10/12pt Times by SPi Publisher Services, Pondicherry, India

In memory of my best friend, Manos Kerpinis.

He was a remarkable engineer and operations manager, always trying to operate at the frontier.

Contents

7

Backtesting superfund portfolio strategies based on frontier-based mutual fund ratings

Olivier Brandouy,[1] Kristiaan Kerstens,[2] and Ignace Van de Woestyne[3]

[1] *IAE – Sorbonne Graduate Business School, Université Paris 1, France*
[2] *CNRS-LEM (UMR 8179), IESEG School of Management, France*
[3] *Hogeschool-Universiteit Brussel, Belgium*

7.1 Introduction

Professional mutual fund managers and individual investors face an increasingly large set of investment opportunities. Among these opportunities, the subset of mutual funds (MF) offers an amazing heterogeneity. For example, Morningstar, one of the best established rating agencies in this business (one can also cite Lipper, Standard & Poor's, Fitch Ratings, EuroPerformance, etc.), offers access to more than 160 000 MF all over the world through its database 'Morningstar Direct' (about 77 000 in Europe and about 26 000 in the United States). Consequently, the ratings offered by these rating agencies, and particularly their reliability, are of major interest for investors considering selecting some of these MF shares for their portfolios.

For example, Blake and Morey (2000) examine the Morningstar rating as a predictor of US domestic equity MF performance: these authors find little evidence that Morningstar's top-rated funds outperform the second- and third-rated funds. More recently, Kräussl and Sandelowsky (2007) offer a much more extensive study: not only US domestic equity MF, but also international equities, taxable bonds, and municipal bonds as well, with not a 5-year but a 10-year horizon. Their sample allows assessing Morningstar's revised rating since July 2002 when ratings got based on 64 categories rather than four broad asset classes till then.

Efficiency and Productivity Growth: Modelling in the Financial Services Industry, First Edition. Edited by Fotios Pasiouras.

The upshot of their study is that Morningstar's rating till July 2002 can predict severe underperformance, but cannot discriminate between 3- to 5-star-rated MF. But, the results of the revised rating system are even worse: there is no significant performance difference in out of sample periods between 1- and 5-star MF. The merits of the 2011 Morningstar announcement of a new so-called analyst rating that intends to look forward rather than backward and that will coexist with the current ratings obviously remains to be evaluated.

Since about two decades, there have been successive attempts to transpose the successful frontier estimation methodologies from production theory to the analysis of financial problems. Sengupta (1989) is probably the first to introduce an explicit efficiency measure into a mean-variance (MV) portfolio model. In a MF rating context, the seminal article proposing some efficiency measure is Murthi, Choi, and Desai (1997). But, it is likely the article by Morey and Morey (1999) proposing both a mean-return expansion and a risk-contraction function that has in fact triggered a series of new developments in the use of efficiency measures in portfolio theory and in MF rating in particular (e.g., Briec, Kerstens, and Lesourd, 2004; Lamb and Tee, 2012; Lozano and Guttiérez, 2008, among others). Anyway, these developments have led to a burgeoning literature of about 40 or so related articles, published in a variety of journals and covering a collection of MF types (ethical MF, hedge funds, pension funds, etc.). Somewhat in parallel, starting with the seminal article of Alam and Sickles (1998) one can mention an emerging literature providing frontier-based asset selection and integration in portfolio models (e.g., Abad, Thore, and Laffarga, 2004 or Nguyen and Swanson, 2009).

Starting from the foundations of Modern Portfolio Theory, an enormous literature on portfolio performance evaluation has been developed using total-risk foundations (e.g., the standard deviation or variance of returns). Among the wide variety of financial performance indexes, one classic is the Sharpe ratio (also known as the reward-to-volatility). A recent survey summarizing these and more recent developments of financial and portfolio performances gauges is found in the book by Bacon (2008).

While a MV utility-maximizing agent should aim at a portfolio with the highest reward-to-risk ratio (a tangency portfolio or maximum Sharpe ratio portfolio), very few operational procedures in finance are in fact available that guarantee such a position on the MV frontier. For instance, market-cap-weighted indices are known to provide inefficient risk–return trade-offs. This has recently led to a stream of articles aimed at making hitherto inefficient benchmarking indices efficient (see, e.g., Martellini, 2008 or Clark, Jokung, and Kassimatis, 2011). Independently of the above frontier-based literature applied to financial topics, in the finance literature some authors have proposed to define a relative performance measure using the portfolio frontier as benchmark. For instance, the work of Cantaluppi and Hug (2000) proposes an efficiency ratio in relation to the MV-efficient frontier and contests the arbitrary and nonfrontier nature of most current proposals that define performance with respect to some other, supposedly relevant, portfolio or index. These authors suggest looking for the maximum performance that could have been achieved by a given portfolio relative to a relevant portfolio frontier. Also Broihanne, Merli, and Roger (2008) stress the importance of measuring performance relative to a portfolio frontier and underline that such an approach avoids the choice of a risk-free rate and a market portfolio as some kind of absolute benchmark. While this review may be incomplete, it is an understatement to conclude that the use of relative performance efficiency measures evaluated relative to some frontier is marginal at best in the current finance literature.

The purpose of this chapter is not to offer a complete overview of these rather recent frontier applications in a MF rating context. For instance, it completely ignores the question of the determinants of MF performance (examples include sector- and country-specific factors, such as different time paths of technological innovations, fluctuations in the macro-economic environment, etc.). Also the question about the precise attribution of the role of the MF manager, traditionally focusing on stock picking and market timing capabilities, is sidestepped. Instead, the goal is to offer the first detailed backtesting analysis of these frontier-based MF ratings compared to the Morningstar rating (representative for the fund rating agencies) and some traditional financial performance measures. While this is not the first backtesting analysis of frontier MF ratings (see, e.g., Matallín, Soler, and Tortosa-Ausina, 2011), it is – to the best of our knowledge – the most extensive backtesting analysis provided in the literature focusing on the relative merits of different backward-looking performance rating tools in predicting future MF performance.

7.2 Frontier-based mutual funds rating models

7.2.1 A taxonomy

Given the rather widespread criticisms of traditional financial performance measures, several authors have recently been introducing nonparametric frontier methods to assess MF performance.[1] Following Tsolas (2011), it is useful to distinguish between four different modeling approaches:

(i) Models directly transposed from production theory;

(ii) Models combining traditional performance measures (e.g., Sharpe ratio) with additional dimensions;

(iii) Models directly transposed from portfolio theory;

(iv) Hedonic price models (a new proposal launched in Kerstens, Mounir, and Van de Woestyne, 2011b).

We discuss each of these modeling approaches in turn to develop our own selection of models put at a test in the Empirical section.

(i) Frontier models can handle multiple dimensions and yield a single efficiency measure with respect to a frontier composed of similar units. Directly transposing these basic models from production theory to financial applications, there seems to be a wide variation of specifications around in the literature. A rather up-to-date and fairly comprehensive review of this literature is offered in Glawischnig and Sommersguter-Reichmann (2010). A mix of traditional and frontier-based performance studies as well as an international perspective is provided in the Gregoriou (2007) book.

Without any ambition of completeness or representativeness, Table 7.1 reviews a limited selection of articles. Several observations can be made. First, apart from a variety of costs,

---

[1] Stochastic frontier models have been more rarely employed for MF rating purposes: an example is Annaert, van den Broeck, and Vander Vennet (2003). However, notice that little of what is developed in this chapter is really conditioned by the use of some specific frontier estimation method.

Table 7.1 Frontier models of MF performance: a selection.

Article	Inputs	Outputs	Model
Basso and Funari (2003)	Subscription and redemption cost Beta Standard deviation returns Standard semi-deviation returns	Return (or excess return) Ethical score	VRS
Choi and Murthi (2001)	Standard deviation returns Expense ratio Loads Turnover ratio	Gross return	VRS
Galagedera and Silvapulle (2002)	1, 2, 3, and 5 year standard deviation returns Sales charges Expense ratio Minimum investment	1, 2, 3, and 5 year gross return	VRS
Glawischnig and Sommersguter-Reichmann (2010)	Mean lower return Lower mean semi-variance Lower mean semi-skewness	Mean upper return Upper semi-variance Upper semi-skewness	VRS
McMullen and Strong (1998)	Standard deviation returns Sales charges Expense ratio Minimum investment	1, 3, and 5 year annual returns	VRS
Murthi, Choi, and Desai (1997)	Standard deviation returns Expense ratio Loads Turnover	Gross excess return	CRS
Wilkens and Zhu (2001)	Standard deviation returns % Negative monthly returns per year	Average return Skewness Minimum return	VRS

there seems to be a need for models accounting for higher-order moments (especially for hedge funds, and the like). Most articles we are aware of include at most up to the third moment (e.g., Glawischnig and Sommersguter-Reichmann, 2010, or Wilkens and Zhu, 2001). Second, some articles use standard moments (e.g., Galagedera and Silvapulle, 2002), while other articles use lower or upper partial moments or some combination of these partial moments (e.g., Glawischnig and Sommersguter-Reichmann, 2010), or even combine standard and partial moments (e.g., Basso and Funari, 2003). Third, there are quite a few attempts to offer assessments over several time horizons simultaneously. For instance, McMullen and Strong (1998) include one-, three-, and five-year annualized returns, but do not apply these same time horizons to the other factors in their model. A more consistent model in this respect is probably the article by Morey and Morey (1999) (see also (iii)) who evaluate both returns and standard deviations over three time horizons simultaneously.

This literature suffers in our view from a variety of problems. First, by defining some frontier on a sample of MF, one in fact seems to target at defining a series of superfunds to assess in turn each of the underlying MF. But, such an approach ignores diversification and interaction effects among moments. It could at best be considered a type of linear approximation of a possibly nonlinear portfolio model. An open question of how good such approximation turns out to be then remains.

Second, these nonparametric frontier-based articles assessing MF raise three crucial specification issues (following Kerstens, Mounir, and Van de Woestyne, 2011b). First, the very notion of returns to scale may not necessarily be directly transposed to the finance context. Second, the inclusion of higher-order moments is sometimes rather arbitrarily stopped at some order and rarely is based on a systematic moment approximation strategy. And third, related to the first problem, the hypothesis of convexity may or may not provide a suitable linear approximation. But, if it offers just an approximation, then an open question is the meaning of the resulting projection points. Depending on the closeness of the approximation, such projection points can or cannot in fact be obtained by an investor. If so, then how can these projection points be of any help to the investor looking for advice from MF ratings?

(ii) Some frontier models combine a traditional financial performance measure with some additional variables. There is some variation of specifications: for instance, Basso and Funari (2001, 2003), Chang (2004), and Sengupta and Zohar (2001) add the Beta coefficient from the Capital Asset Pricing Model (CAPM), while Haslem and Scheraga (2003) include a Sharpe index. In fact, such models can be interpreted as a special case of the first category.

Being just a special case of the first category, these models in our opinion share all of their defects. In addition, there is the risk of one major interpretational problem: what does the efficiency measure in such a frontier, being a combination of a traditional financial performance measure and some additional variables, mean? For instance, Haslem and Scheraga (2003) define an input-oriented frontier model with the Sharpe index as the single output combined with a series of inputs (% cash, % stocks, total assets, expense ratio, price/earnings, and price/book ratios). By contrast, the articles of Basso and Funari (2001, 2003) pay a lot of attention to the interpretation of the resulting efficiency measure.

Specifically for the use of Beta in some of these articles, it is widely recognized that CAPM does a poor job at explaining gross returns (even in a MV world). This has led to alternative multifactor models, some of which are theoretically grounded (e.g., Arbitrage Pricing Theory) and others that are just empirically validated (so-called Fama-French models).[2] Furthermore, CAPM presupposes normality and relates a MF to a market portfolio

[2] See, e.g., Peterson Drake and Fabozzi (2010) for more background information.

benchmark approximated by a factually suboptimal market index. Beta as such measures undiversifiable risk with respect to this supposedly universal market benchmark. Thus, it has little meaning in a nonnormal world that many of these frontier models aim to focus on. Obviously, depending on the number of moments to be integrated in the frontier model, one could eventually consider looking at 3CAPM and 4CAPM (e.g., Jurczenko and Maillet, 2006). The latter possibilities remain to be explored, subject to the caveats already mentioned.

(iii) Starting with Sengupta (1989) and especially Morey and Morey (1999), portfolio models with either return-expansion or risk-contraction efficiency measures have been developed that can handle diversification and interaction effects among eventually multiple moments and/or multiple time horizons. For instance, Zhao, Wang, and Lai (2011) develop a special MV model that includes Beta and that aims at increasing both return and return above the market benchmark, while Lamb and Tee (2012) develop an iterative approximation procedure to deal with the traditional failure of frontier models to cope with diversification.

These efficiency measures have been generalized by Briec, Kerstens, and Lesourd (2004) who propose transposing the shortage function as an efficiency measure into the MV model and who also develop a dual framework to assess the satisfaction of MV preferences. By analogy to other fields, a decomposition of portfolio performance into allocative and portfolio (technical) efficiency is proposed. The key advantage is that this shortage function is perfectly compatible with a general mixed risk aversion preference structure (i.e., a preference for augmenting odd moments and reducing even moments). Therefore, the approach can be extended to higher-dimensional portfolio spaces: for example, MV-skewness (MVS) space (see Briec, Kerstens, and Jokung, 2007) or even higher-order models (see Briec and Kerstens, 2010). Another advantage is that a slight variation on the shortage function (see Kerstens and Van de Woestyne, 2011) can handle negative data that occur naturally in a financial context, while maintaining a proportional interpretation – a convenience for practitioners.

The article by Morey and Morey (1999) on multihorizon return-expansion or risk-contraction MV frontier models has been extended into a more general time-discounted model based on the shortage function in Briec and Kerstens (2009). Brandouy et al. (2010a) propose a Luenberger portfolio productivity indicator to estimate changes in the relative positions of portfolios with respect to the traditional MV frontier, and the eventual shifts of the frontier over time. Analyzing the local changes relative to these MV (or higher-moment) portfolio frontiers, this productivity indicator separates performance changes due to portfolio strategies and performance changes due to the market evolution. Applications of this new portfolio modeling approach include reconstructions of MVS portfolio sets (see Kerstens, Mounir, and Van de Woestyne (2011a) for the general-moment case and Brandouy, Kerstens, and Van de Woestyne (2010b) for the lower partial-moment case), comparisons with alternative portfolio models (e.g., Lozano and Guttiérez, 2008), and applications to hedge fund industry (see Jurczenko and Yanou, 2010).

While the promise of this new multimoment approach to portfolio analysis remains to be thoroughly assessed, the application of such portfolio approach to rate MF could face at least the following two problems. First, there is the choice of the MF universe that is unclear. Should one be comparing MF among themselves in different classes? Or, should one be comparing the MF relative to some supposedly common underlying asset universe for a given class of MF (as if the investor could pick his/her own MF from the underlying assets)? Second, for even small classes of MF, the computational burden is very high when adding higher moments and/or multiple time horizons. For some of the larger MF classes of considerable size, the computational burden could be prohibitive, even with today's computing power. Probably, this makes these models currently unsuitable to implement in practice.

(iv) A new proposal launched in Kerstens, Mounir, and Van de Woestyne (2011b) is to found MF rating on a hedonic pricing model. Based on the characteristics' approach to consumer theory, utility is a function of the characteristics of products and services, not a function of the goods vector itself. The study of market equilibria for heterogeneous goods differing along multiple characteristics derives an implicit price for the vector of observed characteristics aggregating these into a single measure of market value. While the estimation of price–qualities functions and frontiers has become quite common in consumer and marketing applications, the use of this characteristics' approach in the finance literature remains rather rare.[3] In finance, authors like Blake (1990), Dodds (1986), Heffernan (1992), among others, have argued in favor of interpreting several financial products in terms of a variety of characteristics.

In essence, the proposal of Kerstens, Mounir, and Van de Woestyne (2011b) is to consider MF as fee-based (e.g., various loads) financial products characterized by some distributional characteristics as summarized by a combination of moments. Thus, the traditional price–qualities frontier of consumer goods and services simply becomes a fee–moments frontier for MF: the investor pays a series of loads and fees to benefit of some future, uncertain return when he/she resells his/her shares depending on the investment horizon.

This viewpoint disregards any repercussions at the portfolio level. In fact, given investor preferences for higher-order moments and the current lack of any summary performance measure in this multimoment portfolio theory, one could argue that any private or professional investor can only judge the suitability of any financial product, such as a MF, by explicitly testing whether it could improve his/her particular portfolio. The MF rating based on the hedonic pricing model then only has the modest aim to offer a list of potentially suitable candidates of MF to consider for inclusion in any given portfolio.

As argued at length in Kerstens, Mounir, and Van de Woestyne (2011b), there are important issues to tackle when constructing hedonic rating models. Firstly, there is the issue of specifying the nature of returns to scale to be used in the model. From the existing literature, we observe that variable returns to scale (VRS) and constant returns to scale (CRS) are quite common assumptions. Kerstens, Mounir, and Van de Woestyne (2011b) discuss this issue of choosing the proper returns to scale in detail and conclude that VRS is more meaningful from a theoretical perspective. Furthermore, their empirical tests also favor VRS to CRS. Secondly, one can opt for either convex or nonconvex models. Kerstens, Mounir, and Van de Woestyne (2011b) raise the issue of indivisibilities, that is, the impossibility of combining some dimensions of several MF because this ignores diversification effects (see also the discussion in (i)). This leads to a preference for a basic nonconvex model avoiding the interpretation problems raised above that projection points should be

[3] Friedman (1983) offers a detailed survey and critique on the use of the characteristics' approach in various industries (see his ch. 4).

feasible. On occasion, also frontier studies from the first category apply a nonconvex model (see, e.g., Matallín, Soler, and Tortosa-Ausina, 2011).

While such a basic nonconvex model in a production context sometimes leads to a majority of efficient units, yielding a lack of discrimination, the potentially huge amount of observations around in most financial settings creates the hope to mitigate this problem. Anyway, Kerstens, Mounir, and Van de Woestyne (2011b) preliminary empirical tests favor nonconvex compared to convex models with a variety of higher moments.

In summary of these four different modeling approaches, the upshot of this discussion is that especially the first and the fourth approaches show some promise for practical MF rating purposes. Therefore, we consider both the traditional convex and nonconvex VRS models when developing our empirical research strategy.

7.2.2 MF frontier rating models retained

Following this conclusion, we now introduce the convex and nonconvex VRS-based shortage function (sometimes also known under the name directional distance function). Assume the set of n MF under evaluation is indexed by $j=1,\ldots,n$. Each MF is characterized by m input-like values x_{ij}, $i=1,\ldots,m$ and s output-like values y_{rj}, $r=1,\ldots,s$. Assume that MF $o\in\{1,\ldots,n\}$ needs to be evaluated. Then, its inefficiency determined by the convex VRS-based shortage function is obtained by solving a mathematical programming problem:

$$\begin{aligned}\max \lambda \ \textbf{s.t.} \quad & \sum_{j=1}^{n} z_j y_{rj} \geq y_{ro} + \lambda |y_{ro}|, \qquad r=1,\ldots,s,\\ & \sum_{j=1}^{n} z_j x_{ij} \leq x_{io} - \lambda |x_{io}|, \qquad i=1,\ldots,m,\\ & \lambda \geq 0, \forall j=1,\ldots,n : z_j \geq 0, \sum_{j=1}^{n} z_j = 1.\end{aligned} \tag{7.1}$$

The inefficiency determined by the nonconvex VRS-based shortage function can be computed with the help of the following mathematical program:

$$\begin{aligned}\max \lambda \ \textbf{s.t.} \quad & \sum_{j=1}^{n} z_j y_{rj} \geq y_{ro} + \lambda |y_{ro}|, \qquad r=1,\ldots,s,\\ & \sum_{j=1}^{n} z_j x_{ij} \leq x_{io} - \lambda |x_{io}|, \qquad i=1,\ldots,m,\\ & \lambda \geq 0, \forall j=1,\ldots,n : z_j \in \{0,1\}, \sum_{j=1}^{n} z_j = 1.\end{aligned} \tag{7.2}$$

Notice that both models specify the direction vector following the proposal made by Kerstens and Van de Woestyne (2011). This specification allows to have negative data values, which is indeed the case when examining periods involving a financial crisis (e.g., negative returns). Furthermore, both models project the MF o under evaluation in the direction $g=(-|x_{1o}|,\ldots,-|x_{mo}|,|y_{1o}|,\ldots,|y_{so}|)$. Consequently, all output-like values, y_{ro}, $(r=1,\ldots,s)$,

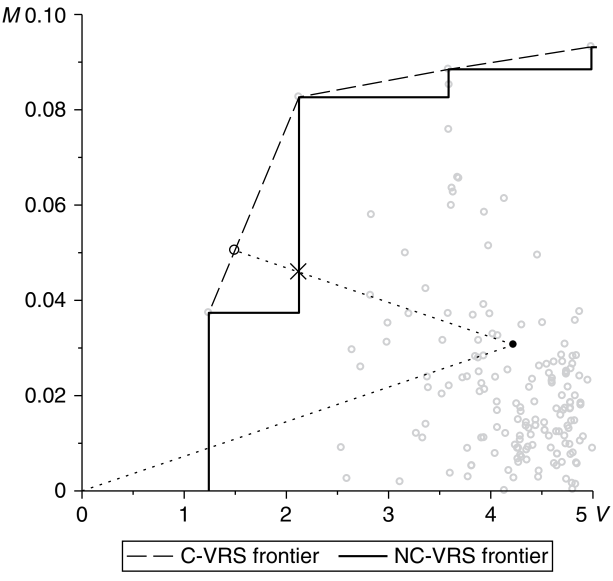

Figure 7.1 Visualization of convex and nonconvex frontiers determined by 814 MF in the main database: Expected return and variance are considered as output and input, respectively.

are increased and the input-like values, $x_{io}, (i=1,\ldots,m)$, are decreased simultaneously in proportion to their initial values. Furthermore, λ measures the amount of inefficiency (i.e., larger values indicate less efficient MF), while an efficient MF has a shortage function value of 0 ($\lambda = 0$). Finally, note that model (7.1) results in a linear programming (LP) problem, while model (7.2) leads to a mixed integer programming (MIP) problem.

As an example, Figure 7.1 visualizes the convex (dashed line) and nonconvex frontiers (solid line) determined by the 814 MF (gray circles) from the main database of this research (see Section 7.3.3). For these funds, the expected return and variance computed over the first three-year period are considered as output and input, respectively. This time window of three years used in these computations is identical to the one employed by Morningstar in its three-year rating. This particular restriction to one input and one output is considered here only for visualization purposes. In the procedure described later (see Section 7.3.1), multiple inputs and/or outputs are considered. In order to focus on the interesting region of the frontiers, Figure 7.1 only shows the area with positive returns and a variance limited to 5. Consequently, not all initial MF are visible here.

To show the projecting capabilities of the shortage function, MF number 502 ('LBBW Dividenden Strategie Euroland R') is located as a black solid circle. For this fund, the expected return is equal to 0.0307 while the variance equals 4.2238. The optimization program determined by model (7.1) returns the inefficiency value of λ_C=0.6457. From this value, the location of the optimal frontier position (visible as the black circled point) can be derived easily following the right-hand sides of the inequality constraints in model (7.1). This resulting optimal point has an expected return of 0.0505 and a variance of 1.4964. Note that the return is increased while the variance is decreased compared to its initial position. Analogously, when considering the nonconvex model (7.2), one obtains an inefficiency value of $\lambda_{NC}=0.4973$ leading to the optimal frontier point with an expected return of 0.0459 and a variance of 2.1235. This point is shown in Figure 7.1 as a diagonal cross. Also here, the expected return is increased, while the variance is reduced. Note that, due to the shape of the nonconvex frontier in relation with that of the convex frontier, the 'distance' from the initial position to the optimal frontier point is larger in the convex case compared with the nonconvex case: this explains the relation $\lambda_C > \lambda_{NC}$.

Intensive comparative testing of these two selected approaches, a selection of more traditional financial performance measures, as well as rating agencies ratings are needed to assess which MF rating methodology turns out to be better. It is exactly to this exercise that we now turn.

7.3 Backtesting setup, data description, and frontier-based portfolio models

In this section, we evaluate several strategies using a backtesting approach. Backtesting consists in simulating financial strategies to assess their potential merit over historical data sets. On the one hand, in doing so one maintains the assumption that their factual implementation during past periods and market conditions would not have had an impact on historical prices. This conception is very neoclassical: it suggests a price taker framework describing how prices emerge in thick financial markets. For example, such a framework is widely used in risk management (e.g., for evaluating Value at Risk models: see Campbell, 2005), or in portfolio management for comparing several investment strategies (for instance, DeMiguel, Garlappi, and Uppal (2009) or Tu and Zhou (2011) compare a naïve diversification strategy against several more or less sophisticated optimization strategies). On the other hand, this neoclassical approach appears to be inconsistent for those claiming that prices may well be influenced by factual implementation of the backtested investment strategies (see, for a general example of a microscopic simulation, Levy, Levy, and Solomon, 1995). In line with this conception, an ecological competition approach is more suitable (see, e.g., Sprott, 2004 or Brandouy, Mathieu, and Veryzhenko, 2012).

In the following, we present a large number of strategies that are backtested in this chapter. Note that we have opted for a simple, direct backtesting approach illustrating the outcomes of different models over a given time window rather than, for example, exploring massive simulations of a single strategy over different time windows. Apart from Matallín, Soler, and Tortosa-Ausina (2011) to which we turn in some detail below, we are unaware of any other frontier study of MF involving backtesting.

7.3.1 Backtesting setup

We now consider two major backtesting strategies. The first strategy is called 'EWP without optimization' (where EWP stands for equally weighted portfolio) and composes for each period considered the EWP of the m best performing MF. For obtaining these best performing funds, a ranking is made according to some performance measure. This can be a particular hedonic rating model, or one of the traditional performance indicators (such as Morningstar's stars or, for example, the Sharpe, Sortino, Treynor, Kappa, or Omega ratios).

The second strategy is called 'EWP with optimization' and starts off by identifying an EWP just like in the first strategy. Rather than using these portfolios directly as a guideline for investment, these EWPs are now considered as starting points for an additional portfolio optimization process based on the position-dependent shortage function. This shortage function then projects the EWP onto an optimal portfolio frontier derived from a universe that is limited to only those MF present in the EWP. Three portfolio models are considered: the traditional MV model, a multimoment MVS model, and a MVS-kurtosis (MVSK) model. Since for each period the universe is restricted to only those MF considered in the EWP, the shortage function ensures that the projected optimal portfolio is also composed of only these same underlying MF. Consequently, this additional optimization process only changes the weights of the MF present in the EWP, not the set of MF considered.

An overview of all detailed backtesting scenarios considered further on in Section 7.3 is found in the first column of Table 7.3. The notation consists of two parts, that is, the parts before and after ‘to’. The first part of the notation indicates what model is used for ranking the MF for selecting the *m* best ones. This can be done using a convex (indicated with ‘C’) hedonic rating model or a nonconvex (indicated with ‘NC’) model. For example, ‘MVS + L-C’ refers to the convex model with expected returns, variance, skewness, and the loads/fees selected. Apart from a hedonic rating system, more traditional indicators are considered as well. In particular, we include the Morningstar rating and the traditional performance measures: Sharpe, Sortino, Treynor, Kappa, and Omega ratios.[4] For example, ‘Stars’ refers to a ranking based on the three-year Morningstar ratings. The second part of the notation refers to one of the three portfolio models additionally used in the case of EWP with optimization (i.e., MV, MVS, or MVSK). Since the first strategy, that is, EWP without optimization, involves no portfolio model in the second stage, this is indicated with ‘NoOpt’ in the second part. Proceeding in this way, in total 48 backtesting scenarios can be identified for further investigation.

Note that in the backtesting scenarios considered in Section 7.3, a selection of the 10, 20, 30, 40, or 50 best open-end funds is considered. In the case of ties (e.g., in the Morningstar ratings or particularly when using nonconvex hedonic models), MF are selected randomly among the tied observations.

The performance of all these backtesting scenarios is tested first and foremost by evaluating and ranking the realized terminal value starting with a capital of unity, with and without transaction costs. In addition, some representative traditional performance measures in finance, such as the Sharpe and Omega ratios, are computed and interpreted as performance gauges. The Sharpe ratio is traditionally conceived as suitable for the MV world, while the Omega ratio is supposedly capable of assessing a nonnormal world.

Notice that Matallín, Soler, and Tortosa-Ausina (2011) essentially follow our ‘EWP without optimization’ strategy. These authors focus on the impact of some robustness parameters in the estimation of robust versions of the nonconvex frontier model (i.e., Free Disposal Hull (FDH), and its order-*m* and order-α robust versions) on some basic backtesting strategies. In fact, three backtesting strategies based on past efficiency are considered: (a) buying past top quantile and selling past bottom quantile; (b) buying past top quantile; and (c) selling bottom quantile. Thus, their study mainly differs from ours in that we also employ traditional convex frontier models and that we also apply an ‘EWP with optimization’ strategy.[5]

[4] For definitions and detailed presentations of these ratios: see, for example, Bacon (2008).

[5] There do exist some backtesting studies using frontier models for asset selection in terms of a financial strength indicator and then assessing the resulting portfolios over historic time (see, e.g., Edirisinghe and Zhang, 2007, 2010).

7.3.2 Frontier-based portfolio models

We now briefly describe the position-dependent shortage function in the portfolio context. This shortage function is first introduced in Briec, Kerstens, and Lesourd (2004) with respect to a MV universe. In Briec, Kerstens, and Jokung (2007), this model is then adapted to the MVS world. For a discussion of methods capable of visualizing the corresponding MVS-frontier, we refer to Kerstens, Mounir, and Van de Woestyne (2011a). Briec and Kerstens (2009) show that this adaptation from MV to MVS can be generalized to an arbitrary multi-moment universe. To save space, we restrict ourselves to only mentioning the position-dependent shortage function with respect to the MVSK universe.

We introduce some basic notations. A portfolio consisting of n MF available in the financial universe can be considered as a vector of weights $x=(x_1,\ldots,x_n)$ indicating the individual proportions of each MF. Obviously, $\sum_{i=1}^{n} x_i = 1$. If short selling is excluded, then the condition $x_i \geq 0$ for all $i \in \{1,\ldots,n\}$ must be satisfied. All MF in the financial universe are characterized by their raw returns registered over a given time window. From this information, the expected return vector, covariance matrix, and the skewness and kurtosis tensors can be derived. Additionally, either directly from the initial raw returns or indirectly from the statistical vectors, matrices, and tensors computed, the expected return, variance, skewness and kurtosis of an individual portfolio x can be computed. We denote these by Ret(x), Var(x), Skew(x), and Kurt(x), respectively.

Consider a portfolio x_o under evaluation. Then, the position-dependent shortage function identifies an inefficiency value λ for x_o obtained from solving the following nonlinear optimization model:

$$\max \lambda \ \textbf{s.t.} \quad \begin{aligned} & \text{Ret}(x) \geq \text{Ret}(x_o) + \lambda|\text{Ret}(x_o)|, \\ & \text{Var}(x) \leq \text{Var}(x_o) - \lambda \text{Var}(x_o), \\ & \text{Skew}(x) \geq \text{Skew}(x_o) + \lambda|\text{Skew}(x_o)|, \\ & \text{Kurt}(x) \leq \text{Kurt}(x_o) - \lambda \text{Kurt}(x_o), \\ & \lambda \geq 0, \forall i = 1,\ldots,n : 0 \leq x_i \leq 1, \sum_{i=1}^{n} x_i = 1. \end{aligned} \tag{7.3}$$

Note that if one needs the MV-based position-dependent shortage function, then the inequality constraints involving skewness and kurtosis must be dropped from model (7.3). Analogously, dropping only the kurtosis inequality constraint in (7.3) leads to the MVS-based position-dependent shortage function.

The position-dependent shortage function results in an inefficiency value λ. Similar with the shortage functions introduced in models (7.1) and (7.2), the portfolio x_o is more efficient if its inefficiency value is closer to 0. Moreover, the optimal portfolio x corresponding with the optimal value of λ for x_o is located on the corresponding MVSK-frontier. Clearly, this optimal portfolio has a higher return and skewness than x_o, while its variance and kurtosis are lower than that of x_o. Put differently: even moments are increased, while odd moments are decreased.

7.3.3 Data description

For the empirical analysis in Section 7.4, we use several data sets extracted from the 'Morningstar Direct' database. First, we extract a set of 814 open-end MF for which we have collected weekly returns from 9 October 2005 to 2 October 2011. This yields a total of 156 weeks. These open-end MF are rather homogeneous in that they all belong to the large caps European universe (Eurozone and the United Kingdom).

Second, running a MF business induces a variety of costs that are linked either to regular operations (brokerage, consultancy, marketing and distribution expenses, etc.), or to specific investors' transactions (mainly purchases and redemptions). The former costs are gathered under the 'Annual Fund Operating Expenses' denominator, while the latter costs are usually gathered under the heading 'Shareholder Fees'. The first cost category is indirectly covered by the fund shareholders with the proceeds of asset transactions decided by the fund manager and cash inflows coming from new investors. The second cost category is directly charged to the investor for each transaction of a fund's share, or on a regular (e.g., yearly) basis. In the MF prospectus, these elements of information are available. Since we need part of this information in the backtesting setup, we have downloaded for the 814 funds the 'Front Load' (entry fee), 'Redemption Load' (exit fee), and the 'Annual Report Net Expense Ratio'. The 'Front Load' and 'Redemption Load' are examples of shareholder fees and are fixed throughout the whole period under evaluation. The 'Annual Report Net Expense Ratio' (which reflects the actual fees charged during a particular fiscal year) proxies the total annual fund operating expenses. Clearly, this information is only available for each particular year. Therefore, it is extracted for the years 2006 till 2010.

Third, since some of the backtesting models involve Morningstar ratings, we extract for the period October 2008 till September 2011 the three-year Morningstar rating, which is available on a monthly basis. Detailing the Morningstar rating system, it is based on the risk-adjusted rate of return of a MF relative to its 'Morningstar Category'. This rating system for each MF is based on the total return in excess of a 90 days T-Bill, adjusted for front-end loads, deferred loads, and redemptions fees, and decreased by a risk penalty that particularly takes into account downward deviations. Thus, the risk considered by Morningstar consists in the lower semi-standard deviation for the excess return (to be explicit: $\sigma[\min((r_i - r_f),0)]$).[6] The best funds obtain 5 stars, the worst ones, a single star. In fact, the distribution of these stars in each category is a priori fixed and nonuniform. In other words, 10% of the best performing funds receive 5 stars, 22.5% 4 stars, 35% 3 stars, 22.5% 2 stars, and 10% 1 star.

This Morningstar rating is calculated on subsets of MF belonging to the same Morningstar category.[7] Furthermore, the composition of these Morningstar categories evolves over time following changes in the asset basket held by the MF. Notice that the time window of three years used in our MF frontier rating computations is identical to the one employed by Morningstar in determining its three-year Morningstar rating.

Fourth, since some of the backtesting models make use of traditional financial performance measures, the values for the Sharpe, Sortino, Treynor, Kappa, and Omega ratios are also extracted. These data are again available on a monthly basis. Therefore, these are extracted for the same period from October 2008 till September 2011 as the one used for the Morningstar ratings data.

Fifth, backtesting also requires prices for the individual MF. Again, these prices are extracted from 'Morningstar Direct' for the selected MF. These data are available on a daily basis from 1 January 2007 till 1 December 2011. Evidently, all price data are expressed in euros and converted when needed at the ongoing exchange rate. However, this data set turns out to be incomplete. Whenever possible, this missing data is completed (e.g., by using the price of the previous transaction day). For 103 MF in the initial sample, completing

[6] One can also refer to Sharpe (1998) for a more detailed discussion of this measure.

[7] These Morningstar categories are similar in nature to the Standard & Poor's 'GICS'.

missing data is impossible. Therefore, these MF are removed from the data set of prices. This reduces the number of MF to a total of 711. When prices are not needed, then computations are done on the data set containing the original 814 MF. Otherwise, the restriction to the 711 MF with full price information applies. Note that due to the monthly frequency with which the backtesting scenarios are executed, in fact we only need monthly prices from this daily prices database. Thus, this implies that monthly prices are used from 5 October 2008 till 4 September 2011: this is a period of 36 months. However, note that all statistics necessary for the hedonic rating models are computed using the available weekly data (see the first point).

In conclusion, somewhat in contrast to Blake and Morey (2000) and Kräussl and Sandelowsky (2007), notice that our backtesting setup is particular in three respects. First, the sample is more homogeneous in that it focuses on a limited set of Morningstar categories. Second, while the other studies focus on testing the internal validity of the Morningstar ratings (across stars), we backtest the external validity of solely the three-year Morningstar ratings compared to various alternatives in selecting top performers. Last but not least, we on purpose create the harshest possible testing environment by computing our ratings over a normal-to-bull market period (2005–2008), while we backtest all strategies over one of the worst financial crisis periods ever (2008–2011).

7.4 Empirical analysis

7.4.1 Descriptive statistics

We analyze the characteristics of the returns distribution for the sample consisting of 814 MF over the whole period from 9 October 2005 to 2 October 2011. Figure 7.2a to d reports some basic descriptive statistics: in particular, the distributions of the expected return, standard deviation, skewness, and excess kurtosis are displayed. The most prominent features are situated at the third and fourth moments: there is a noticeable negative skewness and a substantial excess kurtosis. Both phenomena can be linked to the financial crisis resulting in a negative trend and a high volatility regime during the period under investigation.

Probably for the same reason, the potential diversification effects in this universe are relatively weak due to rather high levels of correlation. Figure 7.3 reports a gray tone mapping of the covariance matrix computed for all 814 MF over the same period. Notice that off-diagonal correlations vary widely between 0.097 (darkest gray tone) and 1 (lightest gray tone). Typically, several MF belonging to the same investment firm have very similar (or even exactly the same) composition and hence correlate very strongly. For example, two fund shares with a close-to-unity correlation coefficient are ‘Vanguard European Stock Idx Inst EUR’ and ‘Vanguard European Stock Idx Inst USD’, the first one being listed in euros while the second in US dollars. These substantial correlation levels may well mitigate the scope for diversification. The extent to which these may compromise the exploitation of higher-order moments and co-moments remains to be explored.

7.4.2 Analysis of both hedonic rating models

Since hedonic frontier rating models are considered in the backtesting procedures, we first compare the efficiency distributions obtained for some of these frontier models over all 814 selected MF. For these comparisons, we use nonparametric Li (1996) tests. This test statistic

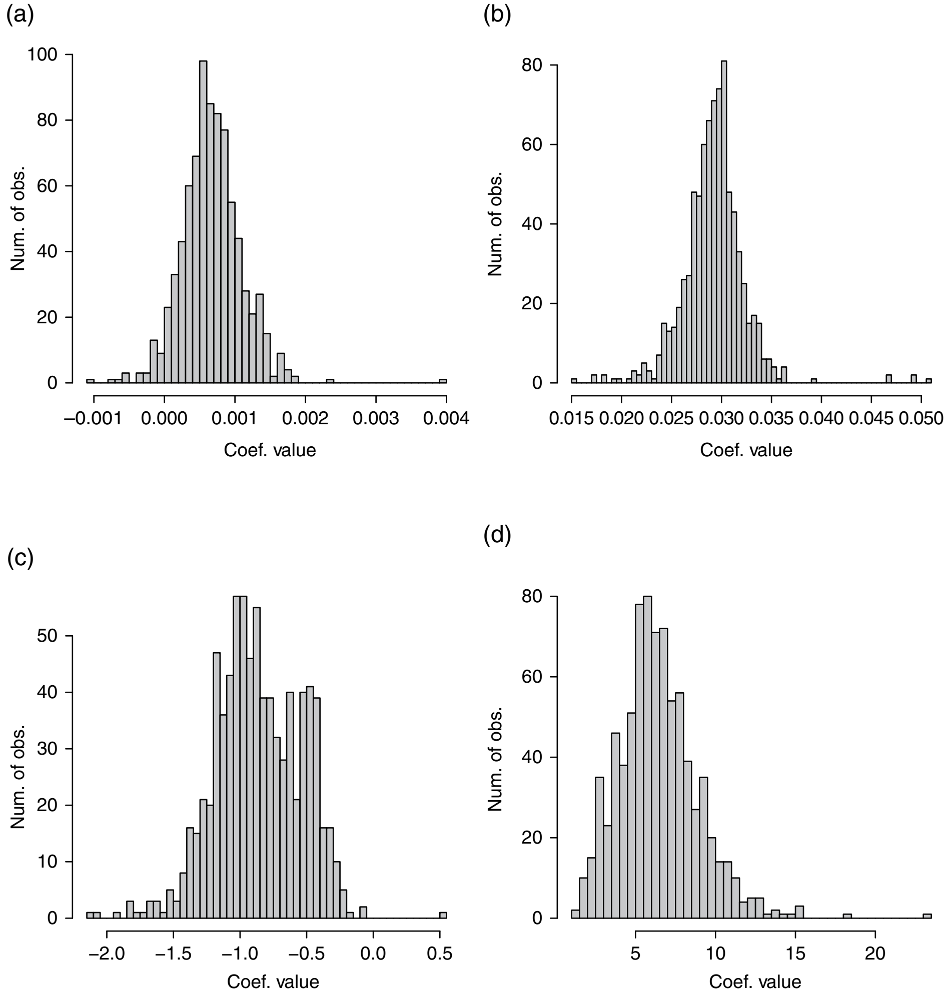

Figure 7.2 Distribution of the first four moments for the 814 MF. (a) Mean, (b) standard deviation, (c) skewness, and (d) excess kurtosis.

basically compares two entire distributions and it is valid for both dependent and independent variables. Notice that independency is not a valid assumption in frontier models where efficiency estimates depend, among others, on the relative size of the sample, the dimensionality of the space under analysis, etc.

Table 7.2 reports the result of in total seven comparisons. The first two (columns 2 and 3) look for the effect of adding an additional moment to the nonconvex model. Similarly, the next two comparisons (columns 4 and 5) demonstrate the effect of adding an additional moment to the convex model. Finally, the last three comparisons (columns 6–8) evaluate the effect of switching between the convex and nonconvex models for an identical number of moments.

Notice that a comparison is made for all 36 months present in the time window under evaluation: each period is represented in a row. In terms of notation, the comparison between

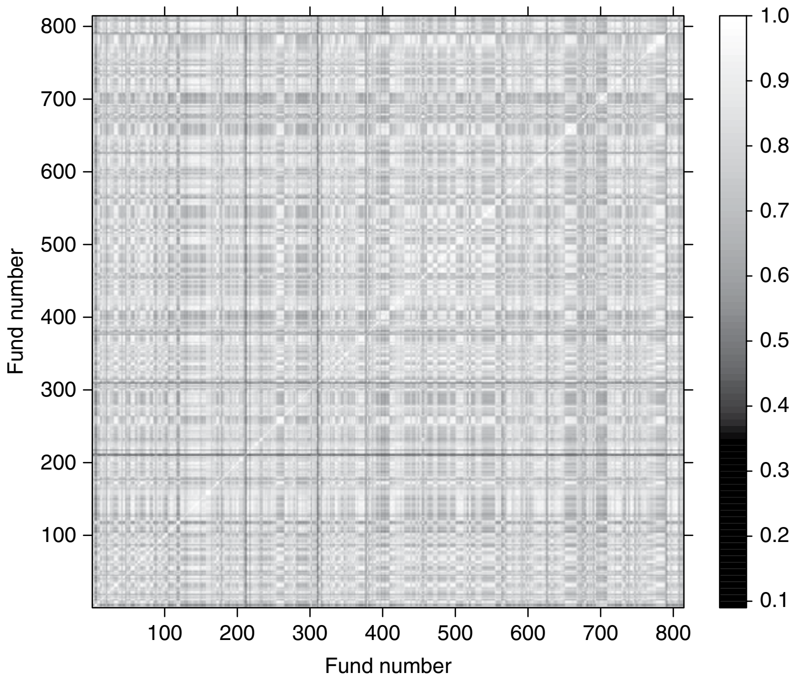

Figure 7.3 Visualization of correlation matrix for the 814 MF.

two models yields an equality sign ('=') if both efficiency distributions can be considered equal at a 5% significance level. However, if the efficiency distributions are significantly different, then the symbol '*' is returned.

We now turn to the interpretation of these 252 (7 × 36) Li (1996) test statistics in Table 7.2. As to the number of moments needed, adding skewness to the nonconvex model has an effect, while adding kurtosis has no effect. For the convex model, adding the skewness only has an effect in some of the periods. Again, no effect is noticeable from adding the kurtosis to this model. Finally, the comparison of the effect of switching between convex and nonconvex models in the last three columns demonstrates that these models are in a large majority of periods significantly different across the three multimoment models. This corroborates the results in Kerstens, Mounir, and Van de Woestyne (2011b).

A second step in the analysis consists in verifying whether the hedonic ratings are correlated with the other performance measures available. To answer this question, we run a nonparametric rank correlation analysis using Kendall's τ statistics. Therefore, consider the rating pairs of rating matrices each of dimension (814×36) (i.e., funds per time windows) containing the efficiency values for all 814 MF and all periods present in the monthly back-testing calendar (i.e., 36 periods). The first rating, hereafter denominated R_1, systematically consists of a hedonic rating. The second rating, denominated R_2 is either the Morningstar rating, or the Sharpe ratio matrix. The correlation estimates are computed over all time periods as follows: $\mathrm{Cor}_{i\in[1,36]}(R_1(.,i), R_2(.,i))$. Literally, this means that we use the set of 814 MF time window after time window (36 times) to assess how our hedonic rankings are in line

Table 7.2 Comparison of different hedonic rating models by means of Li (1996) tests.

Period	MV+L-NC	MVS+L-NC	MV+L-C	MVS+L-C	MV+L-C	MVS+L-C	MVSK+L-C
	MVS+L-NC	MVSK+L-NC	MVS+L-C	MVSK+L-C	MV+L-NC	MVS+L-NC	MVSK+L-NC
1	=	=	=	=	*	*	*
2	=	=	=	=	*	*	*
3	*	=	=	=	*	*	*
4	*	=	=	=	*	*	*
5	*	=	=	=	=	*	*
6	*	=	=	=	=	*	*
7	*	=	*	=	*	*	*
8	*	=	*	=	*	*	*
9	*	=	*	=	*	*	*
10	*	=	*	=	*	*	*
11	*	=	=	=	*	*	*
12	*	=	=	=	*	*	*
13	*	=	=	=	*	*	*
14	*	=	=	=	*	*	*
15	*	=	=	=	*	*	*
16	*	=	=	=	*	*	*
17	*	=	=	=	*	*	*
18	*	=	=	=	*	*	*
19	*	=	=	=	*	*	*
20	=	=	=	=	*	*	*
21	*	=	=	=	*	*	*
22	*	=	=	=	*	*	*
23	*	=	=	=	*	*	*
24	*	=	=	=	*	*	*

(*continued overleaf*)

Table 7.2 (*continued*)

Period	MV+L-NC	MVS+L-NC	MV+L-C	MVS+L-C	MV+L-C	MVS+L-C	MVSK+L-C
	MVS+L-NC	MVSK+L-NC	MVS+L-C	MVSK+L-C	MV+L-NC	MVS+L-NC	MVSK+L-NC
25	*	=	=	=	*	*	*
26	*	=	=	=	*	*	*
27	*	=	=	=	*	*	*
28	*	=	=	=	*	*	*
29	*	=	=	=	*	*	*
30	*	=	=	=	*	*	*
31	*	=	=	=	*	*	*
32	*	=	=	=	*	*	*
33	*	=	=	=	*	*	*
34	*	=	*	=	*	*	*
35	*	=	*	=	*	*	*
36	*	=	*	=	*	*	*

*, Means both efficiency distributions can be considered not equal at 5% significance level.
=, Means both efficiency distributions can be considered equal at 5% significance level.

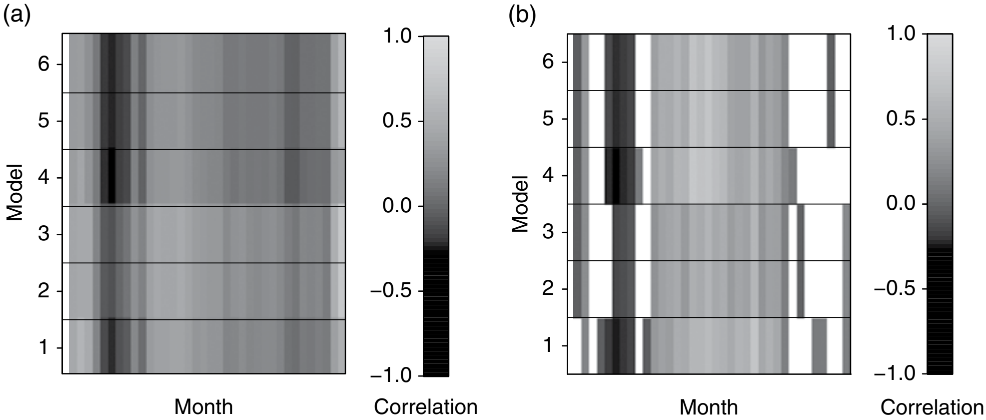

Figure 7.4 Rank correlations by model over time (Models 1: MV+L-NC, 2: MVS+L-NC, 3: MVSK+L-NC, 4: MV+L-C, 5: MVS+L-C, 6: MVSK+L-C). (a) Hedonic versus Morningstar's and (b) Hedonic versus Sharpe.

with other traditional performance ratios. The resulting Kendall's τ statistics are reported in Figure 7.4a and 7.4b. In these figures, a white patch denotes that the correlation is statistically insignificant. The horizontal axis reports the time windows, while the vertical axis indicates with which model the ratings are compared (1: MV+L-NC, 2: MVS+L-NC, 3: MVSK+L-NC, 4: MV+L-C, 5: MVS+L-C, 6: MVSK+L-C).

Starting with Figure 7.4a, most of the time the hedonic ratings delivered by the convex models (models 4–6 in the figure) obtain a negative correlation coefficient when compared to the Morningstar ratings. This result is to some extent expected, since the lower the hedonic ratings the better the MF efficiency levels. Thus, if our approach delivers ratings similar to the Morningstar stars (which increases in performance level), then one should observe a negative correlation coefficient. This is actually the case for most of the periods considered in the backtesting time window, except month 1 (October 2008) and months 35 and 36 (September and October 2011). However, this rather systematic negative correlation is not evident when analyzing nonconvex models. Turning now to Figure 7.4b, that is, the correlation between the hedonic ratings and the Sharpe ratios, no clear relationship can be discerned at all.

Thus, considering the potential impacts of these correlations on the backtesting strategy being tested, one could imagine that significant differences might be observed when using nonconvex models in conjunction with the Morningstar ratings, or when comparing with the Sharpe ratios.

7.4.3 Backtesting results for 48 different strategies

First, we start out by commenting on the first two Tables 7.3 and 7.4 containing terminal values for the different backtesting portfolio strategies in terms of the unity initial wealth without and with transaction costs respectively. Thus, terminal wealth is expressed as a percentage of unity initial wealth. Thereafter, we present some tables evaluating the same strategies in terms of a series of traditional financial performance measures. All of these tables share the same basic structure. The different sample sizes are in the column heading. The different names of the backtesting portfolio strategies following the coding defined in

Table 7.3 Terminal wealth and rank (R) when investing unity for 48 backtesting scenarios without transaction costs.

	MF(10)	R	MF(20)	R	MF(30)	R	MF(40)	R	MF(50)	R
MV+L-C to NoOpt	0.9751	27	0.9999	28	0.9828	47	0.9679	48	0.9600	48
MVS+L-C to NoOpt	1.0050	23	0.9931	30	1.0106	34	1.0035	43	0.9871	47
MVSK+L-C to NoOpt	0.9886	24	1.0135	23	0.9938	43	0.9996	44	0.9892	46
MV+L-NC to NoOpt	1.0151	22	0.9981	29	0.9889	44	0.9978	45	0.9993	43
MVS+L-NC to NoOpt	0.9659	31	0.9821	35	0.9889	45	0.9927	46	0.9983	44
MVSK+L-NC to NoOpt	0.9741	28	0.9751	37	0.9730	48	0.9922	47	0.9932	45
Stars to NoOpt	1.0803	5	1.1344	1	1.0935	1	1.1225	1	1.1094	1
Sharpe to NoOpt	1.0757	6	1.0507	11	1.0518	23	1.0614	14	1.0437	27
Sortino to NoOpt	1.0371	15	1.0630	7	1.0582	20	1.0504	21	1.0391	34
Omega to NoOpt	1.0206	18	1.0155	22	1.0206	27	1.0267	35	1.0238	42
Treynor to NoOpt	1.0583	9	1.0215	20	1.0327	25	1.0279	34	1.0439	26
Kappa to NoOpt	1.0180	21	1.0338	19	1.0265	26	1.0345	32	1.0407	32
MV+L-C to MV	1.0844	3	1.0752	4	1.0680	14	1.0678	12	1.0628	12
MVS+L-C to MV	1.0881	2	1.1038	3	1.0810	6	1.0846	5	1.0723	7
MVSK+L-C to MV	1.1101	1	1.1085	2	1.0815	5	1.0894	2	1.0741	6
MV+L-NC to MV	1.0639	8	1.0505	13	1.0447	24	1.0524	18	1.0568	16
MVS+L-NC to MV	1.0383	14	1.0207	21	1.0798	7	1.0695	10	1.0628	14
MVSK+L-NC to MV	1.0389	13	1.0029	26	1.0899	2	1.0535	17	1.0655	10
MV+L-C to MVS	1.0844	4	1.0752	6	1.0680	14	1.0678	13	1.0628	12
MVS+L-C to MVS	0.9738	30	1.0443	16	1.0636	16	1.0881	3	1.0840	4
MVSK+L-C to MVS	1.0341	16	1.0348	18	1.0605	18	1.0831	7	1.0846	2
MV+L-NC to MVS	1.0510	10	1.0505	13	1.0573	22	1.0524	18	1.0568	16
MVS+L-NC to MVS	0.9659	32	0.9822	34	1.0749	8	1.0696	8	1.0558	18
MVSK+L-NC to MVS	0.9803	26	0.9643	38	1.0892	4	1.0561	16	1.0656	8

MV+L-C to MVSK	1.0754	7	1.0752	4	1.0680	14	1.0678	11	1.0628	12
MVS+L-C to MVSK	0.9738	30	1.0443	16	1.0636	17	1.0881	4	1.0840	5
MVSK+L-C to MVSK	1.0341	17	1.0348	18	1.0605	19	1.0831	6	1.0846	2
MV+L-NC to MVSK	1.0510	10	1.0505	13	1.0573	21	1.0524	20	1.0568	16
MVS+L-NC to MVSK	0.9659	32	0.9822	34	1.0749	8	1.0696	8	1.0558	18
MVSK+L-NC to MVSK	0.9803	26	0.9643	38	1.0892	4	1.0561	15	1.0656	8
Stars to MV	1.0397	12	1.0595	10	1.0700	12	1.0363	31	1.0400	33
Stars to MVS	1.0181	20	1.0598	8	1.0712	11	1.0380	27	1.0425	28
Stars to MVSK	1.0181	20	1.0598	8	1.0712	10	1.0380	28	1.0425	28
Sharpe to MV	0.9576	34	0.9561	44	1.0071	35	1.0257	37	1.0280	41
Sharpe to MVS	0.9569	36	0.9590	43	1.0037	38	1.0202	41	1.0349	38
Sharpe to MVSK	0.9569	36	0.9590	42	1.0037	38	1.0162	42	1.0349	37
Sortino to MV	0.9340	43	0.9500	48	1.0054	36	1.0266	36	1.0307	40
Sortino to MVS	0.9481	38	0.9500	46	1.0023	40	1.0217	39	1.0375	36
Sortino to MVSK	0.9481	38	0.9500	46	1.0023	40	1.0217	40	1.0375	35
Omega to MV	0.9460	39	0.9558	45	0.9834	46	1.0248	38	1.0339	39
Omega to MVS	0.9409	40	0.9614	40	0.9978	41	1.0366	30	1.0409	31
Omega to MVSK	0.9409	40	0.9614	41	0.9978	42	1.0366	29	1.0409	30
Treynor to MV	0.9319	44	1.0017	27	1.0194	30	1.0396	24	1.0448	24
Treynor to MVS	0.9305	48	1.0118	24	1.0195	28	1.0477	23	1.0529	22
Treynor to MVSK	0.9305	48	1.0118	24	1.0195	28	1.0477	22	1.0529	22
Kappa to MV	0.9363	42	0.9755	36	1.0194	31	1.0292	33	1.0444	25
Kappa to MVS	0.9315	46	0.9852	32	1.0118	33	1.0394	25	1.0547	21
Kappa to MVSK	0.9315	46	0.9852	32	1.0118	32	1.0394	26	1.0547	20

Table 7.4 Terminal wealth and rank (R) when investing unity for 48 backtesting scenarios with transaction costs.

	MF(10)	R	MF(20)	R	MF(30)	R	MF(40)	R	MF(50)	R
MV+L-C to NoOpt	0.8447	15	0.9102	7	0.9067	9	0.9022	6	0.8946	7
MVS+L-C to NoOpt	0.8840	8	0.9233	6	0.9523	1	0.9492	3	0.9402	3
MVSK+L-C to NoOpt	0.8308	18	0.9416	4	0.9421	2	0.9572	2	0.9437	2
MV+L-NC to NoOpt	0.9294	5	0.9246	5	0.9223	4	0.9168	4	0.9166	4
MVS+L-NC to NoOpt	0.8146	19	0.8810	12	0.9010	11	0.8948	9	0.9078	5
MVSK+L-NC to NoOpt	0.8436	16	0.8824	11	0.8965	12	0.9025	5	0.9042	6
Stars to NoOpt	0.9531	2	1.0070	1	0.9371	3	0.9575	1	0.9562	1
Sharpe to NoOpt	0.9352	3	0.8254	15	0.8409	20	0.8745	16	0.8597	11
Sortino to NoOpt	0.8853	7	0.8521	13	0.8527	19	0.8629	17	0.8527	17
Omega to NoOpt	0.8473	14	0.8029	19	0.8164	25	0.8345	24	0.8459	18
Treynor to NoOpt	0.8962	6	0.7926	20	0.8332	21	0.8309	25	0.8551	16
Kappa to NoOpt	0.8337	17	0.8257	14	0.8227	24	0.8557	18	0.8585	12
MV+L-C to MV	0.8563	12	0.8079	16	0.8152	27	0.8109	29	0.8023	28
MVS+L-C to MV	0.9557	1	0.9538	3	0.9022	10	0.8789	15	0.8559	15
MVSK+L-C to MV	0.9297	4	0.9563	2	0.9110	6	0.8850	14	0.8645	10
MV+L-NC to MV	0.6961	31	0.8843	9	0.8077	31	0.7693	32	0.7879	32
MVS+L-NC to MV	0.8738	9	0.6788	25	0.8854	13	0.8909	12	0.8793	9
MVSK+L-NC to MV	0.8649	10	0.5999	33	0.8822	16	0.8896	13	0.8842	8
MV+L-C to MVS	0.8627	11	0.8079	18	0.8152	27	0.8109	30	0.8023	30
MVS+L-C to MVS	0.7339	24	0.7798	22	0.8105	29	0.8264	26	0.8383	24
MVSK+L-C to MVS	0.6540	34	0.7655	24	0.7868	32	0.8397	22	0.8393	20
MV+L-NC to MVS	0.7168	29	0.8843	9	0.8233	23	0.7693	32	0.7879	32
MVS+L-NC to MVS	0.7937	22	0.6706	26	0.8671	18	0.8926	10	0.8393	22
MVSK+L-NC to MVS	0.8079	20	0.6277	28	0.8835	14	0.8952	7	0.8562	14
MV+L-C to MVSK	0.8496	13	0.8079	16	0.8152	27	0.8109	28	0.8023	28

MVS+L-C to MVSK	0.7339	24	0.7798	21	0.8105	30	0.8264	27	0.8383	24
MVSK+L-C to MVSK	0.6540	35	0.7655	23	0.7868	33	0.8397	23	0.8393	20
MV+L-NC to MVSK	0.7168	28	0.8843	9	0.8233	22	0.7693	33	0.7879	32
MVS+L-NC to MVSK	0.7937	22	0.6706	27	0.8671	18	0.8926	10	0.8393	22
MVSK+L-NC to MVSK	0.8079	20	0.6277	28	0.8835	14	0.8952	8	0.8562	14
Stars to MV	0.7093	30	0.6222	30	0.9145	5	0.8526	19	0.8418	19
Stars to MVS	0.7197	26	0.6126	32	0.9072	8	0.8487	20	0.8350	26
Stars to MVSK	0.7197	26	0.6126	31	0.9072	7	0.8487	21	0.8350	26
Sharpe to MV	0.6551	33	0.5806	42	0.6055	40	0.6642	44	0.6724	48
Sharpe to MVS	0.6435	38	0.5848	37	0.5893	42	0.6611	47	0.6809	45
Sharpe to MVSK	0.6435	38	0.5848	38	0.5893	42	0.6566	48	0.6809	46
Sortino to MV	0.6095	45	0.5873	36	0.5728	43	0.6645	43	0.6822	44
Sortino to MVS	0.6109	44	0.5840	40	0.5687	44	0.6628	45	0.6906	40
Sortino to MVSK	0.6109	44	0.5840	40	0.5687	44	0.6628	46	0.6906	41
Omega to MV	0.6646	32	0.4865	48	0.5182	48	0.6646	42	0.6744	47
Omega to MVS	0.6491	36	0.4960	46	0.5438	46	0.6782	39	0.6825	42
Omega to MVSK	0.6491	36	0.4960	47	0.5438	46	0.6782	40	0.6825	43
Treynor to MV	0.5937	46	0.5842	39	0.6086	39	0.6992	36	0.7122	38
Treynor to MVS	0.5827	48	0.5939	34	0.6136	38	0.7110	34	0.7210	34
Treynor to MVSK	0.5827	48	0.5939	34	0.6136	38	0.7110	35	0.7210	34
Kappa to MV	0.6336	40	0.5702	45	0.6244	36	0.6765	41	0.7063	39
Kappa to MVS	0.6216	42	0.5787	44	0.6282	35	0.6869	38	0.7177	36
Kappa to MVSK	0.6216	42	0.5787	44	0.6282	34	0.6869	38	0.7177	37

Section 7.3.1 are in the first column. On the right-hand side of each terminal value as percentage of initial wealth one finds a rank in descending order over all possible backtesting strategies within the same column.

Starting with the 'EWP without optimization' strategies, the following tentative conclusions can be drawn. First, the frontier MF ratings are performing rather poorly without transaction costs (rank 22 at best and rank 48 at worst), with a tendency of performance to worsen as the size of the number of included MF increases. Second, the traditional financial performance measures slightly outperform these frontier MF ratings in terms of rankings. In terms of terminal wealth, the results are even more marked: traditional financial performance measures all gain money, while some of the frontier MF ratings lose money overall. Third, the Morningstar ratings systematically outperform all other strategies, except for the smallest sample size (column MF(10)).

In the case of the net terminal value with transaction costs, this picture changes rather drastically. First, the frontier MF ratings are performing rather well in terms of ranks (rank 1 at best and rank 19 at worst). There is now some tendency for performance to improve with the size of the number of included MF. Except for the smallest size classes (MF(10) and MF(20)), the average ranks are very decent to rather exceptional. In particular, the best ranks are obtained by some multimoment (in particular, MVS and MVSK) convex frontier MF ratings: for MF(30), these models occupy ranks 1 and 2. The nonconvex model manages to obtain somewhat lower ranks for the larger size classes (rank 4 and 5 at best). Second, in terms of terminal wealth, all strategies lose money in the end, but traditional financial performance measures now tend to do worse than frontier MF ratings. Third, the Morningstar ratings still tend to outperform all other strategies, except for two sample sizes (i.e., MF(10) and MF(30)). Furthermore, for one sample size (MF(20)), Morningstar is the only strategy generating a positive net terminal gain at all.

One probable explanation for the contrasting results when comparing frontier MF ratings vs traditional financial performance measures is that the former yield relatively more stable portfolios. This seems to result in an edge in terms of economizing transaction costs. Perhaps, due to their multidimensional nature compared to the two-dimensional ratio nature of traditional financial measures, these frontier MF ratings seem intrinsically more stable. This probably requires further exploration.

Switching now to 'EWP with optimization' strategies, one can observe the following. Without transaction costs, the frontier MF ratings are clearly outperforming by large the traditional financial performance measures. There is always a whole series of MF rating strategies that also beat any of the Morningstar-based strategies. Morningstar is now even beaten in a traditional MV world. This basic result holds fundamentally true over all subsamples.

When including transaction costs to look at net terminal wealth, the same basic conclusions can be drawn. However, there are some particular differences worth highlighting. First, the best ranks are obtained by some multimoment (in particular, MVS and MVSK) convex frontier MF ratings for the two smallest size classes (ranks 1–4 at best) and by the same nonconvex models for the two largest size classes (ranks 7–9 at best). Second, one Morningstar rating strategy beats these frontier MF ratings for the class MF(30) solely in a MV context.

Notice that when comparing optimization strategies without and with transaction costs, there is a noticeable difference in the type of portfolio optimization models that seem to perform well. While without transaction costs, some strategies dominate using either MV, MVS, or MVSK portfolio optimization, with transaction costs only MV-based optimization strategies perform well, except for the MF(40) size class where a nonconvex model combined with MVS and MVSK portfolio optimization yields the best ranks (ranks 7–8 at best). A clear reason for this remarkable difference in results is currently wanting. We return to this issue in Section 7.4.4.

Next, we turn to the evaluation of the same basic portfolio strategies in terms of some traditional financial performance measures in Table 7.5 and Table 7.6 reporting the Sharpe and the Omega ratios, respectively. While the Sharpe ratio is a standard MV-oriented performance measure, the Omega ratio by definition aims at capturing substantial parts of the whole return distribution. It can be interpreted as a ratio of gains over losses relative to some loss threshold (see Bacon, 2008 for details). Therefore, it should ideally be able to function in a nonnormal portfolio world. Notice that while the portfolio strategies have been phrased in terms of Sharpe, Sortino, Omega, Treynor, and Kappa ratios, we limit the evaluation of all 48 backtesting scenarios due to lack of space to just a selection of two representative ratios. Fundamentally, the other ratios do not deliver any significant differences in basic conclusions.

Again starting with the 'EWP without optimization' strategies, according to the Sharpe ratio one can make the following observations. First, the Morningstar ratings overall beat the frontier MF ratings, though not in all size classes. In particular, in two size classes, convex frontier MF rating models do better. Second, the traditional financial performance measures do worse than the frontier MF ratings, except perhaps in the first size class. Notice that even some of the nonconvex frontier MF rating models perform decently well, often just behind their convex counterparts. Remark that the strategy based on the Sharpe ratio is beaten by both the Morningstar rating and the frontier MF rating models, except in the smallest size class.

Now considering the 'EWP with optimization' strategies, one immediately can reach the following conclusions. First, these strategies in general do poorer than the nonoptimized strategies, except for the smallest size classes. Second, in general, the frontier MF ratings do a better job than both the traditional financial performance measure and the Morningstar ratings.

Turning now finally to the Omega ratio, starting with the 'EWP without optimization' strategies one can conclude the following. First, the ranks are identical to the ones resulting from the Sharpe ratio, except for the MF(40) size class. Hence, the same basic conclusions can be transposed. One remark to make is that the strategy based on the Omega ratio is beaten by both the Morningstar rating and the frontier MF rating models, except again in the smallest size class for some of the frontier MF rating models. Changing focus to the 'EWP with optimization' strategies, one can quickly reach consensus that the findings very closely follow the ones obtained for the Sharpe ratio previously.

Hence, both traditional financial performance measures yield close to the same conclusions. But, since the strategies constructed on both of these traditional financial performance measures do worse than most of the other strategies, one may wonder to what extent these traditional financial performance measures are really suitable to gauge the performance of such wildly varying portfolio strategies. Perhaps, this question merits thorough further reflection.

7.4.4 Backtesting results for MF rating models: Some plausible explanations

This section aims to shed light on two plausible mechanisms driving some of the empirical results reported in Section 7.4.3. First, one may wonder why frontier MF ratings seem to do rather well compared to well-established financial performance measures that have a long

Table 7.5 Performance backtesting using the Sharpe ratios and rank (*R*) for 48 backtesting scenarios.

	MF(10)	*R*	MF(20)	*R*	MF(30)	*R*	MF(40)	*R*	MF(50)	*R*
MV+L-C to NoOpt	−0.0846	12	−0.0476	7	−0.0495	5	−0.0510	6	−0.0552	7
MVS+L-C to NoOpt	−0.0565	8	−0.0391	6	−0.0244	1	−0.0263	3	−0.0311	3
MVSK+L-C to NoOpt	−0.0795	9	−0.0296	2	−0.0302	2	−0.0222	1	−0.0295	2
MV+L-NC to NoOpt	−0.0359	4	−0.0380	5	−0.0390	4	−0.0426	4	−0.0426	4
MVS+L-NC to NoOpt	−0.0986	18	−0.0627	9	−0.0511	6	−0.0540	7	−0.0476	5
MVSK+L-NC to NoOpt	−0.0809	10	−0.0622	8	−0.0537	7	−0.0506	5	−0.0494	6
Stars to NoOpt	−0.0269	1	0.0039	1	−0.0358	3	−0.0244	2	−0.0250	1
Sharpe to NoOpt	−0.0322	3	−0.0979	14	−0.0859	20	−0.0657	8	−0.0726	8
Sortino to NoOpt	−0.0564	7	−0.0819	13	−0.0796	17	−0.0718	11	−0.0766	11
Omega to NoOpt	−0.0827	11	−0.1102	16	−0.1009	23	−0.0894	19	−0.0831	14
Treynor to NoOpt	−0.0562	6	−0.1192	17	−0.0898	21	−0.0900	20	−0.0758	10
Kappa to NoOpt	−0.0891	14	−0.0993	15	−0.0984	22	−0.0770	16	−0.0750	9
MV+L-C to MV	−0.0945	17	−0.1356	19	−0.1251	28	−0.1273	29	−0.1327	29
MVS+L-C to MV	−0.0292	2	−0.0322	4	−0.0704	12	−0.0861	18	−0.1016	19
MVSK+L-C to MV	−0.0430	5	−0.0305	3	−0.0643	11	−0.0820	17	−0.0957	15
MV+L-NC to MV	−0.2069	43	−0.0796	10	−0.1314	31	−0.1561	32	−0.1432	32
MVS+L-NC to MV	−0.0859	13	−0.2039	27	−0.0736	13	−0.0743	14	−0.0824	13
MVSK+L-NC to MV	−0.0914	16	−0.2608	41	−0.0766	16	−0.0750	15	−0.0791	12
MV+L-C to MVS	−0.0902	15	−0.1356	20	−0.1251	30	−0.1273	30	−0.1327	30
MVS+L-C to MVS	−0.1462	25	−0.1407	22	−0.1215	26	−0.1204	26	−0.1103	22
MVSK+L-C to MVS	−0.1907	32	−0.1469	24	−0.1430	32	−0.1106	24	−0.1105	24
MV+L-NC to MVS	−0.1886	28	−0.0796	12	−0.1193	25	−0.1561	32	−0.1432	32
MVS+L-NC to MVS	−0.1286	23	−0.2034	25	−0.0848	19	−0.0728	13	−0.1120	26
MVSK+L-NC to MVS	−0.1198	20	−0.2374	36	−0.0750	15	−0.0702	9	−0.1002	18

MV + L-C to MVSK	−0.0997	19	−0.1356	18	−0.1251	30	−0.1273	28	−0.1327	28
MVS + L-C to MVSK	−0.1462	24	−0.1407	21	−0.1215	27	−0.1204	27	−0.1103	22
MVSK + L-C to MVSK	−0.1907	33	−0.1469	23	−0.1430	33	−0.1106	25	−0.1105	25
MV + L-NC to MVSK	−0.1886	29	−0.0796	12	−0.1193	24	−0.1561	33	−0.1432	31
MVS + L-NC to MVSK	−0.1286	22	−0.2034	26	−0.0848	18	−0.0728	12	−0.1120	27
MVSK + L-NC to MVSK	−0.1198	21	−0.2374	36	−0.0750	14	−0.0702	10	−0.1002	18
Stars to MV	−0.1943	38	−0.2708	43	−0.0549	8	−0.0938	21	−0.0999	16
Stars to MVS	−0.1906	30	−0.2796	47	−0.0584	10	−0.0951	22	−0.1026	20
Stars to MVSK	−0.1906	30	−0.2796	46	−0.0584	9	−0.0951	23	−0.1026	21
Sharpe to MV	−0.1836	27	−0.2246	30	−0.2278	34	−0.2074	43	−0.2044	47
Sharpe to MVS	−0.1911	34	−0.2225	28	−0.2403	35	−0.2103	47	−0.1980	42
Sharpe to MVSK	−0.1911	34	−0.2225	29	−0.2403	36	−0.2139	48	−0.1980	43
Sortino to MV	−0.2048	40	−0.2326	31	−0.2524	43	−0.2081	44	−0.2020	46
Sortino to MVS	−0.2058	41	−0.2367	34	−0.2556	44	−0.2100	45	−0.1956	40
Sortino to MVSK	−0.2058	42	−0.2367	35	−0.2556	44	−0.2100	46	−0.1956	41
Omega to MV	−0.1824	26	−0.2837	48	−0.3030	48	−0.2056	42	−0.2056	48
Omega to MVS	−0.1928	36	−0.2786	44	−0.2828	46	−0.1957	39	−0.1984	44
Omega to MVSK	−0.1928	37	−0.2786	45	−0.2828	47	−0.1957	40	−0.1984	45
Treynor to MV	−0.2182	46	−0.2406	38	−0.2475	42	−0.1840	36	−0.1784	36
Treynor to MVS	−0.2263	48	−0.2362	32	−0.2443	39	−0.1756	34	−0.1713	35
Treynor to MVSK	−0.2263	47	−0.2362	33	−0.2443	40	−0.1756	35	−0.1713	34
Kappa to MV	−0.1986	39	−0.2629	42	−0.2464	41	−0.2015	41	−0.1882	39
Kappa to MVS	−0.2083	45	−0.2586	39	−0.2440	38	−0.1939	38	−0.1790	37
Kappa to MVSK	−0.2083	44	−0.2586	40	−0.2440	37	−0.1939	38	−0.1790	38

Table 7.6 Performance backtesting using the Omega ratios and rank (*R*) for 48 backtesting scenarios.

	MF(10)	*R*	MF(20)	*R*	MF(30)	*R*	MF(40)	*R*	MF(50)	*R*
MV+L-C to NoOpt	0.7990	12	0.8834	7	0.8797	5	0.8761	5	0.8650	7
MVS+L-C to NoOpt	0.8588	8	0.9029	6	0.9387	1	0.9346	3	0.9229	3
MVSK+L-C to NoOpt	0.8064	10	0.9261	2	0.9248	2	0.9441	1	0.9268	2
MV+L-NC to NoOpt	0.9096	4	0.9044	5	0.9016	4	0.8942	4	0.8947	4
MVS+L-NC to NoOpt	0.7753	19	0.8493	9	0.8745	6	0.8689	7	0.8834	5
MVSK+L-NC to NoOpt	0.8104	9	0.8504	8	0.8669	7	0.8755	6	0.8782	6
Stars to NoOpt	0.9303	1	1.0103	1	0.9108	3	0.9385	2	0.9367	1
Sharpe to NoOpt	0.9193	3	0.7729	14	0.7988	18	0.8437	8	0.8294	8
Sortino to NoOpt	0.8606	7	0.8043	13	0.8118	17	0.8310	13	0.8209	11
Omega to NoOpt	0.8041	11	0.7557	16	0.7663	23	0.7909	20	0.8045	14
Treynor to NoOpt	0.8658	6	0.7292	17	0.7924	21	0.7921	19	0.8216	10
Kappa to NoOpt	0.7955	14	0.7710	15	0.7727	22	0.8189	16	0.8233	9
MV+L-C to MV	0.7891	16	0.7074	18	0.7264	30	0.7214	29	0.7129	29
MVS+L-C to MV	0.9287	2	0.9221	4	0.8372	12	0.8032	18	0.7701	18
MVSK+L-C to MV	0.8926	5	0.9258	3	0.8517	11	0.8123	17	0.7830	15
MV+L-NC to MV	0.5629	45	0.8177	10	0.7171	31	0.6718	32	0.6946	32
MVS+L-NC to MV	0.7929	15	0.5574	25	0.8229	13	0.8284	14	0.8105	13
MVSK+L-NC to MV	0.7807	17	0.4764	42	0.8157	16	0.8249	15	0.8170	12
MV+L-C to MVS	0.7972	13	0.7074	20	0.7264	28	0.7214	30	0.7129	30
MVS+L-C to MVS	0.6836	24	0.6994	22	0.7276	26	0.7351	26	0.7491	26
MVSK+L-C to MVS	0.6168	28	0.6757	24	0.6944	32	0.7550	24	0.7504	22
MV+L-NC to MVS	0.5806	37	0.8177	12	0.7396	25	0.6718	32	0.6946	32
MVS+L-NC to MVS	0.6877	23	0.5518	26	0.7962	20	0.8313	12	0.7498	24
MVSK+L-NC to MVS	0.7094	20	0.5038	38	0.8173	15	0.8351	9	0.7740	16

MV+L-C to MVSK	0.7787	18	0.7074	19	0.7264	28	0.7214	28	0.7129	28
MVS+L-C to MVSK	0.6836	25	0.6994	21	0.7276	27	0.7351	27	0.7491	26
MVSK+L-C to MVSK	0.6168	29	0.6757	23	0.6944	33	0.7550	25	0.7504	23
MV+L-NC to MVSK	0.5806	38	0.8177	12	0.7396	24	0.6718	33	0.6946	31
MVS+L-NC to MVSK	0.6877	22	0.5518	27	0.7962	19	0.8313	11	0.7498	25
MVSK+L-NC to MVSK	0.7094	21	0.5038	38	0.8173	14	0.8351	10	0.7740	16
Stars to MV	0.5751	42	0.4313	46	0.8643	8	0.7789	21	0.7650	19
Stars to MVS	0.5845	36	0.4263	48	0.8557	10	0.7753	22	0.7585	20
Stars to MVSK	0.5845	36	0.4263	47	0.8557	9	0.7753	23	0.7585	21
Sharpe to MV	0.6188	27	0.5336	30	0.5212	34	0.5692	43	0.5635	47
Sharpe to MVS	0.6094	30	0.5360	28	0.5021	38	0.5625	47	0.5721	42
Sharpe to MVSK	0.6094	30	0.5360	29	0.5021	39	0.5558	48	0.5721	43
Sortino to MV	0.5799	39	0.5205	31	0.4871	43	0.5685	44	0.5659	46
Sortino to MVS	0.5792	40	0.5128	35	0.4834	44	0.5636	45	0.5745	40
Sortino to MVSK	0.5792	40	0.5128	36	0.4834	44	0.5636	46	0.5745	41
Omega to MV	0.6220	26	0.4351	45	0.4126	48	0.5693	42	0.5623	48
Omega to MVS	0.6070	32	0.4425	43	0.4264	46	0.5820	39	0.5715	44
Omega to MVSK	0.6070	33	0.4425	44	0.4264	47	0.5820	40	0.5715	45
Treynor to MV	0.5540	46	0.5141	34	0.5015	42	0.6004	36	0.6134	36
Treynor to MVS	0.5452	47	0.5198	32	0.5020	40	0.6126	34	0.6236	35
Treynor to MVSK	0.5452	48	0.5198	32	0.5020	40	0.6126	35	0.6236	34
Kappa to MV	0.5871	34	0.4835	41	0.5073	35	0.5768	41	0.5939	39
Kappa to MVS	0.5747	44	0.4878	39	0.5066	37	0.5855	38	0.6075	37
Kappa to MVSK	0.5747	43	0.4878	40	0.5066	36	0.5855	38	0.6075	38

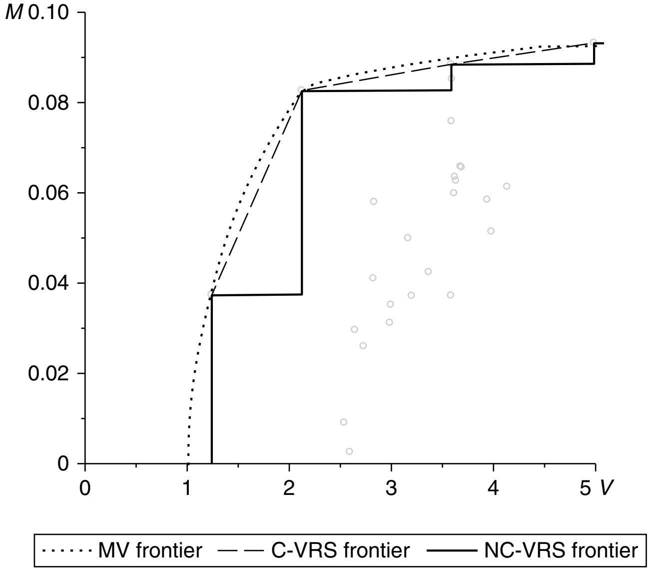

Figure 7.5 Visualization of MV-frontier based on the 40 best MF and the convex and nonconvex frontiers determined earlier in Figure 7.1.

tradition in finance. The second question is why some strategies dominate using either MV, MVS, or MVSK portfolio optimization without transaction costs, while only MV-based optimization strategies seem to perform well when including transaction costs.

The answer to the first question is probably situated in the approximative nature of these frontier estimators compared to some of the underlying portfolio models these are related to in the second stage. For illustrative purpose, Figure 7.5 shows part of the MV portfolio frontier obtained from selecting the 40 best performing MF from the main database. For this example, the MF have been ranked according to the convex shortage function. For comparison, also the convex and nonconvex frontiers obtained earlier using models (7.1) and (7.2) are included in the image.

Clearly, for this example, the MV portfolio frontier and the convex frontier of MF correspond quite well. By contrast, the fit with the nonconvex frontier model is a bit poorer with regard to this convex MV portfolio frontier.

In general, however, the goodness-of-fit of convex and nonconvex frontier MF rating models is an empirical issue. While it is hard or close to impossible to illustrate, one may speculate that a similar argument can be developed to explain why in some cases a multimoment nonconvex rating model can outperform a convex model. While MV portfolio frontiers are convex, it is well-known that MVS and MVSK portfolio frontiers can be highly nonconvex (see, e.g., Kerstens, Mounir, and Van de Woestyne (2011a) for illustrations of the MVS case). In such nonconvex portfolio worlds, the nonconvex rating models may well prove to have better approximation capacity than their currently more widely applied, traditional convex rating models. Obviously, it would be highly desirable to find ways to shed more light on this conjecture.

The answer to the second question is not straightforward. We currently have at best some elements of an answer. First, recall that the period under analysis is characterized by strong

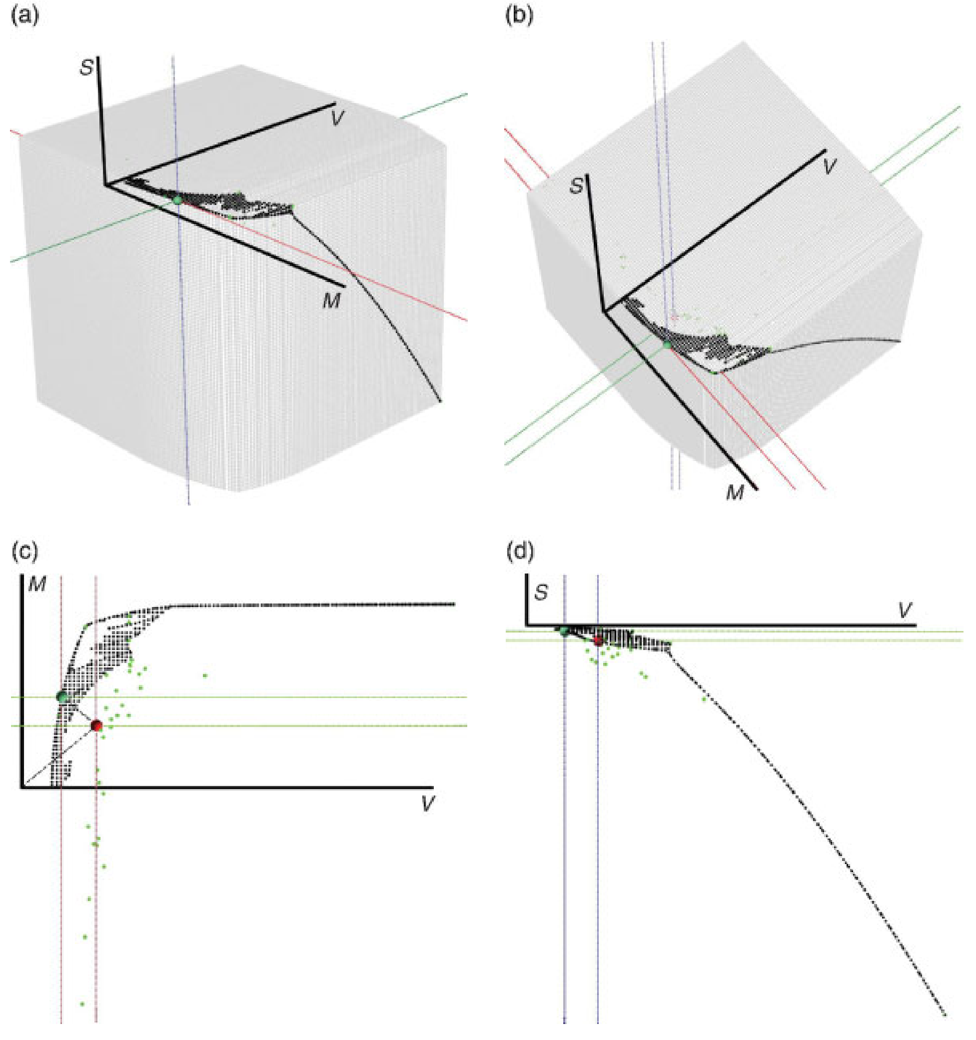

Figure 7.6 Visualization of MVS-frontier based on the 40 best MF. (a) Overview 1, (b) overview 2, (c) MV projection of the MVS model, and (d) VS projection of the MVS model.

negative skewness for most MF (recall Figure 7.2). Second, in periods with strong negative skewness, the way in which the position-dependent shortage function projects relative to multimoment portfolio models seems to lead to optimal frontier portfolios close to the ones situated on a basic MV portfolio frontier. We can illustrate this plausible explanation with the help of Figure 7.6.

Figure 7.6 presents different views of the MVS-frontier corresponding to the same 40 best performing MF used in Figure 7.5. Figure 7.6a and b shows two general views of the MVS-frontier. The black-colored part of the frontier refers to the strongly efficient

MVS-frontier, while the gray-colored part represents the weakly efficient MVS-frontier.[8] Figure 7.6c shows the projection of the MVS-frontier in the MV-plane. The MV-frontier is clearly visible as the boundary of the black-colored area. This is the same MV-frontier as the one visualized in Figure 7.5. Figure 7.6d presents the projection of the MVS-frontier in the VS-plane. It is clearly visible from this figure that the MVS-frontier is situated in the area of negative skewness. As a consequence, projecting a portfolio according to the position-dependent shortage function ends up most likely onto or close to the MV-frontier.

To illustrate this behavior of the position-dependent shortage function, the EWP and its projection onto the MVS-frontier are visualized in all four images of Figure 7.6 as the two big circle-shaped points. These points are best visible in Figure 7.6c. A projection by means of the MV-based shortage function results in exactly the same optimal point. Consequently, in this case adding a skewness constraint to the position-dependent shortage function does not alter the optimal solution obtained by merely optimizing over MV-space.

Thus, the use of a position-dependent shortage function in periods of negative skewness can lead to projections of MVS portfolio models close to the MV-frontier plane, making both models close to indistinguishable in practice. Consequently, the weak or nonexisting gains of using a higher-moment instead of a MV portfolio model are insufficient in combination with the seemingly slightly higher transaction costs to guarantee any strategic advantage. Obviously, this reasoning could benefit from further scrutiny. While it has already been established that the choice of direction vector affects the relative ranking of MV portfolios (see Kerstens, Mounir, and Van de Woestyne, 2012), the current backtesting results are yet another reason to start further exploring alternative choices of direction vector for specific purposes (*in casu*, coping with a period of negative skewness).

7.5 Conclusions

We basically test the traditional convex vs nonconvex frontier MF rating models in an extensive backtesting framework against, on the one hand, traditional financial performance measures and, on the other hand, the three-year Morningstar ratings. The basic conclusions from this first intensive and extensive backtesting analysis can be summarized as follows.

First, without portfolio optimization strategies and ignoring transaction costs, frontier MF ratings are performing poorly compared to traditional financial performance measures. With transaction costs, the previous conclusion reverses. However, in both cases the Morningstar ratings tend to outperform all other strategies, except for some of the smaller sample sizes. In conjunction with portfolio optimization strategies without and with transaction costs, the frontier MF ratings clearly outperform traditional financial performance measures and even do better than the Morningstar-based strategies (except for one particular sample size). Overall, naïve diversification strategies seem to be working better than strategies involving portfolio optimization. However, it is clear that especially MVS and MVSK multimoment portfolio models may currently suffer from the choice of period of analysis that is deeply marked by the financial crisis. Thus, for the time being, our results fall perfectly in line with those of DeMiguel, Garlappi, and Uppal (2009) and others.

Second, convex frontier MF ratings seem to work better than nonconvex ones, though the capability of the latter to better approximate nonconvex higher-moment portfolio models

[8] We refer the reader to Kerstens, Mounir, and Van de Woestyne (2011a) for more information on these notions.

remains to be further investigated in view of the difficulty of multimoment portfolio models to operate properly during the current sample period.

Thus, somewhat in contrast to Blake and Morey (2000) and Kräussl and Sandelowsky (2007), our results do offer some benefit of the doubt to the three-year Morningstar ratings to pick MF shares that prove to perform decently in the nearby future, relatively speaking. But, some new alternative gauges founded on frontier estimation prove to function at least equally well. In particular, there seems to be some promise in combining multimoment frontier-based MF ratings in conjunction with corresponding multimoment portfolio optimization models using the same basic methodology. Obviously, there is plenty of room for further refinements and intensive backtesting to substantiate this preliminary conclusion.

In view of some of the difficulties encountered and conjectures made, there is a series of methodological refinements one can think about. First, while we have focused on standard moments in the entire analysis, the Morningstar rating is at least partially influenced by some elements of a lower partial-moments approach. Thus, the whole analysis may eventually benefit from being redone using partial rather than standard moments.

Second, while the frontier MF ratings as well as the traditional financial performance measures are applied to the whole sample, the Morningstar ratings in fact apply only to the same Morningstar category. There happen to be several Morningstar categories in our sample. Furthermore, these Morningstar categories evolve over time following changes in the asset composition of the MF. To have a more refined comparison, one could envision computing frontier MF ratings and traditional financial performance measures on the same Morningstar categories as they evolve over time. Obviously, apart from complicating the analysis, this may also compromise the nonparametric frontier MF rating estimations depending on the size of these Morningstar categories, because of the curse of dimensionality characteristic for nonparametric estimators.

Third, while an equally weighted portfolio is a standard approach to backtest in a MV world, it remains an open question whether this starting point is useful in a nonnormal MVS or MVSK portfolio approach. This question is somewhat related to the issue of the choice of projection direction developed in Section 7.4.4.

In terms of further refinements of the backtesting setup, one can think of the following issues. First, it could be important to select a sample period that includes both clear boom and bust periods. Second, some of the doubts raised with respect to the traditional financial performance measures in relation to the multimoment portfolio models certainly require further investigation. We have hardly been able to scratch the surface in this chapter. Third, the current results seem to vary somewhat in nonmonotonous ways across sample sizes. This is currently very hard to explain and probably merits some deeper investigation.

To some extent, one could also imagine that this experience could inspire another related literature using frontier models for asset selection purposes rather than MF rating (e.g., Abad, Thore, and Laffarga, 2004; Alam and Sickles, 1998; and Nguyen and Swanson, 2009). It clearly remains an open question whether the current approach attempting to define a superfund starting from some subsets of well-rated candidate MF offers the same scope as composing a simple portfolio based on some well-rated candidate assets. A much more fundamental methodological issue is to what extent the use of two-stage approaches, with some frontier model in the first stage and another kind of portfolio selection model in the second stage, can be justified in view of some criticisms of such approaches in the frontier literature (Simar and Wilson, 2007).

Acknowledgments

We thank the editor of this volume for valuable comments. Ignace Van de Woestyne acknowledges financial support by the National Bank of Belgium.

References

Abad, C., Thore, S.A., and Laffarga, J. (2004) Fundamental analysis of stocks by two-stage DEA. *Managerial and Decision Economics*, **25** (5), 231–241.

Alam, I. and Sickles, R. (1998) The relationship between stock market returns and technical efficiency innovations: evidence from the US airline industry. *Journal of Productivity Analysis*, **9** (1), 35–51.

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.

Bacon, C. (2008) *Practical Portfolio Performance Measurement and Attribution*, 2nd edn, John Wiley & Sons, Ltd, Chichester.

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.

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

Blake, D. (1990) Portfolio behaviour and asset pricing in a characteristics framework. *Scottish Journal of Political Economy*, **37** (4), 343–359.

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

Brandouy, O., Kerstens, K., and Van de Woestyne, I. (2010b) Exploring bi-criteria versus multi-dimensional lower partial moment portfolio models. *International Journal of Technology, Modelling and Management*, **1** (1), 25–39.

Brandouy, O., Mathieu, P., and Veryzhenko, I. (2012) Optimal portfolio diversification? A multi-agent ecological competition analysis, in *Highlights in Practical Applications of Agents and Multi-Agent Systems. 10th International Conference on Practical Applications of Agents and Multi-Agent Systems*, vol. **156** of Advances in Intelligent and Soft Computing (eds J. Bajo Pérez, J. Corchado, E. Adam, A. Ortega, M. Moreno, E. Navarro, B. Hirsch, H. Lopes Cardoso, V. Julián, M. Sánchez, and P. Mathieu), Springer, Berlin, pp. 323–332.

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, 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., and 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.

Broihanne, M.H., Merli, M., and Roger, P. (2008) On the robustness of mutual funds ranking with an index of relative efficiency. *Bankers, Markets & Investors*, **94**, 32–43.

Campbell, S. (2005) A review of backtesting and backtesting procedures. Finance and Economics discussion series no. 2005-21. Federal Reserve Board, Washington.

Cantaluppi, L. and Hug, R. (2000) Efficiency ratio: a new methodology for performance measurement. *Journal of Investing*, **9** (2), 1–7.

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.

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

Clark, E., Jokung, O., and Kassimatis, K. (2011) Making inefficient market indices efficient. *European Journal of Operational Research*, **209** (1), 83–89.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009) Optimal versus naive diversification: how inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, **22**, 1915–1953.

Dodds, C. (1986) Portfolio modelling and the characteristics approach. *Managerial Finance*, **12** (3), 16–18.

Edirisinghe, N. and Zhang, X. (2007) Generalized DEA model of fundamental analysis and its application to portfolio optimization. *Journal of Banking & Finance*, **31** (11), 3311–3335.

Edirisinghe, N. and Zhang, X. (2010) Input/output selection in DEA under expert information, with application to financial markets. *European Journal of Operational Research*, **207** (3), 1669–1678.

Friedman, J. (1983) *Oligopoly Theory*, Cambridge University Press, Cambridge.

Galagedera, D. and Silvapulle, P. (2002) Australian mutual fund performance appraisal using data envelopment analysis. *Managerial Finance*, **28** (9), 60–73.

Glawischnig, 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. (ed.) (2007) *Performance of Mutual Funds: An International Perspective*, Palgrave, New York.

Haslem, J. and Scheraga, C. (2003) Data envelopment analysis of Morningstar's large-cap mutual funds. *Journal of Investing*, **12** (4), 41–48.

Heffernan, S. (1992) A computation of interest equivalences for non-price features of bank products. *Journal of Money, Credit and Banking*, **24** (2), 162–172.

Jurczenko, E. and Maillet, B. (2006) The four-moment capital asset pricing model: between asset pricing and asset allocation, in *Multi-Moment Asset Allocation and Pricing Models* (eds E. Jurczenko and B. Maillet), John Wiley & Sons, Ltd, Chichester, pp. 113–163.

Jurczenko, E. and Yanou, G. (2010) Fund of hedge funds portfolio selection: a robust non-parametric multi-moment approach, in *The Recent Trend of Hedge Fund Strategies* (ed. Y. Watanabe), Nova Science, New York, pp. 21–56.

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

Kerstens, K., Mounir, A., and Van de Woestyne, I. (2011a) Geometric representation of the mean-variance-skewness portfolio frontier based upon the shortage function. *European Journal of Operational Research*, **210** (1), 81–94.

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

Kerstens, K., Mounir, A., and Van de Woestyne, I. (2012) Benchmarking mean-variance portfolios using a shortage function: the choice of direction vector affects rankings! *Journal of the Operational Research Society*, **63** (9), 1199–1212.

Kräussl, R. and Sandelowsky, R.M.R. (2007) The predictive performance of Morningstar's mutual fund ratings (August 17, 2007). http://dx.doi.org/10.2139/ssrn.963489; http://papers.ssrn.com/sol3/papers.cfm?abstract_id=963489. Accessed on 20 December 2012.

Lamb, J. and Tee, K.H. (2012) Data envelopment analysis models of investment funds. *European Journal of Operational Research*, **216** (3), 687–696.

Levy, M., Levy, H., and Solomon, S. (1995) Microscopic simulation of the stock market: the effect of microscopic diversity. *Journal de Physique I (France)*, **5**, 1087–1107.

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

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

Martellini, L. (2008) Towards the design of better equity benchmarks: rehabilitating the tangency portfolio from modern portfolio theory. *Journal of Portfolio Management*, **34** (4), 34–41.

Matallín, J., Soler, A., and Tortosa-Ausina, E. (2011) On the informativeness of persistence for evaluating mutual funds performance using partial frontiers. Valencia, Instituto Valenciano de Investigaciones Económicas, Technical report. *IVIE WP-EC 2011-08*. http://www.ivie.es/downloads/docs/wpasec/wpasec-2011-08.pdf. Accessed on 20 December 2012.

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

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

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

Nguyen, G. and Swanson, P. (2009) Firm characteristics, relative efficiency, and equity returns. *Journal of Financial and Quantitative Analysis*, **44**, 213–236.

Peterson Drake, P. and Fabozzi, F. (2010) *The Basics of Finance: An Introduction to Financial Markets, Business Finance, and Portfolio Management*, John Wiley & Sons, Inc, Hoboken.

Sengupta, J. (1989) Nonparametric tests of efficiency of portfolio investment. *Journal of Economics*, **50** (1), 1–15.

Sengupta, J. and Zohar, T. (2001) Nonparametric analysis of portfolio efficiency. *Applied Economics Letters*, **8** (4), 249–252.

Sharpe, W.F. (1998) Morningstar's risk-adjusted ratings. *Financial Analysts Journal*, **54** (4), 21–33.

Simar, L. and Wilson, P. (2007) Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, **136** (1), 31–64.

Sprott, J. (2004) Competition with evolution in ecology and finance. *Physics Letters A*, **325**, 329–333.

Tsolas, I. (2011) Natural resources exchange traded funds: performance appraisal using DEA modeling. *Journal of CENTRUM Cathedra*, **4** (2), 250–259.

Tu, J. and Zhou, G. (2011) Markowitz meets Talmud: a combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, **99**, 204–215.

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

Zhao, X., Wang, S., and Lai, K. (2011) Mutual funds performance evaluation based on endogenous benchmarks. *Expert Systems with Applications*, **38** (4), 3663–3670.